AI for Scientific Discovery: From Algorithmic Tools to Integrated Research Platforms

Dong-Shan Jian（Peking University）

jiandongshan@hotmail.com

AI 智能体赋能核物理研究研讨会, 中国·上海, 2026/07/04

Outline

I. AI for established fields: current status

II. AI for frontier fields: challenges

a. Complexity of frontier research tasks
b. The fundamental research–coding gap
c. Accessibility for the broader research community

III. Aether: an integrated vibe research platform

IV. Summary and outlook

Primary focus: reasoning

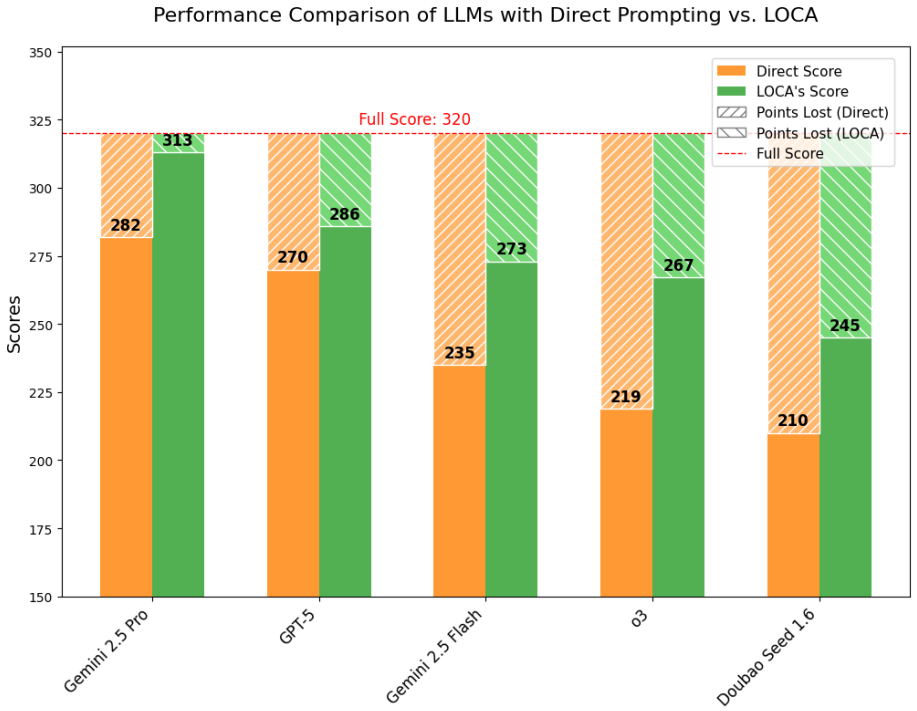
- Primary focus: 2023 ~ 2025
- LLMs hallucinate in complex (long) scientific reasoning tasks
- Scientific research has extremely stringent requirements for accuracy
- Inherent complexity of natural science
  - Modeling of real-world scenarios
  - Understanding of fundamental laws
  - Long, multi-step and unstructured reasoning
  - Implicit constraints
  - Deterministic & probabilistic, precise & approximate
  - …

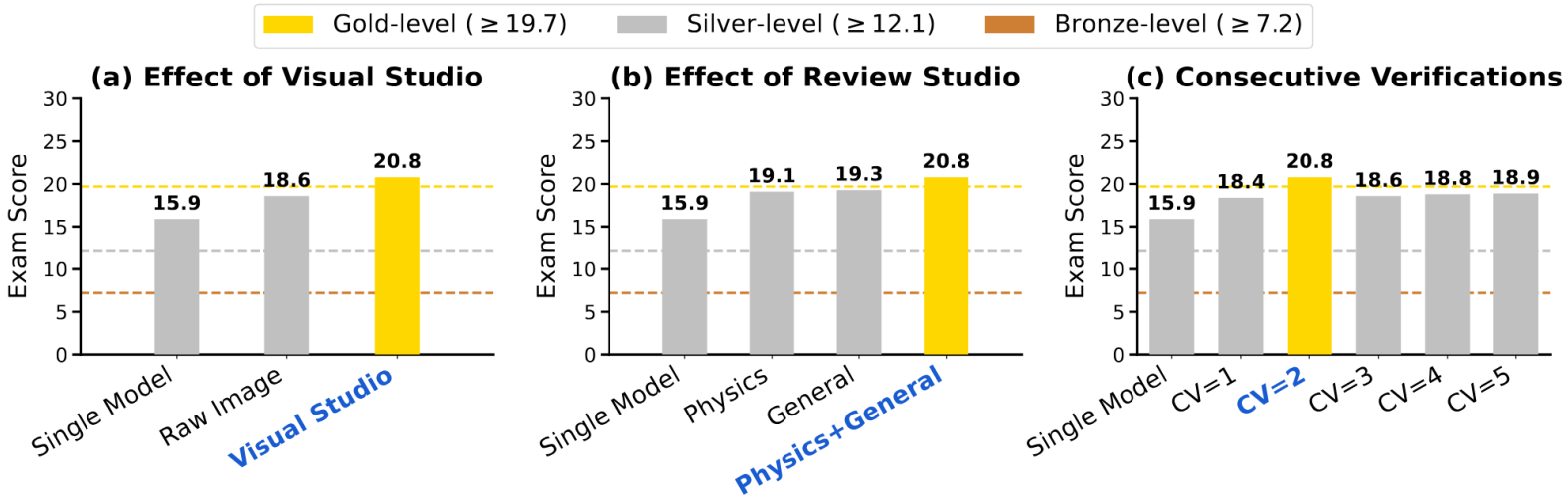
Various approaches

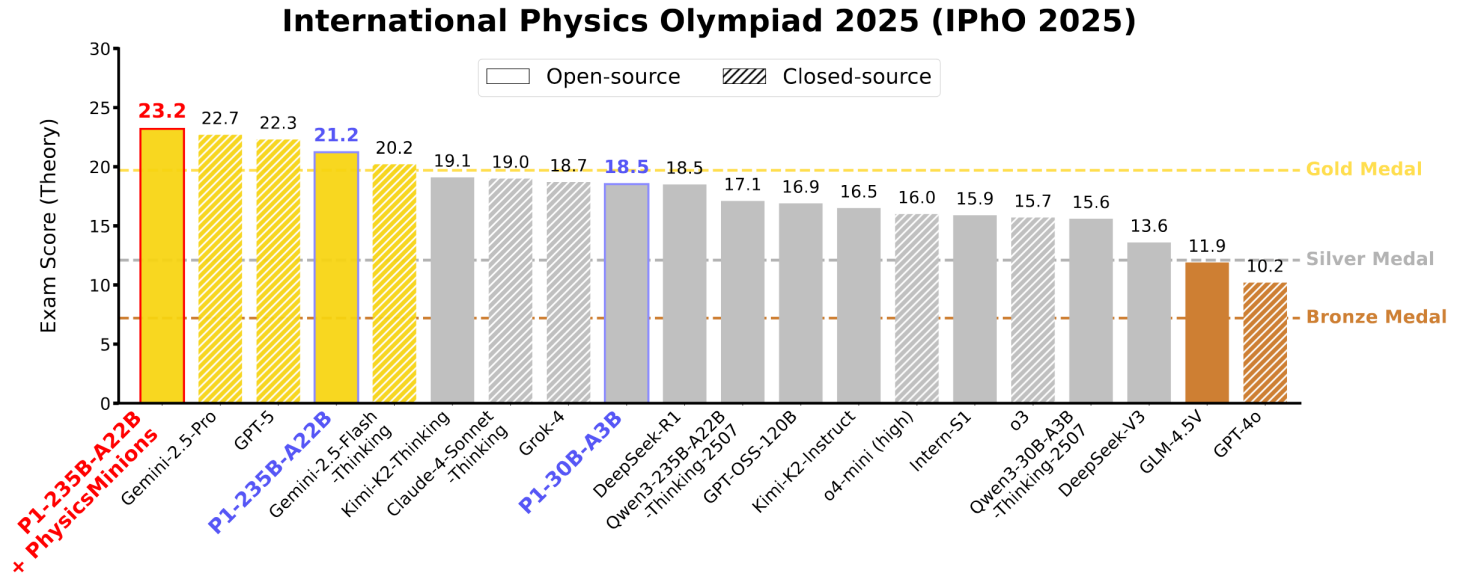
- **Training-based: improve the model's intrinsic reasoning**
  - Scaling
  - Fune-tuning
  - Reinforcement learning
- **Training-free: augment reasoning with external mechanisms**
  - Prompt engineering
  - Tool calling
  - External knowledge
  - Workflows / agents (early-stage)
  - Domain-specific design
  - E.g., CoT, ToT, GoT, self-refine, MAD, Physics Supernova, LOCA…
- **Breakthrough progress, significantly surpassing humans**

Achievements

- **E.g., on Olympiad-level physics problems**







Outline

I. AI for established fields: current status

II. AI for frontier fields: challenges

- a. Complexity of frontier research tasks
- b. The fundamental research–coding gap
- c. Accessibility for the broader research community

