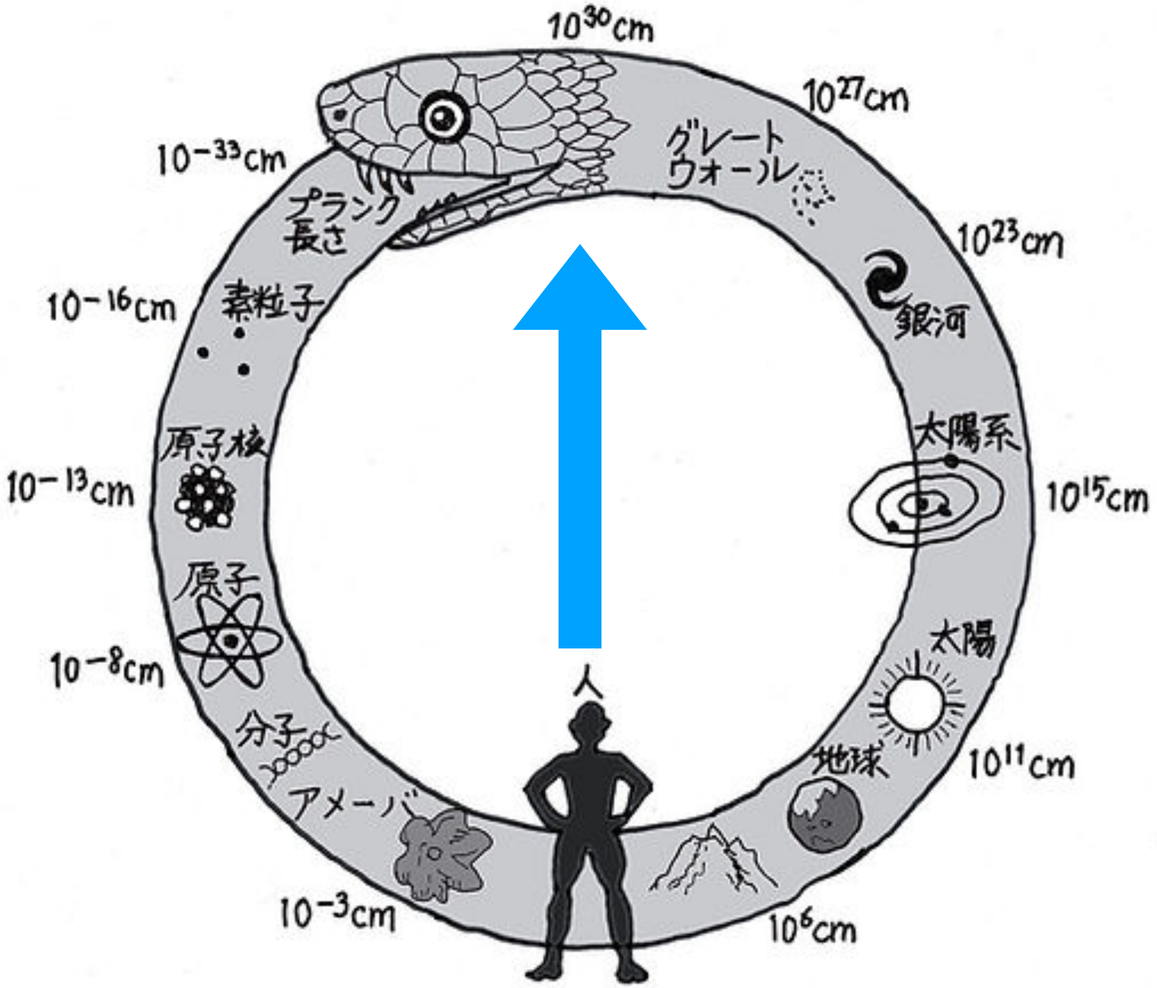
Towards a Physical Theory of Agents

Hua Xing Zhu
Peking University

In collaboration with: **Zeyu Cai**, Qing-Hong Cao, **Shaoyang Guo,** Tie-Jiun Hou, Zeyu Li, Chang Liu, Xiaohui Liu, Ming-xing Luo, Jichen Pan, **Shi Qiu**, **Yunbo Sun**, **Zhuo-Yang Song**, **Jiashen Wei,** Tong-Zhi Yang, **Yixuan Yin,** Xinbo Yuan, Shu-tao Zhang

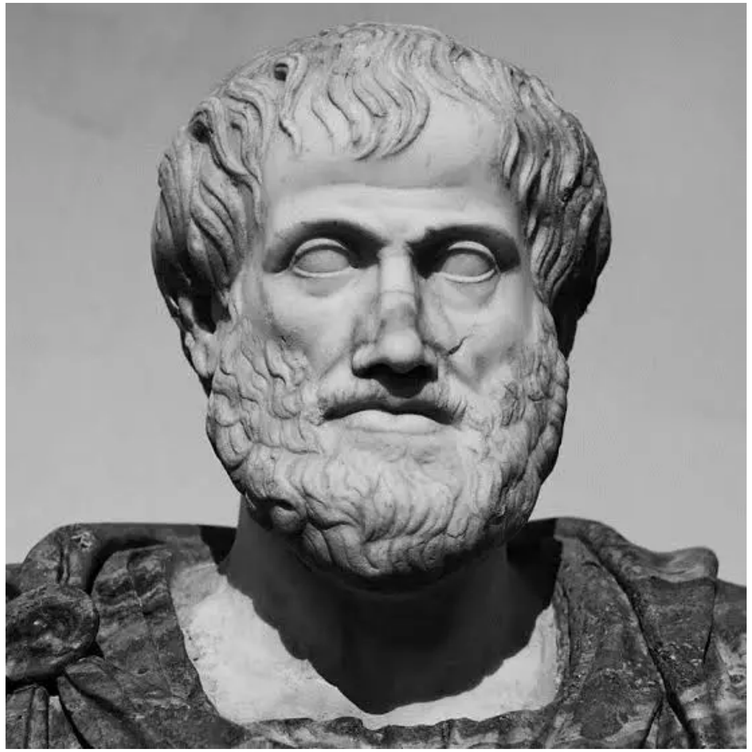
2026年新物理前沿与交叉科学研讨会
2026年5月17日，南京

A physicist's view of the universe



- Elementary particles are building blocks of the universe
- Each scale comes with its own physics: effective theory
- Yet the smallest and largest are intricately connected: UV <->IR

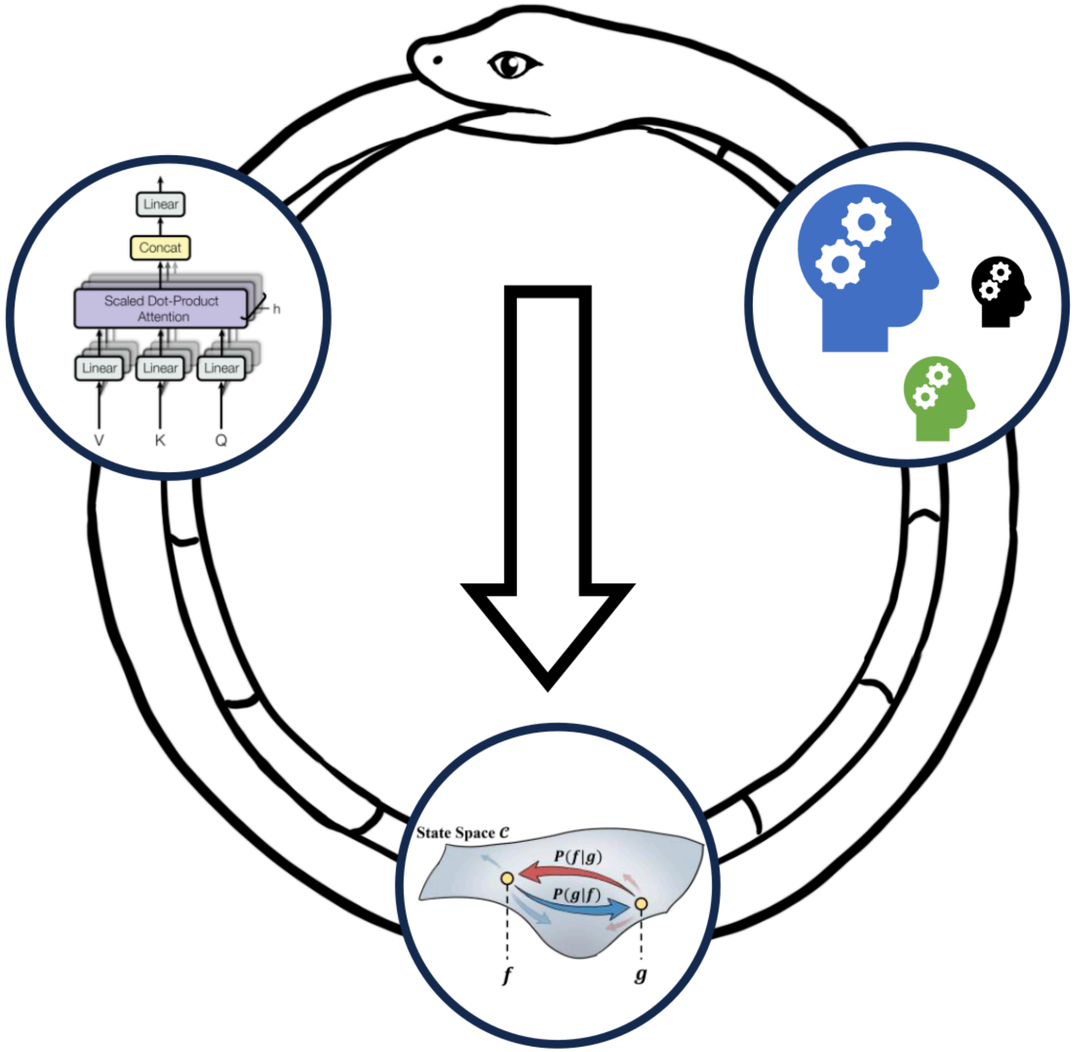
The study of human mind and intelligence



He argues that some parts of the soul, **the intellect**, can exist without the body, but most cannot.

Intelligence is never merely a measure of cognitive or academic ability; true intelligence is the practical ability to understand **human relationships, cultivate virtue, and contribute to social harmony**.

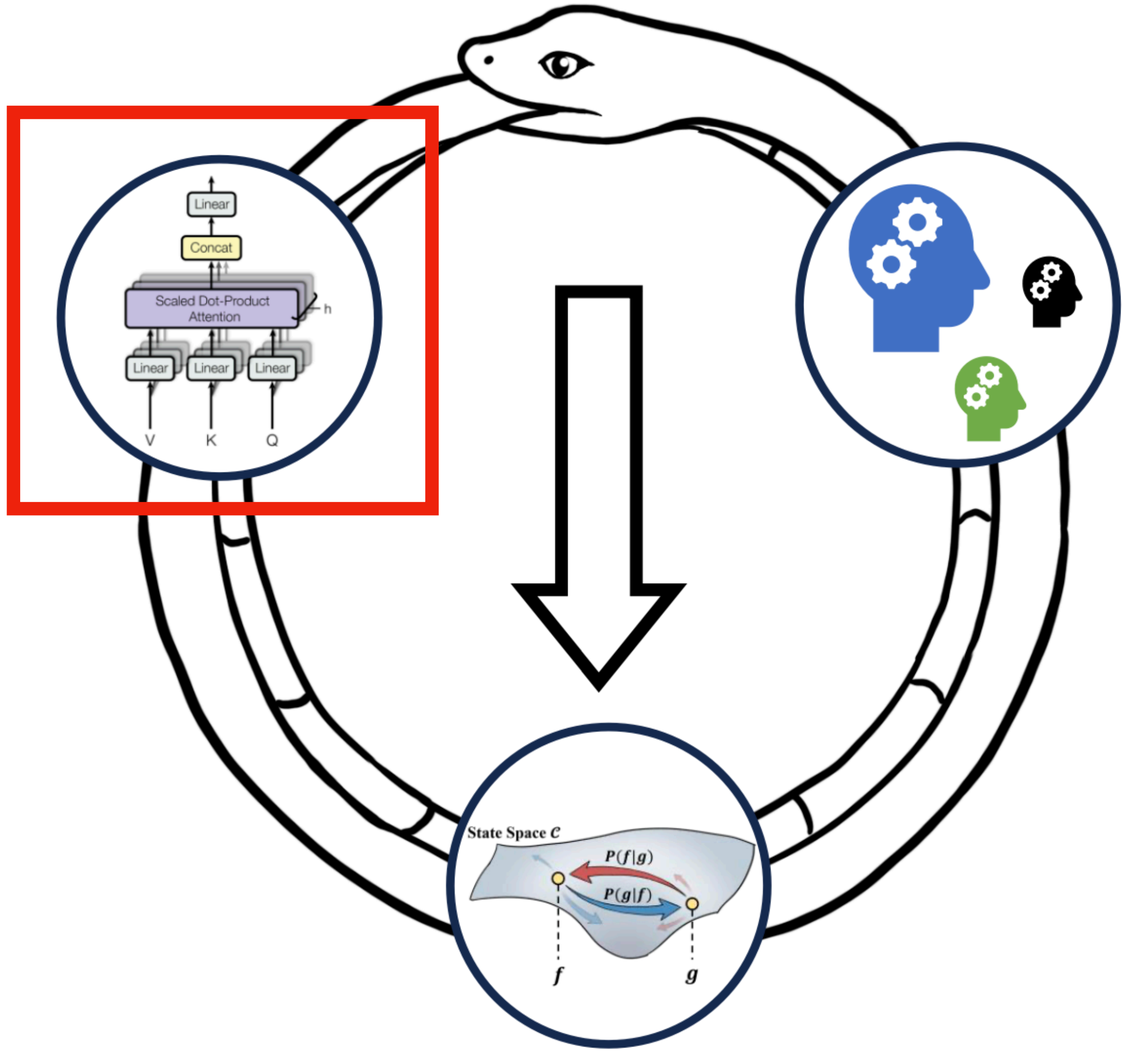
Towards a physicist's view of artificial intelligence



