

# AI Autonomous Scientific Discovery: From Newton's Second Law to Future Scientific Paradigms

Dong-Shan Jian (Peking University)

[jiandongshan@hotmail.com](mailto:jiandongshan@hotmail.com)

Spring 2026 IHEP ML Workshop, Beijing, 2026/04/15



北京大學



# Outline

---

## **I. Background and motivation**

## **II. From data to model**

- a. Symbolic regression
- b. FunSearch
- c. AI-Newton

## **III. From model to data**

- a. Established field
- b. Frontier field

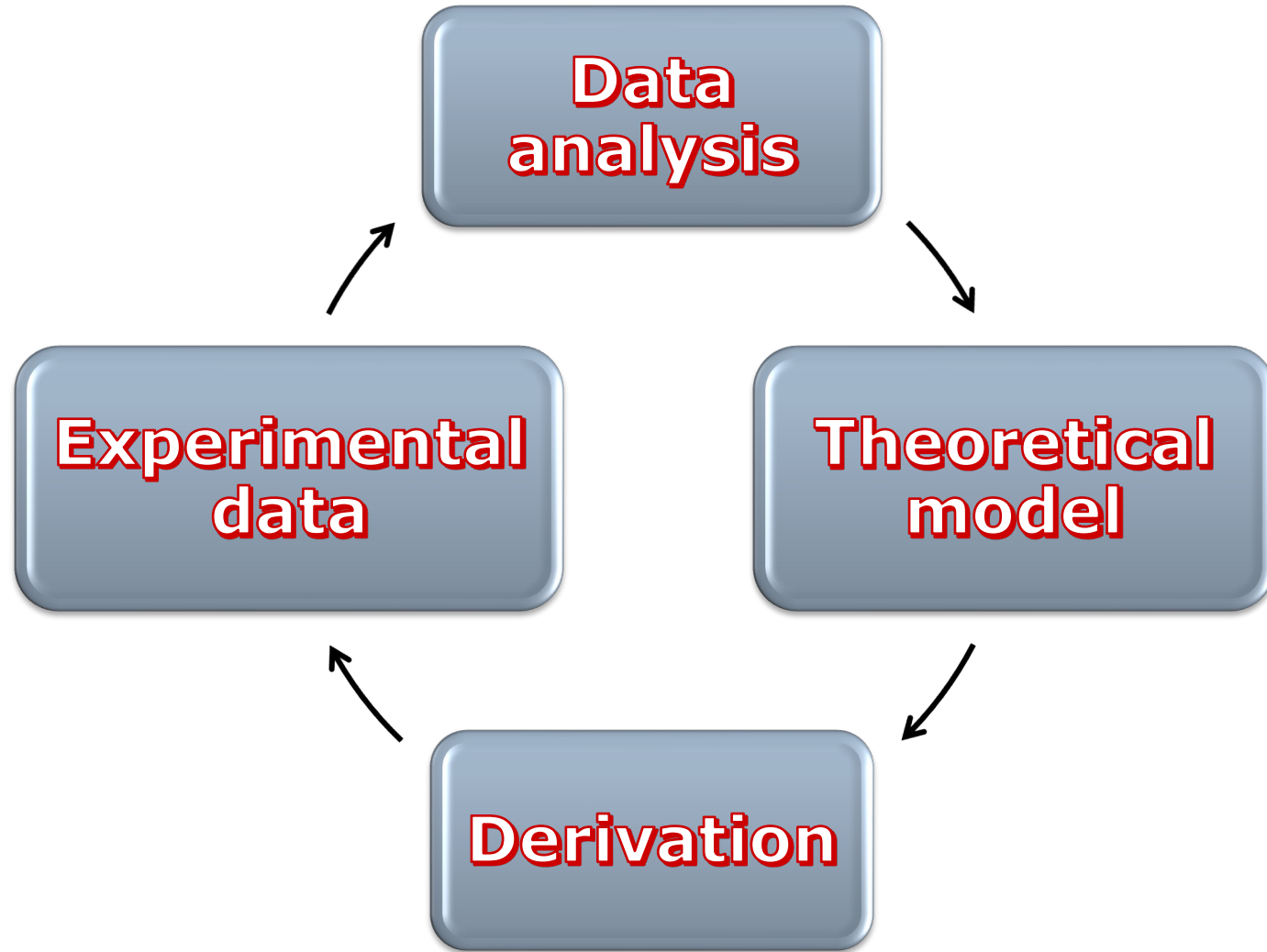
## **IV. Aether: an integrated vibe research platform**

## **V. Summary and outlook**

# The circle of scientific discovery

---

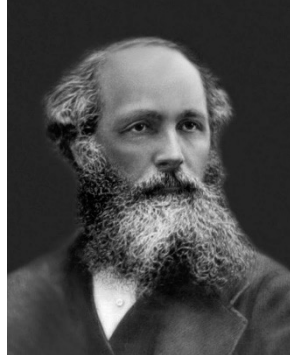
**From data to model**



**From model to data**

# Human scientific discovery

## ➤ Fundamental natural laws: human contributions



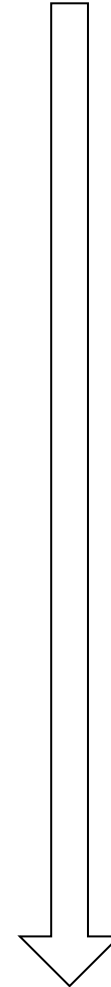
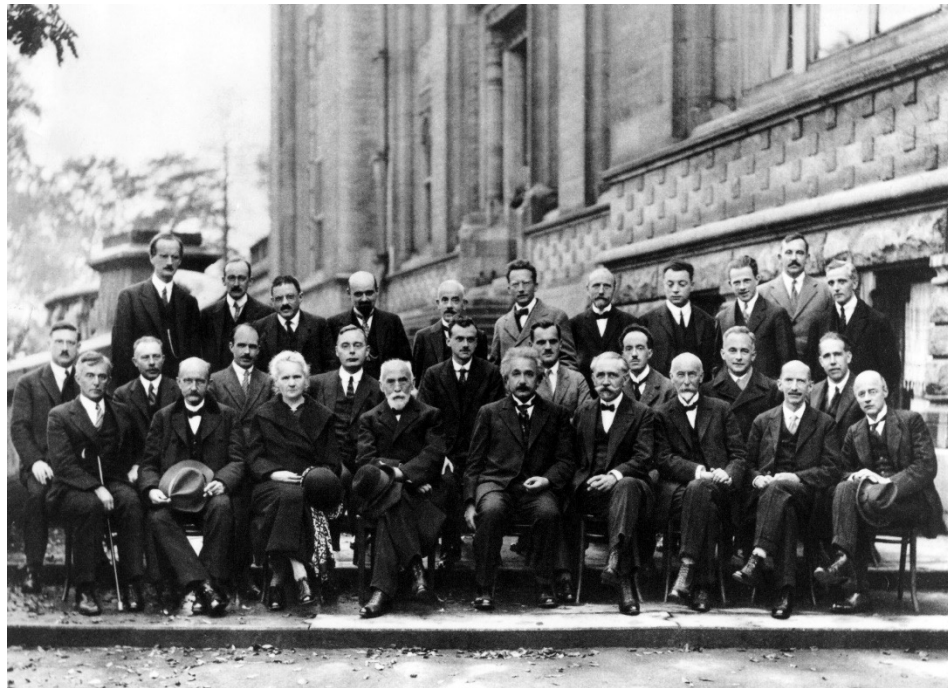
Galileo's  
laws of motion

Newton's  
laws of motion

Maxwell's  
electromagnetic theory

Theory of relativity

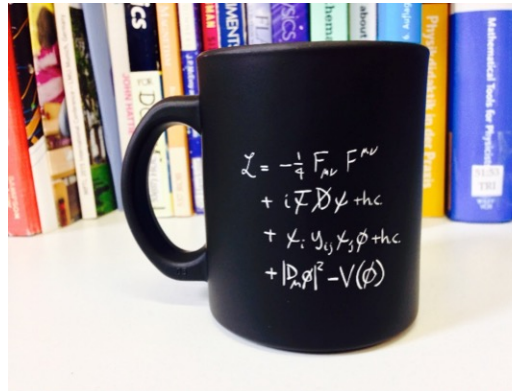
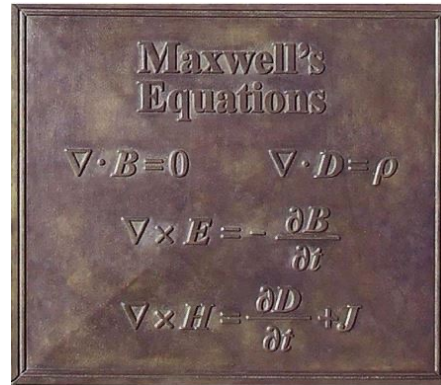
Quantum theory



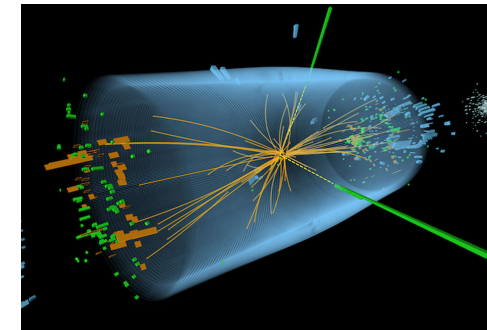
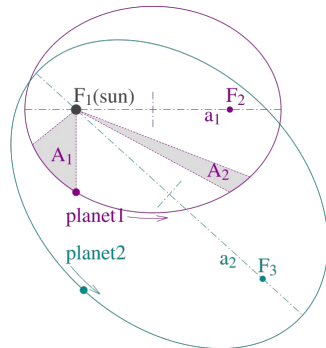
# Reflection