III. Aether: an integrated vibe research platform

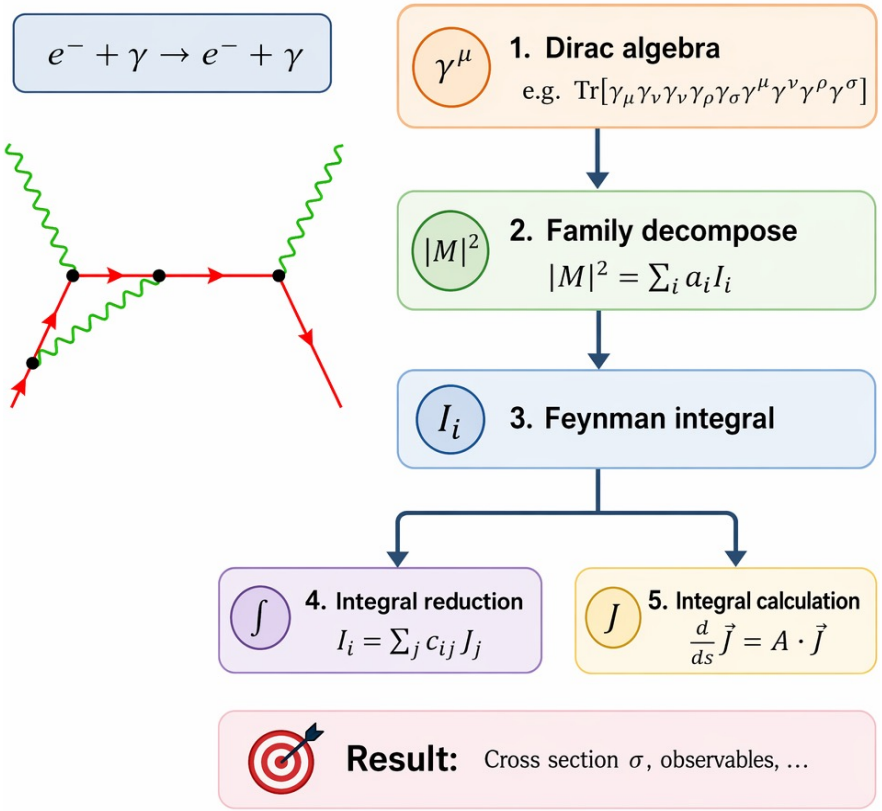
IV. Summary and outlook

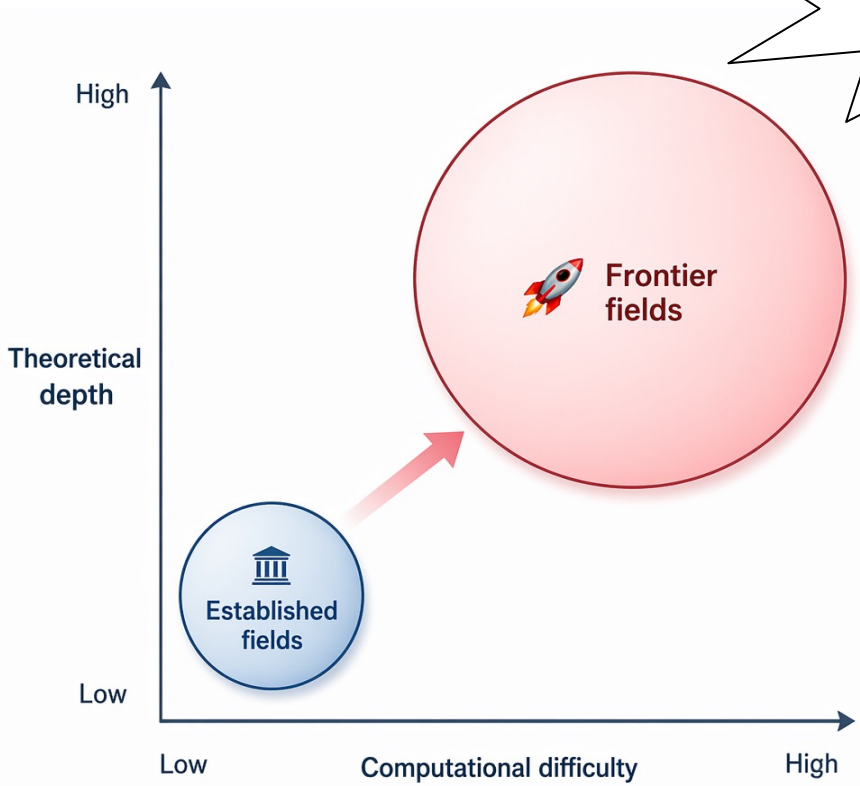
Challenges for frontier fields

- **Great uncertainty**

  Highly complex domain-specific methods

- **Example: perturbative quantum field theory**





Solution 1: prompt engineering

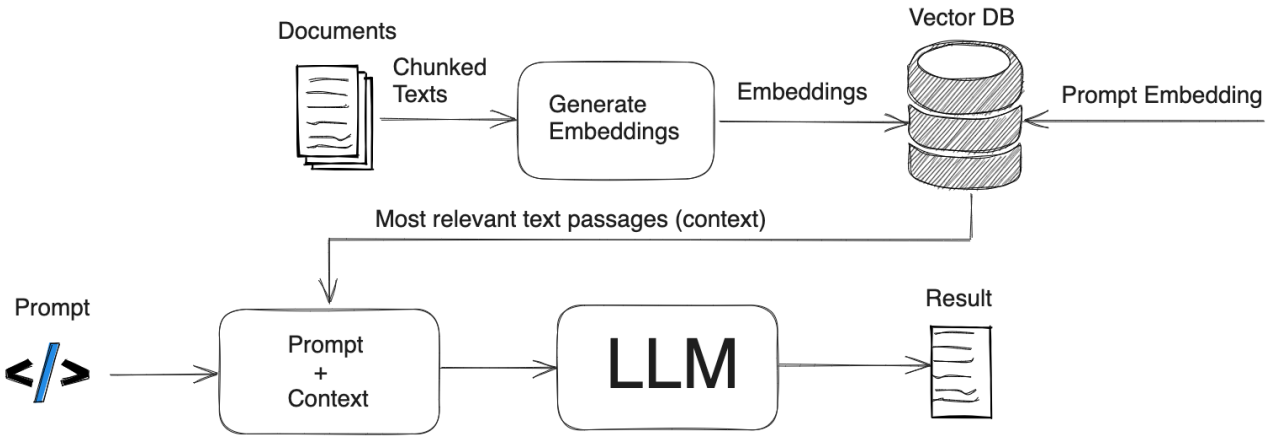
➢ Guide AI's reasoning behavior

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}

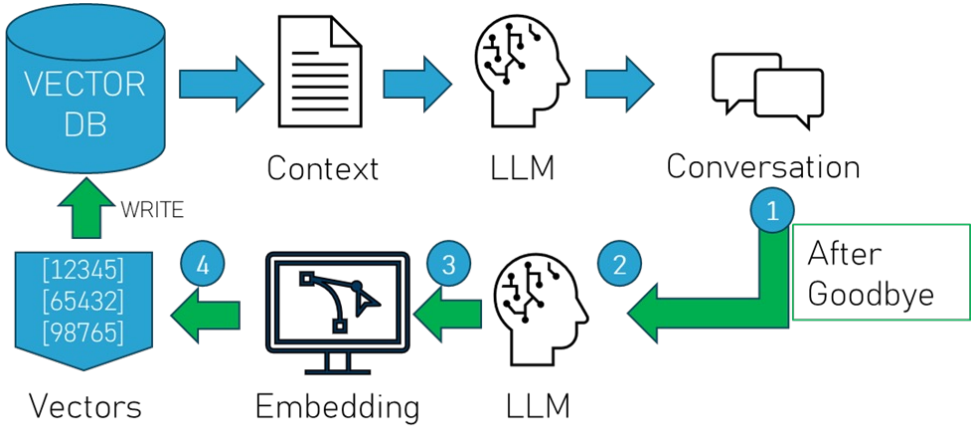
https://doi.org/10.1038/s42005-025-01956-y

Solution 2: RAG / memory

- **Provide AI with external knowledge that it hasn't encountered**
- **RAG(Retrieval-Augmented Generation)**



- **Memory(accumulated experience)**



- **Difference**

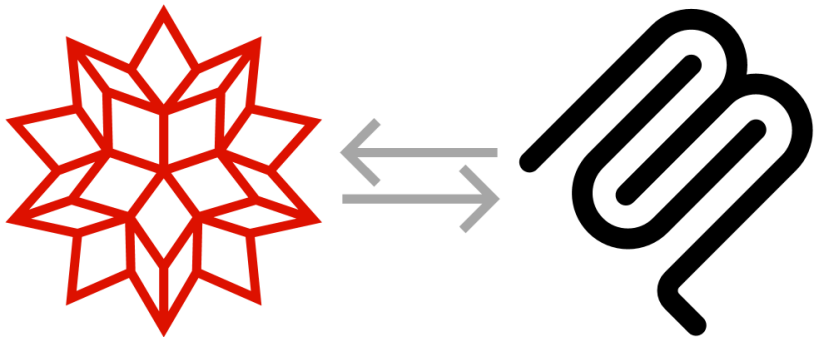
	Memory	RAG
Data updates	atomic, frequent	batch, infrequent:
Primary use	personalized, dynamically evolving facts	general-purpose, relatively static domain knowledge

Solution 3: tool calling/MCP

➢ Tool calling

Considering absolute accuracy and efficiency

Leave professional work to professional tools

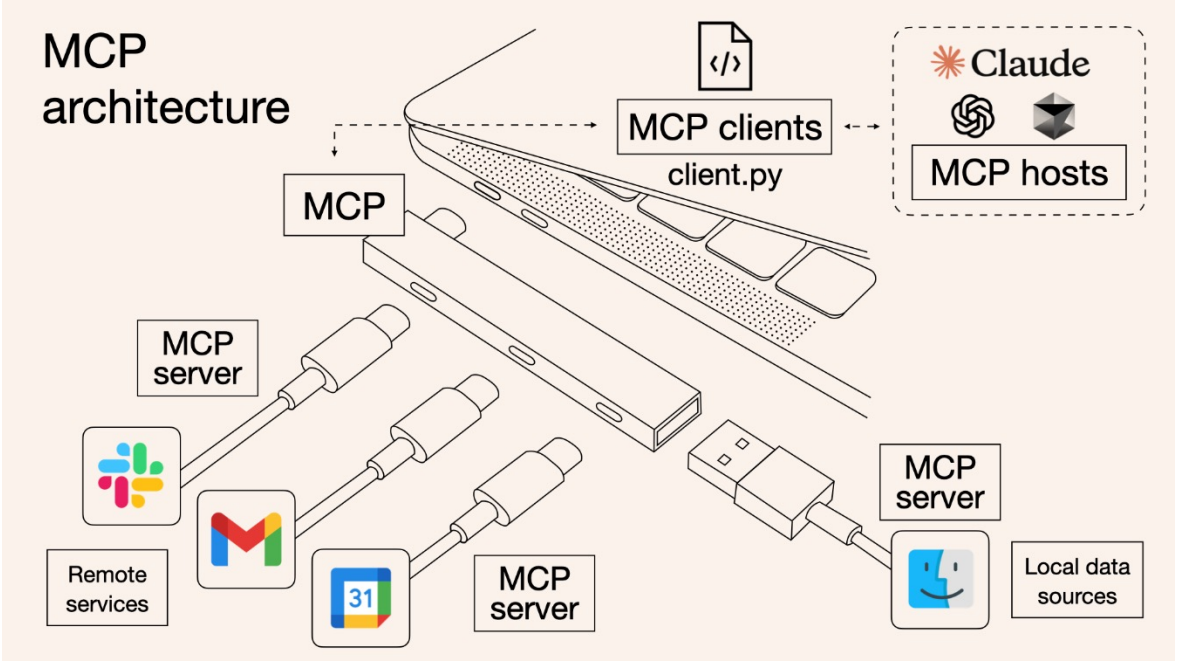


➢ MCP(Model Context Protocol)