- Microscopic: token dynamics
- Macroscopic: benchmark in general sense
- **mesoscopic: ensemble/states**

Undergraduate thesis:
https://sonnynondegeneracy.github.io/

Towards a physicist's view of artificial intelligence



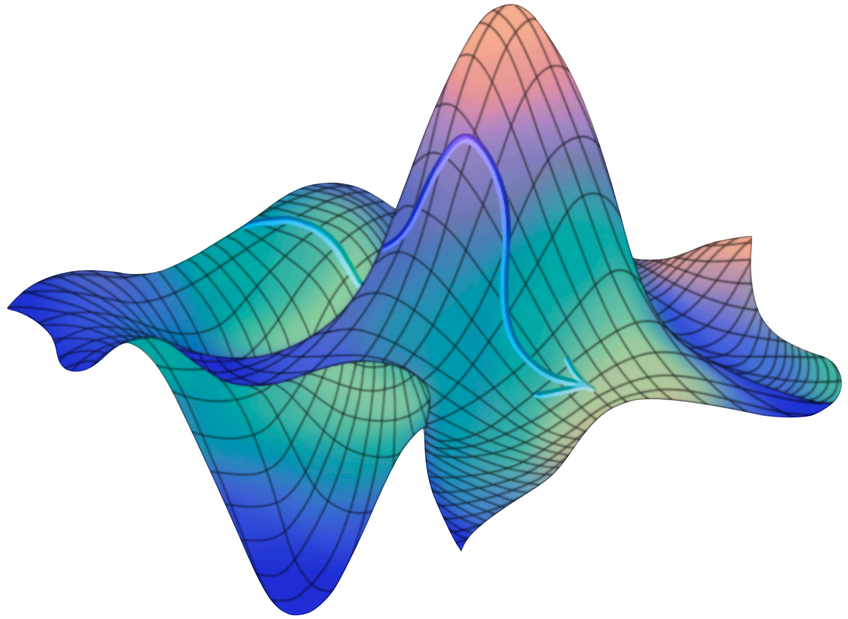
- Microscopic: token dynamics
- Macroscopic: benchmark in general sense
- **mesoscopic: ensemble/states**

Undergraduate thesis:
https://sonnynondegeneracy.github.io/

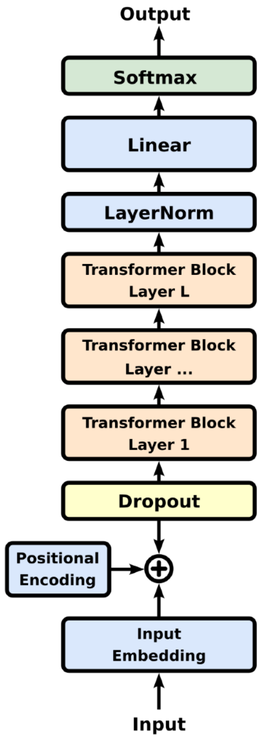
The manifold hypothesis

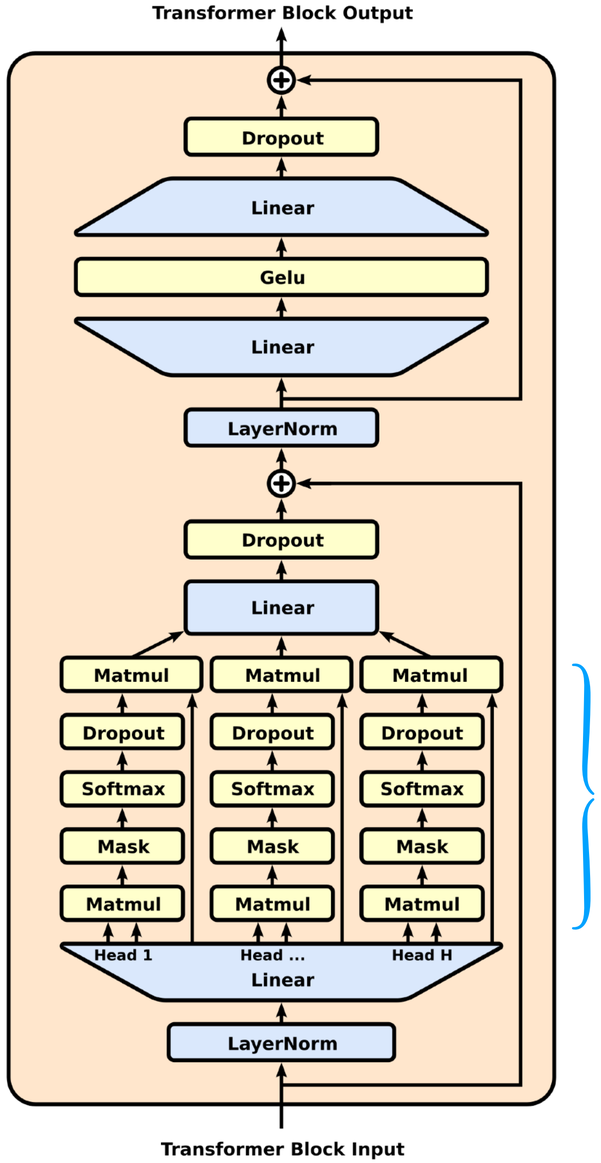
- A typical image dimension
  512x512x3 ~ 7 x 10^5
- For a typical latent diffusion model, VAE has dimension
  16, 000 dimension
- Large Language Model (LLM) model human intelligence
- Does intelligence has intrinsic manifold? Aksenov, Bodnia, Freedman, Mulligan 2603.20396
- How to measure the dimension of manifold?

https://link.springer.com/article/10.1007/s11263-022-01598-5?utm_source=researchgate.net&utm_medium=article



Transformer as an operator





attention layer

Formalization of transformer:

$$\text{Emb}: \quad x_i \to \boldsymbol{t}_i \in \mathbb{R}^{d_{\text{embed}}}$$

$$\boldsymbol{t}_i \in \mathcal{E}_0 \subset \mathbb{R}^{d_{\text{embed}}}$$

transformer layer: $\mathcal{T}_\xi: \quad \mathcal{E}_\xi \to \mathcal{E}_{\xi+1}$

$$\boldsymbol{t}_{i,\xi} \to \mathcal{T}_\xi \boldsymbol{t}_{i,\xi} = \boldsymbol{t}_{i,\xi} + F_\xi \left(\sum_{j,k,l} \hat{\Gamma}_{ijkl,\xi} \left(\hat{\boldsymbol{t}}_{j,\xi} \otimes \hat{\boldsymbol{t}}_{k,\xi} \right) \cdot \hat{\boldsymbol{t}}_{l,\xi}, \phi(i) \right)$$

Z.Y. Song, Z.Y. Li, Q.H. Cao, M.X. Luo, HXZ, 2503.22547

Physics observable on the manifold

Token correlation $$E(\xi) = \frac{\sum_{i \neq j} \langle t_{i,\xi} \cdot t_{j,\xi} \rangle}{\left\langle \|t_{k,\xi}\|^2 \right\rangle}$$

Sum of dot product of two token in embed space

$$t_{i,\xi+1} = t_{i,\xi} - \|t_{i,\xi}\| \sum_{\alpha} \left(\hat{t}_{i,\xi} \cdot n_\alpha\right) n_\alpha$$ layer transformer as a projection

$$E(\xi) \cdot (d-1) = \text{ Constant}$$

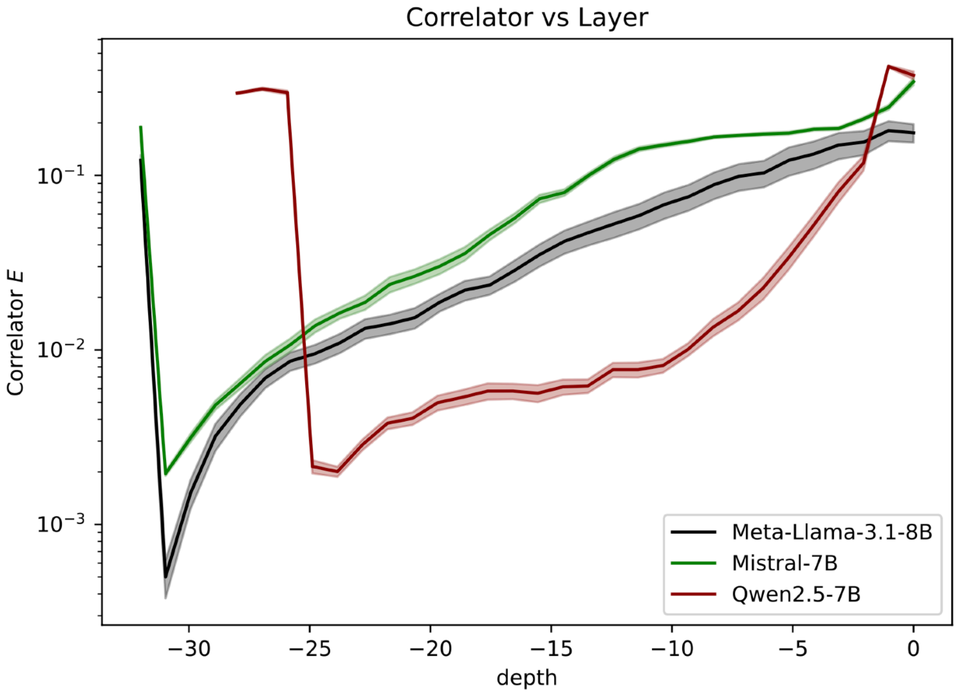
Token correlation is inverse to the intrinsic dimension

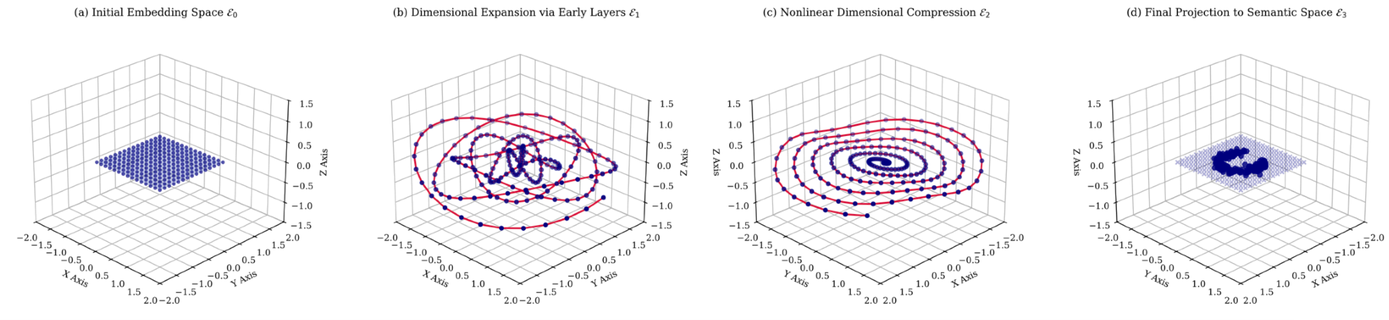
Intrinsic dimension can be measured form token correlation!