## ➤ Human exploration of natural laws:

- **Advantages:** interpretability, conciseness, **universality**

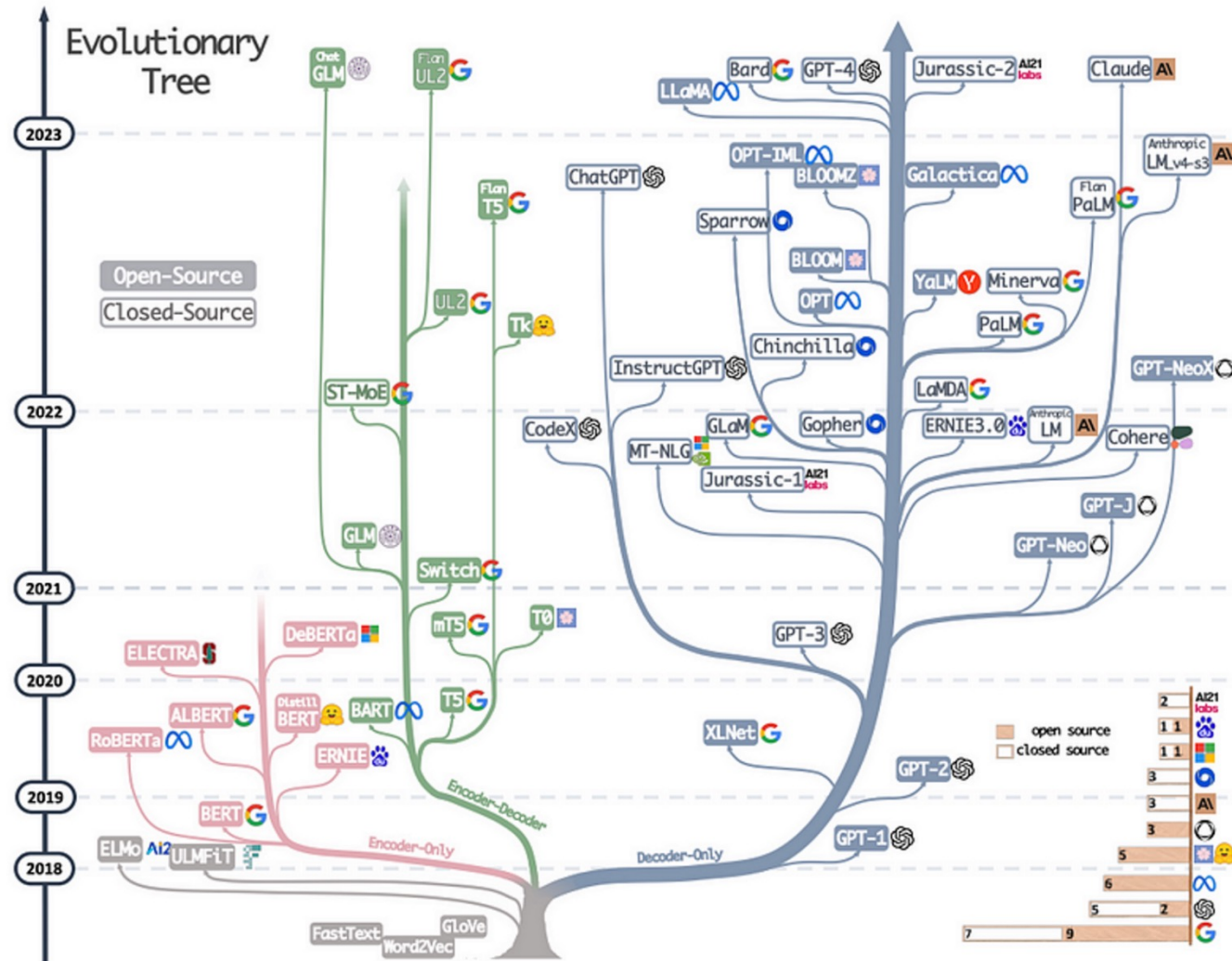


- **Disadvantages:** long period, preconceived notion, insufficient ability to handle complex problems



# Advancement in AI

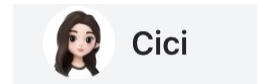
## ➤ AI: rapid advancement in recent years



# Advancement in AI

## ➤ AI (especially LLMs) and their derivative products excel in **general domains**

- Chat client



- Translation



- Coding agent



OpenAI Codex



- Research



# AI-driven scientific discovery

---

Tools/algorithms for assisting in solving specific scientific problems ✓

Continuous and autonomous scientific discovery?

General theories? Scientific insights?

**Still an open issue!**

Reddy and Shojaee, 2412.11427

# Outline

---

## I. Background and motivation

## II. From data to model

### a. Symbolic regression

### b. FunSearch

### c. AI-Newton

## III. From model to data

### a. Established field

### b. Frontier field

## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

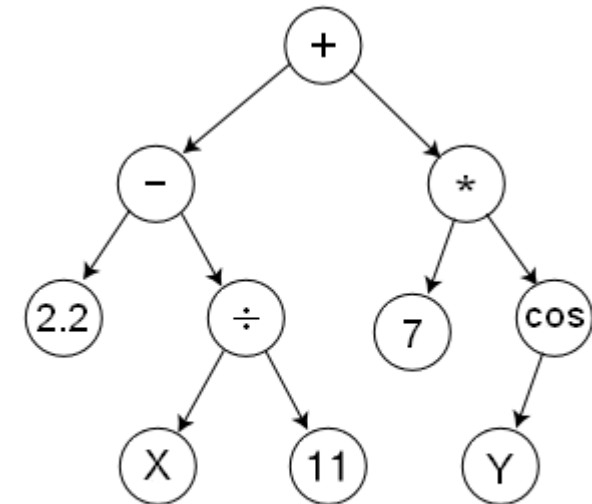
# Symbolic regression

- Conventional regression:  $y = f(x; a_1, a_2, \dots)$  to describe a given dataset
- Symbolic Regression (SR): A functional expression describing variables in a given dataset
- Without prescribing the specific form of the function



Kepler

From Manual to Automated



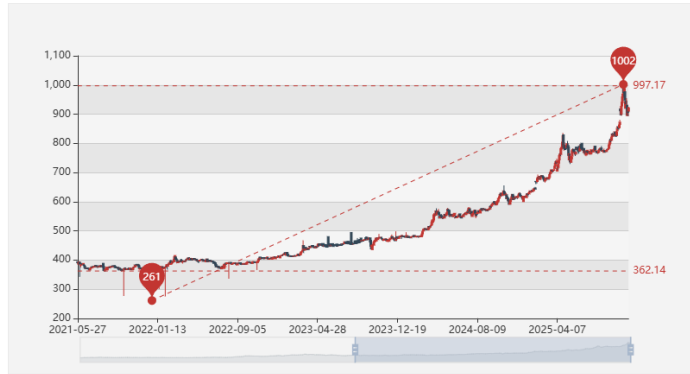
$$\left(2.2 - \left(\frac{X}{11}\right)\right) + (7 * \cos(Y))$$

Genetic programming SR

# Applications of SR

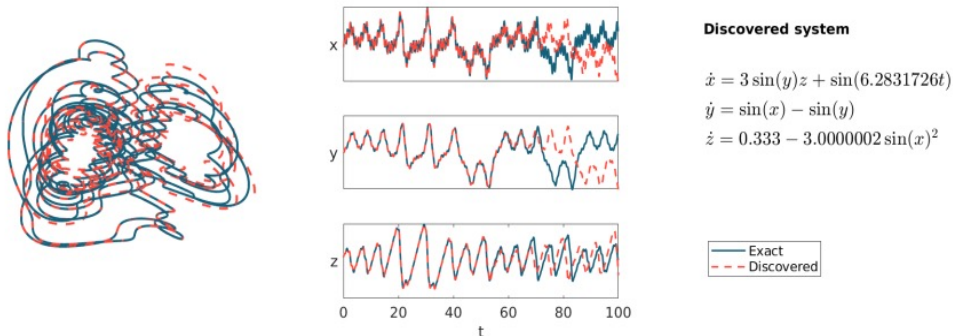
## ➤ Applications in various fields

### Predict prices



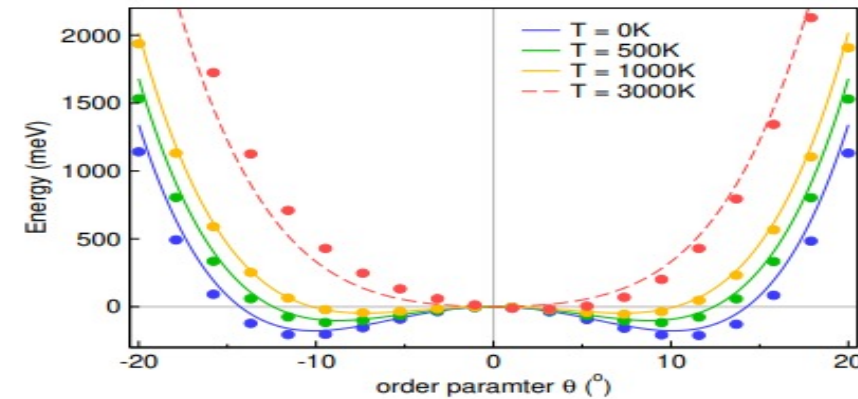
[Forecasting Selected Commodities' Prices with the Bayesian Symbolic Regression](#)

### Nonlinear dynamic equations



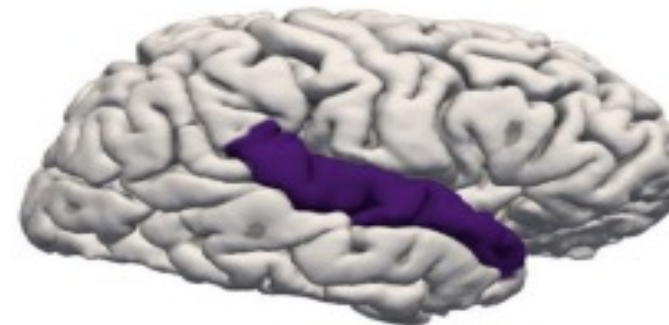
[AI-Lorenz: A physics-data-driven framework for black-box and gray-box identification of chaotic systems with symbolic regression](#)

### Free energy functions



[Symbolic regression in materials science](#)

