High Energy THeory Forum (HETH-Forum)
高能理论论坛 / 中国科学院高能物理研究所理论室

AI+ 科研：一年的实践

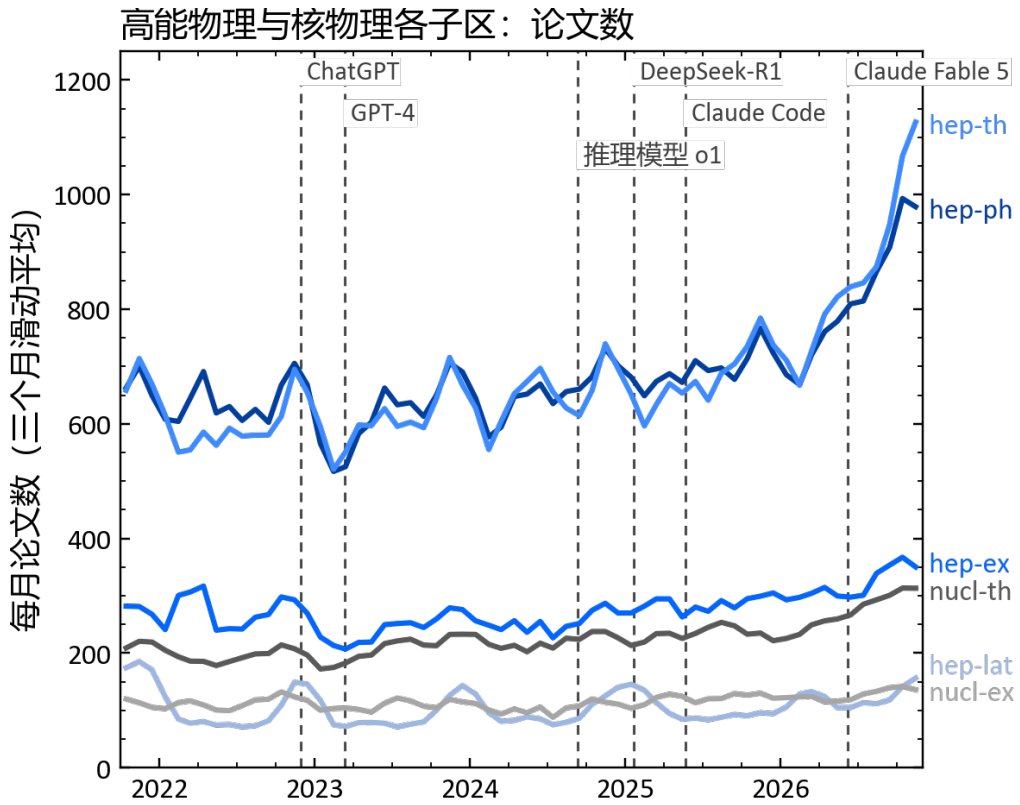
郭奉坤

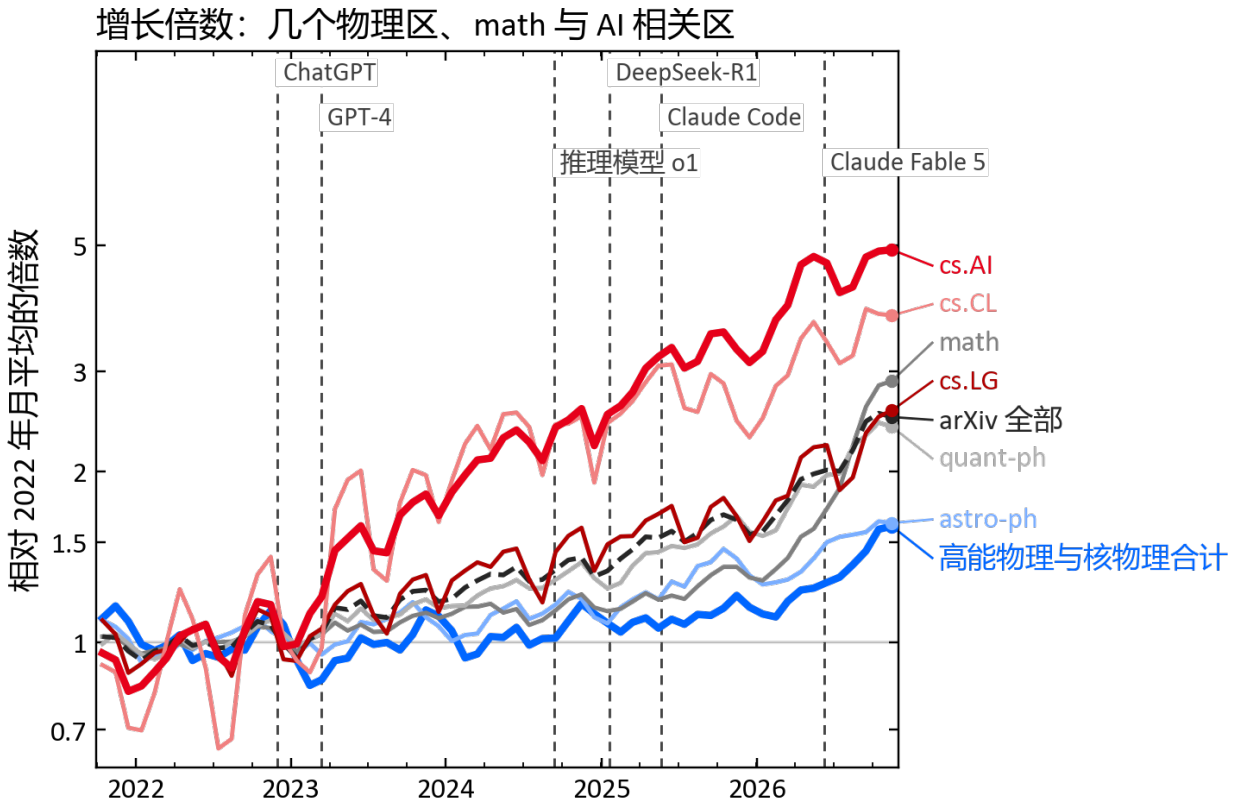
中国科学院理论物理研究所

2026 年 11 月 25 日

arXiv 投稿量

- 2025 年相对 2022 年的月均论文数：cs.AI 约 3 倍，cs.CL 约 2.6 倍
- 2026 年 1–10 月较上年同期：hep-th 增长 24%，hep-ph 增长 17%
- 2026 年 10 月 arXiv 全站新投稿 42,152 篇，图中各区含交叉列入