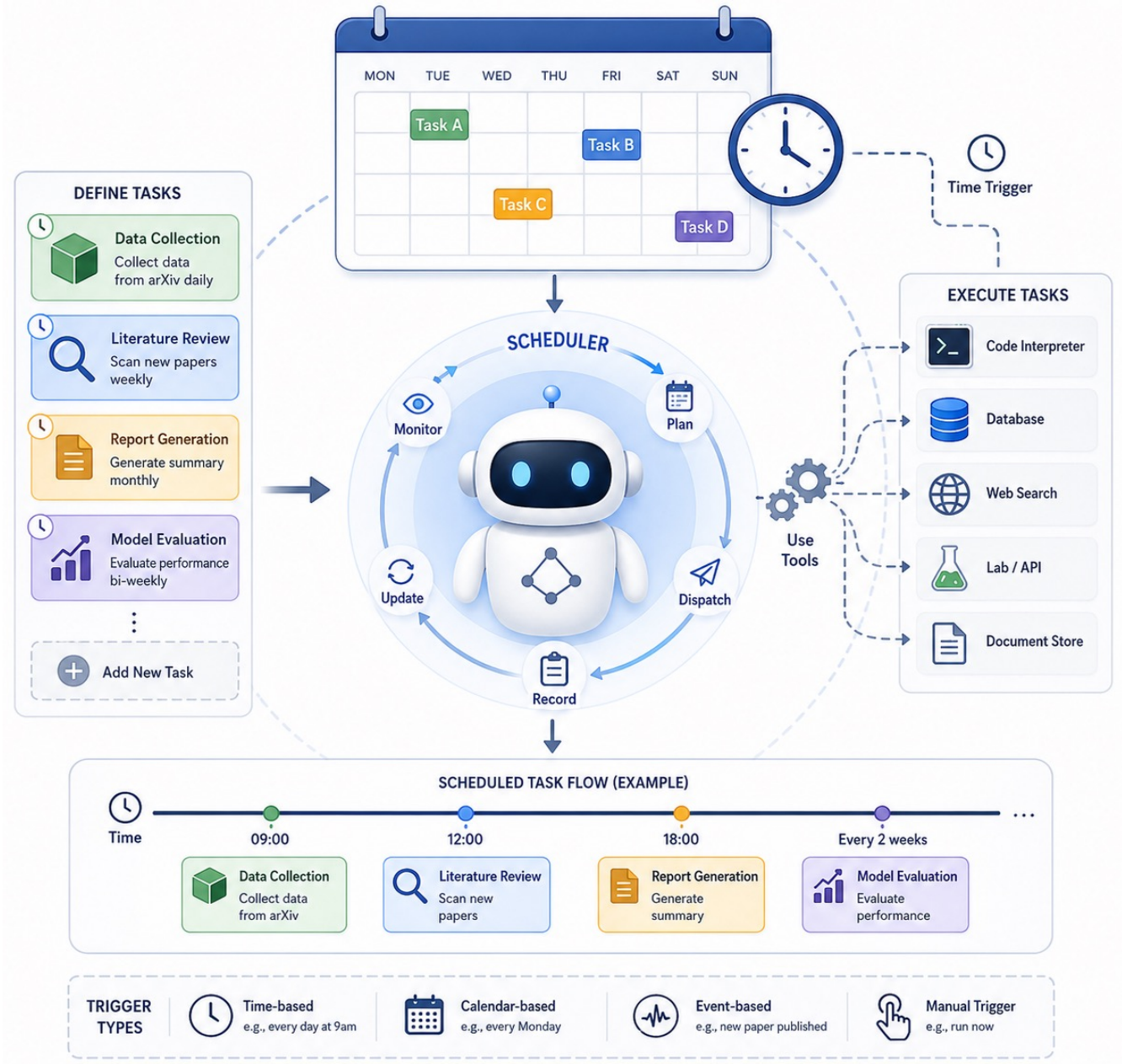
https://resources.wolframcloud.com/PacletRepository/resources/Wolfram/MCPServer/

Standard protocol for communication between LLMs and tool services

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

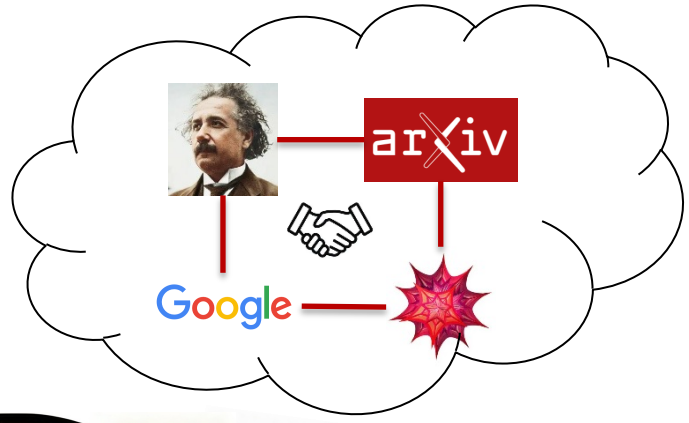


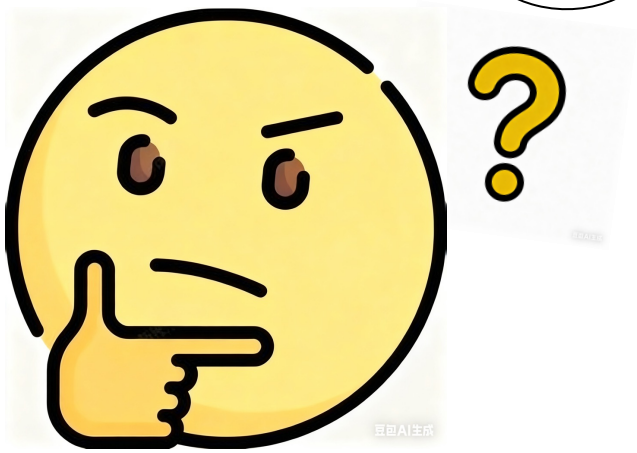
Solution 4: scheduling



Solution 5: multi-agent workflow

- **Single-LLM belike:**



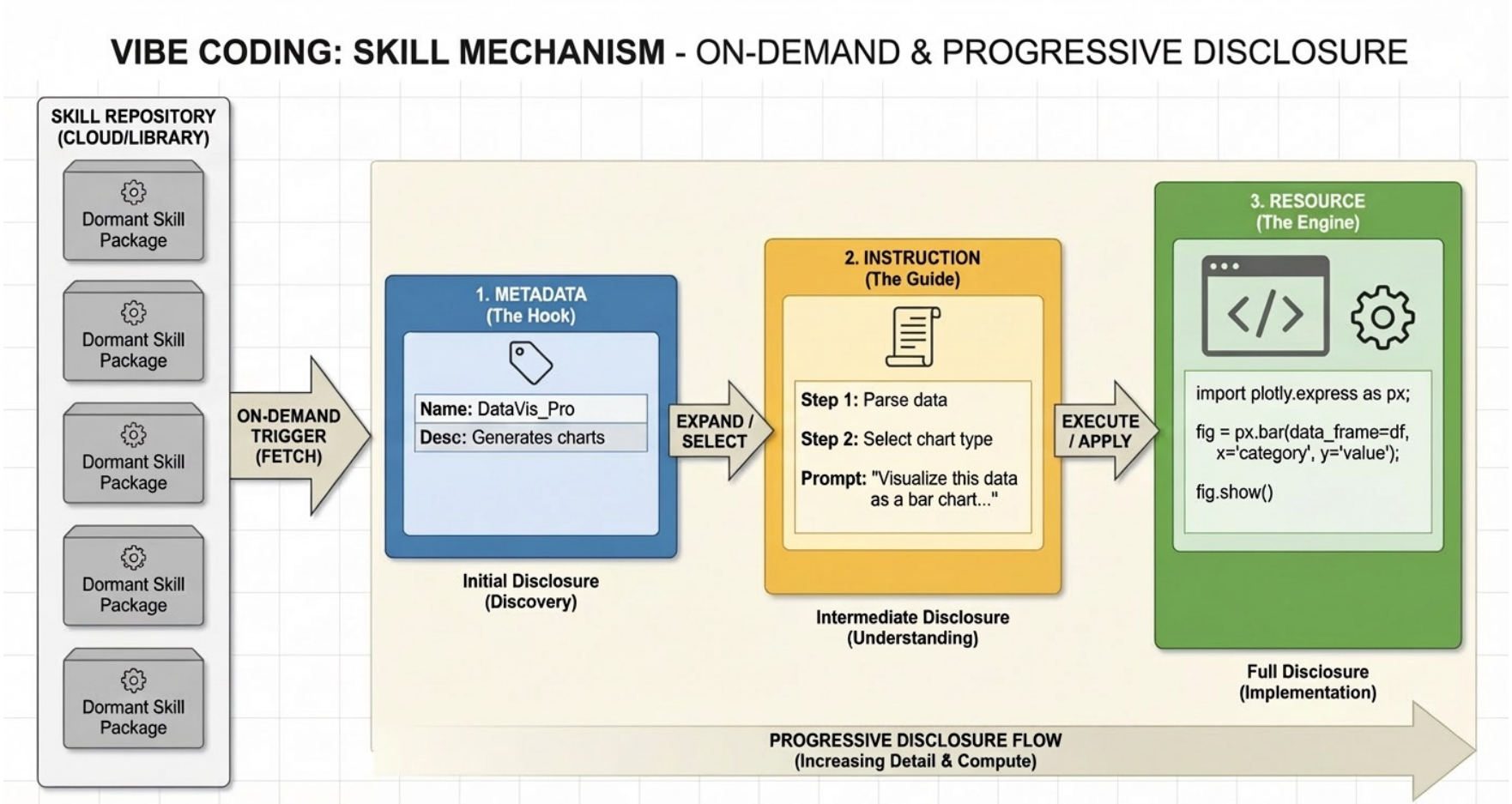


- **Multi-agent belike:**

- **For truly complex, multi-step problems**
- **Agents with different roles collaborate like humans**

Solution 6: skills

- **Proposed by Anthropic in 2025**
- **Program written in natural language**



Outline

I. AI for established fields: current status

II. AI for frontier fields: challenges

a. Complexity of frontier research tasks

b. The fundamental research–coding gap

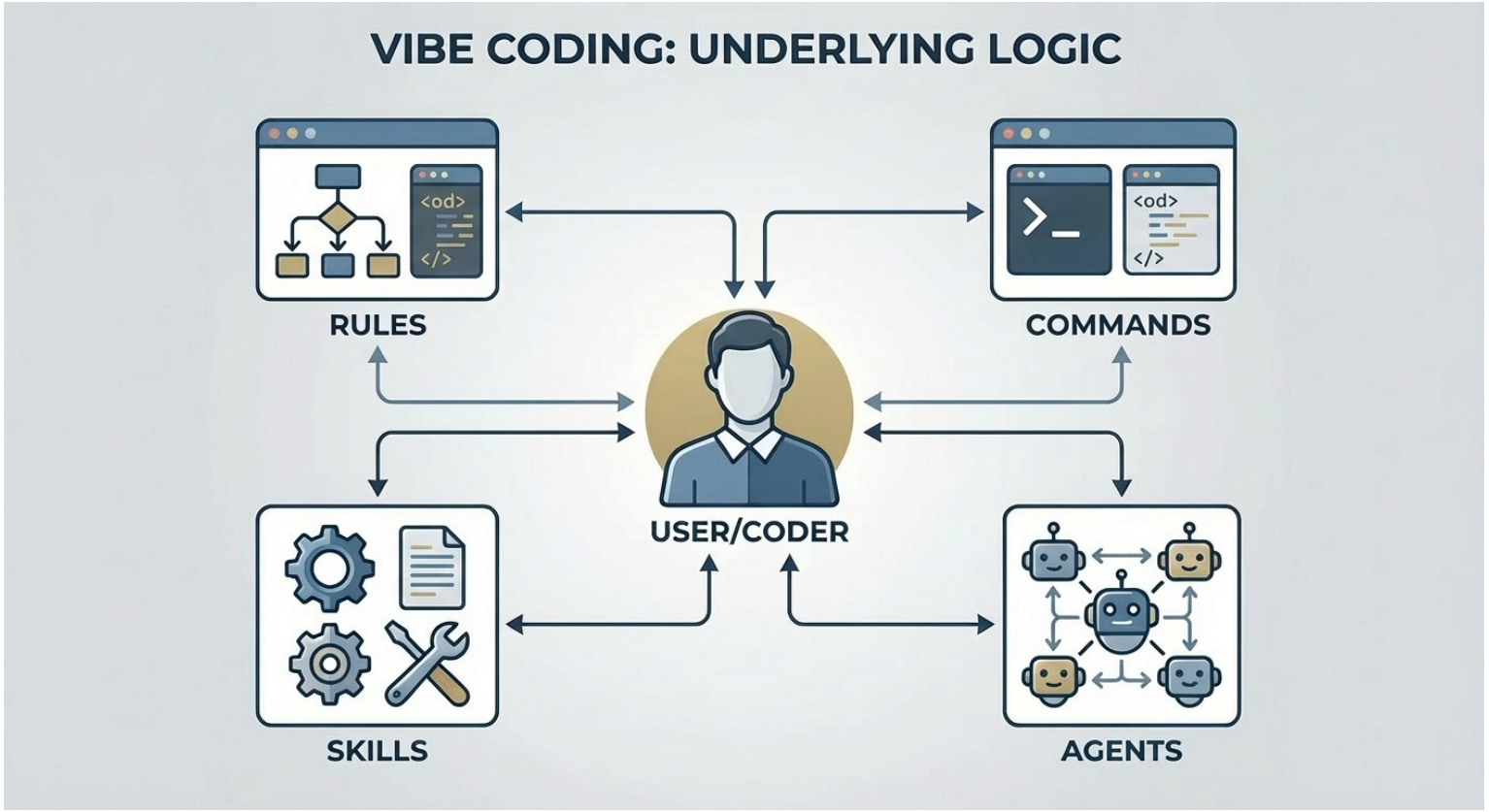
c. Accessibility for the broader research community