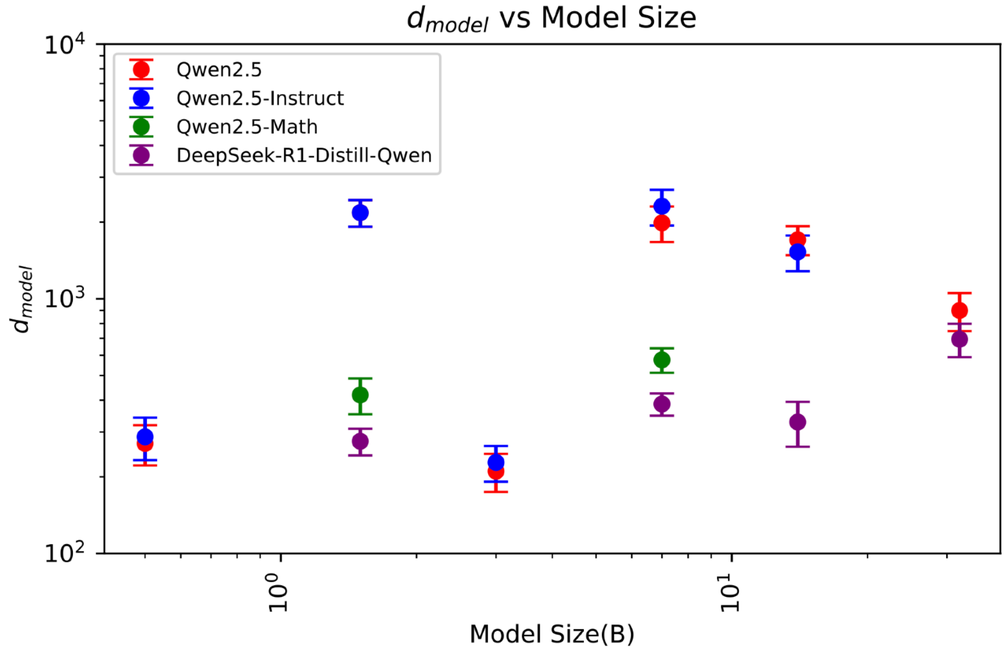
Z.Y. Song, Z.Y. Li, Q.H. Cao, M.X. Luo, HXZ, 2503.22547

Layer-wise evolution of correlation and dimension

$$E(\xi) \cdot (d-1) = \text{ Constant}$$

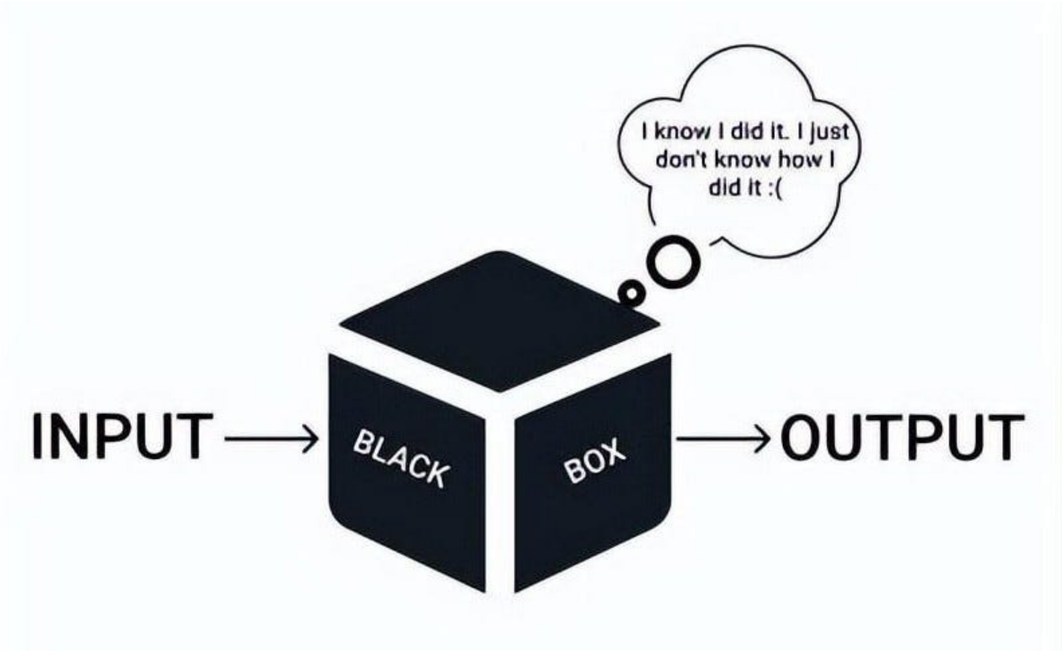






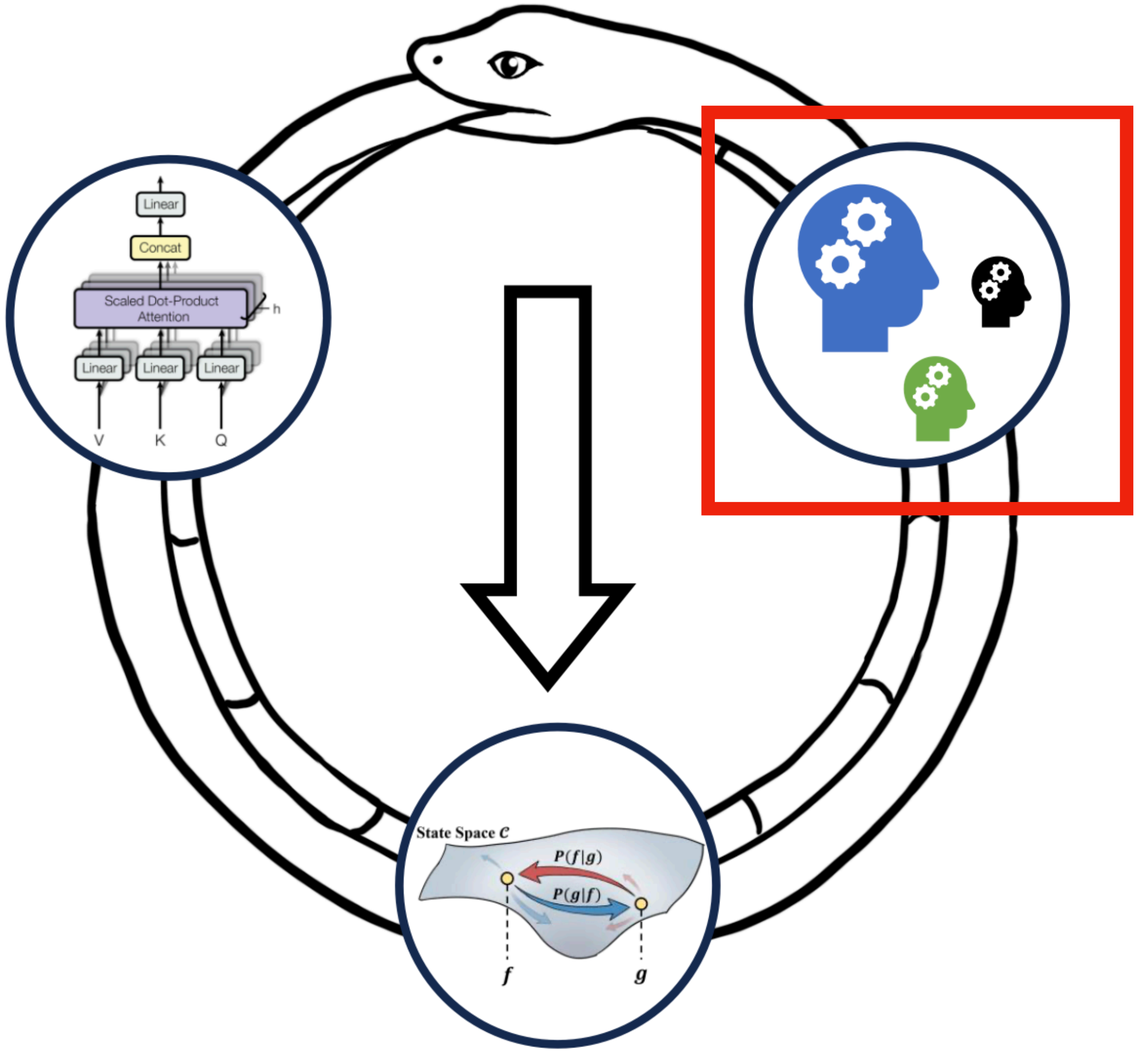
Z.Y. Song, Z.Y. Li, Q.H. Cao, M.X. Luo, HXZ, 2503.22547

AI is still a blackbox



- How to **exploit** the scientific ability of LLM/Agent
- How to **improve** the scientific ability of LLM/Agent
- Ultimately, understanding the blackbox is crucial for aligning AI with human interests

Towards a physicist's view of artificial intelligence



- Microscopic: token dynamics
- Macroscopic: benchmark in general sense
- **mesoscopic: ensemble/states**

Undergraduate thesis:
https://sonnynondegeneracy.github.io/

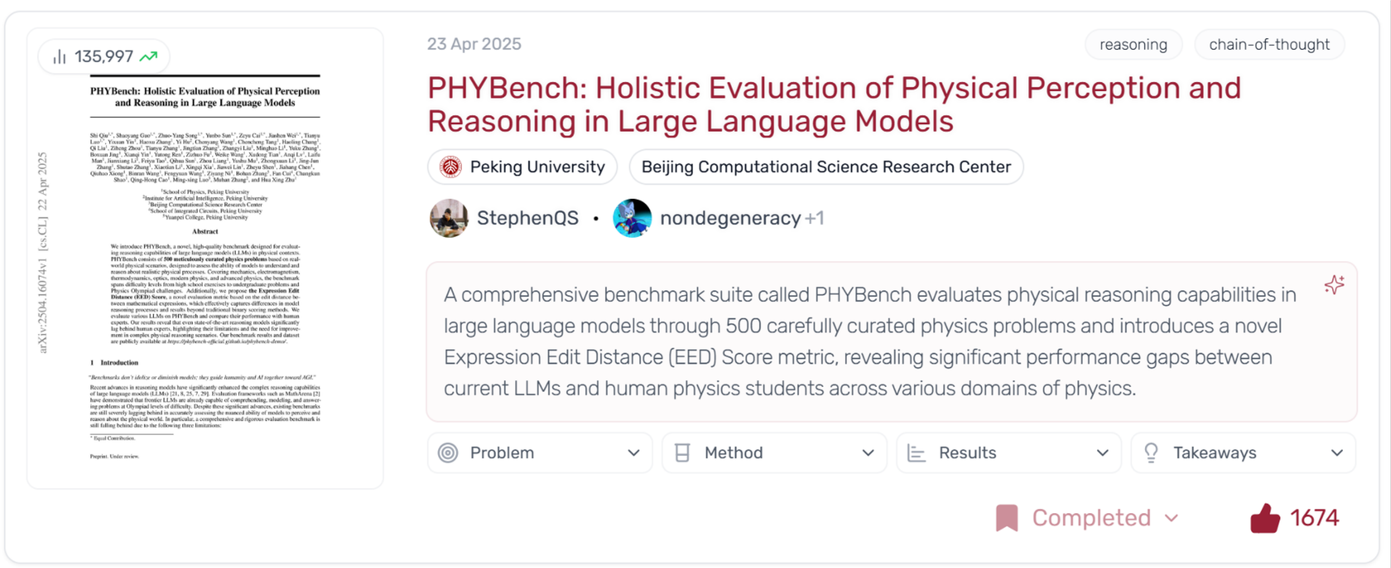
Benchmark as a collaboration between science and industry community

Shaping the Future of Mathematics in the Age of AI

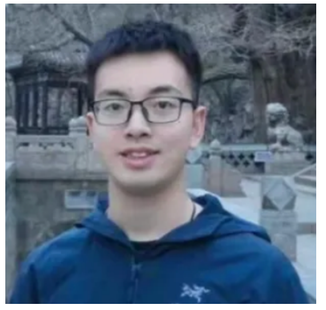
Johan Commelin* Mateja Jamnik† Rodrigo Ochigame‡ Lenny Taelman§
Akshay Venkatesh¶

Another proposal being pursued by multiple groups is the creation of community-owned benchmarks. Good benchmarks can signal community values to those developing AI systems, while also helping mathematicians understand realistic use cases for, and limitations of, the technology. For example, benchmarks that focus on open-ended problems can counter overstated claims based on narrow task suites.

多个团队正在推进的另一项提议是创建社区共有的基准测试。优秀的基准测试不仅能向人工智能系统的开发者传递社区的价值取向，还能帮助数学家理解该技术的实际应用场景及其局限性。例如，聚焦于开放式问题的基准测试可以有效反驳那些基于狭隘任务套件而产生的夸大言论。



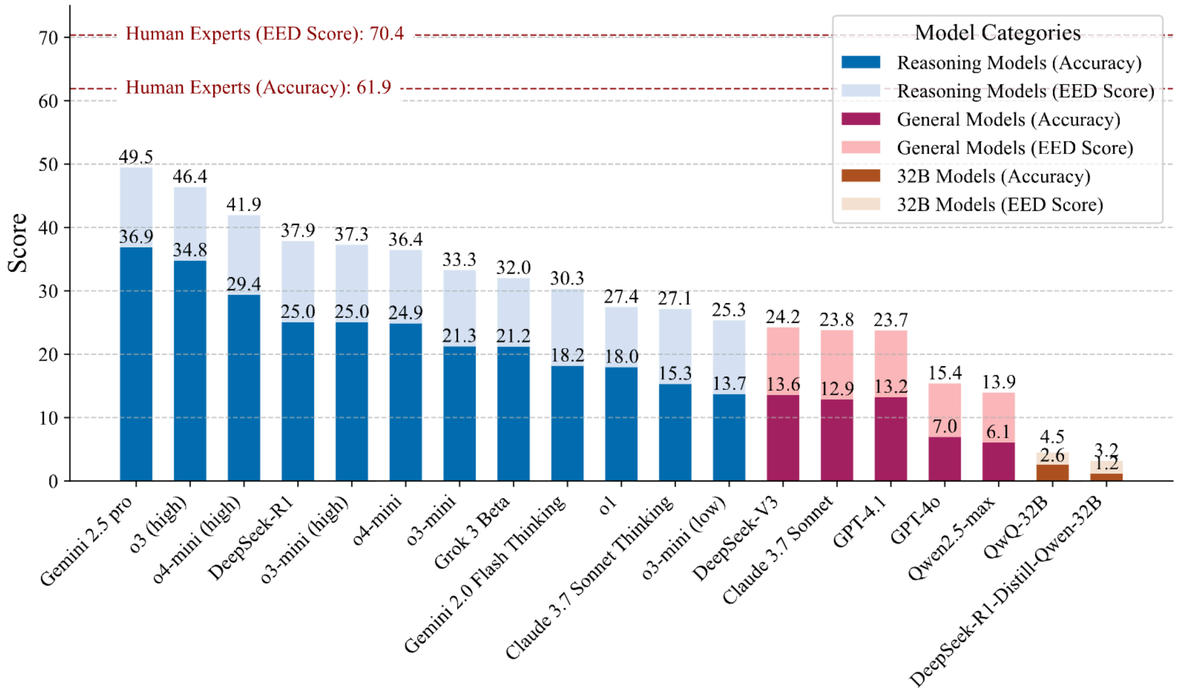
An Olympiad level physics competition benchmarks, curated by undergraduate students from Peking University