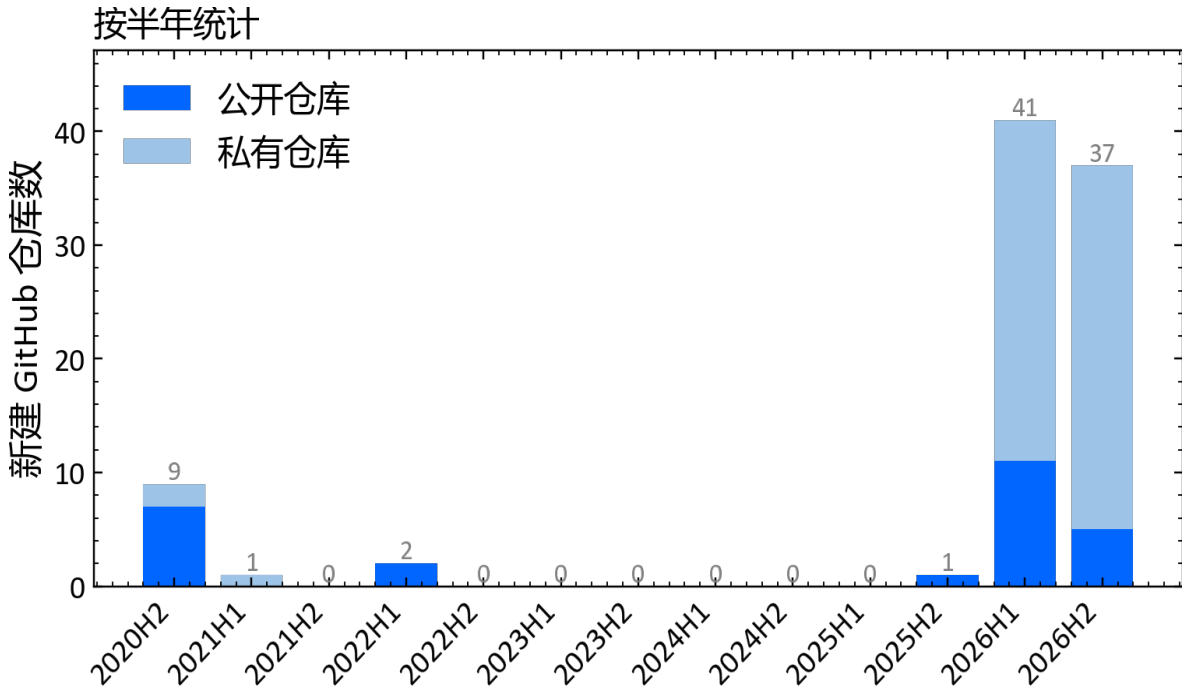




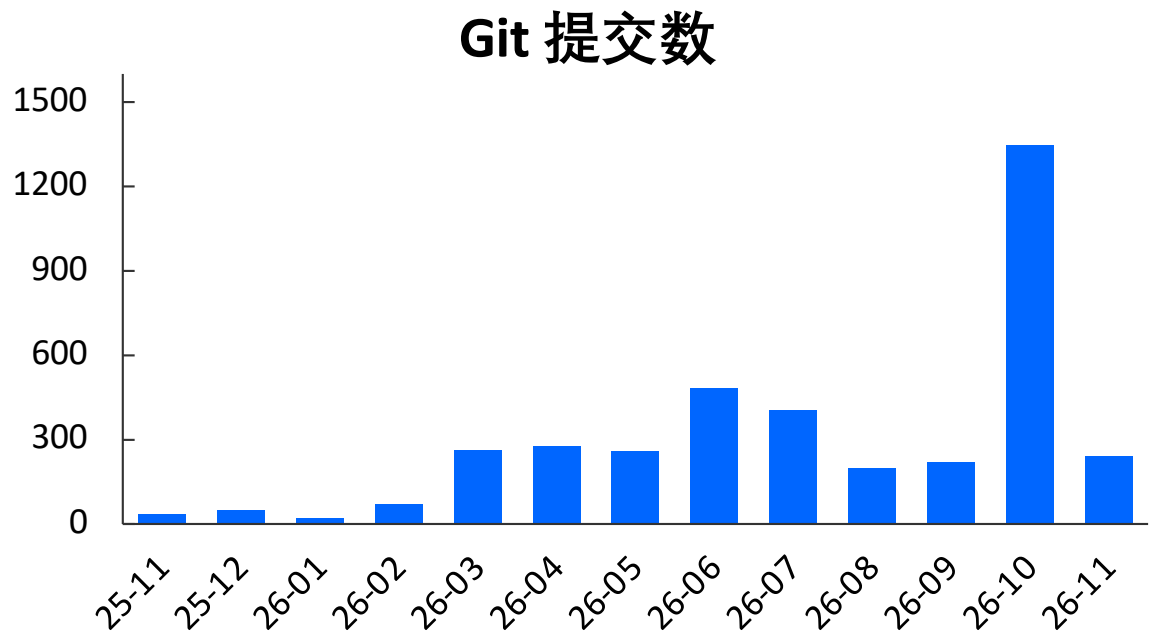
数据：arXiv API 与全站月投稿统计。2026 年 11 月按前 24 天折算。

解放生产力的革命性工具

- 2025 年 11 月 至今新建 79 个 GitHub 仓库， 2020 年 7 月至 2025 年 9 月共 12 个

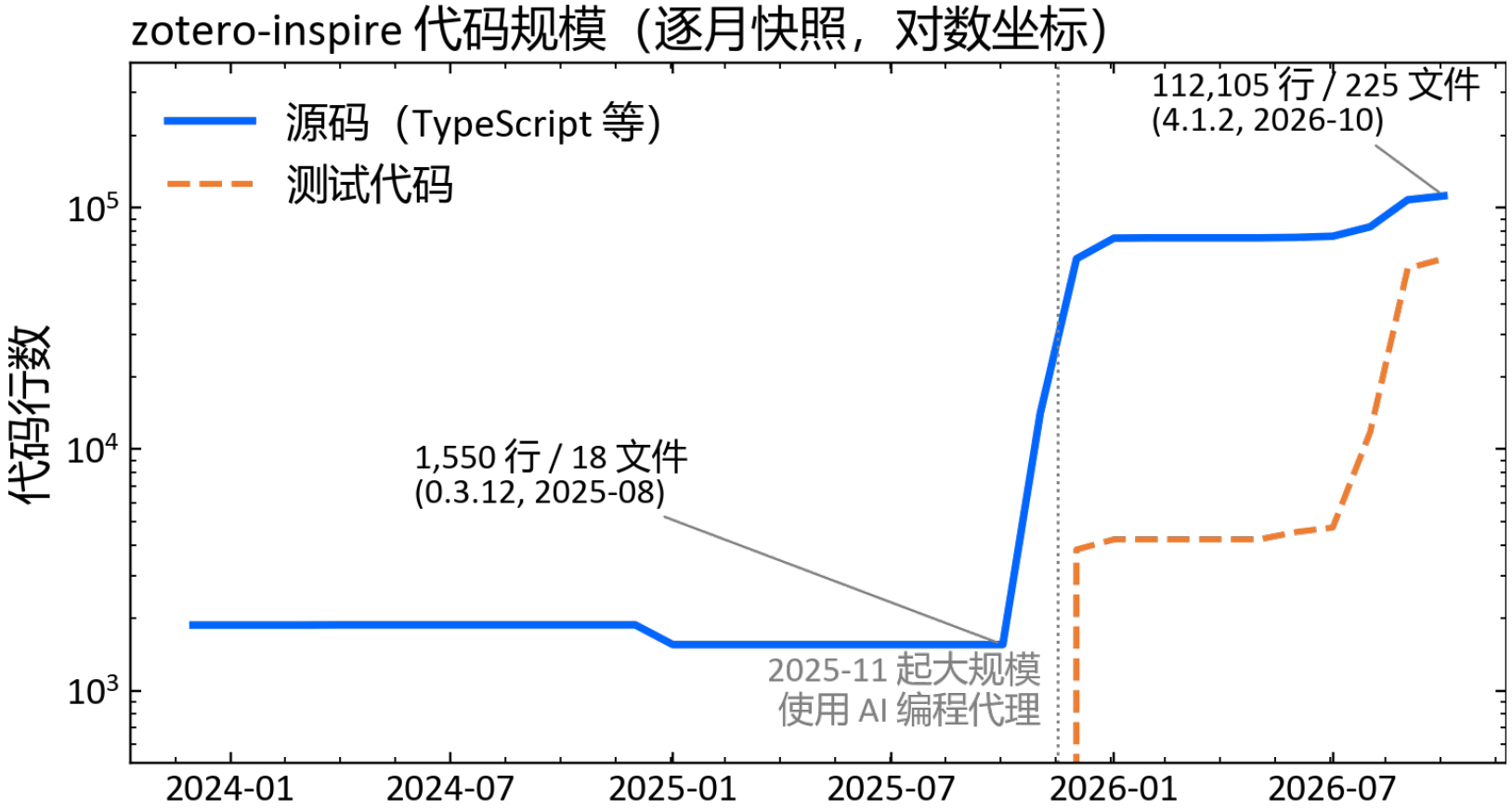


- 九个主要代码库的每月提交数



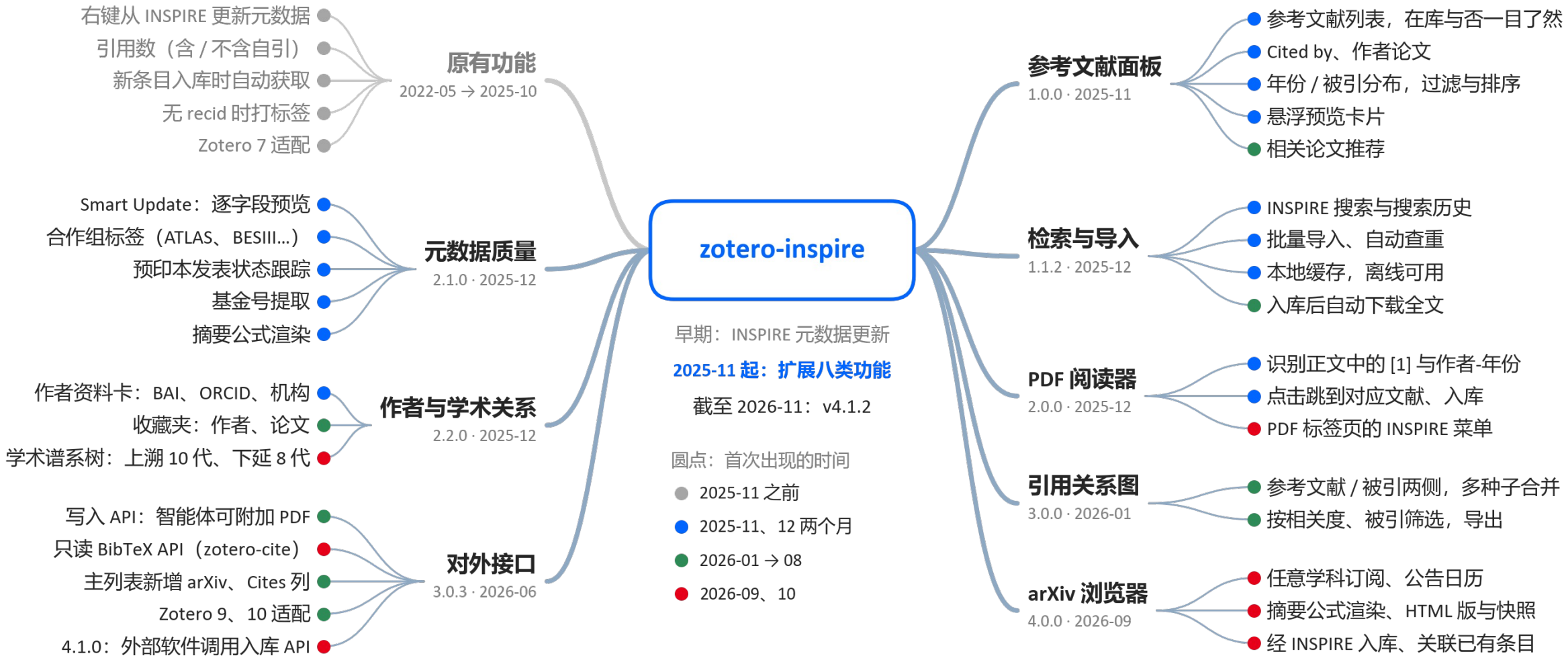
zotero-inspire 的开发

- 2025 年 11 月开始用 AI 编程智能体扩充 zotero-inspire 的功能
- 0.3.12 源码 1,550 行。4.1.2 源码 112,105 行，测试 61,636 行