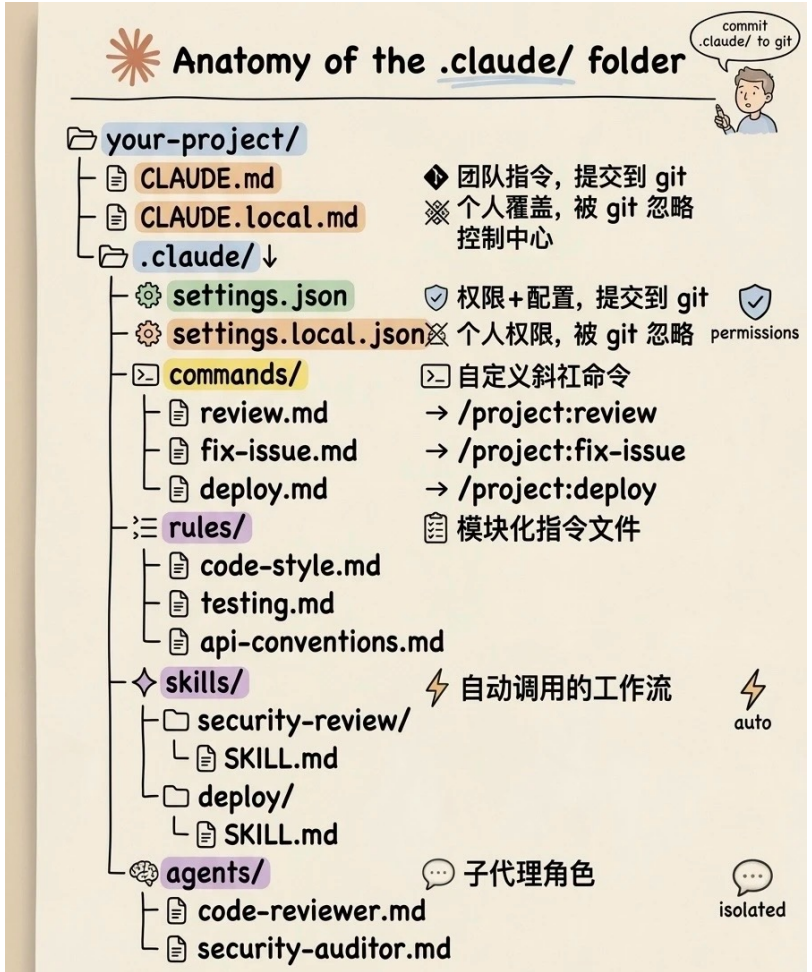
III. Aether: an integrated vibe research platform

IV. Summary and outlook

From vibe coding to vibe research

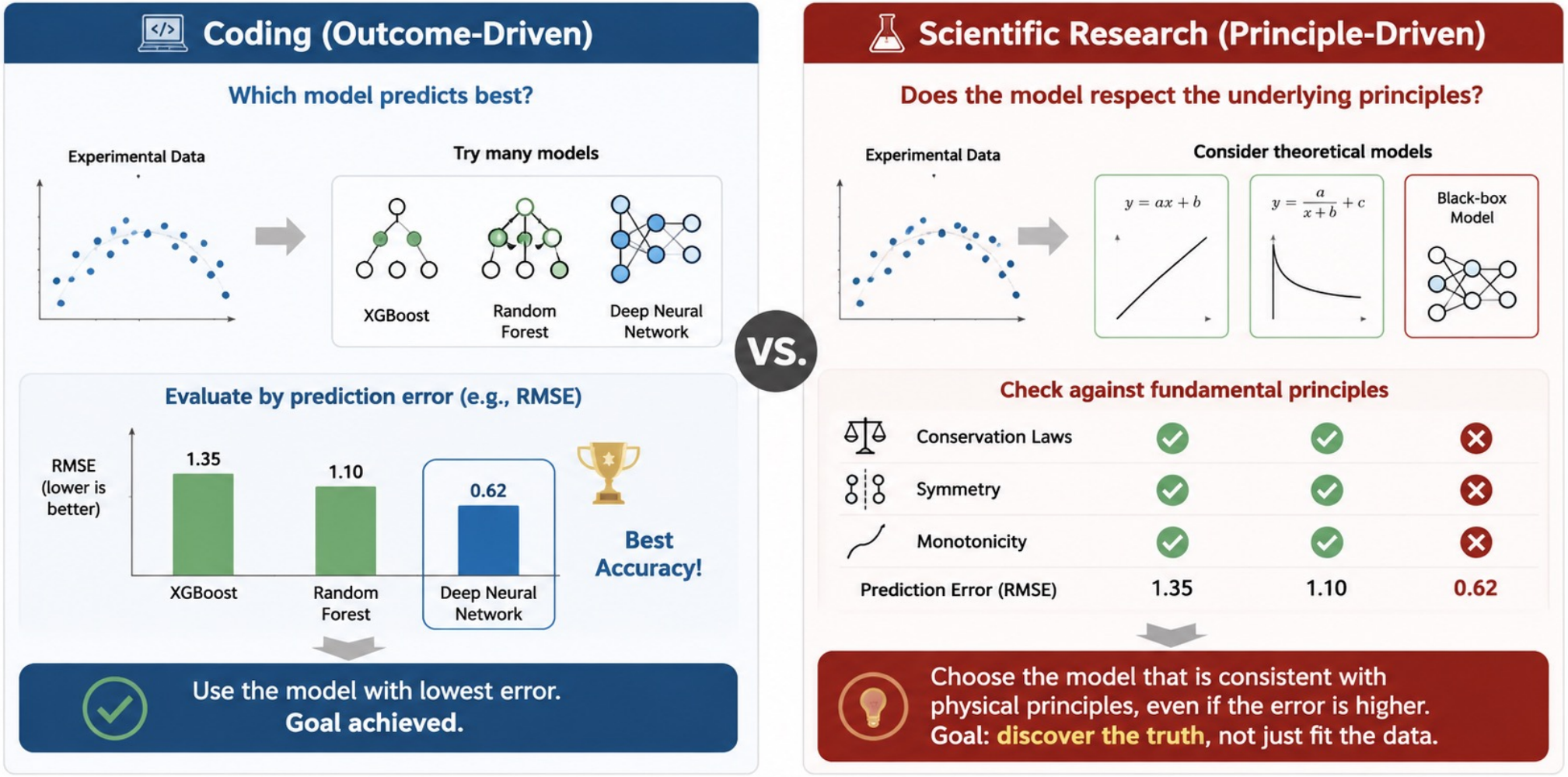
- Inspiration from vibe coding: a high level of automation
- Coding vs. research: fundamental differences in underlying logic





The fundamental research-coding gap

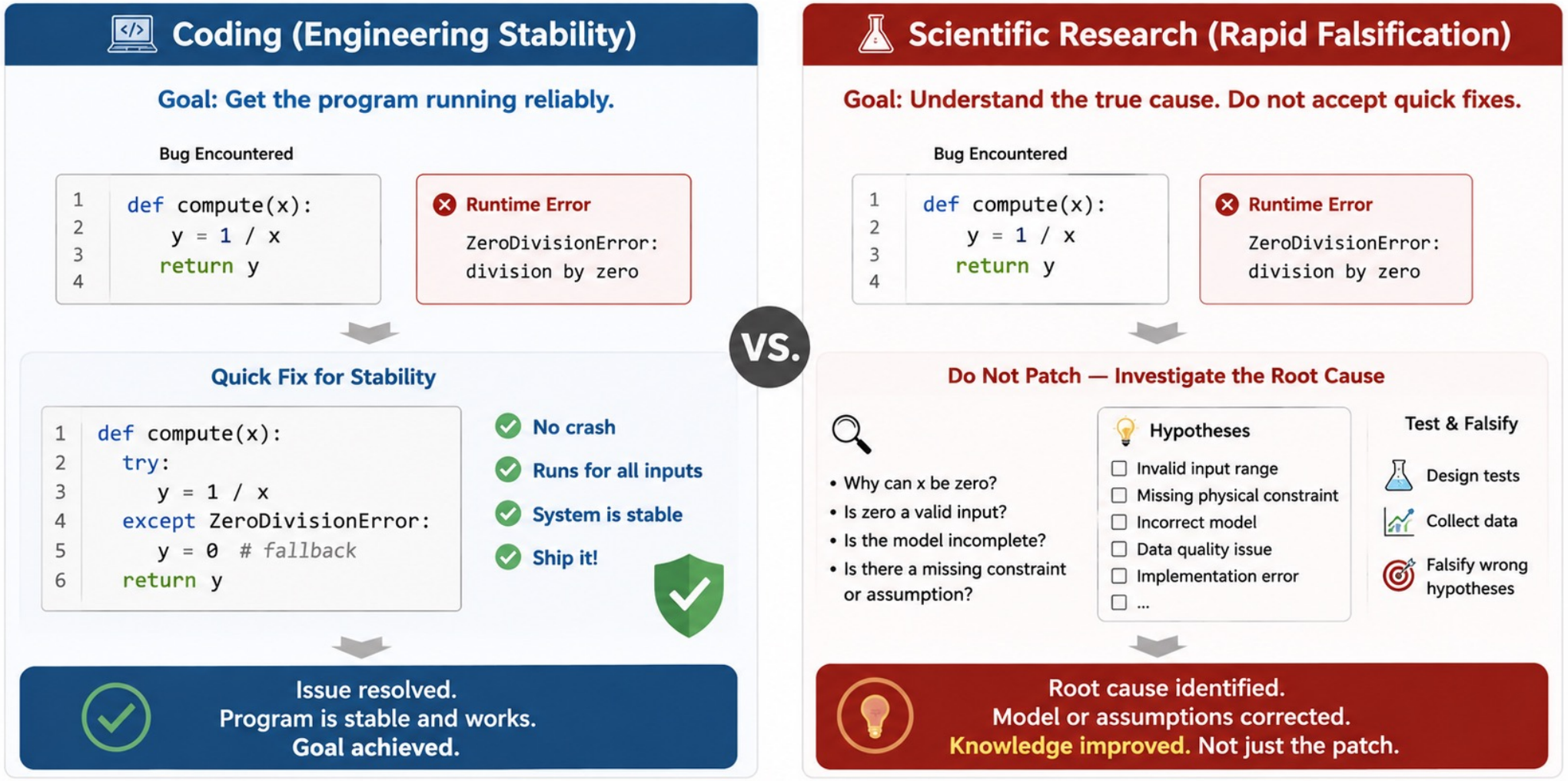
➢ Outcome-driven vs. principle-driven



Key Difference: Engineering optimizes for the best prediction. Science seeks models that are true to reality.