NIPS 2025

Shi Qiu, et al., , 2504.16074

Progress in model capability as measured from PHYbench



April 2025

ByteDance 字节跳动 | Seed

Seed2.0 正式发布

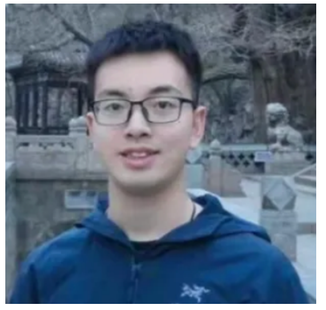
日期
2026-02-14

分类

模型发布

Seed2.0 正式发布

Capability	Benchmark	Seed2.0 pro	Seed2.0 lite	Seed2.0 mini	GPT-5.2 High	GPT-5 mini High	Claude Sonnet 4.5
STEM	GPQA Diamond	88.9	85.1	79	**92.4**	82.1	
	Superchem (text-only)	51.6	48	16.2	58	34.8	
	BABE	50	50.2	40.4	**58.1**	49.2	
	Phybench	74	73	56	74	60	
	FrontierSci-research	**25**	18.3	3.3	**25**	18.3	
	FrontierSci-olympiad	74	70	44	**75**	69	



Feb 2026

NIPS 2025

Shi Qiu, et al., , 2504.16074

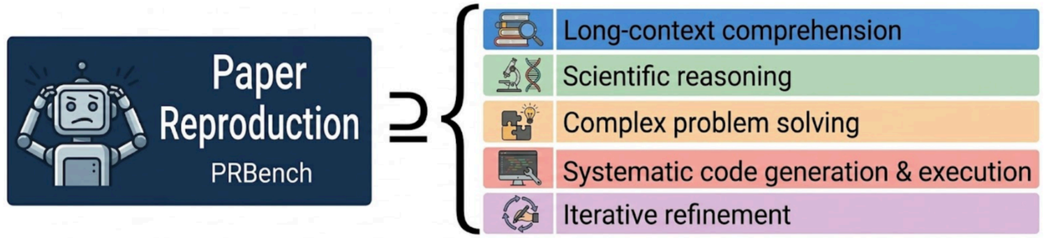
PRBench: ultimate benchmark for AI intelligence

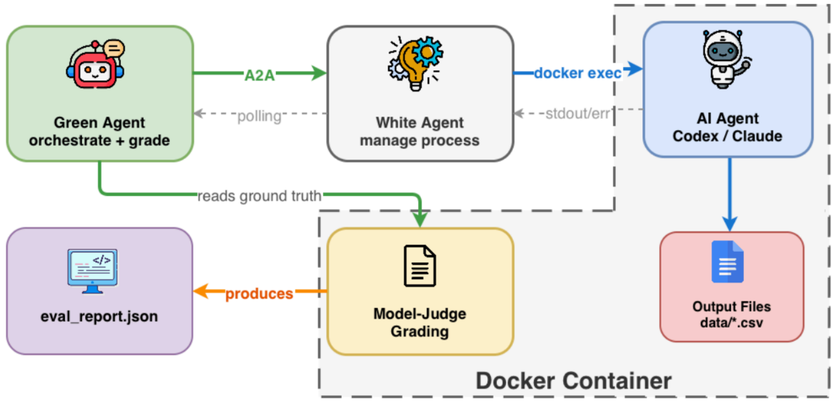
PRBench: End-to-end Paper Reproduction in Physics Research

Shi Qiu[1], **Junyi Deng**[1], **Yiwei Deng**[1], **Haoran Dong**[1], **Jieyu Fu**[1], **Mao Li**[1], **Zeyu Li**[1], **Zhaolong Zhang**[1], **Huiwen Zheng**[1], **Leidong Bao**[1], **Anqi Lv**[1], **Zihan Mo**[1], **Yadi Niu**[1], **Yiyang Peng**[1], **Yu Tian**[1], **Yili Wang**[1], **Ziyu Wang**[1], **Zi-Yu Wang**[1], **Jiashen Wei**[1], **Liuheng Wu**[1], **Aoran Xue**[1], **Leyi Yang**[1], **Guanglu Yuan**[1], **Xiarui Zhan**[1], **Jingjun Zhang**[1], **Zifan Zheng**[1], **Pengfei Liu**[1], **Linrui Zhen**[1], **Kaiyang Li**[1], **Qichang Li**[1], **Ziheng Zhou**[1], **Guo-En Nian**[1], **Yunwei Xiao**[1], **Qing-Hong Cao**[1], **Linjie Dai**[1], **Xu Feng**[1], **Peng Gao**[1], **Ying Gu**[1], **Chang Liu**[1], **Jia Liu**[1], **Ming-xing Luo**[2], **Yan-Qing Ma**[1], **Liang-You Peng**[1], **Huichao Song**[1], **Shufeng Wang**[1], **Chenxu Wang**[1], **Tao Wang**[1], **Yi-Nan Wang**[1], **Chengyin Wu**[1], **Pengwei Zhao**[1], and **Hua Xing Zhu**[1,†]

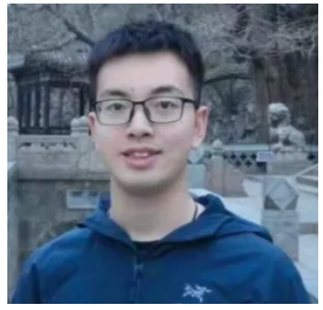
[1]School of Physics, Peking University, China
[2]Beijing Computational Science Research Center, China

March 31, 2026





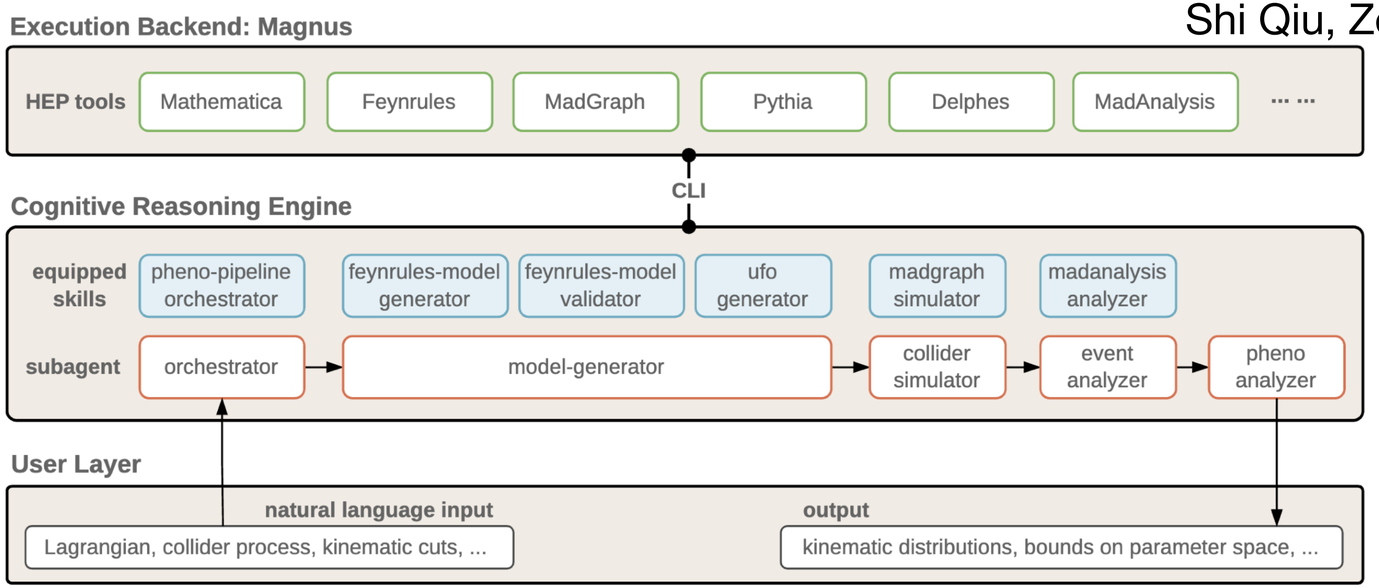
Agent	Methodology	Code	Data Acc.	Instruction	Overall
OpenAI Codex					
GPT-5.3-Codex	**78.00**	**43.00**	**21.00**	**92.00**	**34.00**
OpenCode-based Agents					
GPT-5.3-Codex	72.00	36.00	16.00	90.00	28.50
Kimi K2.5 (1T)	61.90	22.00	11.40	80.60	20.57
DeepSeek V3.2 (671B)	63.20	19.30	8.90	84.20	18.50
Minimax 2.7 (230B)	55.60	16.30	10.00	86.00	17.97
GLM-5 (744B)	50.50	18.80	10.60	67.00	17.87
End-to-End Callback Rate			0.0		