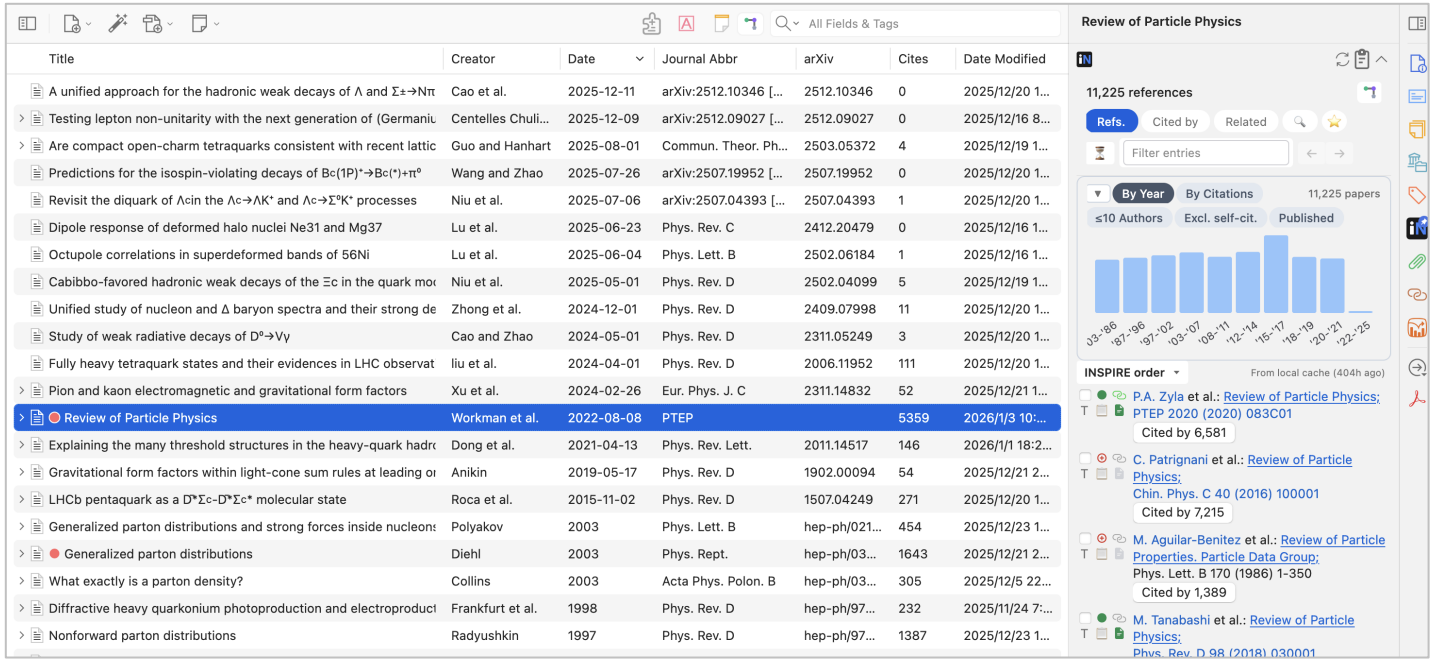
从更新 INSPIRE 条目，扩展到 arXiv 浏览、引文检索和数据接口。

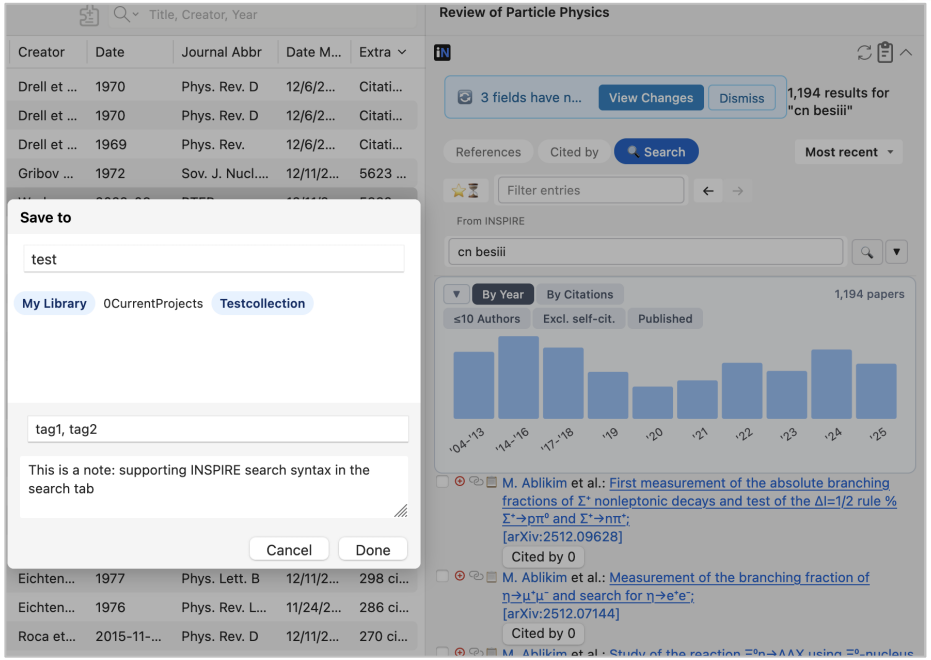
zotero-inspire 的功能



日常使用中发现的问题和 GitHub issue 用于改进插件。

zotero-inspire： 参考文献、引用与检索

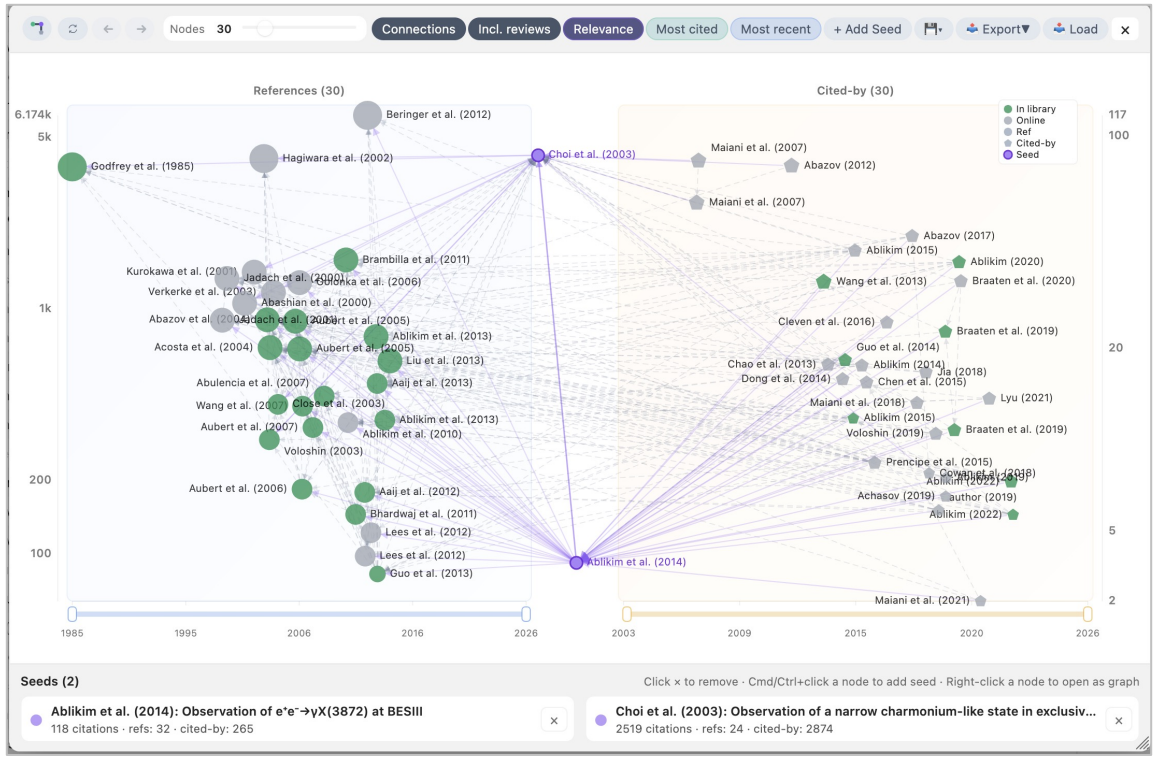




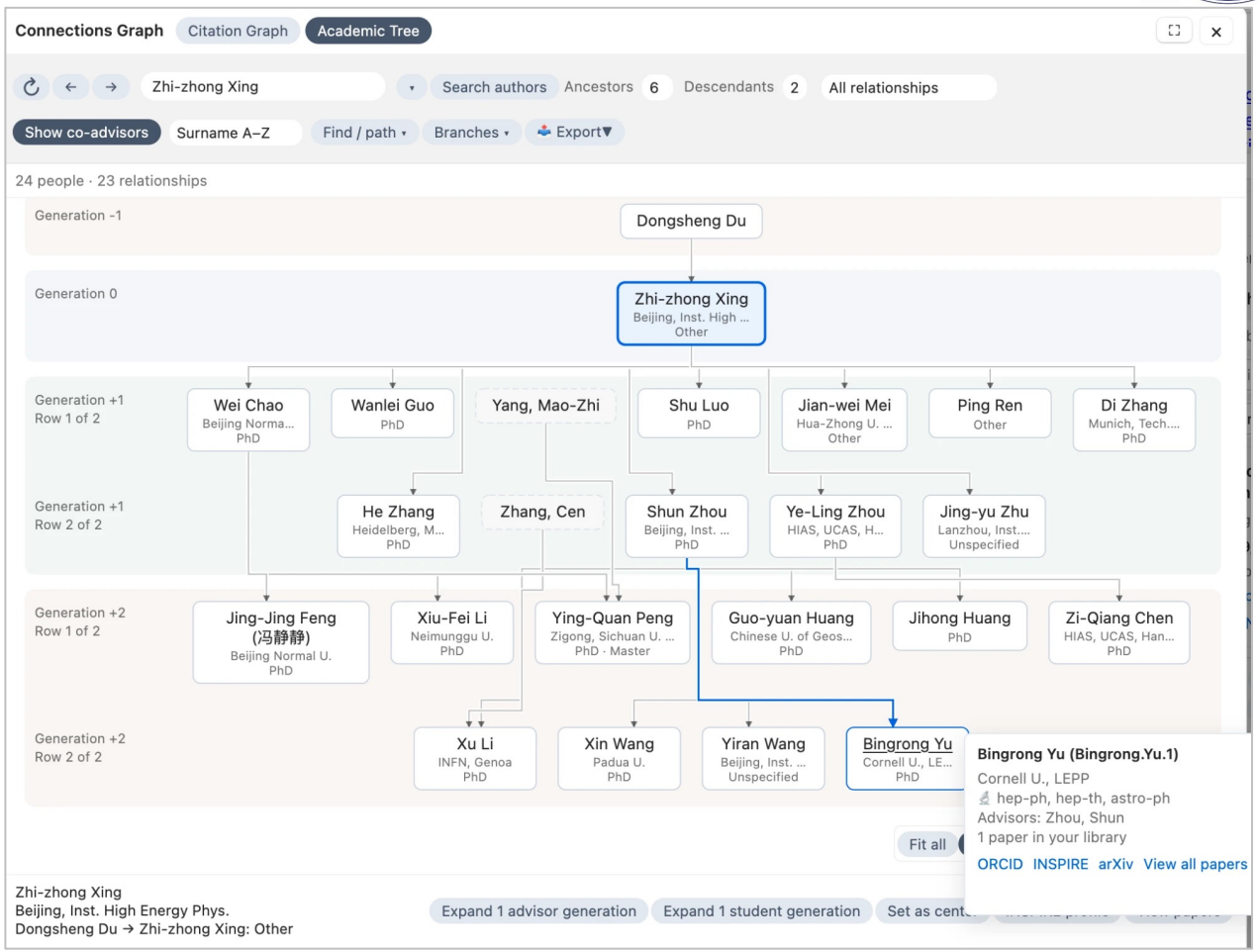
- 左图侧栏： PDG 综述的全部 11,225 条参考文献
- 绿点表示已在库，红色加号表示可一键入库
- 年份 / 被引分布；可只看 ≤10 位作者、去自引、已发表
- 主列表新增 arXiv 号与引用数两列

- 右图（2025-12 界面）： 面板内用 INSPIRE 检索语法，"cn besiii" 得到 1,194 篇
- 多选批量入库：选合集、加标签与笔记，自动查重