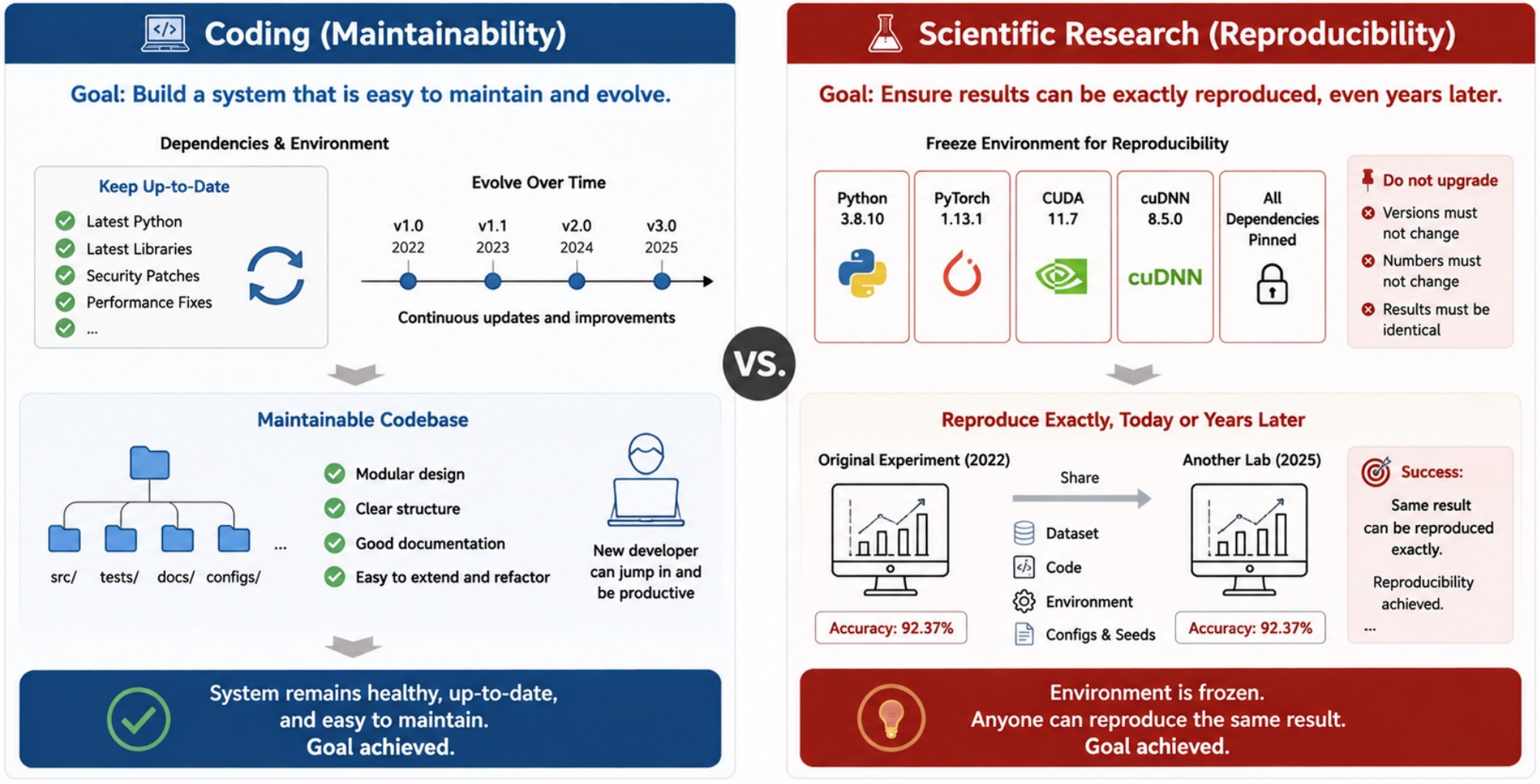
The fundamental research-coding gap

➢ Engineering stability vs. rapid falsification



The fundamental research-coding gap

➢ Maintainability vs. reproducibility



Outline

I. AI for established fields: current status

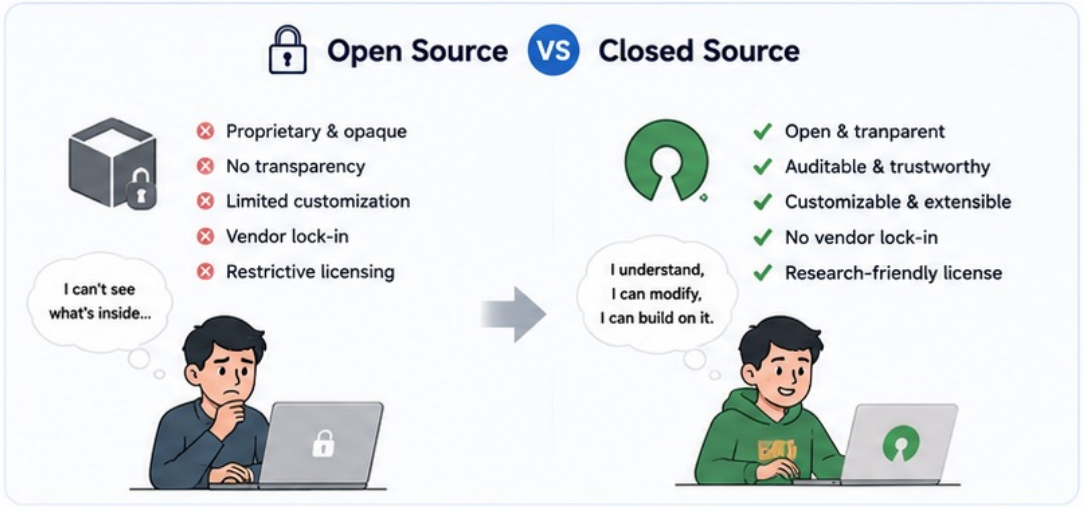
II. AI for frontier fields: challenges

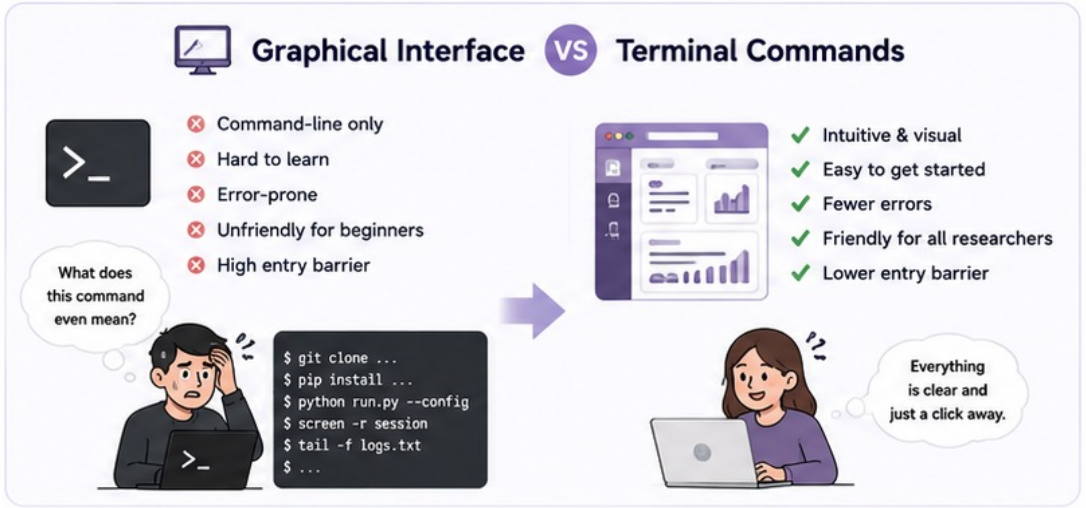
a. Complexity of frontier research tasks
b. The fundamental research–coding gap
c. Accessibility for the broader research community

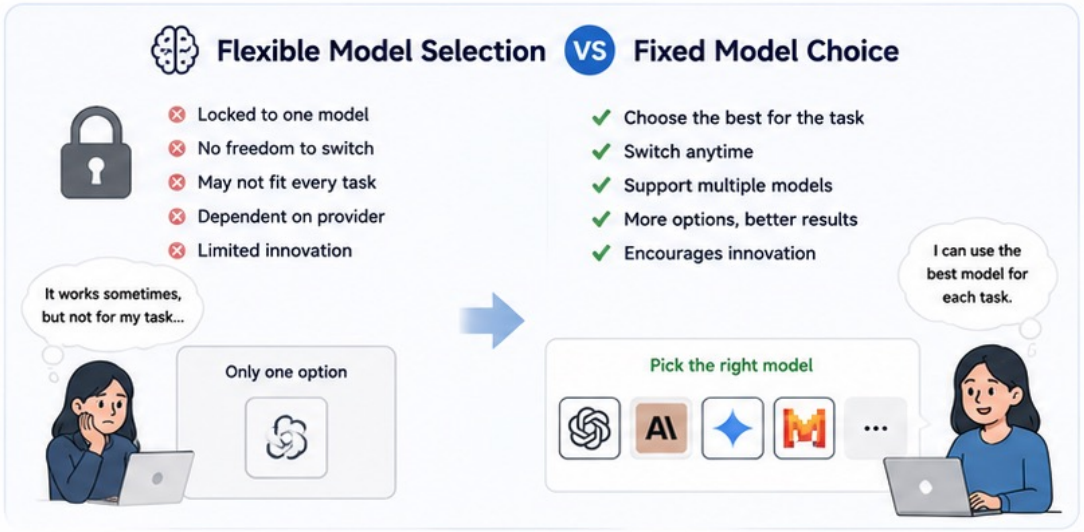
III. Aether: an integrated vibe research platform