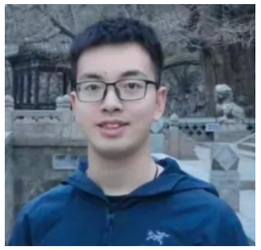


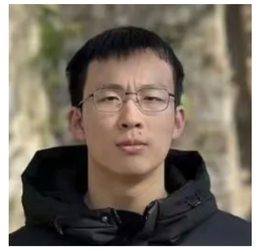
Shi Qiu, et al., , 2603.27646

Dedicated agent framework for science discovery

Shi Qiu, Zeyu Cai, Jiashen Wei, Yixuan Yin, et al., , 2603.14553

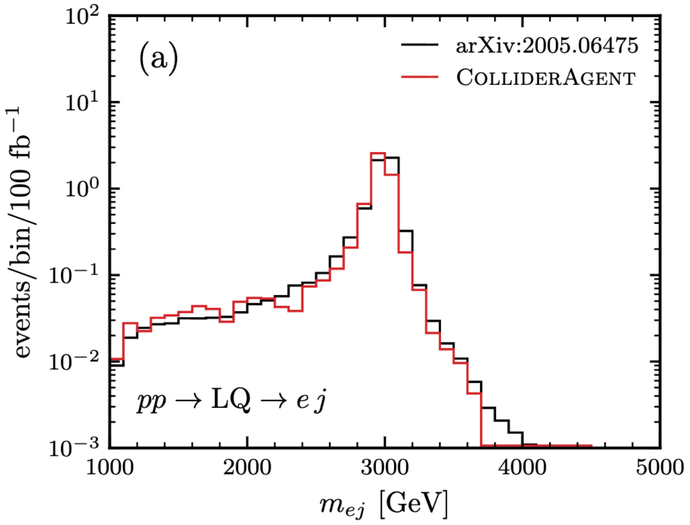


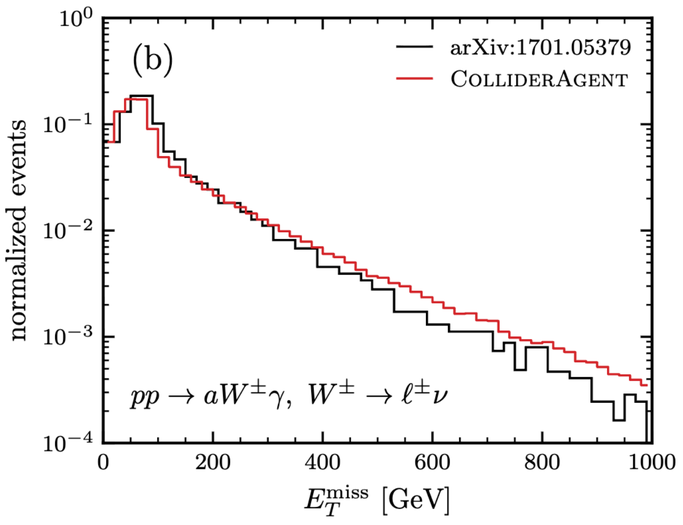


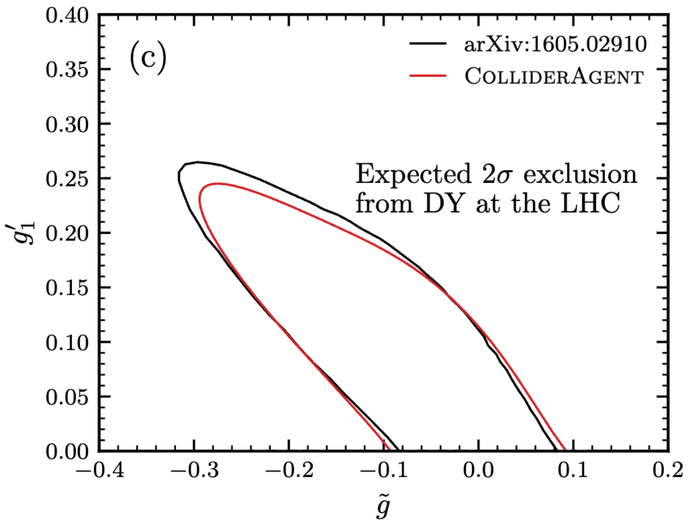


ColliderAgent

Coding Agent Magnus

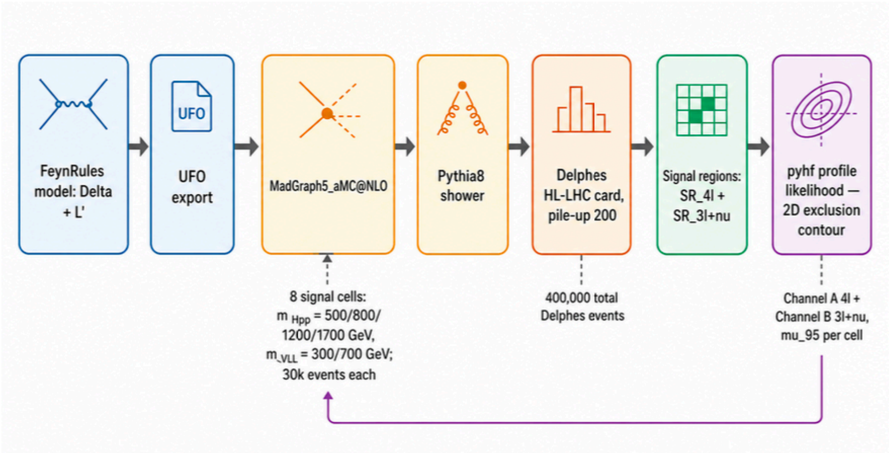






Autoresearchclaw 2.0

$$
\begin{aligned}
\mathcal{L}_{\mathrm{BSM}} \;=\; & (D_\mu \Delta)^\dagger (D^\mu \Delta) - V(H,\Delta) \\
& + i\,\overline{L'}\gamma^\mu D_\mu L' \;-\; m_{L'}\,\overline{L'}L' \\
& - \Big[\, f_{\ell\ell'}\,\overline{L_L^c}\, i\sigma_2\, \Delta\, L_L \\
& \quad + y_E\,\overline{L'_L}\,\Delta\, E_R \;+\; \mathrm{h.c.}\,\Big].
\end{aligned}
$$



Joint $4\ell + 3\ell{+}\nu$ HL-LHC search for a Type-II seesaw triplet coupled to a vector-like-lepton doublet

ARC-Agent collaboration[1]

[1] *Compiled by the ARC-Agent paper-author from a closed end-to-end pipeline; data and code openly available at the run label* `hep_t2seesaw_vll_joint_20260515`.

(Dated: May 16, 2026)

The most recent Type-II seesaw robustness study [1] singles out a coupling to a vector-like-lepton (VLL) sector as the physically best-motivated extension capable of shifting the HL-LHC bound on the doubly-charged scalar mass $m_{H^{\pm\pm}}$ by ± 150 GeV, yet the joint two-dimensional $(m_{H^{\pm\pm}}, m_{\mathrm{VLL}})$ exclusion reach has never been computed. We build the first joint $\Delta \oplus L'_{\mathrm{doublet}}$ FeynRules-generated UFO with the gauge-invariant Yukawa portal $y_E\,\overline{L'_L}\,\Delta E_R + \mathrm{h.c.}$, generate the eight-point grid $m_{H^{\pm\pm}} \in \{500, 800, 1200, 1700\}$ GeV $\times$ $m_{\mathrm{VLL}} \in \{300, 700\}$ GeV in MadGraph5_aMC@NLO + Pythia8 + Delphes (HL-LHC Phase-2 card, $\langle\mu\rangle = 200$), define orthogonal 4ℓ Drell-Yan and $3\ell{+}\nu$ cascade signal regions, and combine them in a `pyhf` profile-likelihood fit at $\sqrt{s} = 14$ TeV, $3\,\mathrm{ab}^{-1}$. The combined 95% CL exclusion at $m_{\mathrm{VLL}} = 300$ GeV reaches $m_{H^{\pm\pm}} = 916.70$ GeV versus the 4ℓ-only standalone reach of 917.18 GeV — an extension of -0.48 GeV, falsifying the pre-registered ≥ 150 GeV success threshold by 150.6 GeV. The mechanism is transparent: the cascade $\sigma \times \mathrm{BR}$ is suppressed by a factor ≈ 5500 relative to the Drell-Yan channel at $m_{H^{\pm\pm}} = 500$ GeV / $m_{\mathrm{VLL}} = 300$ GeV (0.000815 fb vs 4.509 fb) , dominated by the $\mathrm{BR}(N_1 \to Z\ell) \cdot \mathrm{BR}(Z \to \ell\ell) \approx 0.027$ multi-step chain (factor ~ 125) compounded by a ~ 45-fold near-threshold 4-body phase-space squeeze. The result is a publishable, mechanism-level null that reorients future Type-II $\times$ VLL phenomenology toward shorter-chain cascades that avoid the Z-pole, larger v_Δ where the 4ℓ channel is BR-diluted, and gauge-bridged non-Yukawa cascades requiring non-degenerate triplet components. The reported background totals include WZ, ZZ and $t\bar{t}Z$ at full HL-LHC normalization (the $t\bar{t}Z$ sample landed via Magnus job `f81a5ff26a00cfdf`, $\sigma_{t\bar{t}Z} = 2.222$ fb, 5×10^5 Delphes events; see Section VII B) and the b-veto in $\mathrm{SR}_{3\ell+\nu}$ together with the (planned) b-veto in $\mathrm{SR}_{4\ell}$ bound the resulting per-SR $S/\sqrt{B}$ impact at $< 2\%$, an order of magnitude below the $\sim 5500\times$ cascade-vs-DY $\sigma \times \mathrm{BR}$ gap that drives the verdict.

I. INTRODUCTION

The Type-II seesaw mechanism is the textbook minimal extension of the Standard Model that generates Majorana neutrino masses through a Yukawa coupling to a hypercharge-one $SU(2)_L$-triplet scalar Δ. The triplet's doubly-charged component $H^{\pm\pm}$ is a prized collider target because its pair-production proceeds at hadron colliders through a clean Drell-Yan diagram and, in the prompt-leptonic regime $v_\Delta \lesssim 10^{-4}$ GeV [2], decays essentially with $\mathrm{BR}(H^{\pm\pm} \to \ell^\pm\ell^\pm) \approx 1$ — yielding the spectacular four-lepton final state $pp \to H^{++}H^{--} \to \ell^+\ell^+\ell^-\ell^-$. ATLAS, in its full Run-2 same-sign + multilepton search, set the present best direct bound $m_{H^{\pm\pm}} > 1080$ GeV in the flavor-democratic prompt-leptonic regime [3], and a dedicated 2022 HL-LHC projection from the same selection class extrapolated this to a standalone four-lepton reach of ≈ 1.1 TeV at $3\,\mathrm{ab}^{-1}$ [4].

Independently, vector-like leptons (VLLs) — anomaly-free Dirac partners of the SM lepton doublet — are independently motivated by neutrino-mass UV completions, by the muon $(g-2)$ anomaly, and by their role as natural co-portals in Higgs-LFV [5, 6]. Recent LHC multilepton recasts already constrain a flavorful VLL doublet to $m_{\mathrm{VLL}} \gtrsim 300$ GeV [6], while HL-LHC projections of single-channel VLL pair production reach $m_{\mathrm{VLL}} \sim 1$ TeV. The two sectors have been searched for independently — but gauge invariance permits a Yukawa "portal" $y_E\,\overline{L'_L}\,\Delta E_R + \mathrm{h.c.}$ that mixes the two and opens a cross-talk decay $H^\pm \to E^\pm \bar{\nu}_E$ feeding a $3\ell{+}\nu$ final state via the subsequent VLL cascade $E^\pm \to Z\ell^\pm$, $Z \to \ell^+\ell^-$. The 2026 robustness study of Englert, Mitra, Naskar and Saha [1] makes the case explicit: such extensions can shift the minimal-model $H^{\pm\pm}$ exclusion by ± 150 GeV, and the

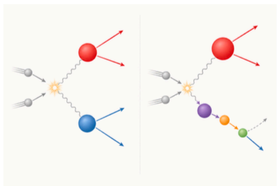


FIG. 1. Conceptual overview of the joint Type-II seesaw triplet Δ + vector-like-lepton doublet $L' = (N, E)$ portal at the HL-LHC: the doubly-charged scalar $H^{\pm\pm}$ decays leptonically (Channel A, 4ℓ Drell–Yan), while the singly-charged $H^\pm$ cascades through the cross-talk Yukawa $y_E\overline{L'_L}\Delta E_R$ into a VLL-mediated $3\ell + \nu$ final state (Channel B).

7

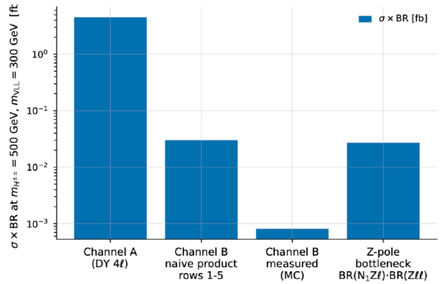


FIG. 7. Decomposition of the $\sigma \times \mathrm{BR}$ gap between Channel A (4ℓ DY) and Channel B ($3\ell + \nu$ cascade) at the most-favoured mass cell $(m_{H^{\pm\pm}}, m_{\mathrm{VLL}}) = (500, 300)$ GeV. The five bars show, on a log scale, Channel A σ, the naive cascade product (applying the three analytic BR-bottleneck factors), the measured Channel B σ, and the dominant $\mathrm{BR}(N_1 \to Z\ell) \cdot \mathrm{BR}(Z \to \ell\ell) \approx 0.027$ factor isolated. The cumulative suppression is $\approx 5500\times$.

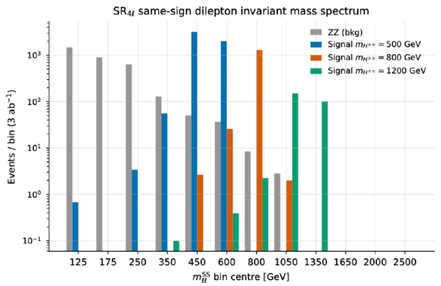
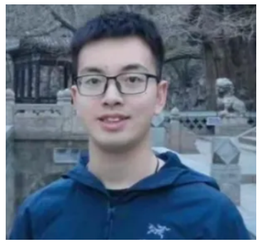
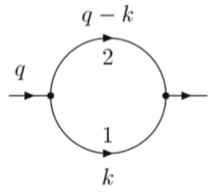


FIG. 8. Stacked-style same-sign dilepton invariant mass distribution in $\mathrm{SR}_{4\ell}$, at HL-LHC integrated luminosity $3\ \mathrm{ab}^{-1}$. Three signal cells $(m_{H^{\pm\pm}}, m_{\mathrm{VLL}}) = (500, 300)$, $(800, 300)$, $(1200, 300)$ GeV are overlaid on the ZZ background (the WZ background is structurally zero in $\mathrm{SR}_{4\ell}$ owing to the $S_T > 600$ GeV and exclusive-4ℓ cuts). The signal peaks at $m_{\ell\ell}^{\mathrm{SS}} \approx m_{H^{\pm\pm}}$ as expected.