- Smart Update： 更新前逐字段预览

zotero-inspire：引用关系图与学术谱系树

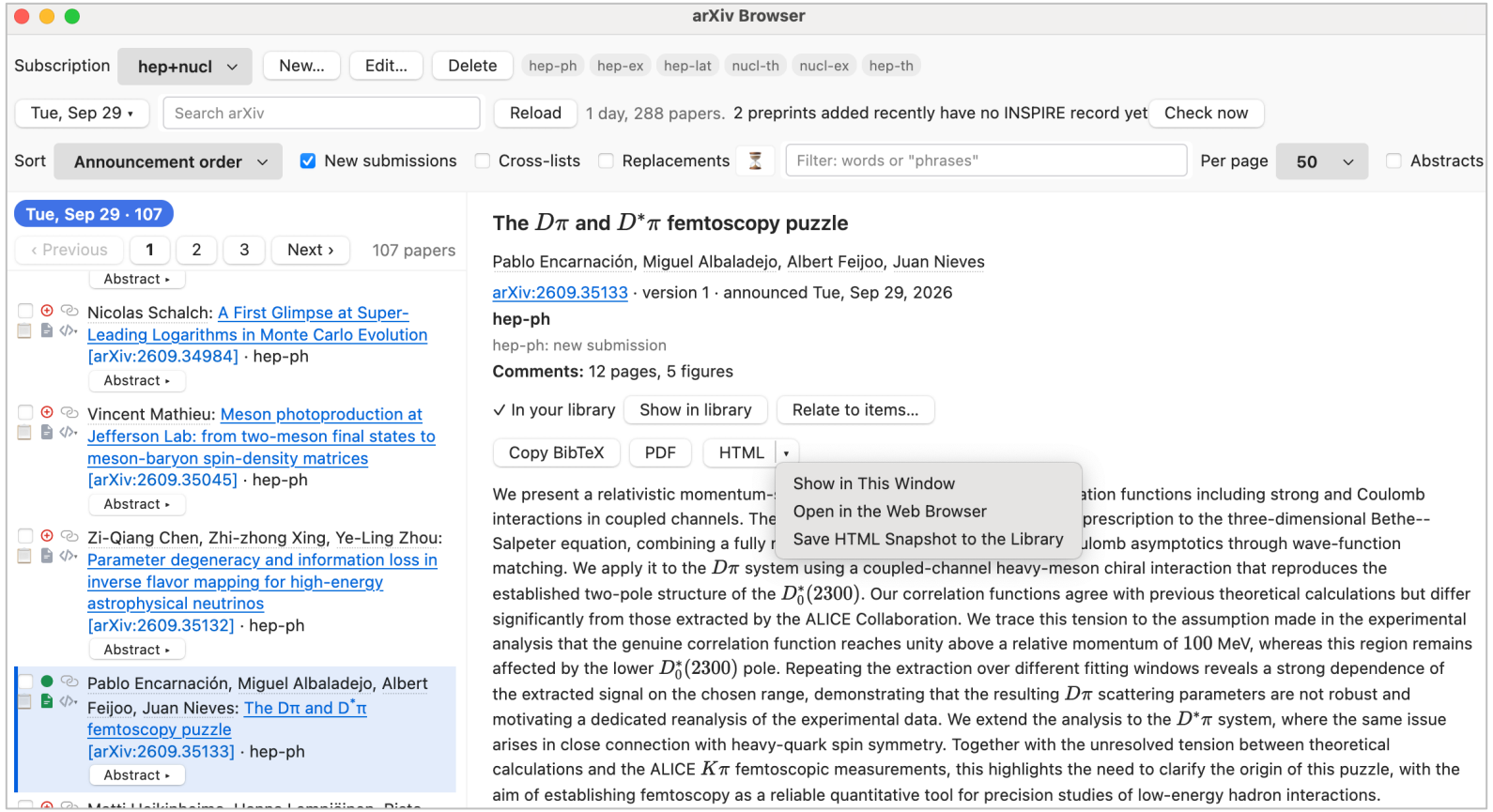


- 左为参考文献、右为被引，纵轴为被引数；图中以 X(3872) 的发现论文（Belle 2003）与 BESIII 2014 年的论文合并为种子
- 按相关度、被引、时间筛选；可导出图与 BibTeX



- 导师—学生关系取自 INSPIRE 作者记录（上溯 10 代、下延 8 代）
- 图中以邢老师为中心，下延两代学生

zotero-inspire： arXiv 浏览器与对外接口

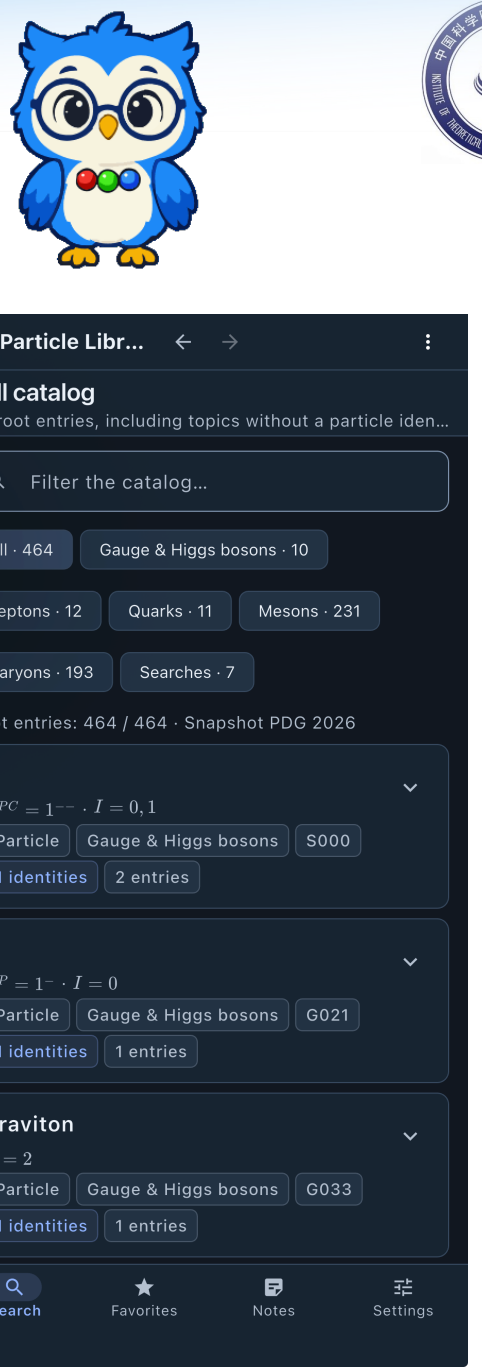


- 4.0.0（2026-09-29）：Zotero 里的独立窗口，可订阅任意学科
- 按公告日浏览；摘要里的公式直接渲染
- 入库时优先用 INSPIRE 的记录，没有再用 arXiv 的数据
- 可关联到库中已有条目；HTML 版可在窗口内阅读、存快照