### Diffusion-reaction model of proteins

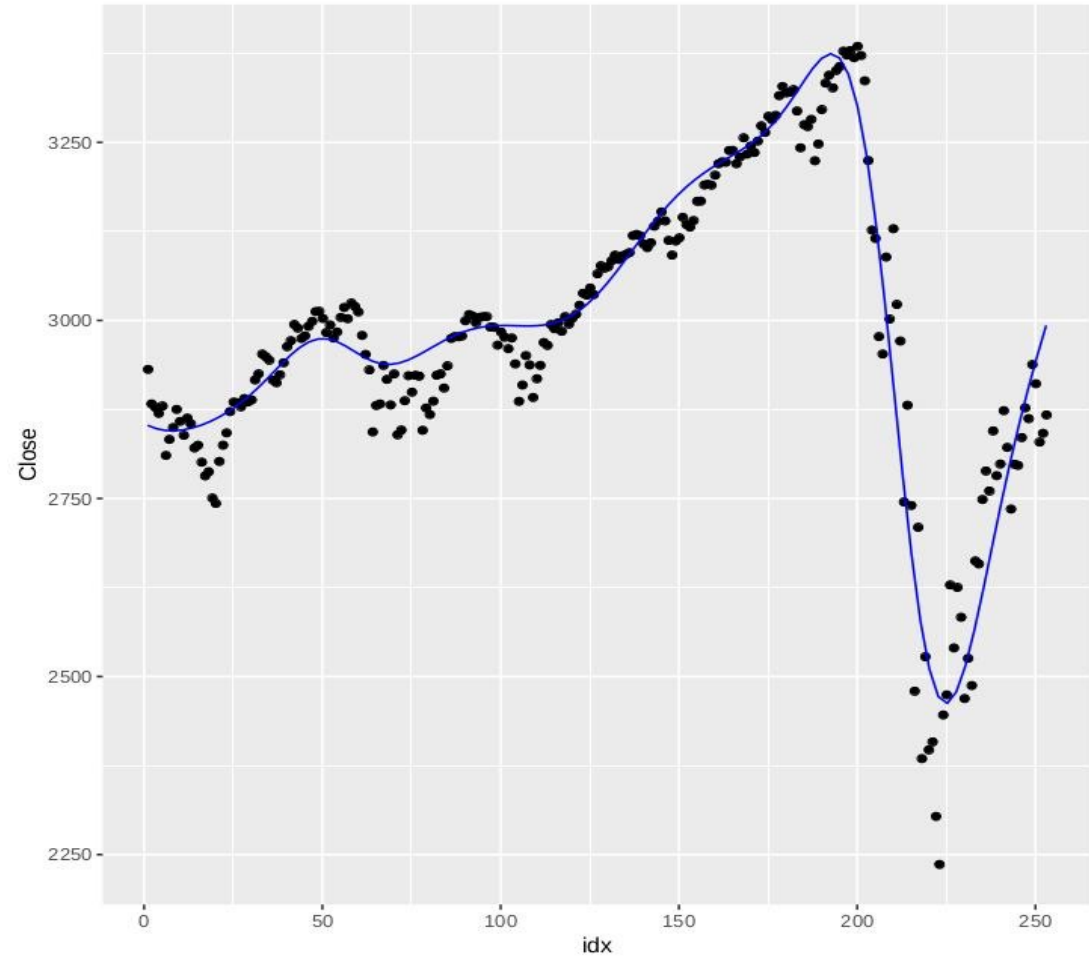


[Discovering a reaction-diffusion model for Alzheimer's disease by combining PINNs with symbolic regression](#)

# SR summary

## ➤ Symbolic regression:

- Automatic (function) discovery
- Interpretability
- Limited expressive capability
- Search space explosion
- Limited cross-problem transferability



# Outline

---

## I. Background and motivation

## II. From data to model

a. Symbolic regression

b. FunSearch

c. AI-Newton

## III. From model to data

a. Established field

b. Frontier field

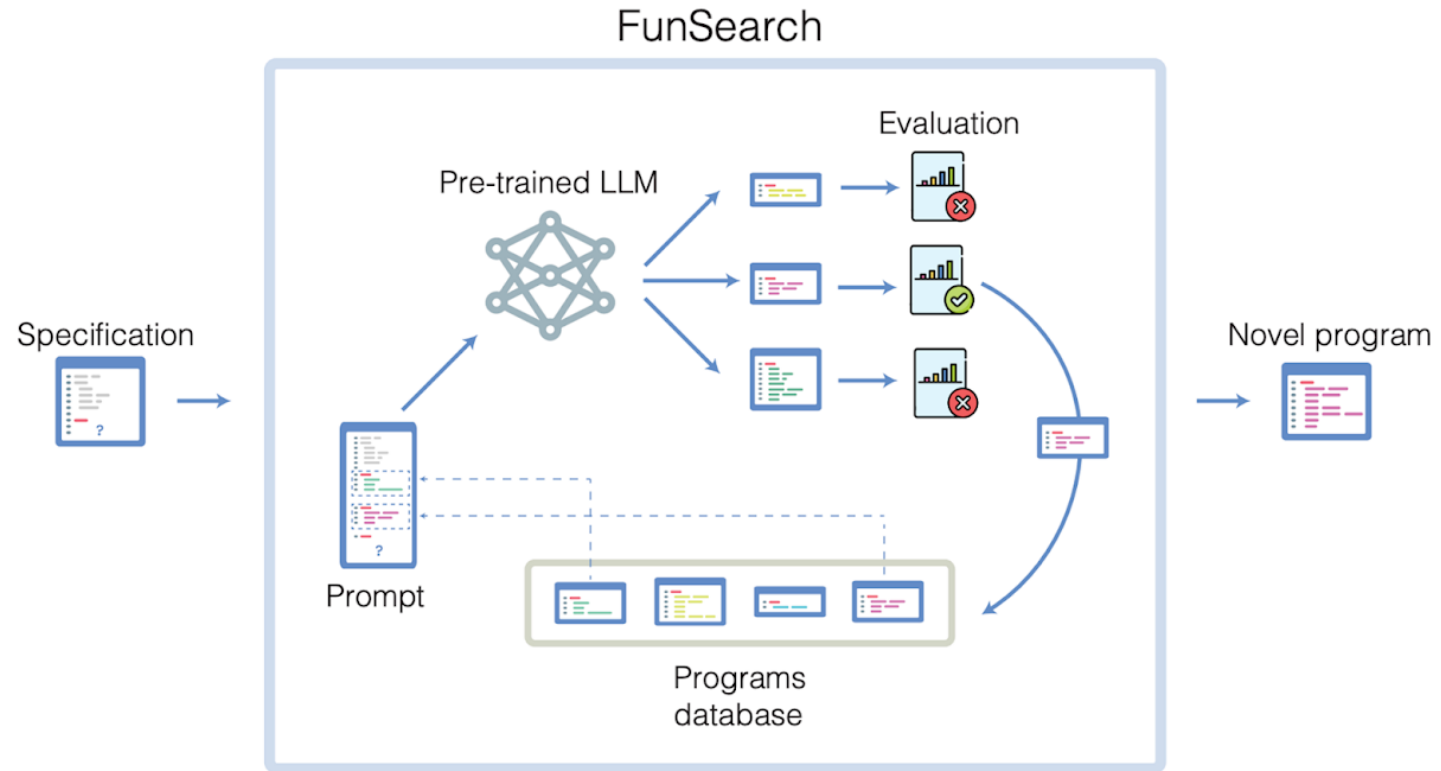
## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

# FunSearch

## ➤ Funsearch: find programs instead of symbolic functions

- FunSearch is an iterative evolutionary procedure
- At each iteration, the LLM is shown a selection of the best programs it has generated so far (retrieved from the program database).
- LLM generates new programs, which are automatically evaluated.
- The best one are added back to the programs database for the next iteration.



Romera-Pareds, et al., Nature 625, 468–475 (2024)

FunSearch = LLM + Evaluator + Evolutionary Search

# FunSearch Summary

---

- Automatic (function) discovery
- Interpretability
- Flexible expressive capability
- Search space explosion
- Limited cross-problem transferability

# A challenge for discovery

---

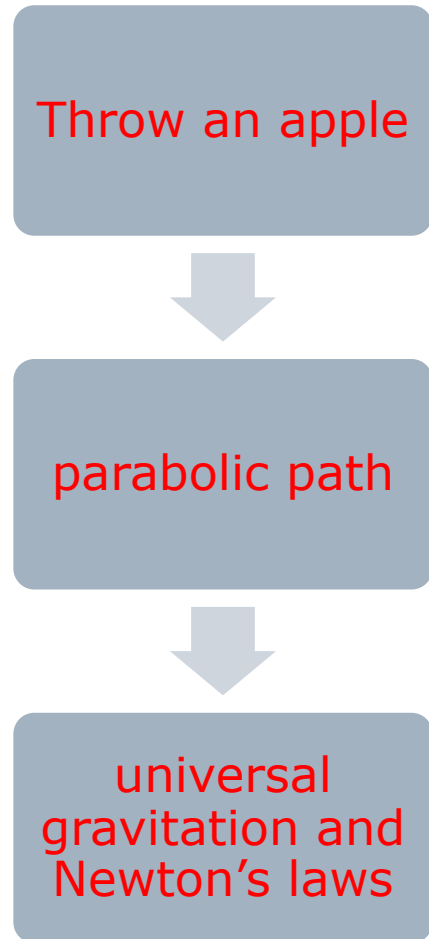
## ➤ Specific model for one experiment

- Explore relations for a set of data
- Symbolic regression, funSearch, ...

## ➤ General model for a large set of experiments

- How to define and explore relations between specific models?