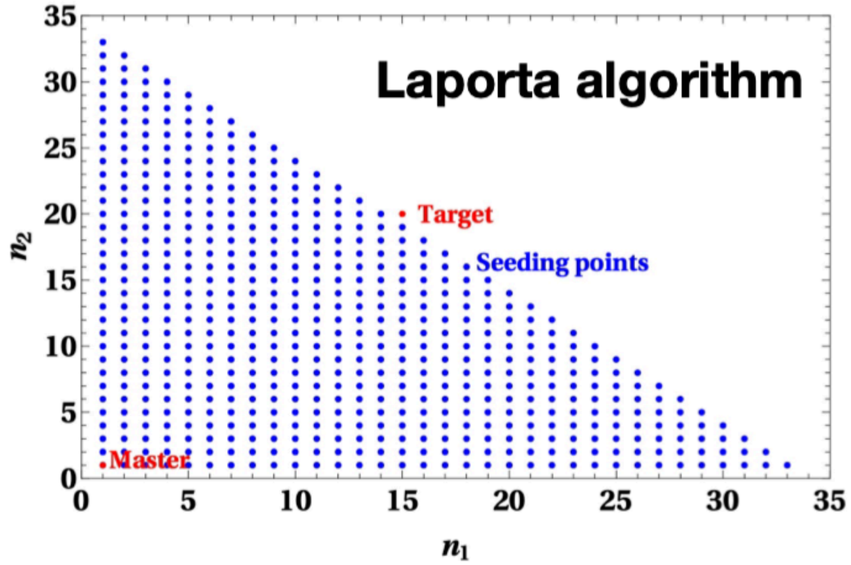
Work in progress

Language Model as a mutation operator

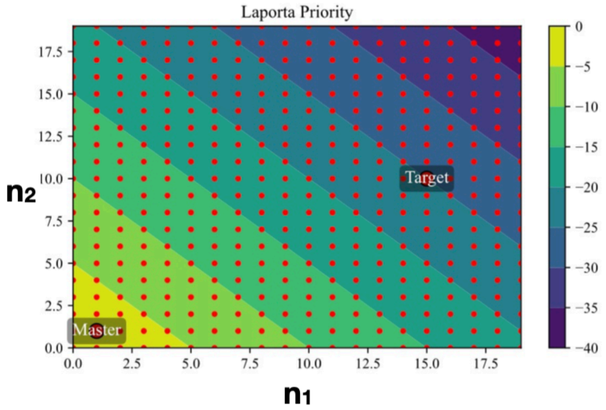


$$I(n_1,n_2)=\int \frac{d^dk}{(2\pi)^d}\frac{1}{(k^2)^{n_1}((q-k)^2)^{n_2}}$$

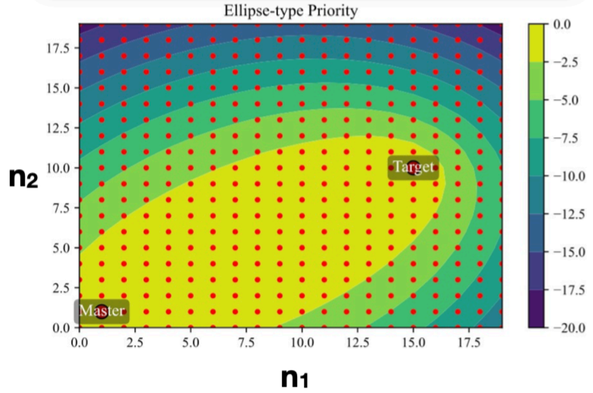
$$0=(d-2n_1-n_2)\,I(n_1,n_2)-n_2 I(n_1-1,n_2+1)+n_2 I(n_1,n_2+1)$$
$$0=n_1 I(n_1+1,n_2-1)-n_1 I(n_1+1,n_2)-n_2 I(n_1-1,n_2+1)$$
$$+n_2 I(n_1,n_2+1)+(n_2-n_1)\,I(n_1,n_2)$$




```
# Epoch 9: Score = 152.2
# Code
import numpy as np
def priority(node: tuple[int, int],
    node_target: tuple[int, int]) ->
    float:
     return -np.sum(node)
```




```
# Epoch 4775: Score = 188.4
# Code
import numpy as np
def priority(node: tuple[int, int],
    node_target: tuple[int, int]) ->
    float:
     x, y = node
     a, b = node_target
     d = ((x - a) ** 2 + (y - b) ** 2) **
         0.5
     p = (x ** 2 + y ** 2) ** 0.5 - (a **
         2 + b ** 2) ** 0.5
     return - (d + p)
```



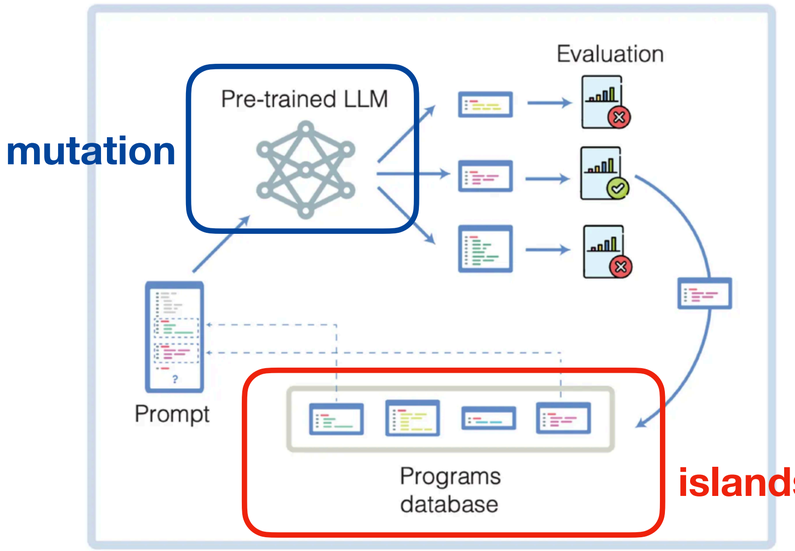
- Rediscover classic Laporta algorithm in 9 Epoch
- An Ellipse-type algorithm achieve 1000 times improvement

LLM as a mutation operator in evolution algorithm

Zhuo-Yang Song, T.Z. Yang, Q.H. Cao, M.X. Luo, HXZ, JHEP accepted

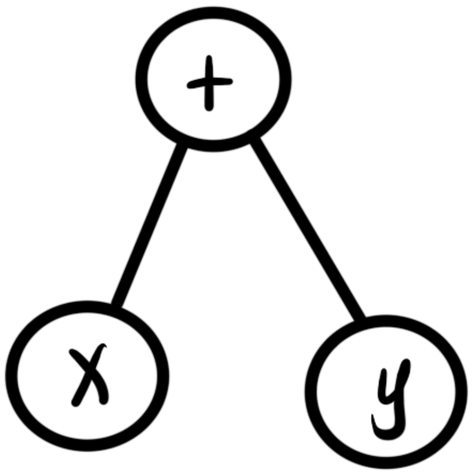
LLM-Driven Symbolic Regression

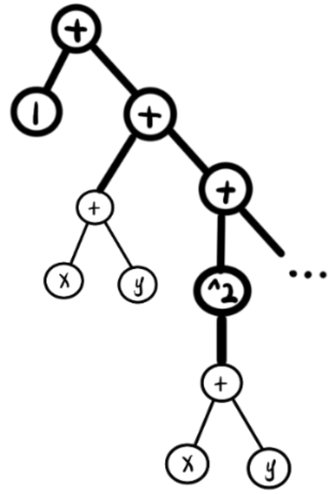
DeepMind, Nature, 2024 FunSearch



LLM prompt and response:
An example is that: [x+y], now I need to find out a better expression to fit the data: I think a good attempt is that `exp(x+y)`, but I find there is no `exp` operator, so I expand the expression to the NNLO and thus my result is \n
[1+(x+y)+(x+y)2/2]**

$$x + y \longrightarrow 1 + (x + y) + \frac{(x + y)^2}{2!} + \cdots$$



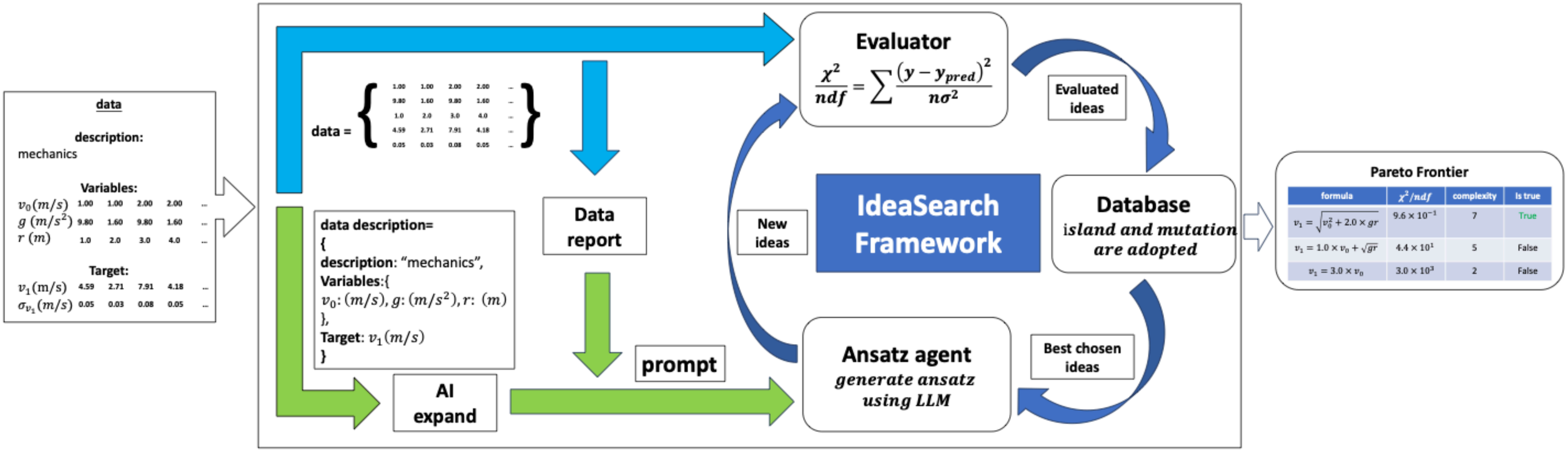


- Combine genetic programming and LLM mutation
- Explore with state + conceptual space
- Directional exploration within an finite space of logic

related works: 2404.18400, 2503.07994, 2505.07956, 2510.08317

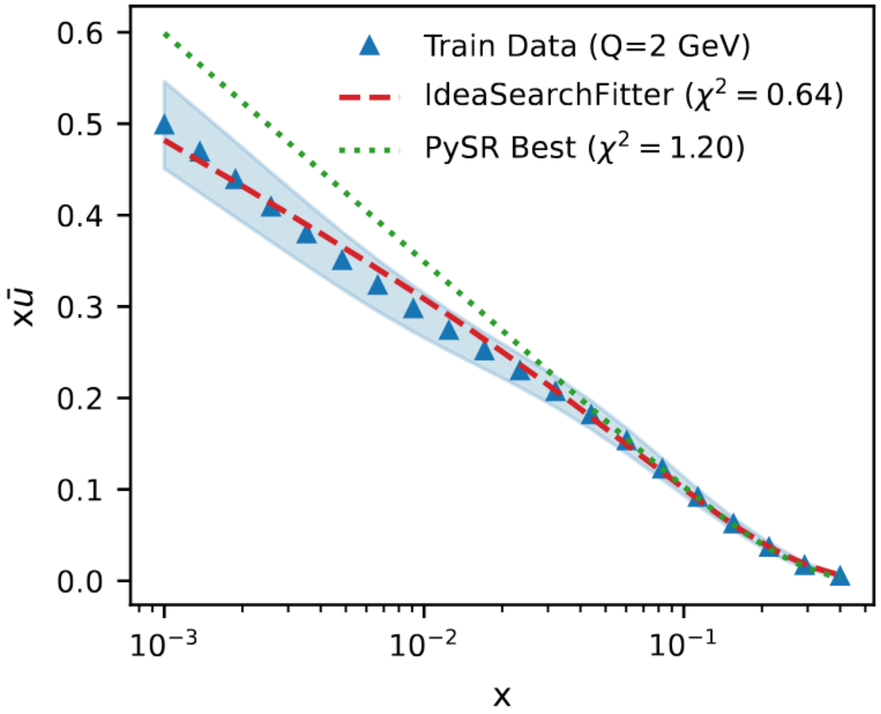
IdeaSearch.Fitter

https://www.ideasearch.cn/
2510.08317

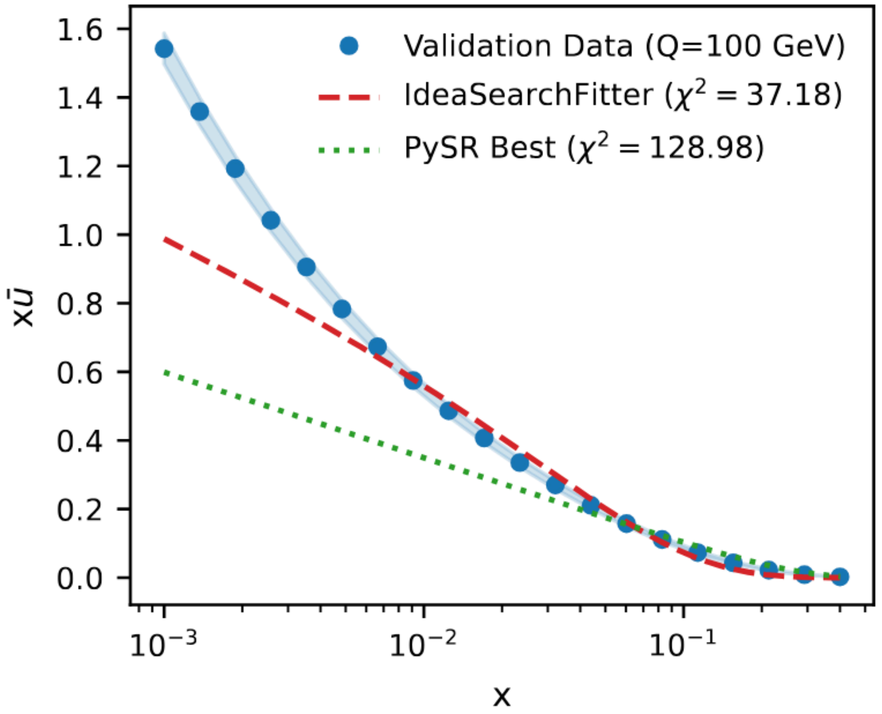


Workflow for Symbolic Regression

Application on PDF fitting



(a) $\bar{u}$ at $Q = 2\,\mathrm{GeV}$



(b) $\bar{u}$ at $Q = 100\,\mathrm{GeV}$

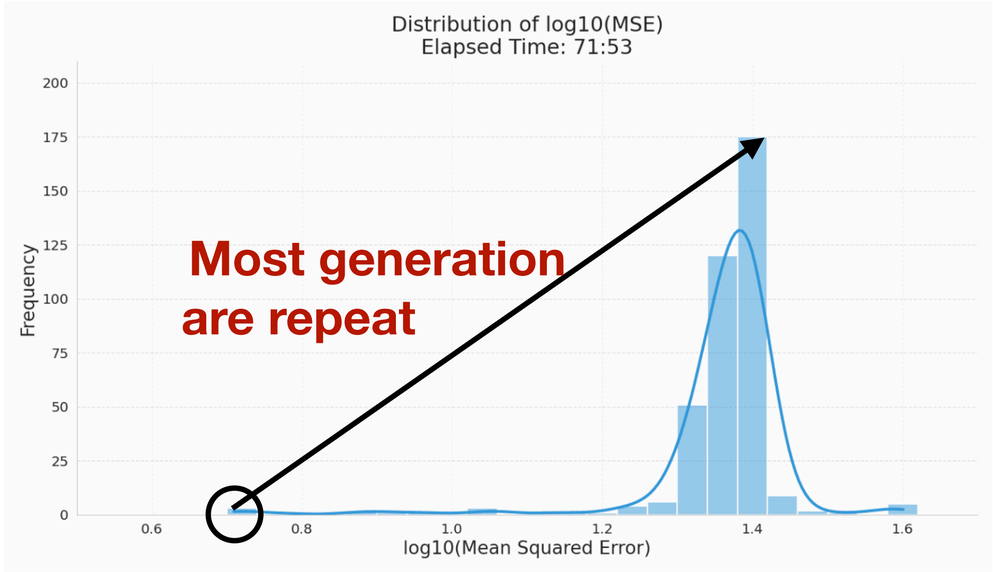
$$x f_{\bar{u}}(x, Q) = p_1 \log\left(1 + p_2 \log(Q)\right) \log\left(\frac{1}{x}\right) (1 - x)^{p_4 + p_5 \log Q}$$