- 对外接口
  - ❑ 写入 API（3.0.3）供 hep-mcp 给条目附 PDF
  - ❑ 只读 BibTeX API（3.1.0）供 VS Code 插件 zotero-cite 从 INSPIRE 取 BibTeX
  - ❑ 4.1.0 新增入库 API，Particle Library 等外部软件可以调用它把文献加入 Zotero

Particle Library

- 离线查阅 PDG 数据。用 DeepSeek Harness 与 DeepSeek-V4.1-Flash-1017，在一个周末完成初版
- 0.2.4（11 月 15 日）：查粒子数据、衰变与历次测量



模型、智能体与工作流框架

- 大语言模型（LLM）根据已有上下文预测下一词元。对话界面也可以连接搜索和计算工具
- 智能体（agent）读文件、查数据库、运行程序，根据返回结果继续工作
- 工作流框架（harness）负责工具调用和执行循环，并管理权限、上下文、记忆和任务
- 同一模型可以用在不同框架里。比较实际表现时，要看它能调用什么工具

Harness（工作流框架）：Claude Code、Codex、OpenCode、Cursor …

Agent（智能体） = 模型 + 工具 + 循环

LLM（大语言模型）

只会"说"：按上下文预测下一个词；知识有截止日期
没有工具就只能凭记忆——"编造文献"的根源
GPT、Claude、Gemini、DeepSeek、Qwen、Kimi、GLM …

工具：读写文件、运行代码、检索文献与数据库、浏览网页、调用其他程序
循环：计划 → 执行 → 观察结果 → 修正，直到完成或预算耗尽

记忆与规则文件（AGENTS.md） 权限与沙箱 子智能体与并行
skills（把经验写成文件） MCP（把数据库接成工具） 后台任务、成本与停止条件

Chatbot（对话界面）

网页 / App 里的 LLM：
ChatGPT、豆包、Kimi …

同一个模型，外壳不同：
是否联网、搜索、跑代码
由厂商决定

多数人对 AI 的印象
来自这一层

agent 指"模型 + 工具 + 循环"这个模式；工具与循环本身由 harness 的程序实现，所以 harness 既承载也实现 agent

从对话到智能体

- 对话用于讨论问题。需要读写项目和运行程序时，使用编程智能体
- 可以从一个经常需要的小程序入手，边试用边修改
- 同一任务可请不同模型独立完成，比较结果并追查分歧

层次	常用工具	适合做什么	开始使用时间
对话	ChatGPT、 Kimi、 DeepSeek、 千问	讨论物理问题、初筛文献、问"这一步怎么来的"	2023 年起
编辑器里的 AI	Cursor、 Windsurf、 Qoder、 Trae、 VS code 中的各种 AI 插件	在代码旁边逐处修改，改动看得清	2025-11（zotero-inspire）
终端里的智能体	Codex、 Claude Code、 OpenCode、 Kimi Code、 qwen-code	读整个项目、跑程序、改文件、看结果再改	2025-12 Codex； 2026-05 Claude Code
规则、 skills 与数据接口	AGENTS.md、 skills、 MCP； nullius	把纠正写成规则，接上 INSPIRE 等数据库，复核有章可循	2026-02 起

常用的智能体框架

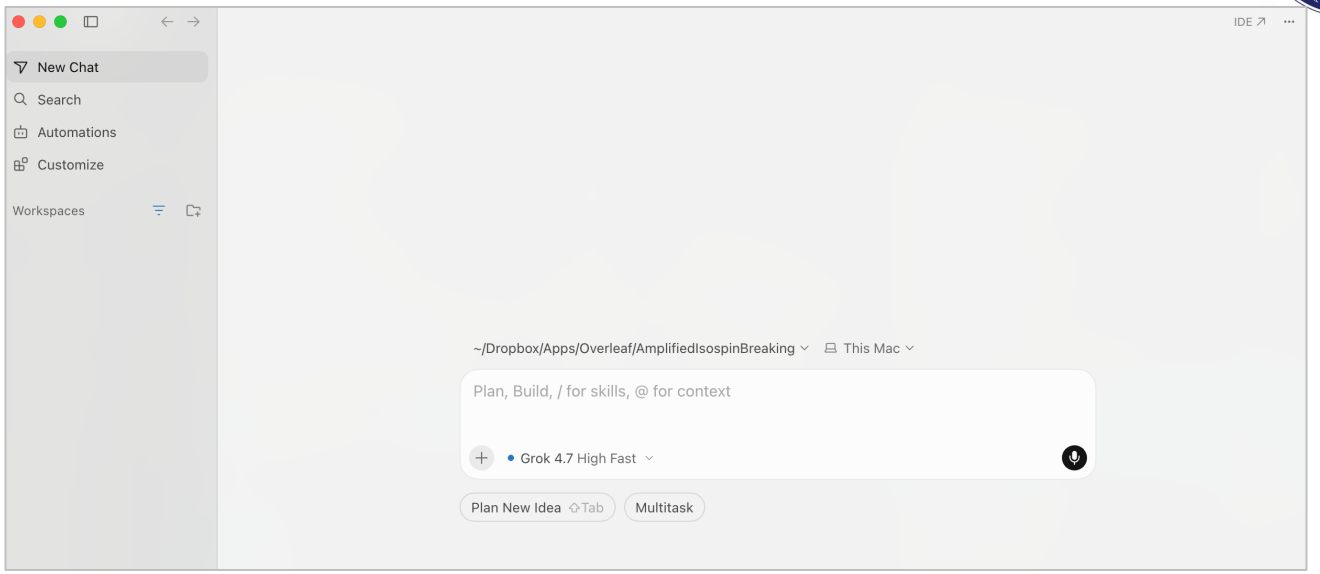
- Claude Code 和 Codex 用于日常研究
- 用其他模型独立复核

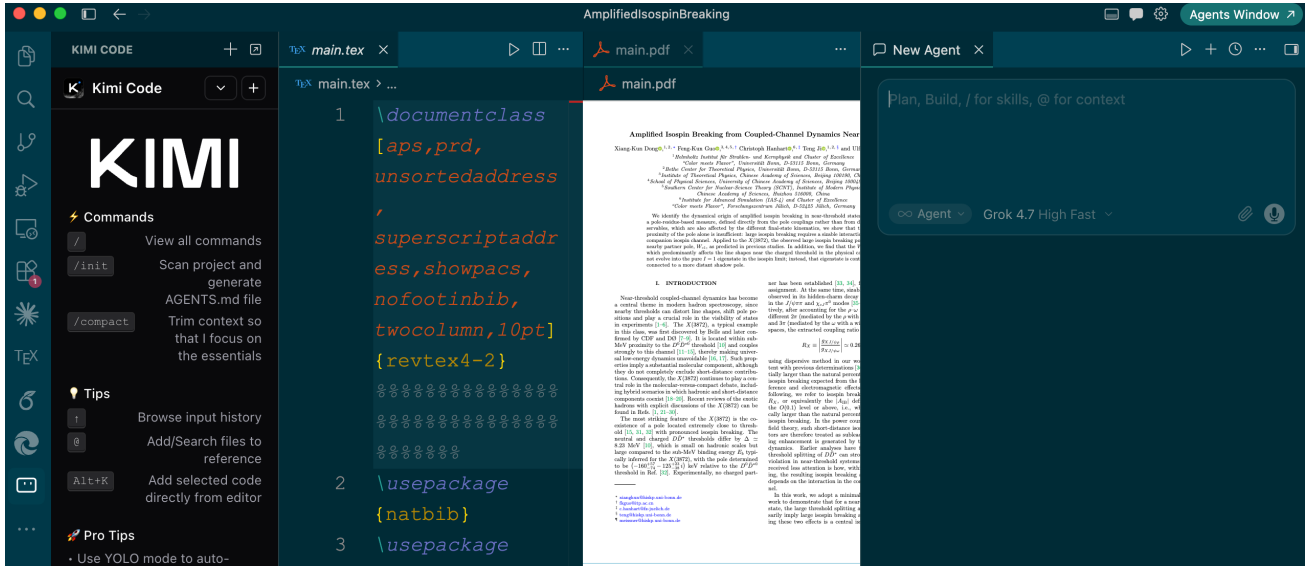
产品	厂商	形态	可接模型	用途
ChatGPT / Codex	OpenAI	网页、桌面、终端	GPT 系	主力之一：文献初筛、协调、实现、复核
Claude Code（CLI / 桌面）	Anthropic	终端 + 桌面	Claude 系	主力之一：子智能体、skills、记忆文件
Cursor	Anysphere	编辑器 + 命令行 + 云端	多家	zotero-inspire 早期（2025-11）；配 LaTeX 插件写论文
OpenCode（开源）	社区	终端	多家 + 开放模型	接 GLM 做只读复核。免费模型和数据使用条款见官方 Console
DeepSeek Harness（2026-08 开源）	DeepSeek	终端 / 桌面 / 本机网页	默认 DeepSeek，也可接其他厂商	两天完成 Particle Library 初版
Kimi Code	月之暗面	桌面 + 终端 + 编辑器插件	Kimi + 开放模型	做第三家族复核
Qoder ｜ Qwen Code	阿里	桌面 ｜ 终端	各自模型	未深度使用
Gemini CLI；Trae；CodeBuddy；Copilot	Google / 字节 / 腾讯 / GitHub	终端 / 编辑器	各自模型	未深度使用

工具文档：developers.openai.com/codex；code.claude.com/docs；cursor.com/docs；opencode.ai/docs/console/models
DeepSeek Harness：github.com/deepseek-ai/deepseek-harness；Agent Skills：agentskills.io；MCP：modelcontextprotocol.io