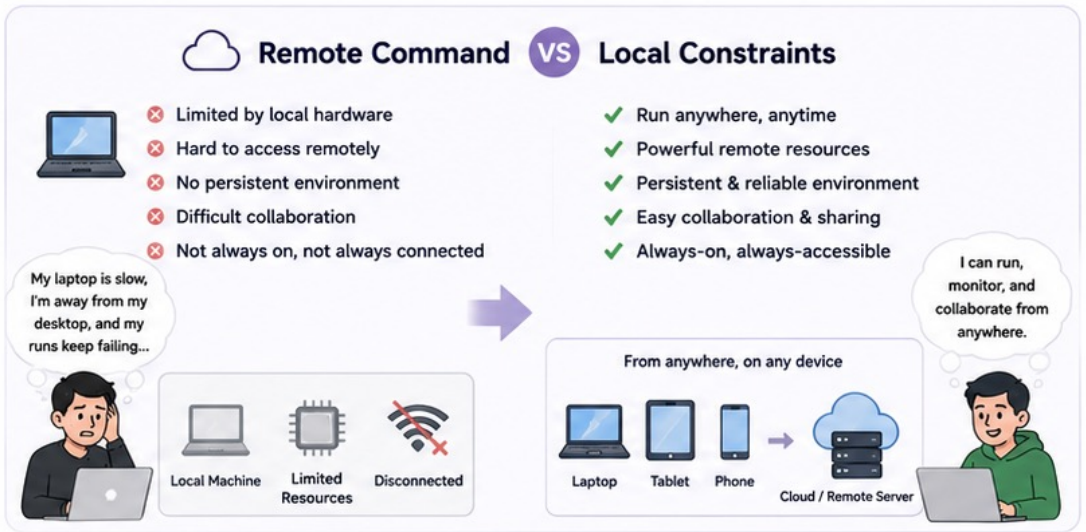
IV. Summary and outlook

Accessibility for the broader research community









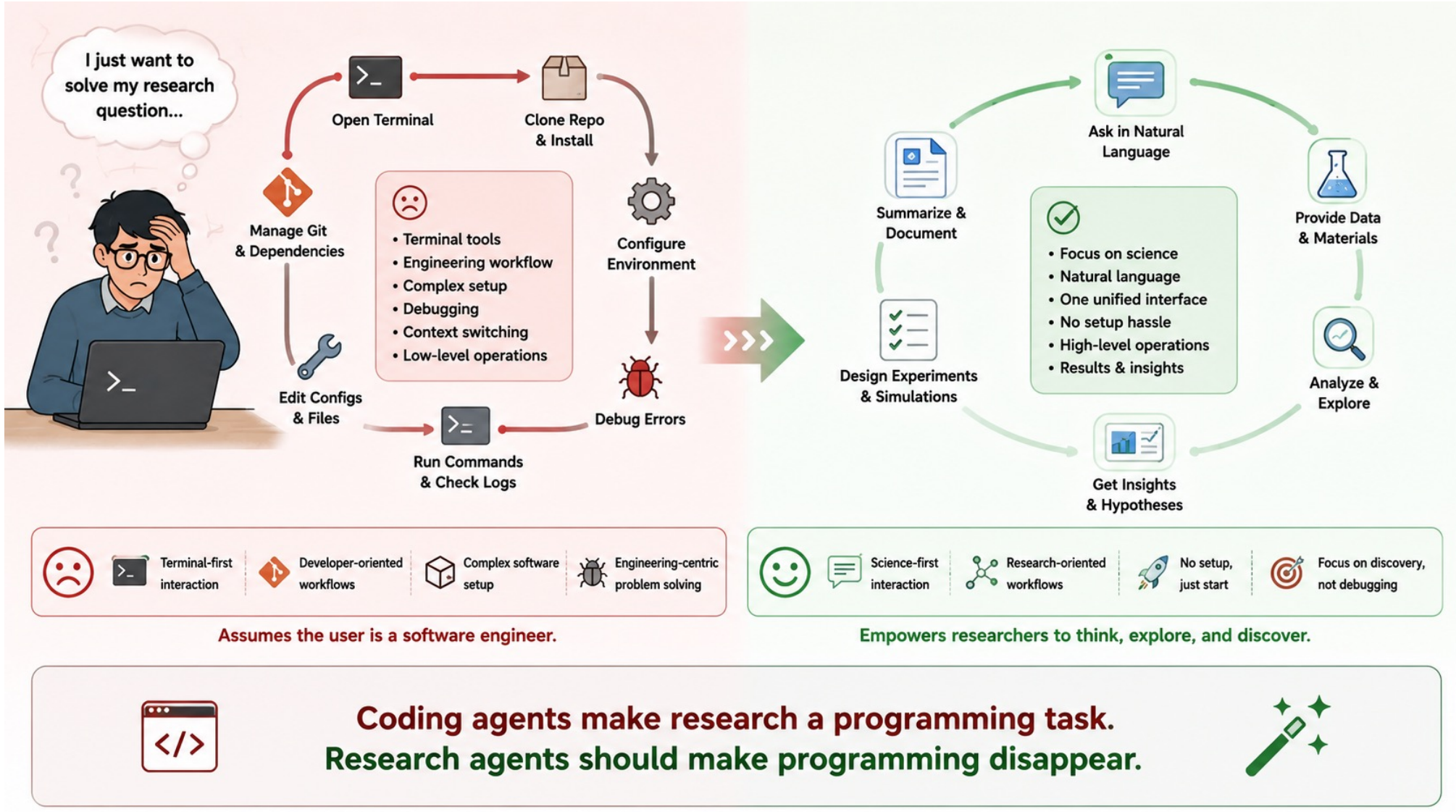
Accessible, Flexible, and Powerful — For Every Researcher.

Open and transparent. Easy to use. Flexible and intelligent. Anywhere, anytime.

AI for research should be accessible to all, not just a few.

Accessibility for the broader research community

- **Not all researchers are skilled programmers**



Outline

I. AI for established fields: current status

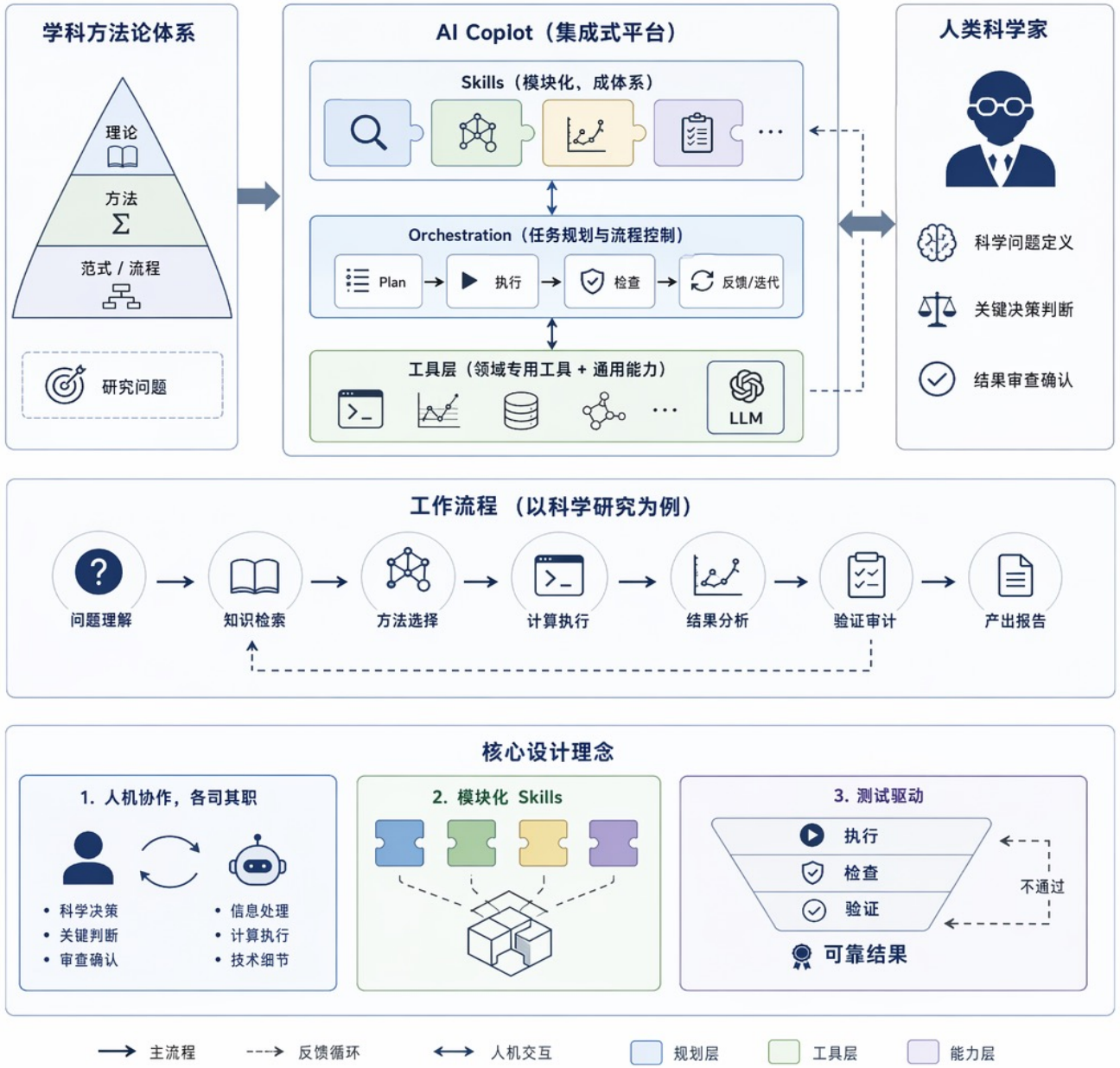
II. AI for frontier fields: challenges

a. Complexity of frontier research tasks
b. The fundamental research–coding gap
c. Accessibility for the broader research community