## ➤ Is it possible for AI to reproduce human's theories?



# Outline

---

## I. Background and motivation

## II. From data to model

- a. Symbolic regression
- b. FunSearch
- c. AI-Newton

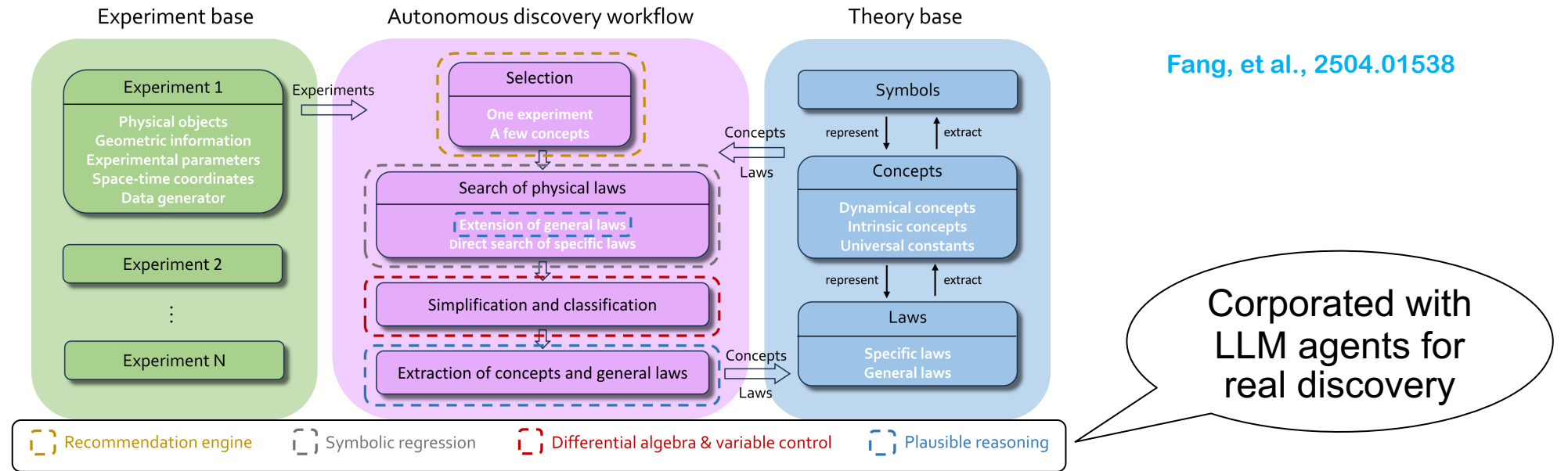
## III. From model to data

- a. Established field
- b. Frontier field

## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

# AI-Newton's architecture

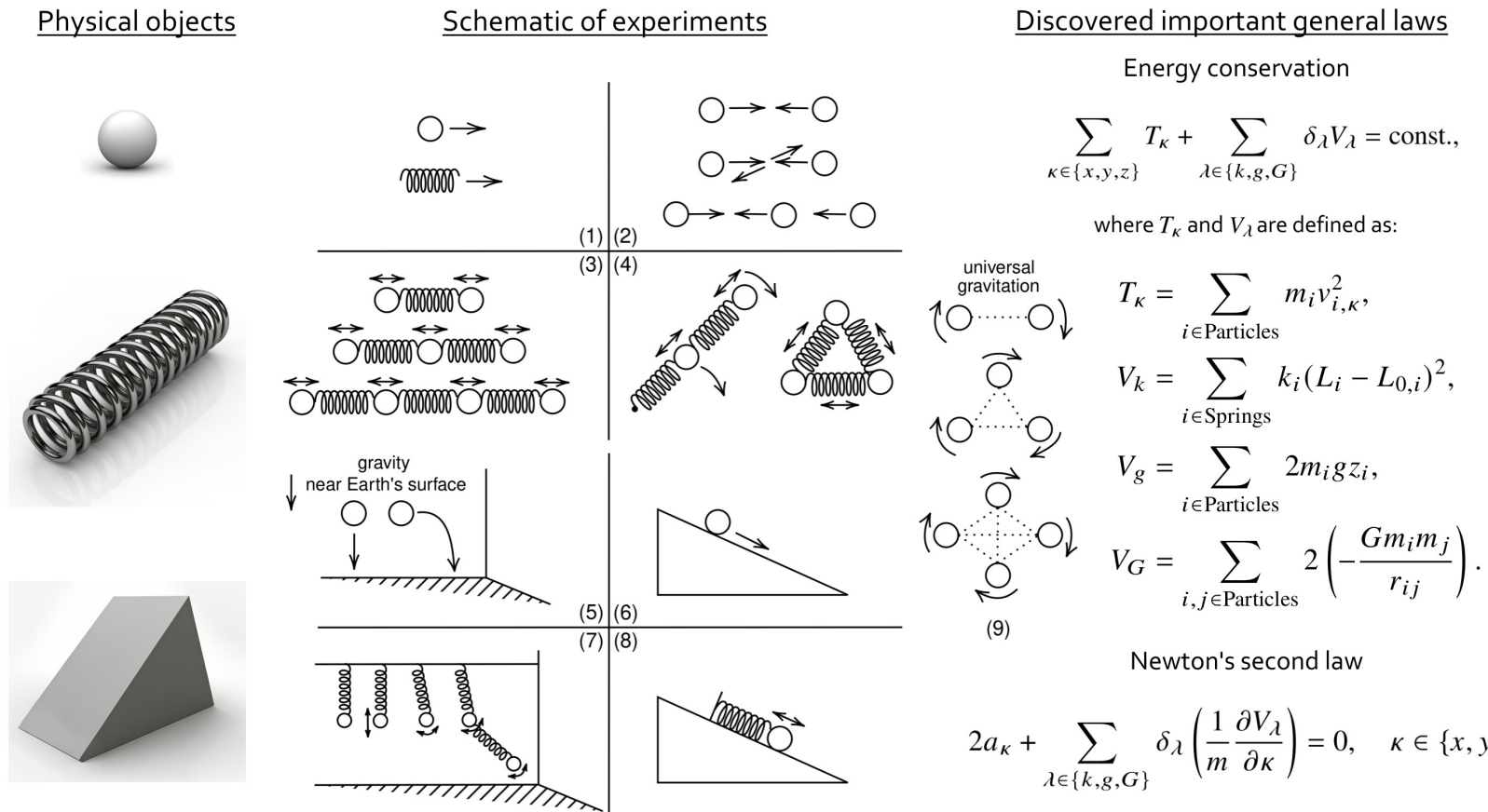


- **Knowledge base (experiment + theory):**  
stores and manages structured knowledge
  - **Knowledge representation:**  
employs a physical domain specific language(DSL)
  - **Autonomous discovery workflow:**  
continuously explores physical laws,  
collaboratively updates both general and specific knowledge
- core: physical concepts
- core: plausible reasoning
1. Effectively represent knowledge
  2. Reduced search space

# Tests and results

- Based on noisy data, important natural laws are discovered!
- Unsupervised! Without prior physical knowledge!

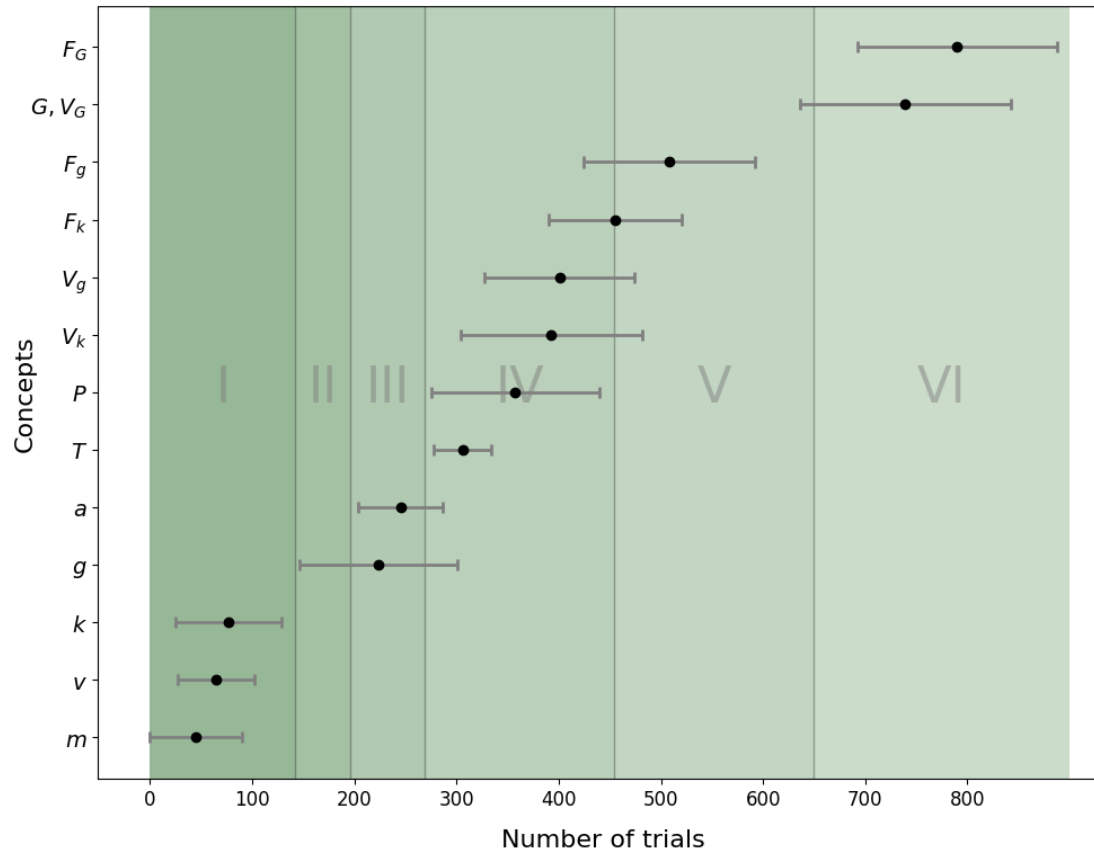
Fang, et al., 2504.01538



( $\delta_{\lambda} = 0$  or  $1$ , determined spontaneously during instantiation as specific laws in experiments)

# Tests and results

## ➤ Statistical analysis of concept discovery timing:



Fang, et al., 2504.01538

(Roman numerals for era numbering)

## ➤ Incremental progression, diversity

# Outline

---

## I. Background and motivation

## II. From data to model

- a. Symbolic Regression
- b. Funsearch
- c. AI-Newton

## III. From model to data

- a. Established field
- b. Frontier field

## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

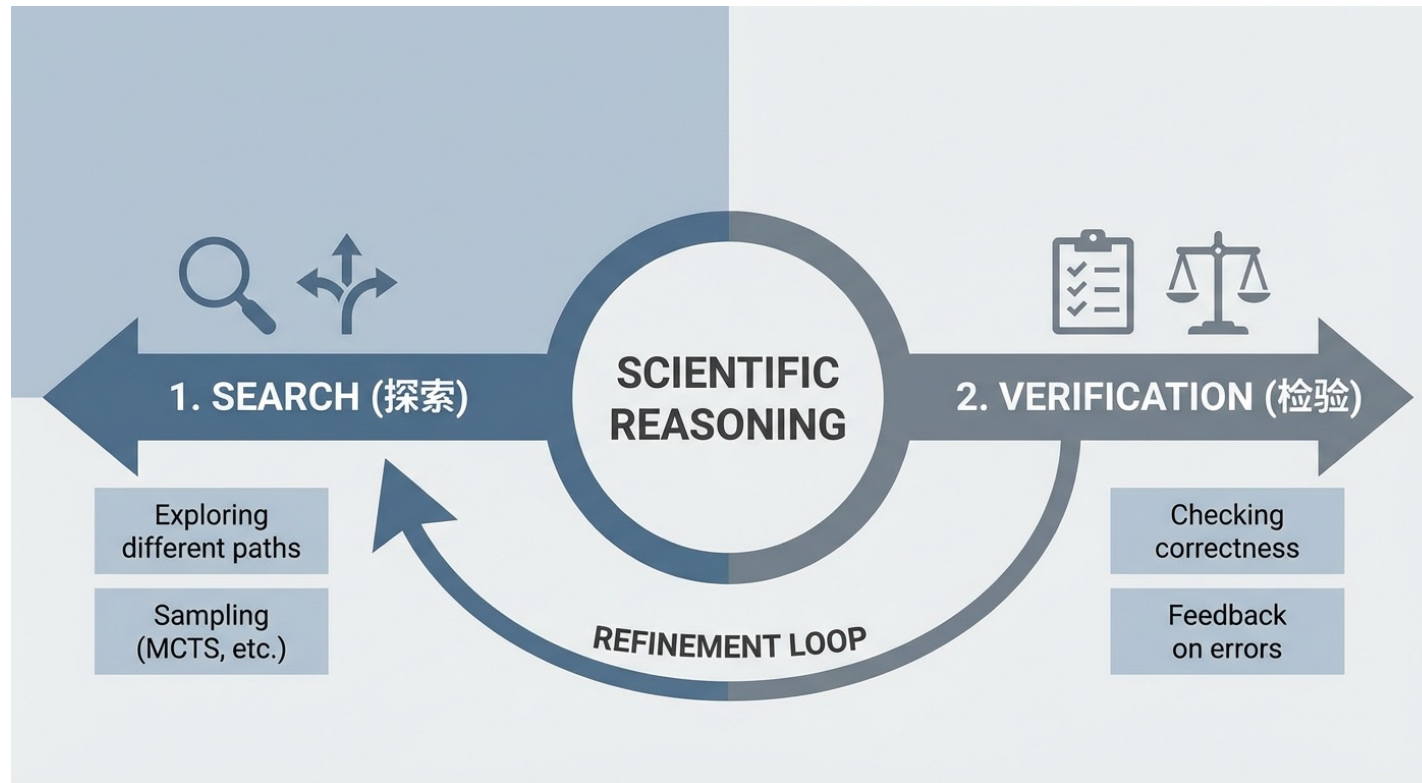
# Challenges for scientific AI

---

- LLMs' reliability often drops in scientific problem-solving
- Prioritize **perfect performance** over **cost control**
- Caused from the inherent complexity of natural sciences
  - Long, multi-step and unstructured reasoning
  - Modeling of real-world scenarios
  - Understanding of fundamental laws
  - Implicit constraints
  - Deterministic & probabilistic, precise & approximate
  - ...