Cursor

- 最初是基于 VS Code 的编辑器，现在也是智能体平台：编辑器里的 Agent、命令行、在云端虚拟机里跑的 Cloud Agents、代码审查 Bugbot；可选各家模型
- 在编辑器里，改动以差异显示在代码旁边，看清审核再接受
- 写论文：装 LaTeX Workshop 或 Overleaf Workshop 插件，在 Cursor 里编辑、编译，或直接同步 Overleaf 项目；AI 逐段修改，逐处审阅
- 2025 年 11 月用 Cursor 为 zotero-inspire 加入参考文献面板（1.0.0）
- 入门：先用它改自己的小程序





ChatGPT 与 Codex

- ChatGPT：讨论物理问题、初筛文献、读论文时追问；免费档可用。凭记忆回答的引用要另查
- Codex：OpenAI 的编程智能体，有终端、桌面应用和编辑器插件；在本地读代码、跑程序、调研文献、改文件，…
- 项目规则写在 AGENTS.md，每次会话自动读入；会话可接着做
- 2025 年 12 月起用于文献检索、代码实现和独立复核

```
$ codex                    进入交互会话
$ codex exec "…"           一次性任务（不交互）
$ codex resume --last      接着上次的会话做
$ codex review             审查当前的改动
/init                      生成项目的 AGENTS.md
```



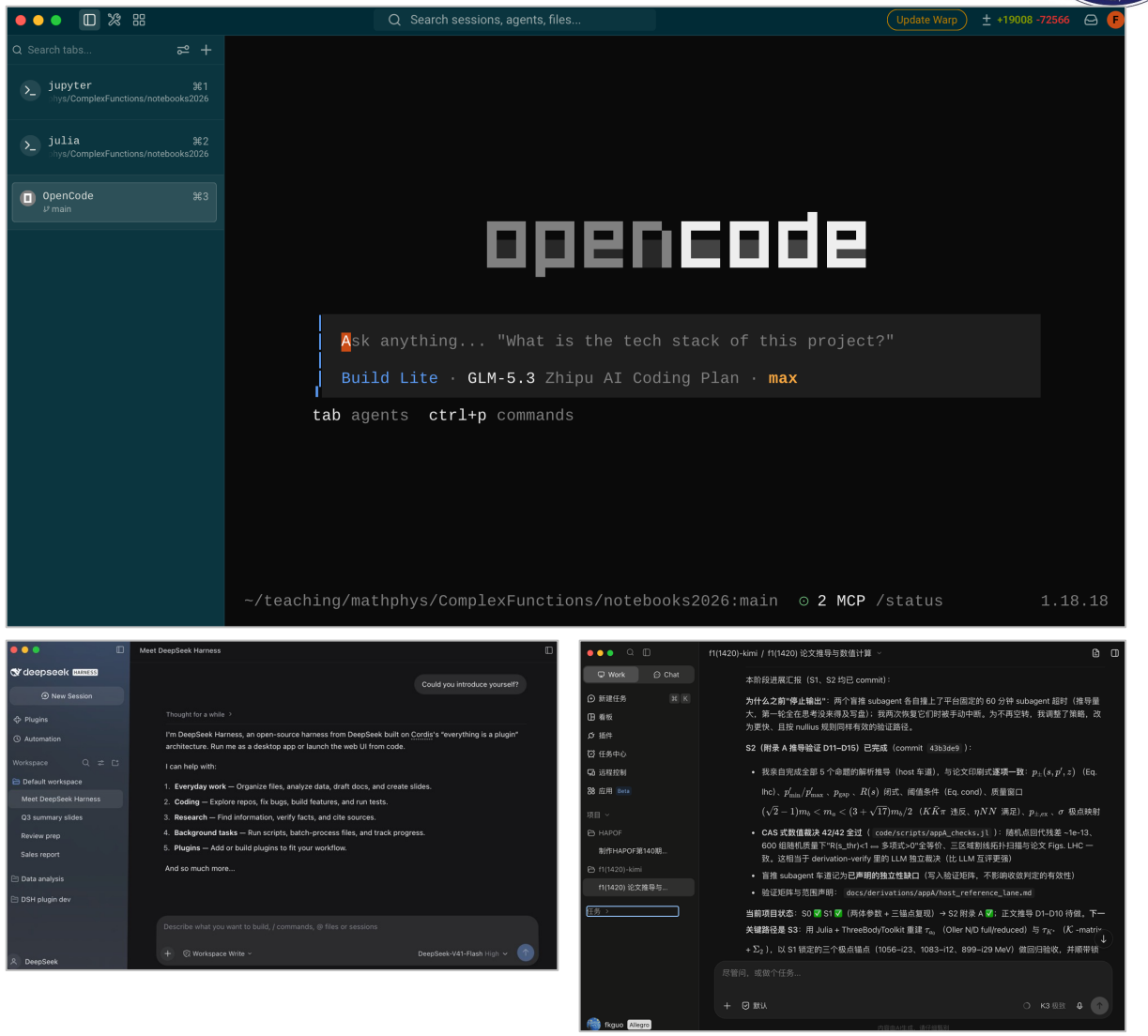
ChatGPT 桌面应用（工作模式）。输入框下方选择项目目录、模型与权限，左栏按项目列出会话

- Claude code, gemini

OpenCode、Kimi Code、DeepSeek Harness

- OpenCode 可连接多家模型提供方。左下是当时本机的模型列表
- 用 OpenCode 调用 GLM，也用 Kimi Code，请不同模型独立复核
- DeepSeek Harness 用于写 Particle Library

```
$ opencode models | grep free
opencode/exo-free
opencode/ling-3.1-flash-free
opencode/longcat-2.5-preview-free
opencode/mimo-v2.6-flash-free
opencode/nemotron-3-ultra-free
opencode/step-5-preview-free
…（本机共 10 个免费模型）
```



上：OpenCode；左下：DeepSeek Harness 桌面应用；右下：Kimi Code

项目规则文件

- 全局规则适用于常见任务，项目规则写本项目的输入、约定和检查
- Codex、OpenCode、Cursor 可读 AGENTS.md。Claude Code 可用 CLAUDE.md 引入同一份文件，现在也支持 AGENTS.md
- 每次遇到可以防止的错误，把具体做法补到规则里

全局规则文件节选：

"复现实现绝不得 include / import 被检验的内核；……分歧靠追踪定位，绝不数票。"

"复核者只读：不得编辑被审的活动树。"

"大 job 先估资源：先跑单点 dry-run 读峰值内存并设上限。"

"委托计算须写死预算与停止条件：数值容差上限、硬时间盒、范围负面清单。"

"任何将被看 / 被用的量进入展示面之前必须完成其验证门"

AGENTS.md (nullius init 生成的项目规则)

```
## Read order
1) project_index.md
2) AGENTS.md
3) project_charter.md
4) research_plan.md
5) research_contract.md
```

Skills

- 一个 skill 是一个文件夹。SKILL.md 写什么时候用、怎么做；可附脚本与参考资料
- 开始只加载名称与描述；任务相关时再读全文，按需读入上下文
- 放在 ~/.claude/skills、~/.agents/skills 等目录；用符号链接让几种工具共用一份
- 用法：直接点名一个 skill，或让智能体按任务选用。开放规范：agentskills.io
- 可以根据一定规模的材料产生skill, 通过实际使用的反馈进一步修改
  - 例如：通过给几十个我的报告文件制作一个 `my-powerpoint-slides` skill



本报告 slides 的初版

pdg-lookup / SKILL.md（节选）

```
---
name: pdg-lookup
description: "Teaches the PDG (Particle Data Group) tool
chain for looking up particle properties, measurements,
decays, and references from the local PDG SQLite
database.\n"
---


…

## Step 1: Check availability
pdg_info {}

## Step 2: Find the particle
pdg_find_particle { query: "top quark" }

## Step 3: Get properties or measurements
pdg_get_property { pdgid: "S008M", particle: "t" }
…
```

MCP 与科研数据库

- MCP（Model Context Protocol）：AI 应用连接外部数据与工具的统一接口，可类比仪器的通用接口
- 一个 MCP 服务提供一组工具；智能体选择工具，外部程序执行查询，结果由数据库返回
- 装好后用中文提要求即可。前一页的 skill 则告诉智能体：这些工具按什么顺序使用

服务	数据来源	可以做什么
hep-mcp	INSPIRE、arXiv 源码、PDG、HEPData、OpenAlex、Zotero	查文献、粒子性质、实验数据
nds-mcp	AME2020、NUBASE2020、ENSDF、EXFOR、JENDL-5、CODATA	查核质量、能级、核反应数据

github.com/fkguo/nullius、nds-mcp

```
# Claude Code (-s user: 对所有项目生效)
$ claude mcp add -s user hep-mcp -- nullius runtime mcp hep-mcp
$ claude mcp add -s user nds-mcp -- npx -y nds-mcp

# Codex
$ codex mcp add nds-mcp -- npx -y nds-mcp
$ claude mcp list    # 查看 Claude Code 已连接的服务
```

nullius

- 2025 年 12 月先开发 hep-mcp，用于查文献和数据。随后加入独立复核与项目记录 ⇒ 我自己的 harness 框架
- 通过 Codex 或 Claude Code 调用 skills 和 MCP 工具

任务	工具	材料
查文献和数据	hep-mcp：INSPIRE、arXiv、PDG、HEPData、OpenAlex、Zotero	论文源码、数据库记录与引用
独立推导、数值检验、引文核对	derivation-verify、numerical-reliability-gate、claim-grounding、review-swarm 等	独立计算、误差或差异的检查记录
开始或继续一个项目	nullius init 建项目；nullius status 恢复计划、进度与对应结果	研究计划、科研笔记、计算输入与输出