III. Aether: an integrated vibe research platform

IV. Summary and outlook

From disciplinary methodology to AI copilot



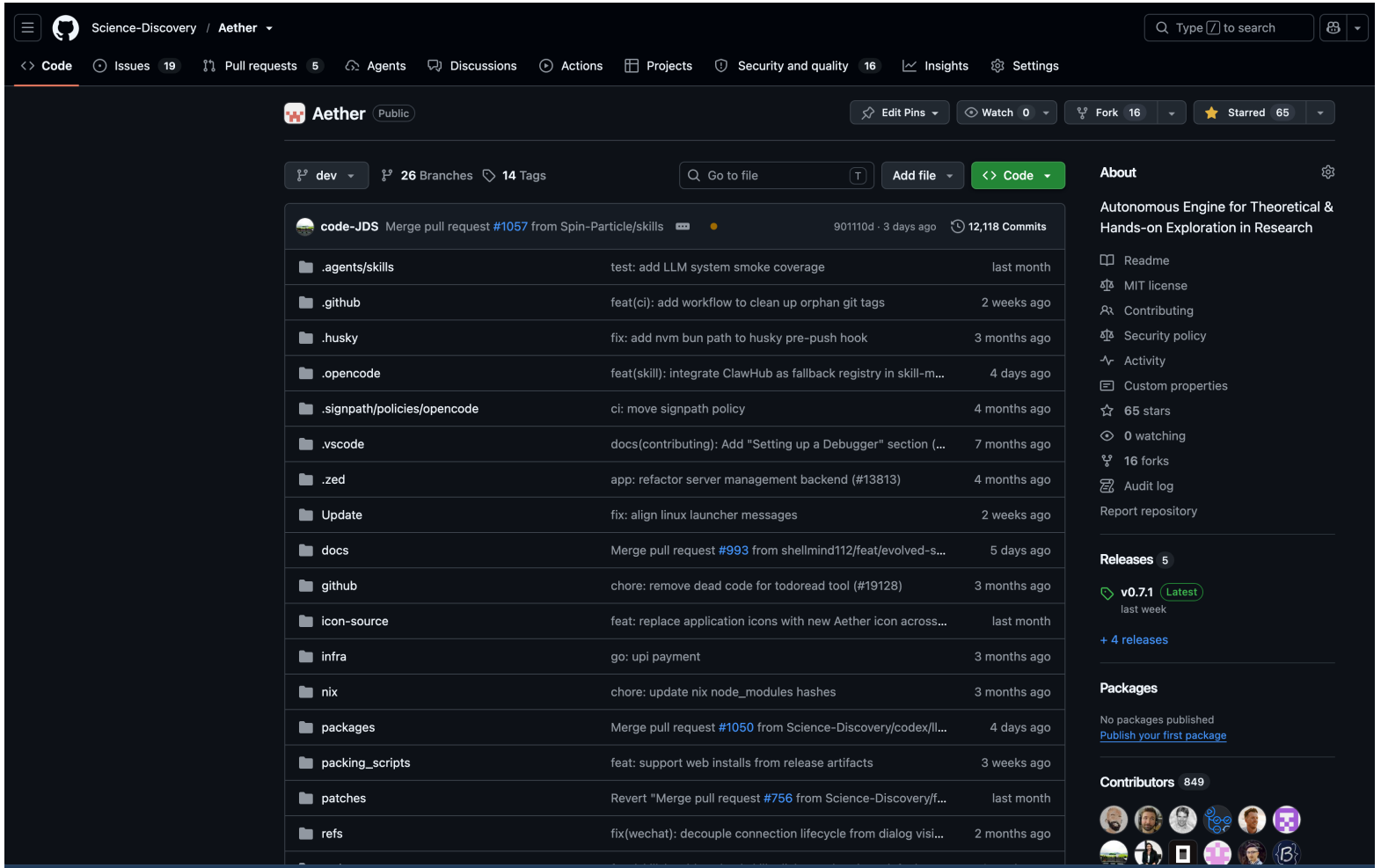
Aether

- **A**utonomous **E**ngine for **T**heoretical & **H**ands-on **E**xploration in **R**esearch
- To achieve a more suitable kind of vibe research
- Based on OpenCode

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

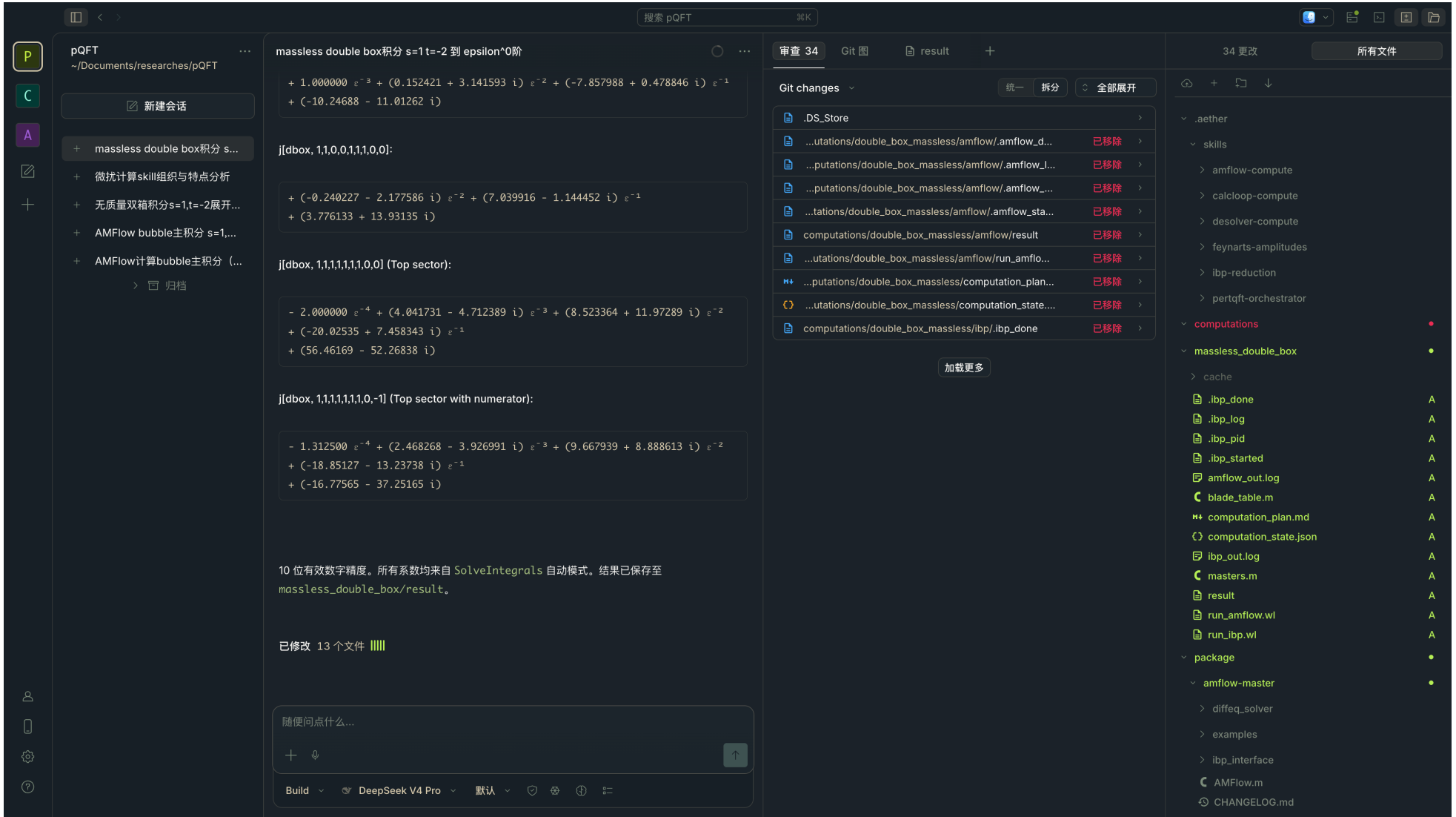
Highlights: inherited from OpenCode

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



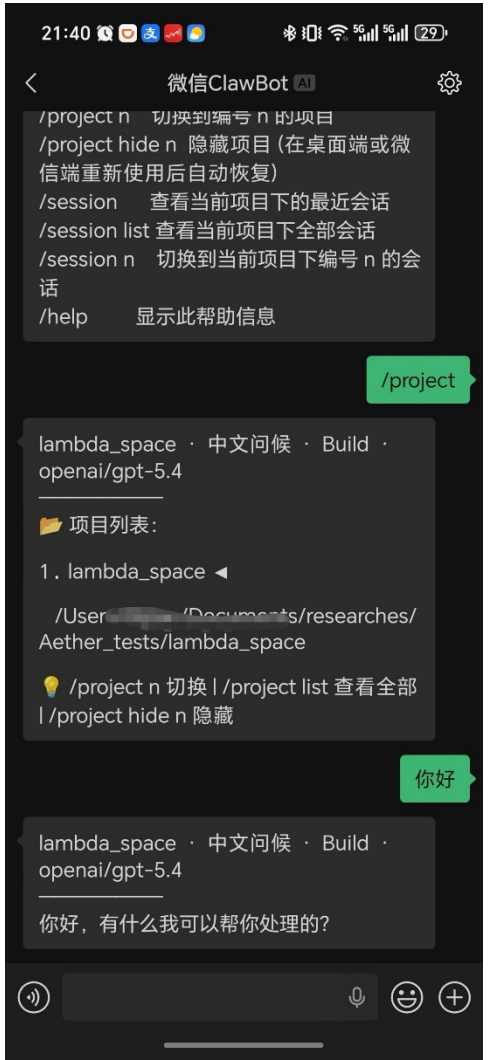
Highlights: highly enhanced UI

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



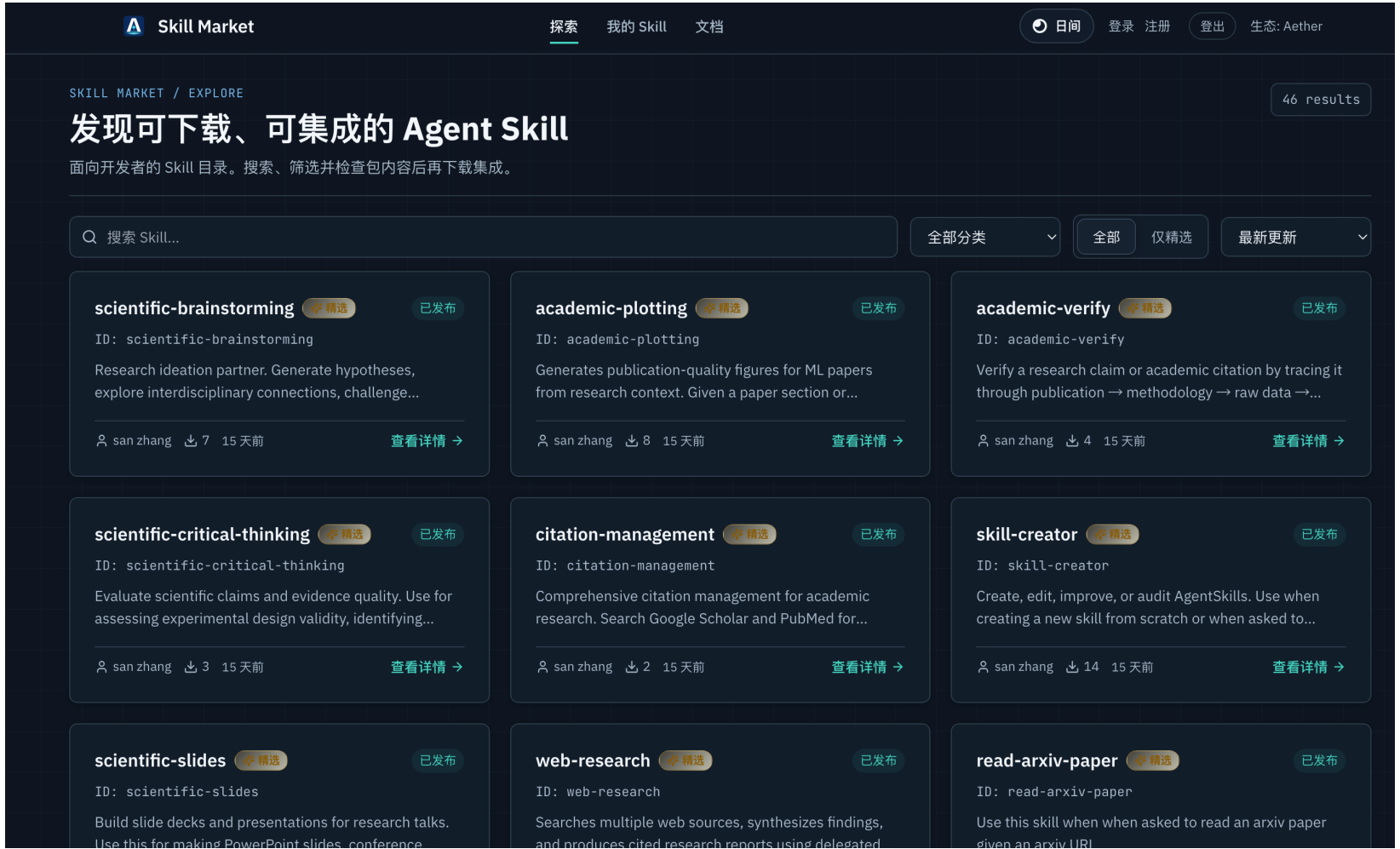
Highlights: mobile devices integration

- **WeChat, QQ and FeiShu integrated, supporting remote work and voice input**



Highlights: research adaptation

- Integrated professional skills/tools for scientific discovery (general & domain-specific)