https://www.ideasearch.cn/

2510.08317

Problem of limited exploration

- LLMs-agent explore in conceptual space
- However, LLMs exploration is directional; insufficient exploration
- Example: in Symbolic Regression, we observe repetitive sampling within the training set
- We need to develop a theory for how LLM explore its conceptual space

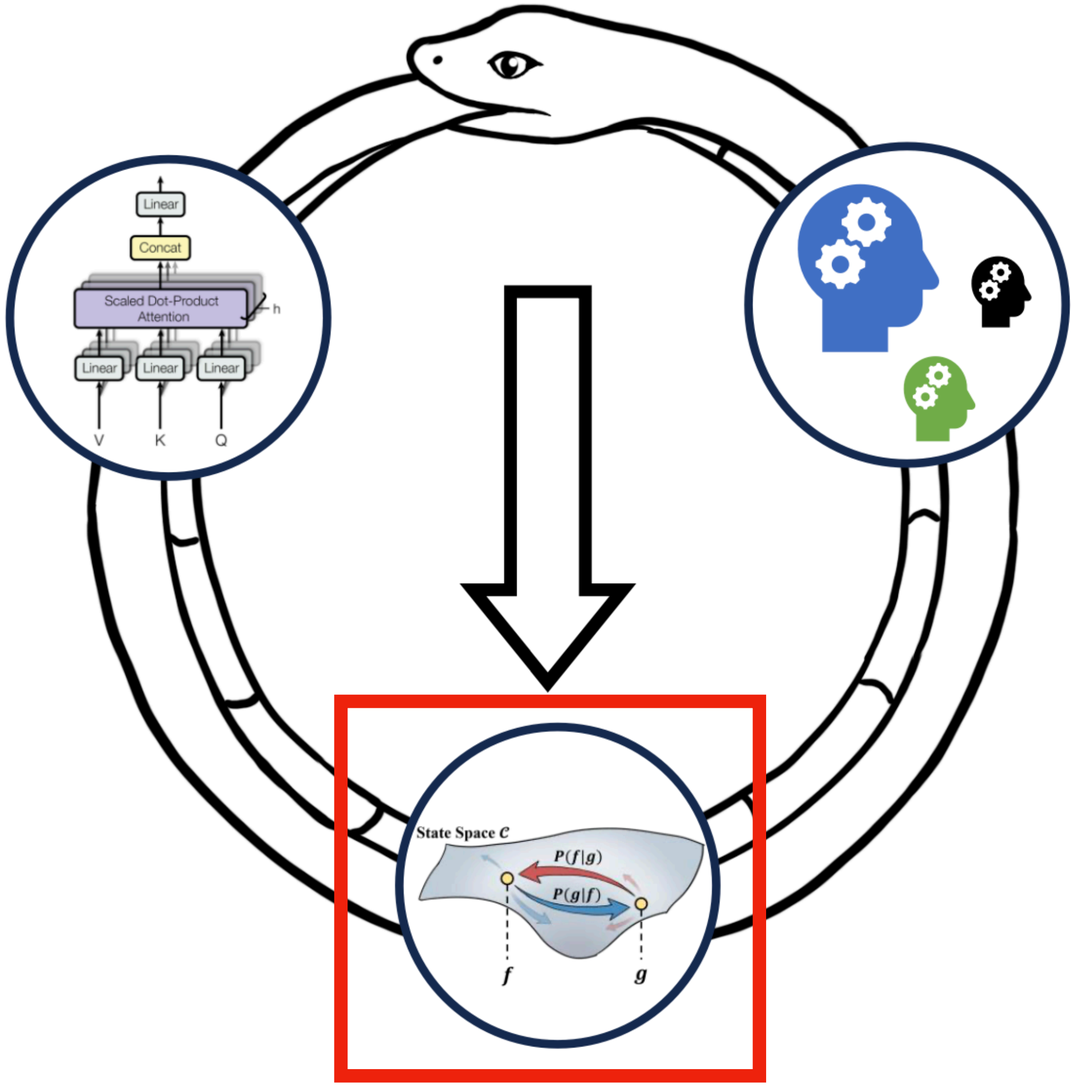


High score but low potential

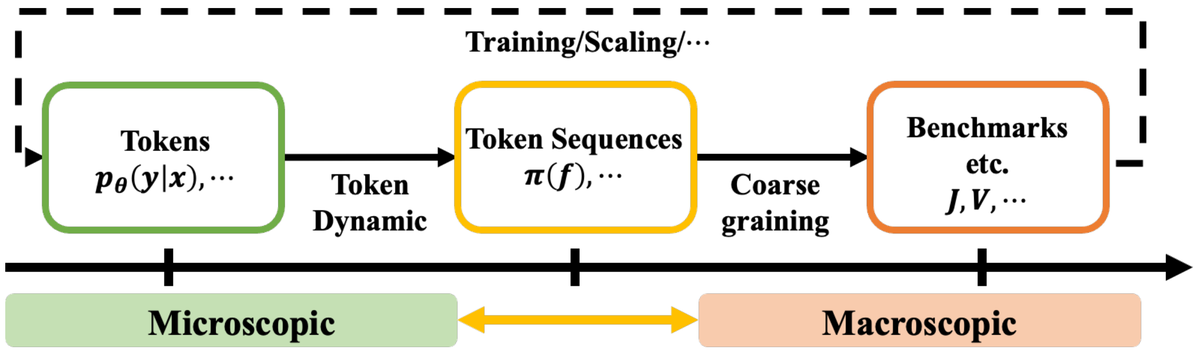
我们发现，AlphaEvolve 擅长发现那些在当前数学框架内本已可及、却尚未被发现的构造。这类构造之所以未被找到，往往是因为需要耗费大量时间与精力去尝试各种标准思想的恰当组合，以适用于特定问题。另一方面，对于那些需要真正新颖且深刻的洞见才能取得进展的问题，AlphaEvolve 很可能并非合适的工具。

Terrence Tao et al.

Towards a physicist's view of artificial intelligence

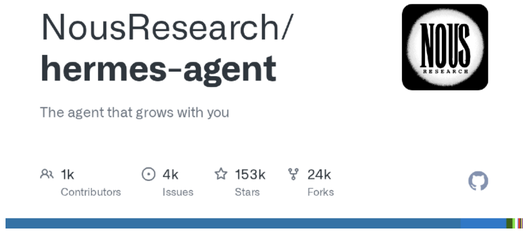


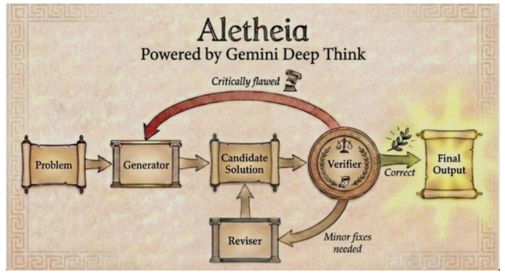
- Microscopic: token dynamics
- Macroscopic: benchmark in general sense
- **mesoscopic: ensemble/states**

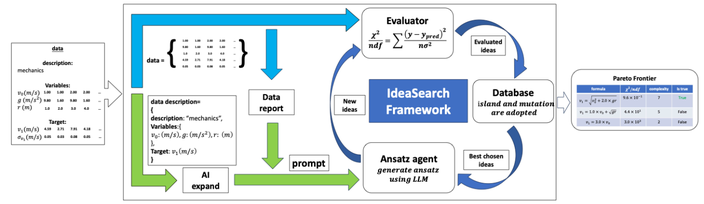


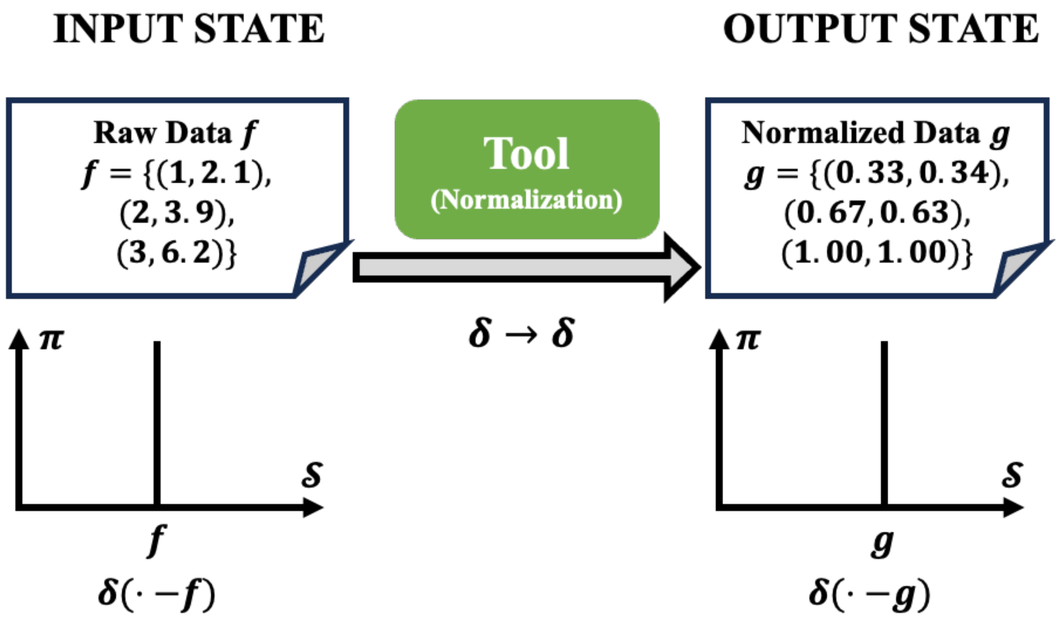
Undergraduate thesis:
https://sonnynondegeneracy.github.io/