公开仓库含 36 个 skill 目录（2026-11-24）。安装和配置见 nullius Quickstart。

nullius：粲偶素谱的研究计划

- 在 Claude Code 或 Codex 中用中文布置任务，智能体调用 nullius 的 skills 和命令
- 给出几篇相关论文，先拟研究计划，修改确认后再计算
- 在项目文件中保存研究计划和进度，换模型或工具时继续使用

在 Claude Code 里说：
这几篇是相关论文：arXiv:hep-ph/0505002、arXiv:hep-ph/0701208、arXiv:0903.5506。用 nullius 在这个目录开一个新的研究项目：用非相对论势模型计算粲偶素谱，并与 PDG 比较。先读这几篇论文，写一份研究计划给我审核；我修改确认之后再开始计算。

它做了什么（6 分半）：
· 调用 research-harness skill；nullius init 建好项目文件
· 经 hep-mcp 取回三篇论文的 LaTeX 源码，逐篇写阅读笔记
· 经 hep-mcp 查 PDG 2026：计划中用到的态都在库里
· 写研究计划：两种势模型、两种解法、复现基准、7 个待定问题
· nullius decision pending：登记"计划待审核"，然后停下

我在这个目录新建了 nullius 项目，读完了三篇论文，写好了研究计划草案，等你审核。没有做任何计算。……

计划要点

· 两种模型：A 是 BGS 的 Coulomb+线性势，B 是 Li–Chao 的屏蔽势。两者用同一组 PDG 2026 输入、同一套数值方法重新拟合。……

· 数值方法：用两种互不共享代码的解法（有限差分和 Gaussian 基展开）互相核对，质量收敛到 0.1 MeV 即停止加密。

· 回归基准：先用两篇论文自己的参数重现它们的谱表，偏差应不超过约 2 MeV，然后才做新的拟合。

· 不做的内容：耦合道效应、S–D 混合、相对论修正和衰变宽度。这些只作为局限写进结论。

需要你决定的问题：D1 模型范围、D2 主拟合集、……、D7 文献扩展（每项附我的建议）

Claude Code（Opus 5.5）+ nullius，本机实际运行（2026-10-12）；右框为回复的节选

nullius：检索 BESIII 论文

- Codex 通过 nullius 的 hep-mcp 查询 INSPIRE
- INSPIRE 提供 recid、卷页与引用数。第三篇 2013 年上 arXiv，2014 年刊出
- 也可查 arXiv 源码、PDG 粒子数据和 HEPData 数据表，或将论文导入 Zotero

在 Codex 里说：
查 BESIII 2013 年发表的关于类粲偶素的论文，按引用数列出前三篇，给出 INSPIRE recid 和 arXiv 号。

```
它调用的工具（hep-mcp，1 分 41 秒）：

hep_health
  检查 INSPIRE 是否可用
inspire_search ×3
  cn:BESIII and date:2013 and t:charmoniumlike …
inspire_literature
  取第三篇（recid 1247675）的详细记录
```

论文内容	引用数	INSPIRE recid	arXiv 号	期刊发表
在 $e^+e^- \to \pi^+\pi^- J/\psi$ 中发现带电类粲偶素结构 $Z_c(3900)$	1327	1225275	1303.5949	PRL 110, 252001 (2013)
发现 $Z_c(4020)$，并在 $e^+e^- \to \pi^+\pi^- h_c$ 中寻找 $Z_c(3900)$	593	1253362	1309.1896	PRL 111, 242001 (2013)
在 $e^+e^- \to (D^*\bar{D}^*)^{\pm}\pi^{\mp}$ 中发现 $Z_c(4025)$	445	1247675	1308.2760	PRL 112, 132001 (2014)

若严格要求期刊发表于 2013 年，则只有前两篇符合；第三篇虽于 2013 年上传 arXiv，但期刊发表年份是 2014 年。

Codex（GPT-6.1 Sol）+ nullius 插件，本机实际运行（2026-10-12）；右表与下句是它回答的原文

粲偶素和类粲偶素谱：任务说明

- 查 PDG 隐粲介子态和首次观测的论文，绘制质量谱，与夸克模型比较

用 hep-mcp 的 PDG 工具和 INSPIRE 检索，整理粲偶素与类粲偶素（隐粲）态的质量谱。

1. 从 PDG 2026 取出所有隐粲介子态（包括 I=1 的 $T_{c\bar{c}}$ 和含奇异夸克的 $T_{c\bar{c}\bar{s}}$）：名称、I^G(J^PC)、质量、宽度（带误差）、主要衰变道。

2. 发现年份与实验：在 INSPIRE 找到首次观测该态的论文，给出年份、实验和 recid。PDG 里最早的文献只作线索；拿不准的标"待核"，不要猜。

3. 用 matplotlib 画质量谱：横轴按 $0^+(0^{-+})$、$0^-(1^{--})$、$0^-(1^{+-})$、$0^+(0^{++})$、$0^+(1^{++})$、$0^+(2^{++})$、$0^-(2^{--})$、$0^-(3^{--})$、量子数未定、I=1、I=1/2 分列。每个态画一条短横线，半透明带表示 $M \pm \Gamma/2$；2003 年以前发现的用红色，之后的用蓝色，旁边用深灰色标发现年份。

4. 用黑线画 Godfrey–Isgur 模型的预言，数值取自 Barnes、Godfrey、Swanson (2005) 的表（读 arXiv 源码）；用灰色虚线画 $D^0\bar{D}^0$、$D^{*0}\bar{D}^0$、$D^{*0}\bar{D}^{*0}$ 阈值，I=1/2 列画 $D_s^-D^{*0}$、$D_s^{*-}D^{*0}$ 阈值，质量取自 PDG。

5. 所有数都必须来自工具返回，不能凭记忆。查询结果存成 states.csv，表格存成 states.md，作图代码 make_spectrum.py、图 spectrum.png、核对清单 checks.md 都放在当前目录。

粲偶素和类粲偶素谱：作图过程

- 第一轮用时 10 分钟。查 PDG 和 INSPIRE，从论文源码取夸克模型的谱表，编程作图
- 有一个态的首次观测拿不准，单独标出并说明原因
- 右侧为作图的补充要求，包括态名、年份和 χcJ(1P) 的处理

步骤	调用的工具	次数
PDG 质量、宽度、衰变	hep-mcp 的 PDG 工具	67
量子数、转录自查	直接读同一个 PDG 数据库	—
首次观测论文	INSPIRE 检索（3 个子智能体）	47
Godfrey–Isgur 预言	arXiv 取 LaTeX 源码	1
写代码、作图、自查	Python 与 sqlite	26
用时（日志）	第一轮 10 分钟；后六轮 4、40、15、2、3、3 分钟	

第二轮：图上每条横线旁边标出态的名称（如 ψ(4040)、χ_c1(3872)、T_cc̄1(3900)，用 mathtext 排上下标），年份放在名称后面；字号放大到 13 以上，用无衬线字体。I=1 列里的态比较挤，左右错开，标注不要互相重叠。数据不变，只改图。

第三轮：这张图要整页放进 16:9 的幻灯片：画布改为 12.4 × 6.1 英寸，所有文字在这个尺寸下不小于 10 pt，标注仍不能重叠；放不下时可以把年份写在态名下面一行。数据不变。

第四轮：年份一律改用 arXiv 首次公开的年份（arXiv 公布即算发表）；没有 arXiv 预印本的老论文用期刊年份。表、CSV 和图一起改。另外图上的字要更大：在 12.4 × 6.1 英寸下不小于 12 pt，放不下就把年份放到态名下一行。