# Scientific reasoning

## ➤ Assuming complete knowledge



## ➤ Hard to detect mistakes due to **logical leaps** in the reasoning

- Both human and AIs alike tend to omit steps they consider "obvious"

# How to address this issue?

---

## ➤ Logical Chain Augmentation (LOCA)

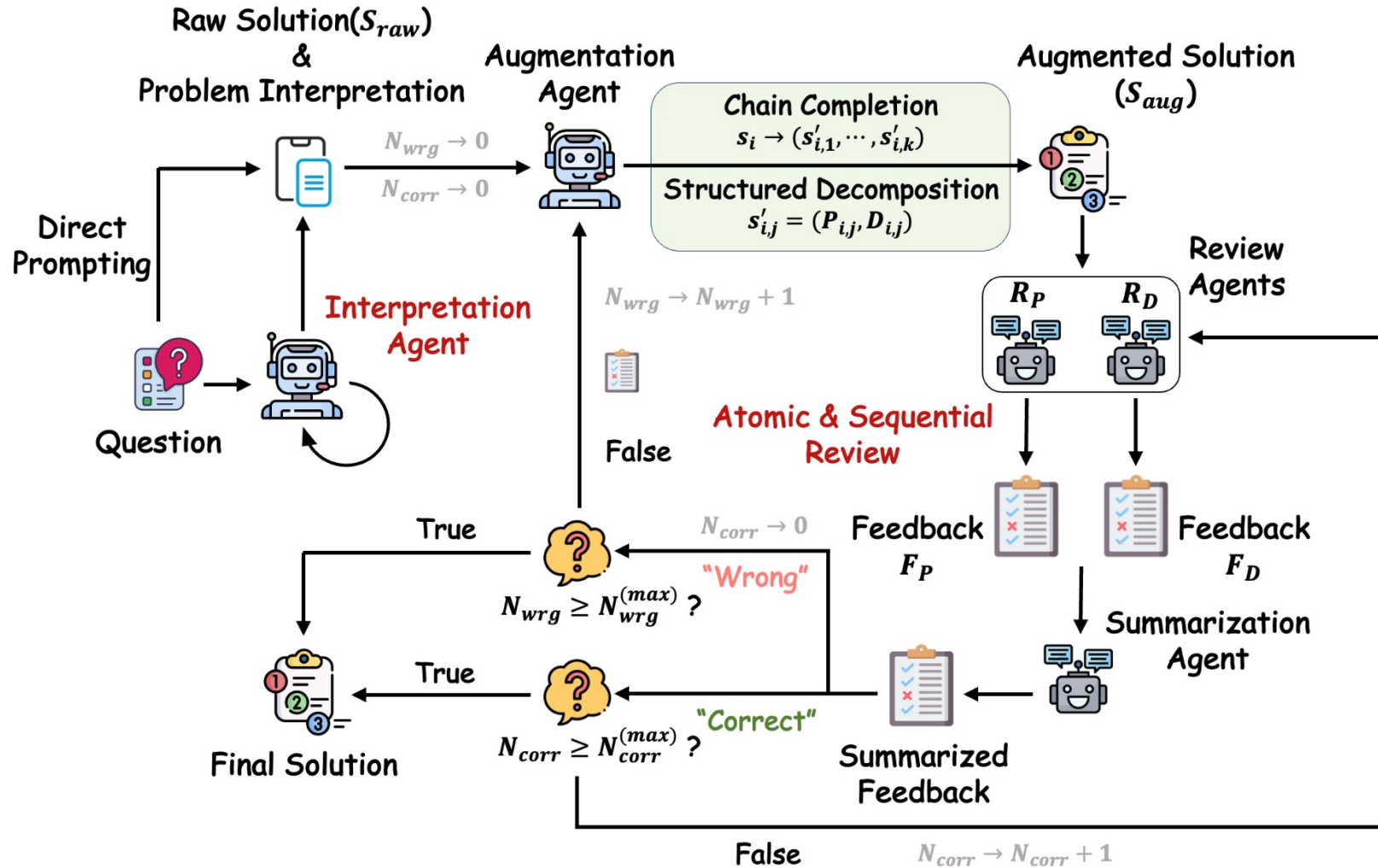
- Enforcing logical completeness and structurally decomposing reasoning steps into verifiable principles and derivations, within an **augment-and-review** loop

Fang et al., 2510.01249

Jian et al., 2511.10515

# Overview of LOCA's pipeline

Jian et al., 2511.10515



# The CPhO: a challenging testbed

---

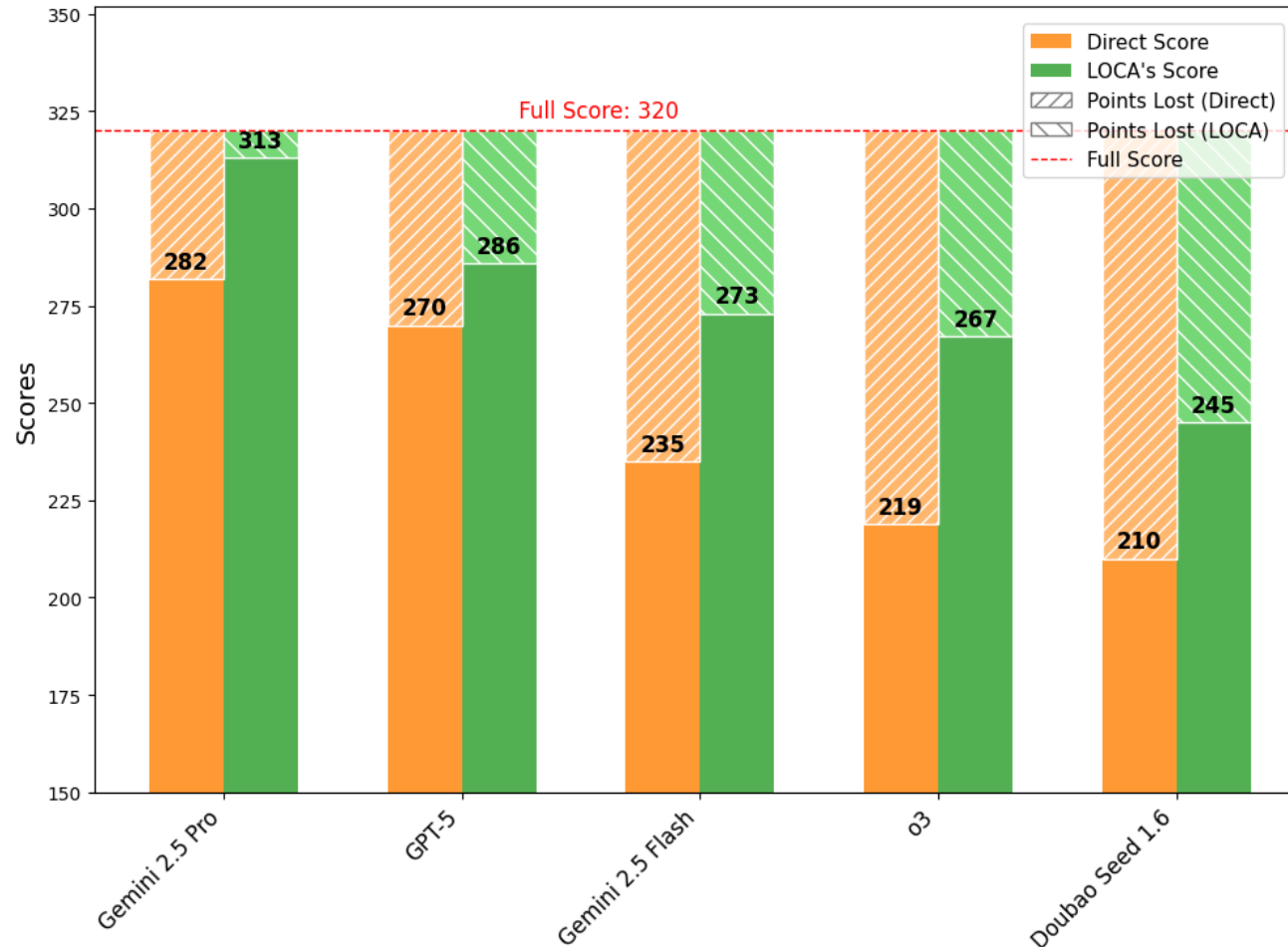
- The Chinese Physics Olympiad (CPhO): a premier national physics competition organized annually in China
- For our evaluation, we focus on the **theory** examination of the 42<sup>nd</sup> CPhO 2025 (final round held in Fuzhou at the end of October)
  - Demands of long, multi-step reasoning
  - Multimodal problems
  - No data contamination issue

# Overall performance across various base LLMs

Jian et al., 2511.10515

## ➤ Overall performance of LOCA on four mainstream LLMs

Performance Comparison of LLMs with Direct Prompting vs. LOCA



- Total scores for th. Problem: 320
- Highest score of human: 207

# Comparison across more baseline methods

Jian et al., 2511.10515

## ➤ Comparison of LOCA and more baselines

Method	1	2	3	4	5	6	7	Total Score	Error Rate
Human's highest	-	-	-	-	-	-	-	204	36%
Direct Prompting	45	41	45	33	39	39	40	282	12%
Zero-Shot-CoT	45	37	45	45	45	38	40	295	7.8%
Few-Shot CoT	45	45	45	41	45	42	39	302	5.6%
ToT	45	45	45	41	45	40	39	300	6.3%
GoT	45	34	20	36	45	39	39	258	19%
MAD	45	33	42	43	45	44	40	292	8.8%
Self-refine	45	43	45	35	39	41	40	288	10%
PSN	45	32	39	43	45	43	45	292	8.8%
LOCA (ours)	45	45	45	45	45	43	45	<b>313</b>	<b>2.2%</b>