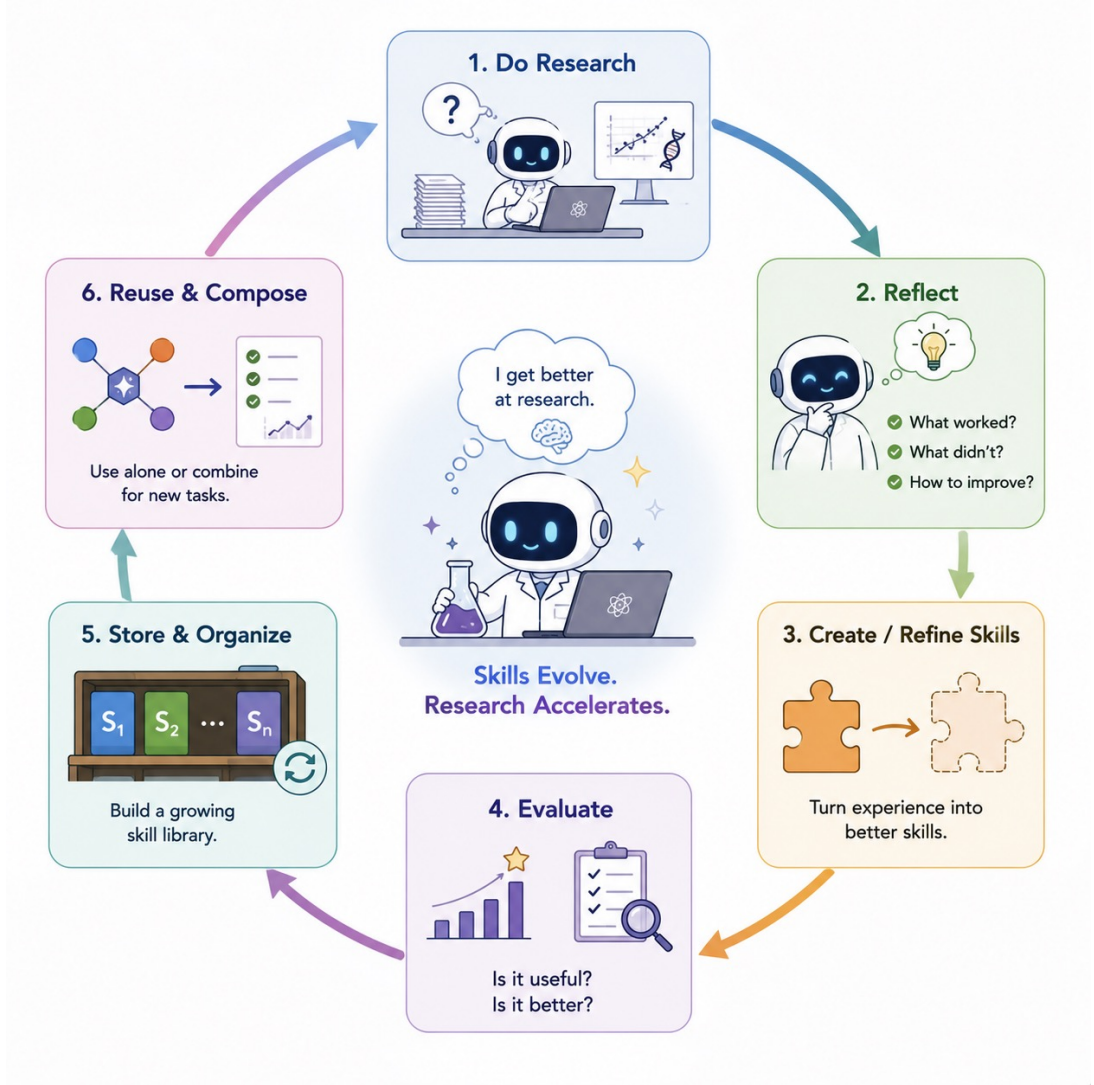


Highlights: research adaptation

- **User-level memory mechanism; the Cron scheduler**

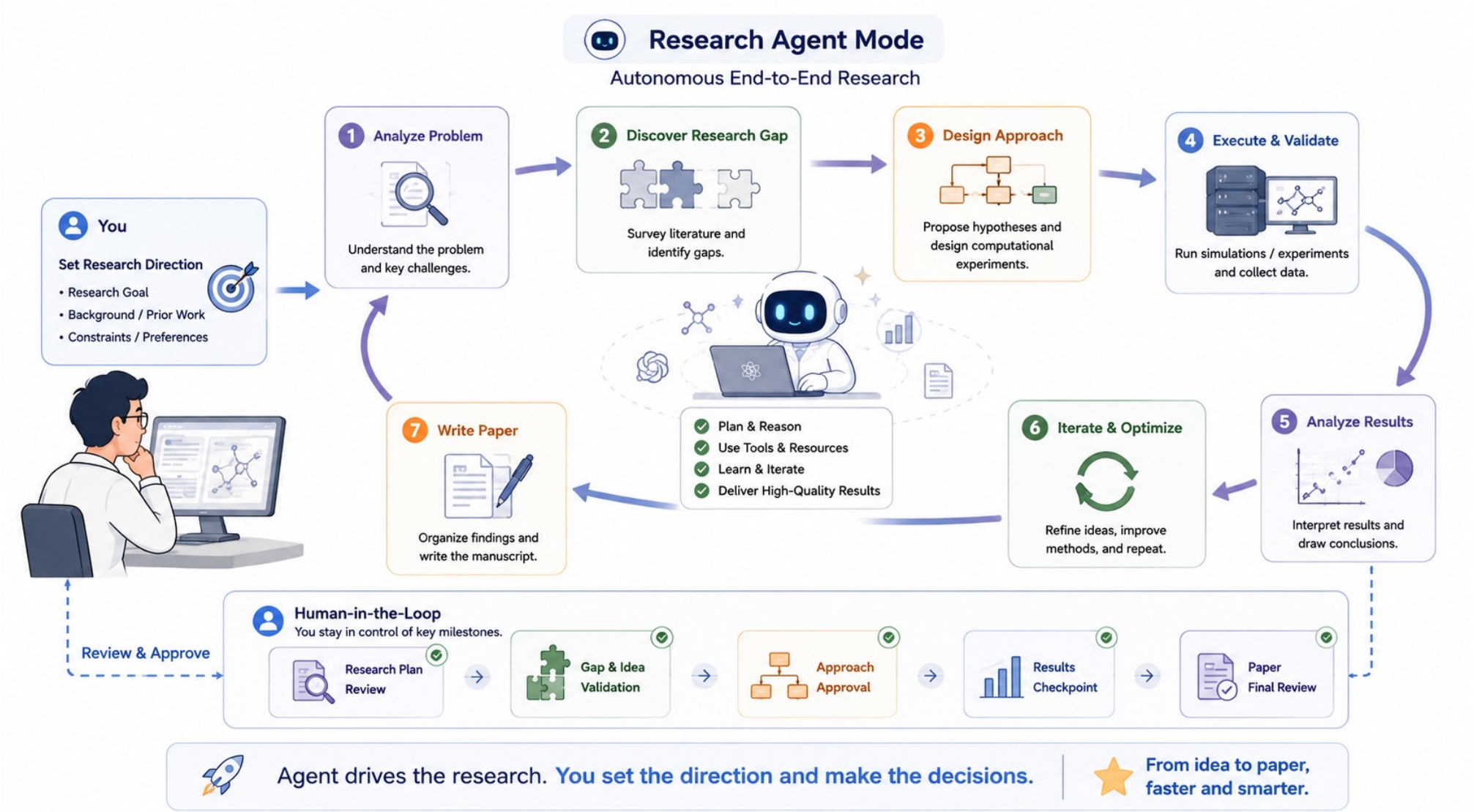
Highlights: research adaptation

➢ Self-evolving scientific research skills



Highlights: research adaptation

➢ **Research agent mode (upcoming public release)**



E.g.1: pQFT skills for high-energy physics

E.g.1: pQFT skills for high-energy physics

E.g.2: research agent mode

费曼积分的FBI-DCT-NIS数值方法

🔬 研究计划确认 — FBI-DCT-NIS 联合费曼积分数值计算方法

前三个阶段已完成。以下是研究计划的摘要，请确认后再进入执行阶段：

📋 研究问题 (4个可证伪问题)

ID	研究问题	可证伪标准
RQ1	FBI-DCT的ε展开流水线能否与NIS自适应重要性采样在统一数值框架中有效结合?	2个2圈基准积分上，30k次DCT评估内|Δ|/AMFlow < 1e-4，方差缩减因子 > 2×
RQ2	dct_nis_python代码库存在哪些架构/算法/鲁棒性不足？需要哪些改进?	识别≥3个实质性缺陷，每个有量化预期收益
RQ3	FBI-DCT-NIS vs AMFlow的逐ε阶精度/效率量化关系	ε^0阶|Δ|/AMFlow < 1e-3，ε^1阶 < 1e-2
RQ4	NIS精度天花板？RQSpline升级是否显著提升?	30轮训练后σ/|I| < 1e-2 且RQSpline改进 < 2× 则已达上限

✅ 可验证的Claims

1. **可行性与精度**: FBI-DCT-NIS在2圈积分上可行，方差缩减 > 2×，ε^0阶精度 < 1e-4
2. **代码审查**: dct_nis_python存在≥3个实质性缺陷，改进后可支撑实用化
3. **精度对比**: ε^0阶|Δ|/AMFlow < 1e-3，与成熟工具可比
4. **误差量化**: NIS统计误差与DCT截断误差可通过误差传播公式分离

Outline

I. AI for established fields: current status

II. AI for frontier fields: challenges

a. Complexity of frontier research tasks
b. The fundamental research–coding gap
c. Accessibility for the broader research community

III. Aether: an integrated vibe research platform

IV. Summary and outlook

Summary and outlook

- AI for scientific discovery is evolving from isolated algorithmic tools toward integrated research platforms
- The next generation of AI research agent should augment the entire scientific discovery process, rather than individual research tasks
- Because scientific research fundamentally differs from coding, dedicated research agents—not generic coding agents—are essential for effective AI-assisted scientific discovery
- General and domain-specific research skills will continue to be developed by experts, ultimately becoming valuable training resources for future foundation models

Thank you !