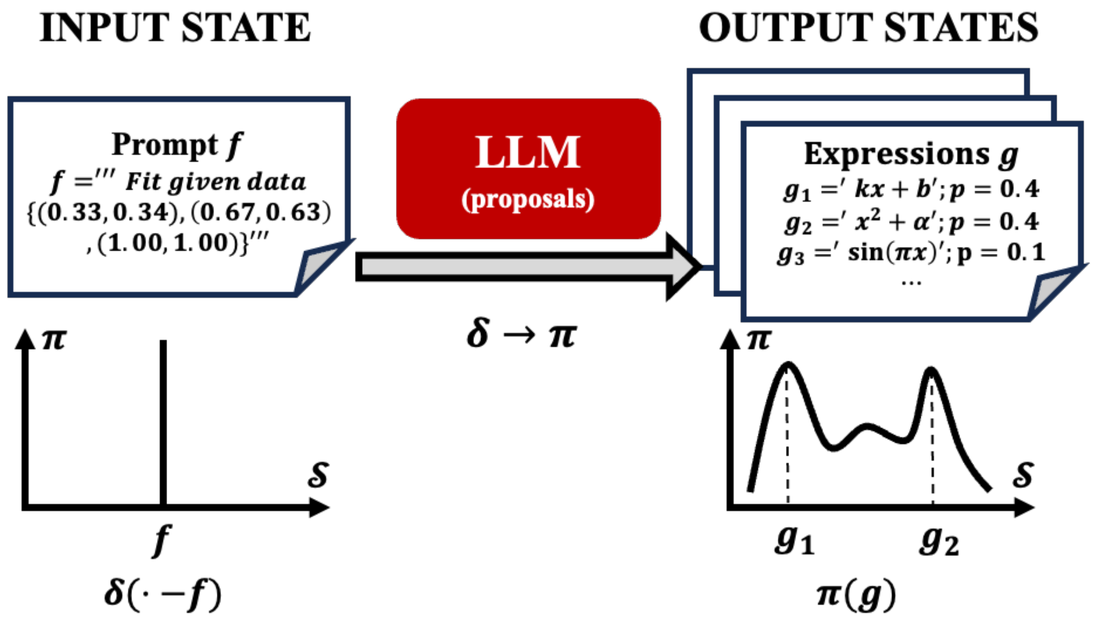
A mathematical framework for agent







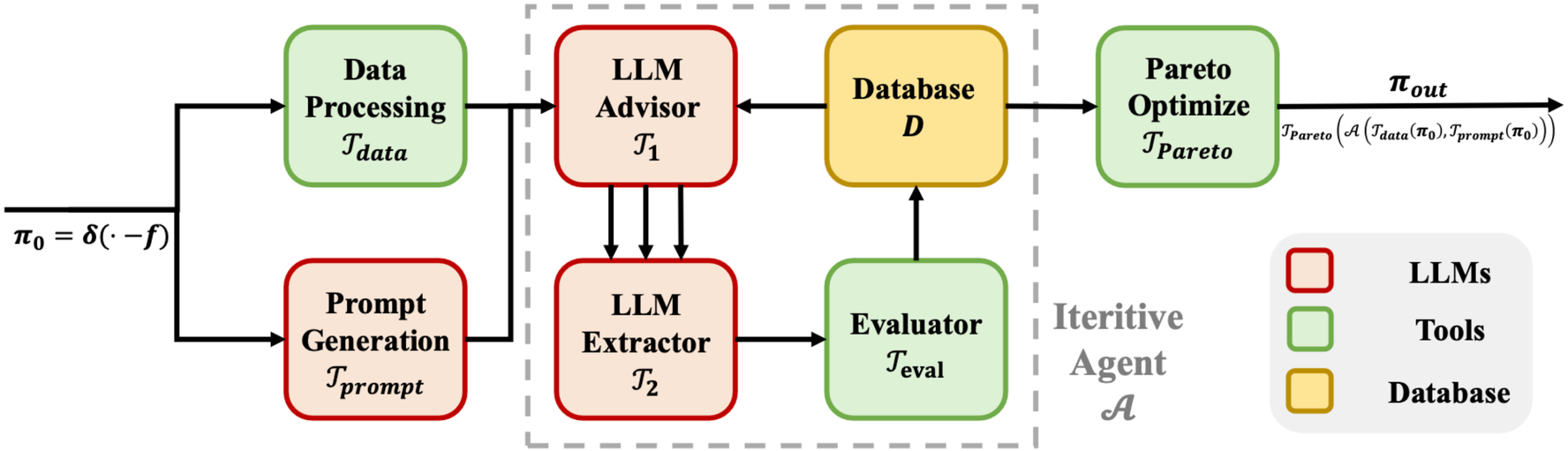




https://sonnynondegeneracy.github.io/

Operator diagram for agent

An operator diagram example for IdeaSearch.Fitter



$$\mathrm{Gen}(D) = \mathcal{T}_{\mathrm{eval}} \circ \mathcal{T}_2 \circ \mathcal{T}_1 \left(\mathrm{Sel}(D); \mathcal{T}_{\mathrm{data}}\ (\pi_0), \mathcal{T}_{\mathrm{prompt}}\ (\pi_0)\right)$$

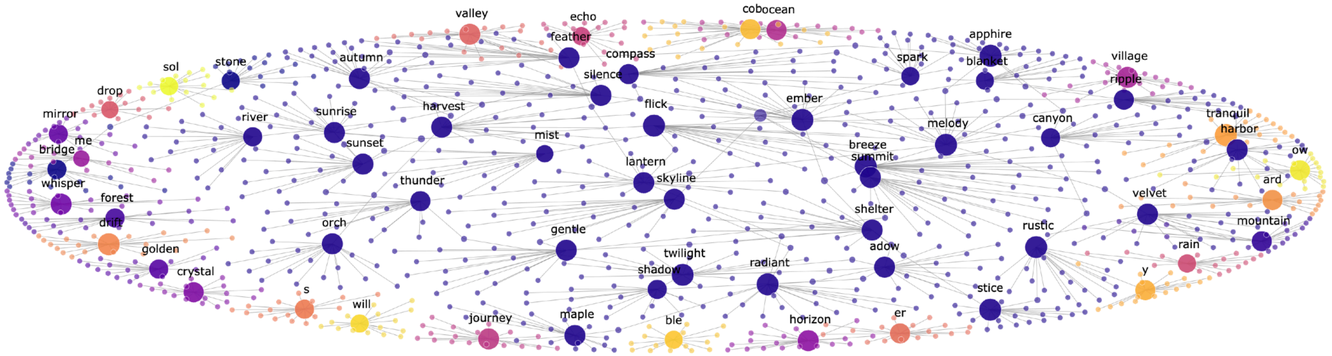
recursive refinement

$$\begin{aligned} D_1 &= \mathrm{Gen}\left(D_0\right) \\ D_2 &= \mathrm{Gen}\left(D_0, D_1\right) \\ D_3 &= \mathrm{Gen}\left(D_0, D_1, D_2\right) \end{aligned}$$

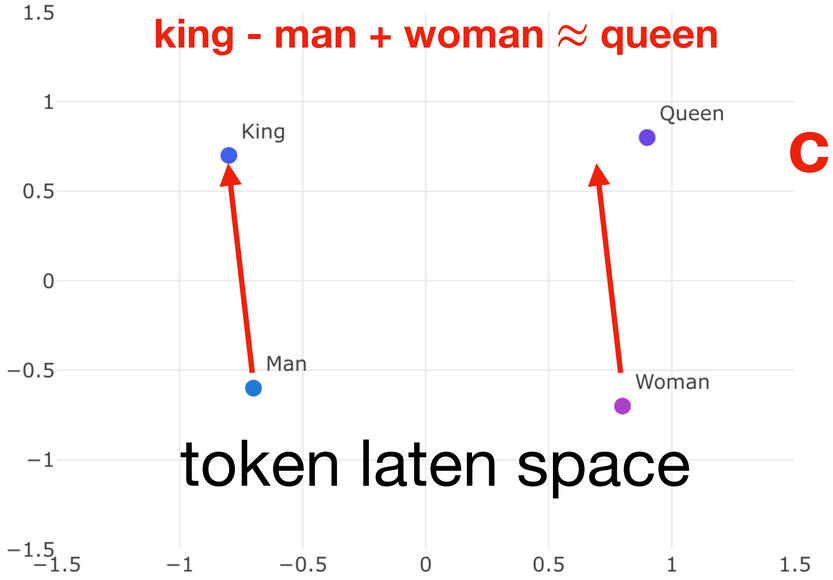
$$\pi_{\mathrm{out}} = \mathcal{T}_{\mathrm{Pareto}}\ \left(\mathcal{A}\left(\mathcal{T}_{\mathrm{data}}\ (\pi_0), \mathcal{T}_{\mathrm{prompt}}\ (\pi_0)\right)\right)$$

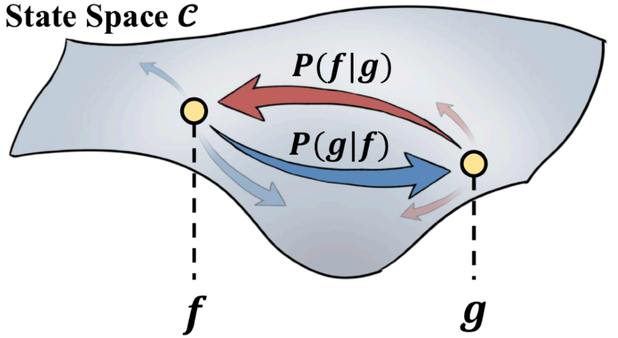
https://sonnynondegeneracy.github.io/

Space of exploration: from token to concept



Source: Huggingface

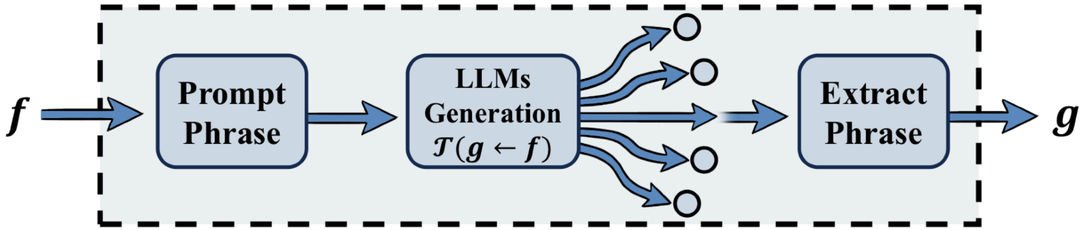




N0: total # of sample

N(g<-f): # of g generated

$$\mathcal{T}(g \leftarrow f) \approx \frac{N(g \leftarrow f)}{N_0(f)}$$



corse-graining

token → concept

```
param1 * tanh(param2 * x + param3) + param4
param1 - (param2 / (x + param3))
param1 * x / (1 + param2 * log(x + 1))
param1 * tanh(param2 * x) + param3
param2 + param1 * (1 - exp(-x))
```

example from Symbolic Regression

Zhuo-Yang Song, Q.H. Cao, M.X. Luo, HXZ, 2512.10047

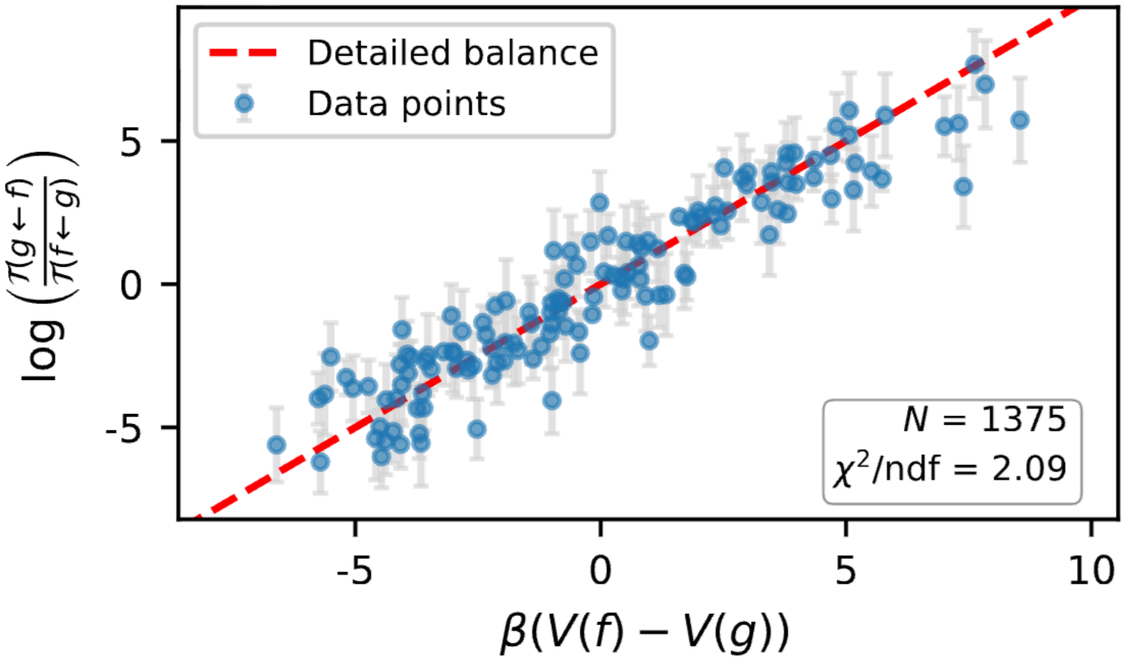
Are LLM-agent near equilibrium system?

- We discover there exist a scaling law in LLM exploration
- We interpret it as a consequence of detailed balance

$$\pi(g)\mathcal{T}(f \leftarrow g) = \pi(f)\mathcal{T}(g \leftarrow f)$$

$$\log \frac{\mathcal{T}(g \leftarrow f)}{\mathcal{T}(f \leftarrow g)} = \beta V(f) - \beta V(g)$$

$$\mathcal{T}(f \leftarrow g) = \mathcal{T}(g \leftarrow f) e^{-\beta(V_{\mathcal{T}}(f) - V_{\mathcal{T}}(g))}$$



Measured in the example of symbolic regression

```
param1 * tanh(param2 * x + param3) + param4
param1 - (param2 / (x + param3))
param1 * x / (1 + param2 * log(x + 1))
param1 * tanh(param2 * x) + param3
param2 + param1 * (1 - exp(-x))
```

- Can we apply more tools from equilibrium physics to study LLM agent?

Zhuo-Yang Song, Q.H. Cao, M.X. Luo, HXZ, 2512.10047

Summary

- LLM/Agents provide rich opportunities for PHY x AI
- Dynamics at microscopic token level: manifold hypothesis and manifestation in generation dynamics
- Macroscopic characterization of LLM/Agents capability:
  - Benchmark
  - LLM/Agents as a new kind of scientific methods

  } alignment

- Better engineering realization calls for better understanding of the blackbox
  - **mesoscopic: state space, transition kernel, thermodynamics limit**