# Outline

---

## I. Background and motivation

## II. From data to model

- a. Symbolic regression
- b. FunSearch
- c. AI-Newton

## III. From model to data

- a. Established field
- b. Frontier field

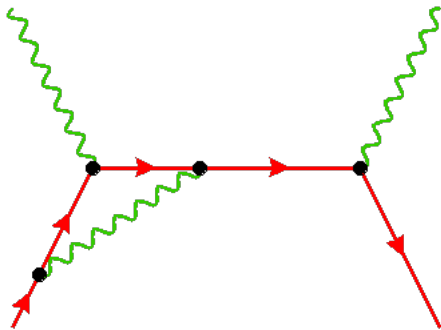
## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

# Challenge for advanced theoretical calculation

## ➤ Example: perturbative quantum field theory

$$e^- + \gamma \rightarrow e^- + \gamma$$



Dirac algebra: e.g.  $\text{Tr}[\gamma_\mu \gamma_\nu \gamma_\rho \gamma_\sigma \gamma^\mu \gamma^\nu \gamma^\rho \gamma^\sigma]$

Family decompose:  $|M|^2 = \sum_i a_i I_i$

Feynman integral

Integral reduction  $I_i = \sum_j c_{ij} J_j$

Integral calculation  $\frac{d}{ds} \vec{J} = A \cdot \vec{J}$

$$I_{(1,1,1)} = \int \left( \prod_{i=1}^2 \frac{d^D \ell_i}{i\pi^{D/2}} \right) \frac{1}{(\ell_1^2 - 1)\ell_2^2(\ell_1 + \ell_2)^2},$$

# Solution 1: prompting

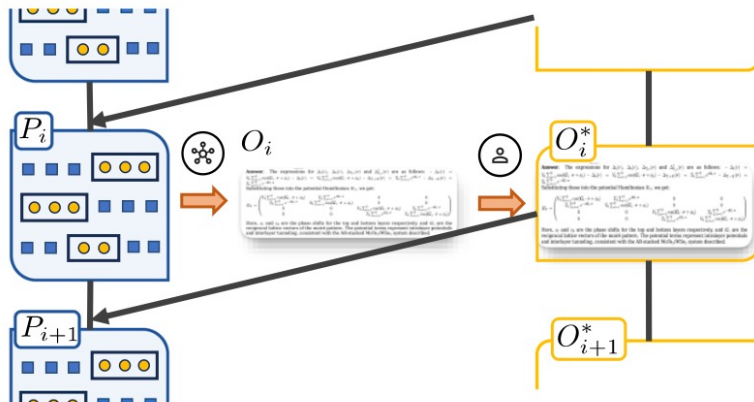
## ➤ Guide LLM to solve heavy reasoning problem step-by-step

You will be instructed to convert the noninteracting Hamiltonian  $\{nonint\_symbol\}$  in the second quantized form from the basis in real space to the basis in momentum space. To do that, you should apply the Fourier transform to  $\{real\_creation\_op\}$  in the real space to the  $\{momentum\_creation\_op\}$  in the momentum space, which is defined as  $\{definition\_of\_Fourier\_Transformation\}$ , where  $\{real\_variable\}$  is integrated over all sites in the entire real space. You should follow the EXAMPLE below to apply the Fourier transform. [Note that hopping have no position dependence now.] You should recall that  $\{nonint\_symbol\}$  is  $\{expression\_nonint\}$ . Express the total noninteracting Hamiltonian  $\{nonint\_symbol\}$  in terms of  $\{momentum\_creation\_op\}$ . Simplify any summation index if possible.

Use the following conventions for the symbols (You should also obey the conventions in all my previous prompts if you encounter undefined symbols. If you find it is never defined or has conflicts in the conventions, you should stop and let me know):  $\{definition\_of\_variables\}$

<https://doi.org/10.1038/s42005-025-01956-y>

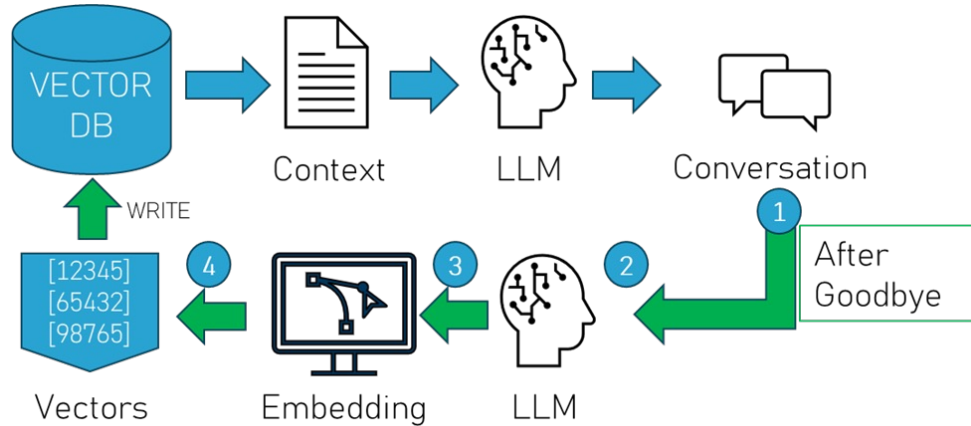
## ➤ Correct errors



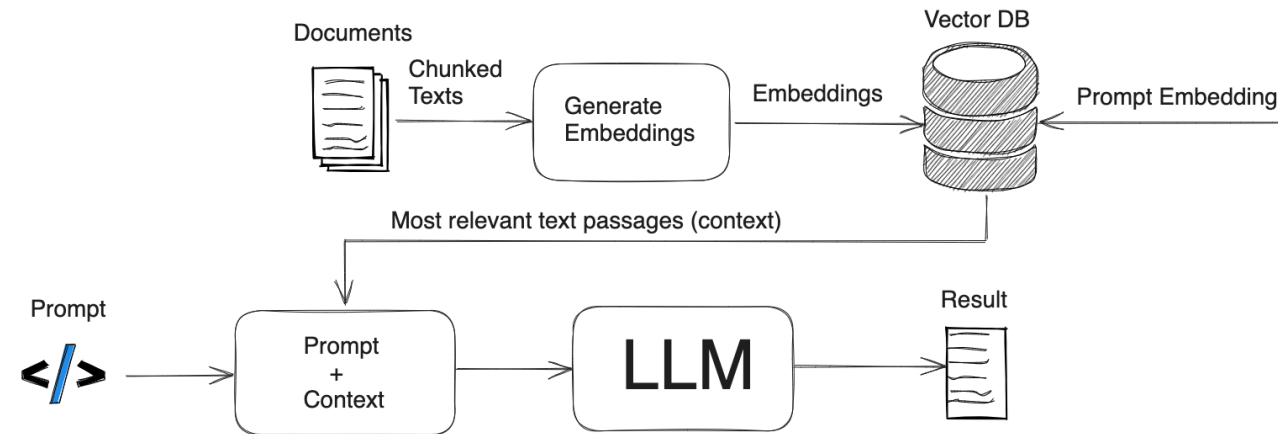
- **Evaluator** for judging (in these paper, a human evaluator is used)
- **Reviewer** for extracting errors and explanation

# Solution 2: memory/RAG

## ➤ Memory



## ➤ RAG (Retrieval-Augmented Generation)



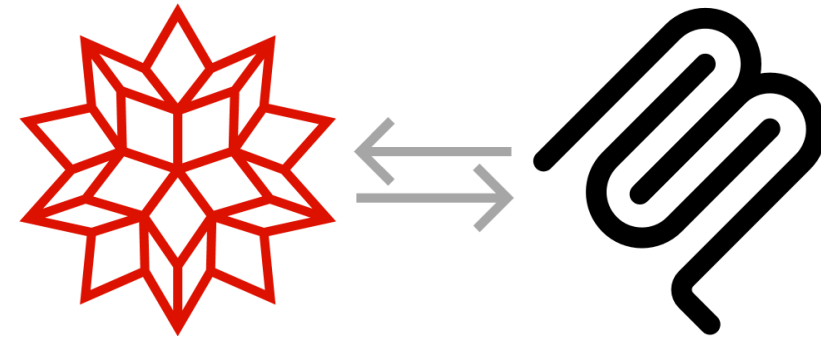
## ➤ Difference

	Memory	RAG
<b>Data Updates</b>	<b>Atomic, Frequent</b>	<b>Batch, Infrequent:</b>
<b>Primary Use</b>	personalized, dynamically evolving facts	general-purpose, relatively static domain knowledge

# Solution 3: tool calling/MCP

## ➤ Tool calling

Leave professional work to professional tools

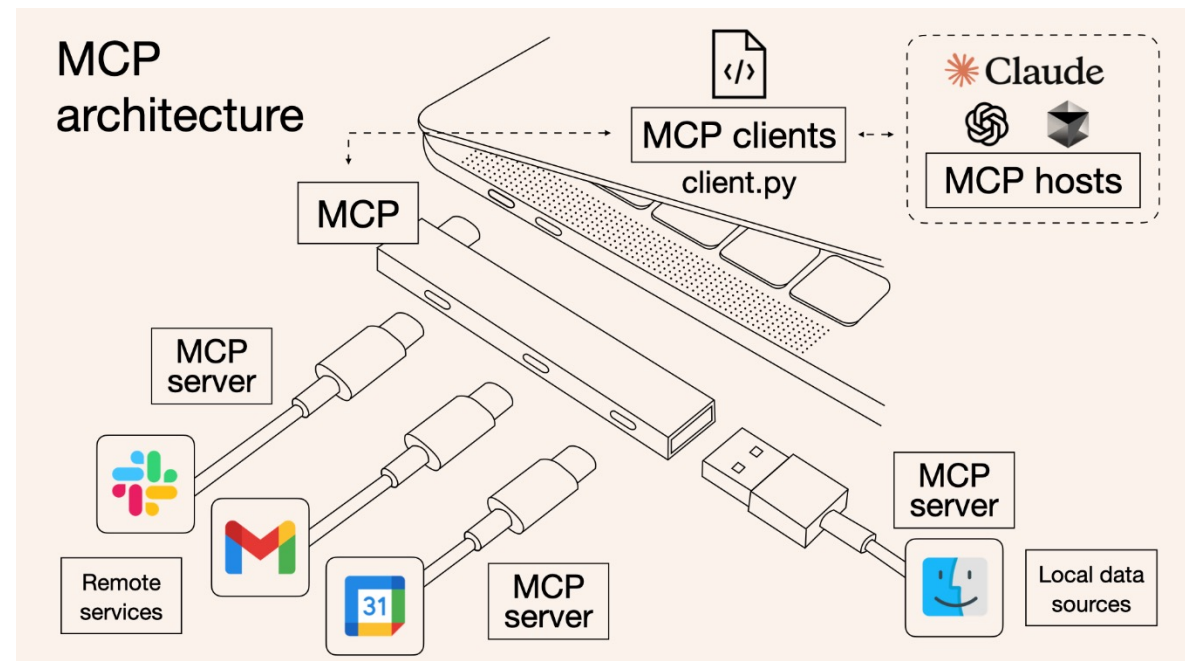


<https://resources.wolframcloud.com/PacletRepository/resources/Wolfram/MCPServer/>

## ➤ MCP(Model Context Protocol)

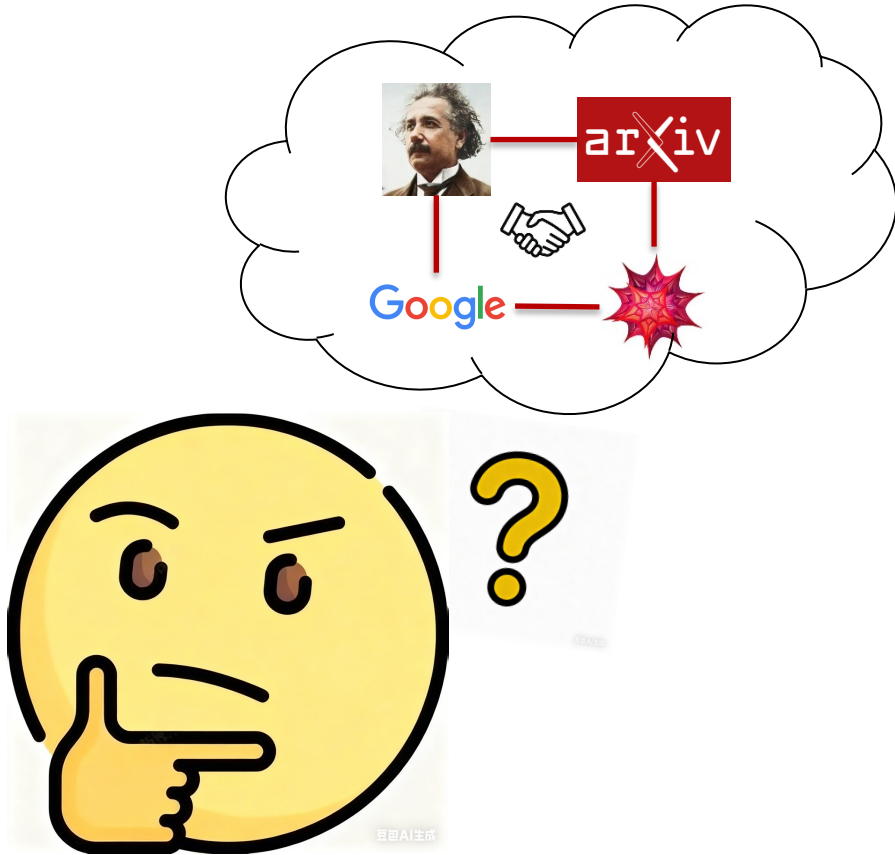
Standard protocol for communication between LLM and Services

- What tools?
- How to use?
- When to call?



# Solution 4: multi-agent framework

➤ Single-LLM belike:

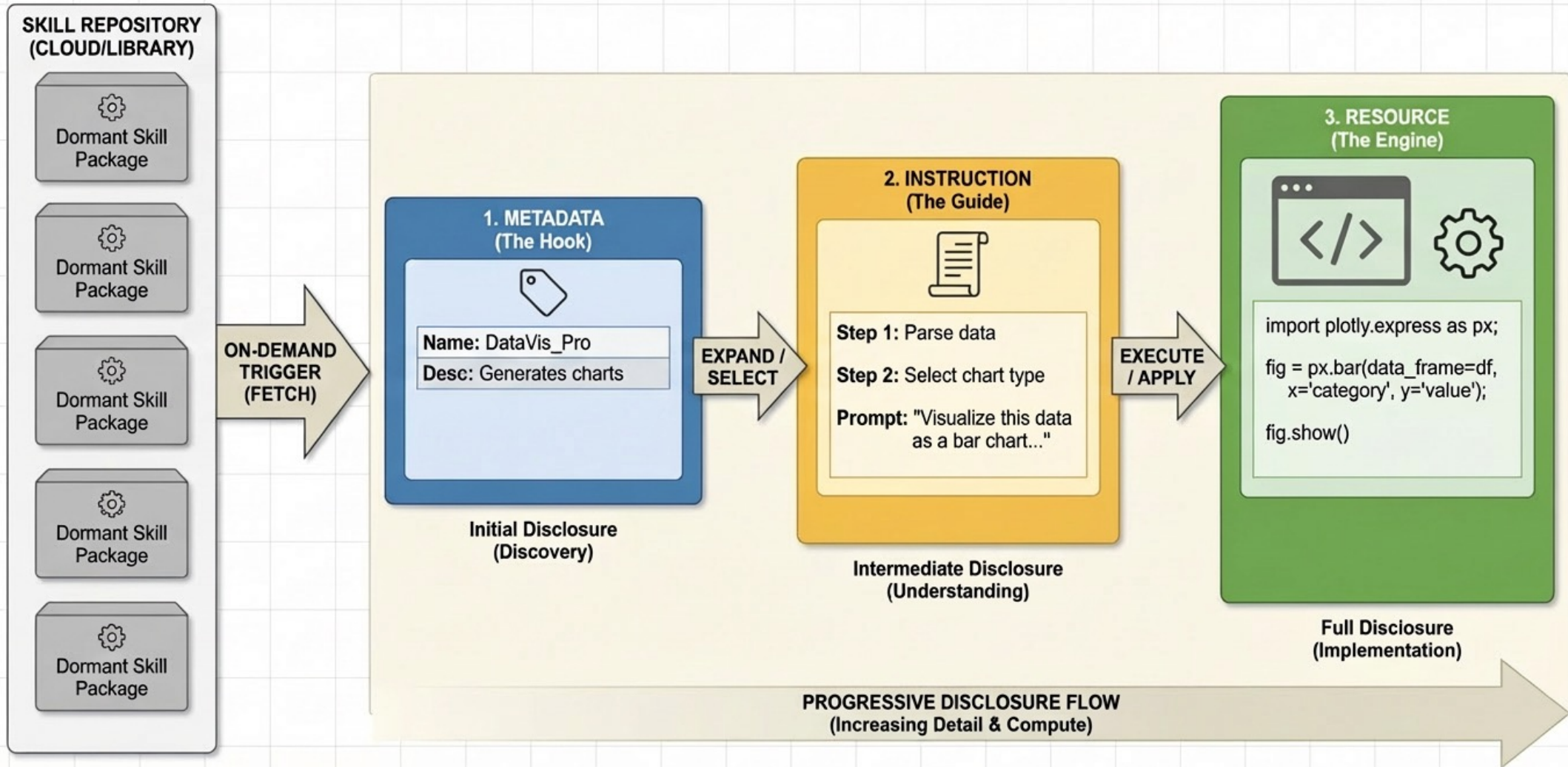


➤ Multi-agent belike:



# Solution 5: skills

## VIBE CODING: SKILL MECHANISM - ON-DEMAND & PROGRESSIVE DISCLOSURE



# Outline

---

## I. Background and motivation

## II. From data to model

- a. Symbolic Regression
- b. Funsearch
- c. AI-Newton

## III. From model to data

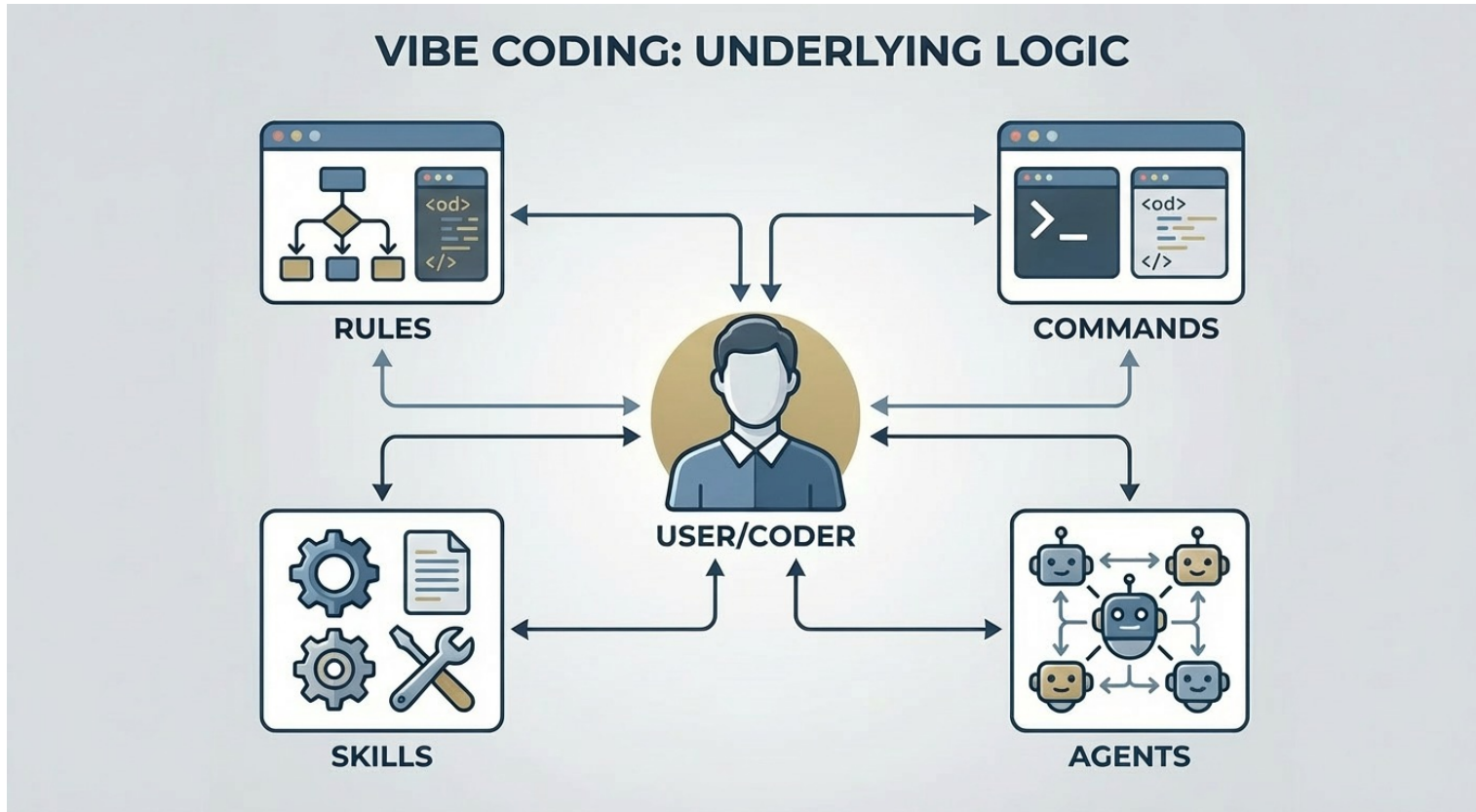
- a. Established field
- b. Frontier field

## IV. Aether: an integrated vibe research platform

## V. Summary and outlook

# From vibe coding to vibe research

## ➤ Inspiration from vibe coding



**Anatomy of the `.claude/` folder** commit .claude/ to git

- `your-project/`
  - `CLAUDE.md` 团队指令, 提交到 git
  - `CLAUDE.local.md` 个人覆盖, 被 git 忽略
  - `.claude/` ↓ 控制中心
  - `settings.json` 权限+配置, 提交到 git
  - `settings.local.json` 个人权限, 被 git 忽略
  - `commands/` 自定义斜杠命令
    - `review.md` → /project:review
    - `fix-issue.md` → /project:fix-issue
    - `deploy.md` → /project:deploy
  - `rules/` 模块化指令文件
    - `code-style.md`
    - `testing.md`
    - `api-conventions.md`
  - `skills/` 自动调用的工作流
    - `security-review/`
      - `SKILL.md`
    - `deploy/`
      - `SKILL.md`
  - `agents/` 子代理角色
    - `code-reviewer.md`
    - `security-auditor.md`

Everything Claude needs to know about your project lives right here.

# Aether

- **A**utonomous **E**ngine for **T**heoretical & **H**ands-on **E**xploration in **R**esearch
- **B**ased on **O**pen**C**ode

<https://aether.aiphys.cn/>  
<https://github.com/Science-Discovery/Aether>

Aether

快速开始 视频教程 项目介绍 核心能力 Skill 广场 GitHub

## Aether: 专为科研人员定制的 AI助手

覆盖 AI 辅助阅读、文献处理、批量转换、项目管理、终端编码与微信协作, 兼容多模型与多 Provider, 桌面端和 Web 端均可开箱即用。

[立即查看 GitHub](#) [下载客户端](#) [查看微信连接教程](#)

[查看文档](#) →

如有问题或建议, 请联系 [aethersci@163.com](mailto:aethersci@163.com)

### LATEST NOTES

#### 最新动态

版本 v0.3.1  
发布日期  
2026-04-09

**发布**

项目目录发现、迁移恢复与微信协作体验继续补强

v0.3.1 重点打磨真实使用中的“恢复”和“切换”流程: 最近项目改为基于实际会话记录和路径前缀进行筛选, 空搜索时可直接展示根目录, 目录选择与过滤更可靠; 旧版本数据库迁移新增更稳健的校验、归档与兜底修复流程, 升级后更容易找回历史会话; 微信协作侧解耦连接生命周期与弹窗显示, 默认优先回到最近活跃项目, 并统一消息头与时间显示, 移动端远程接力更顺手。

版本 v0.3.0  
发布日期  
2026-04-04

**发布**

AI 阅读完成度、更新安装链路和项目同步继续增强

版本 v0.2.2  
发布日期  
2026-04-03

**发布**

项目管理体验更清晰

[展开更多动态](#)

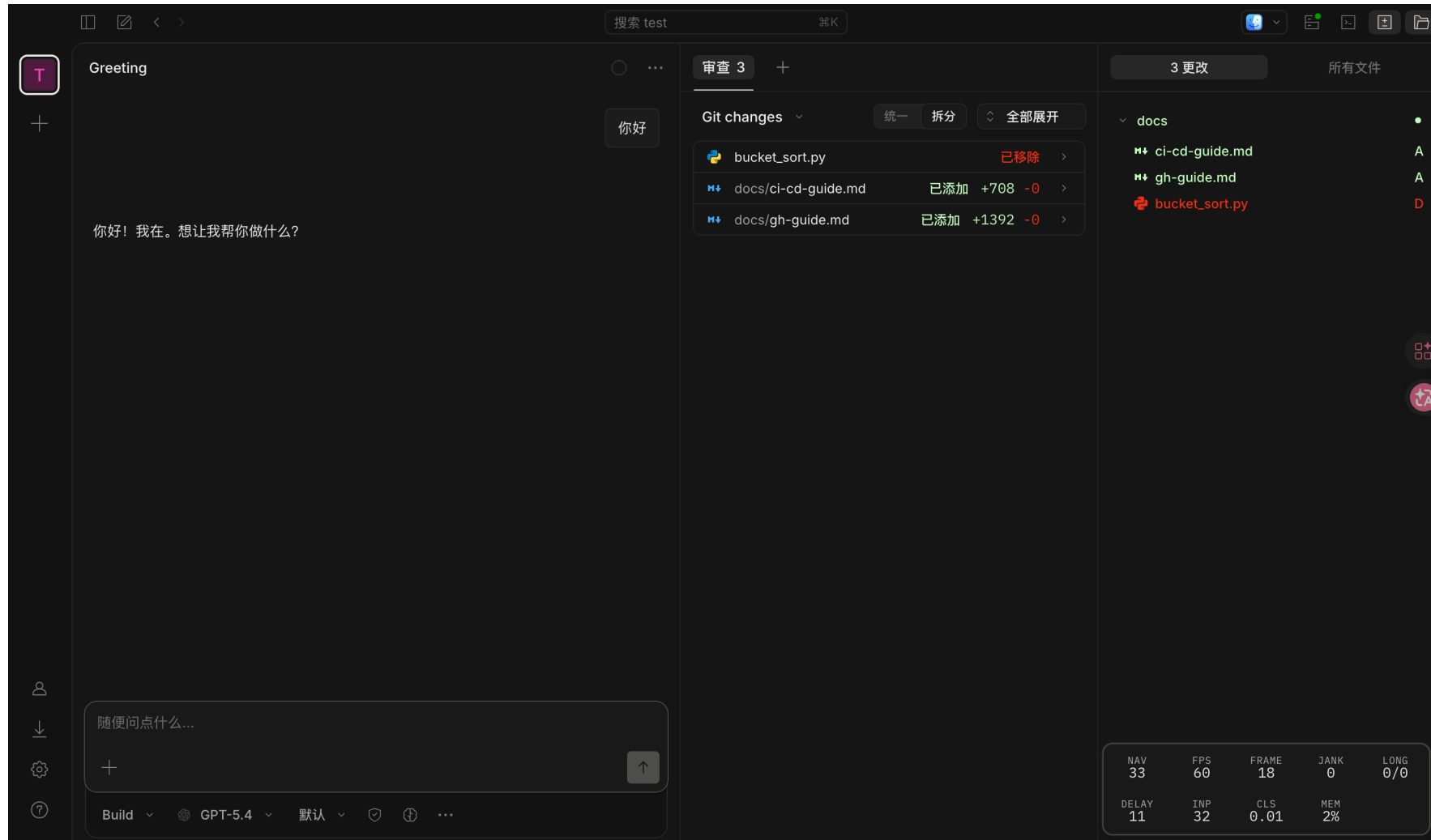
# Highlights: inherited from OpenCode

---

- **Open source**
- **Fully utilize the foundational capabilities of vibe coding**

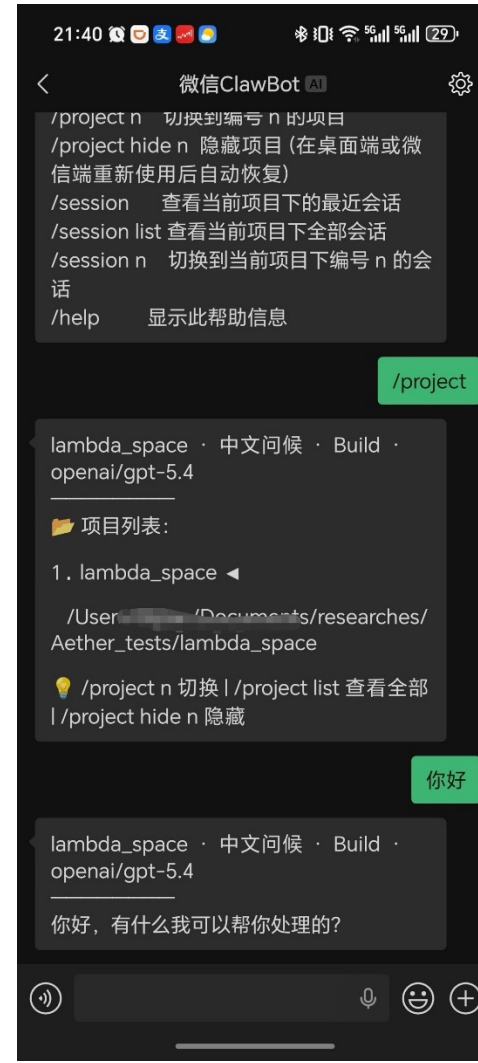
# Highlights: highly enhanced UI

➤ User-friendly, ready to use out of the box



# Highlights: WeChat & FeiShu integration

## ➤ WeChat & FeiShu integrated, supporting remote work and voice input



# Highlights: research adaptation

---

- **Agent and skill architecture for scientific research scenarios**
- **Integrated professional skills/tools for scientific discovery**
- **User-level memory mechanism**
- **Self-evolving skills**

# Highlights: research adaptation

两圈四点doublebox主积分在s=1,t=10展开计算



请帮我计算两圈四点doublebox拓扑的主积分在运动学量s=1, t=10下的结果, 展开到维数正规化因子epsilon的正1阶。

已探索 4 次读取, 1 次搜索

Shell Checks AMFlow environment status

已探索 1 次搜索

我先检查了本机环境: [Mathematica](#) 和 [AMFlow](#) 都能找到, 但当前 [Blade](#) 的路径配置需要先修一下, 否则没法真正跑两圈积分。

我这里有一个关键歧义需要你确认, 我建议默认这样做:

- 按“无质量平面四点两圈 [double box](#)”标准族来算;
- 外腿满足  $p_1^2=p_2^2=p_3^2=p_4^2=0$ ,  $(p_1+p_2)^2=s=1$ ,  $(p_2+p_3)^2=t=10$ ;
- 结果采用 AMFlow 的归一化: 每圈  $1/(i\pi^{(D/2)})$ ;
- 并且我自动修复本机 [Blade](#) 配置后再计算。

如果你同意这个默认设定, 我就直接继续算到  $\epsilon^1$ ; 如果不是这个族/归一化, 请直接给我:

- 传播子列表;
- 你说的“主积分”具体是哪一组;
- 归一化约定。



# Summary and outlook

---

- Human scientific discovery necessitates a new research paradigm, AI may help
- From data to theory:
  - Plenty of ideas proposed
  - Surpasses human experts in many topics
- From theory to data:
  - AI surpasses human experts for established fields
  - AI improves fast for frontier fields
- Integrated wide research platform will become a crucial tool for AI-empowered scientific research.
- AI for scientific discovery: remains in its infancy, but very promising

***Thank you!***