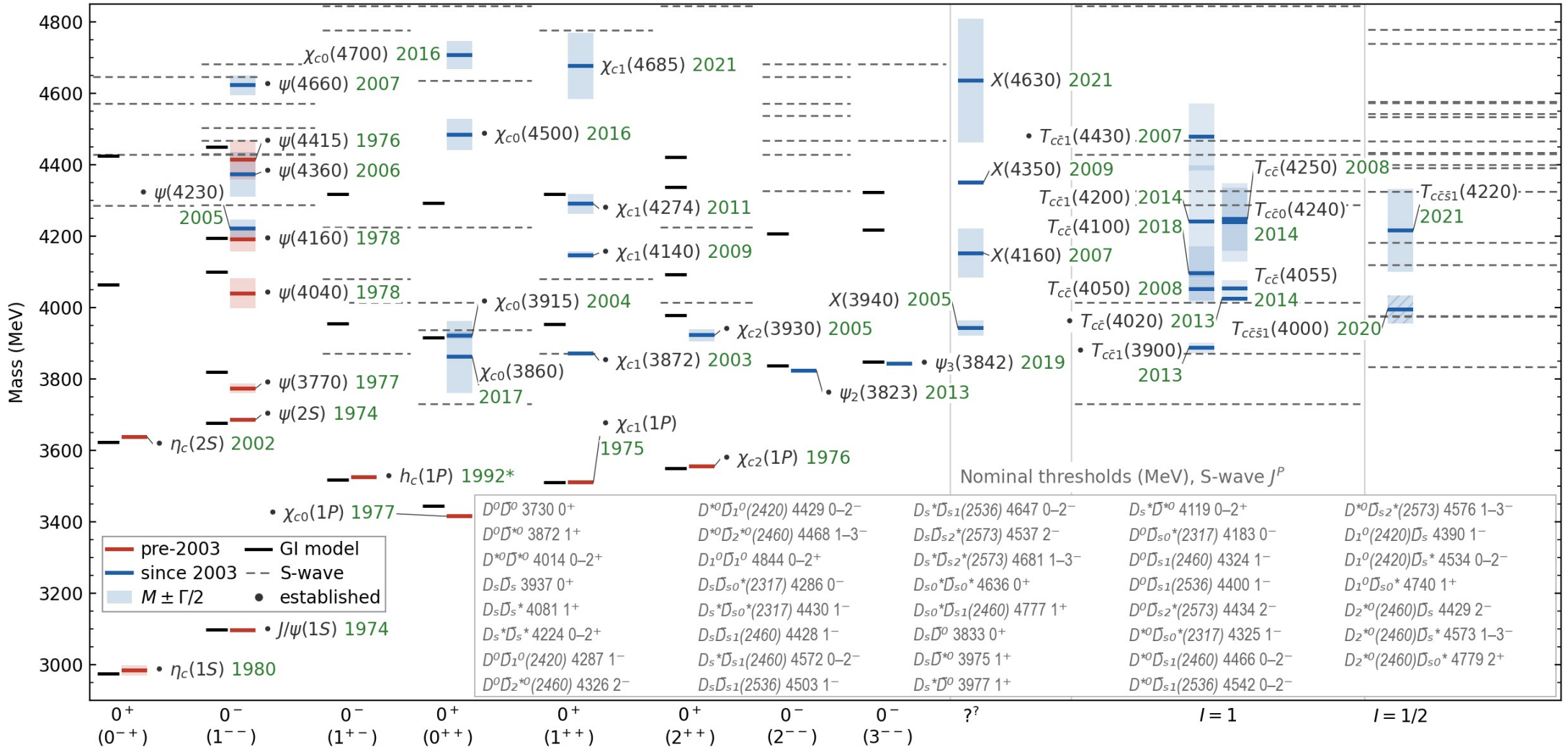
第五轮：图上的字号改为 11 pt，其他不变。

第六轮：χ_c0(1P)、χ_c1(1P)、χ_c2(1P) 的首次观测按 PDG 中列出的最早论文为准；图上的字号改为 10 pt。表、CSV 和图一起改，其他不变。

第七轮：年份改用绿色；图例加一个不透明的背景框。其他不变。

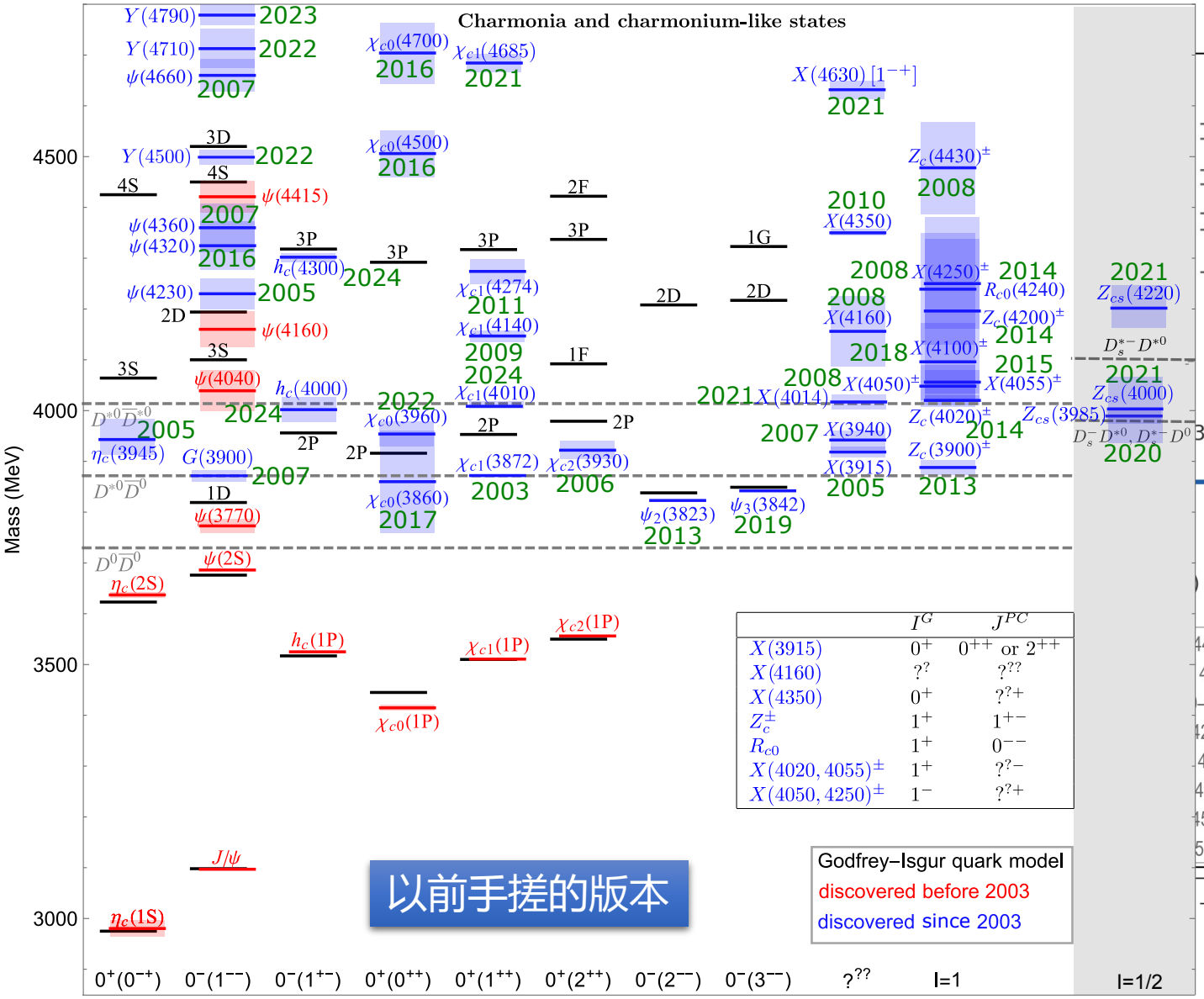
粲偶素与类粲偶素质量谱

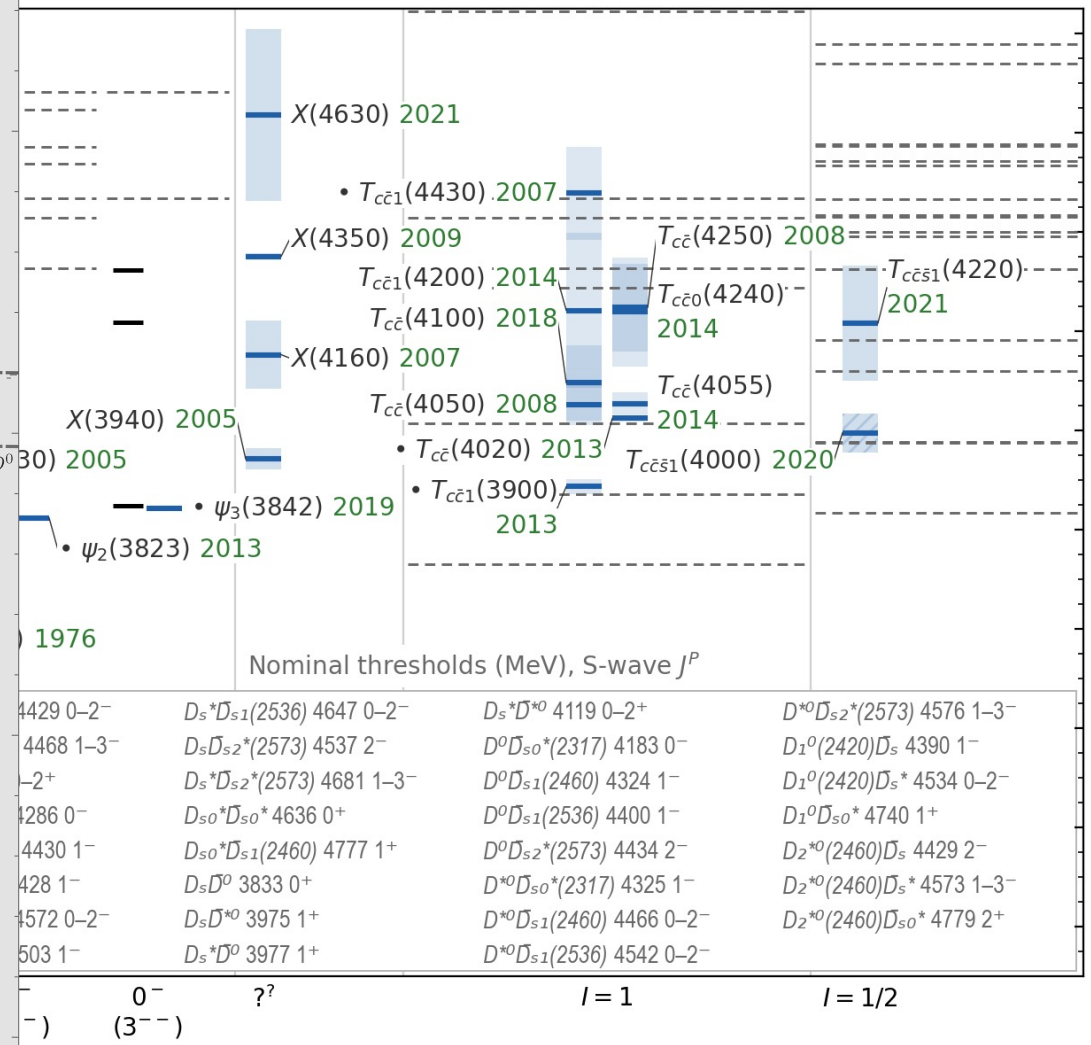
T. Barnes et al., PRD 72, 054026 (2005)　F. Takahashi et al., IJMPA 41, 2630011 (2026)



圆点：PDG 标为 established 的粒子。红：2003 年以前发现；蓝：2003 年及以后。黑线：Godfrey–Isgur 模型。

粲偶素与类粲偶素质量谱

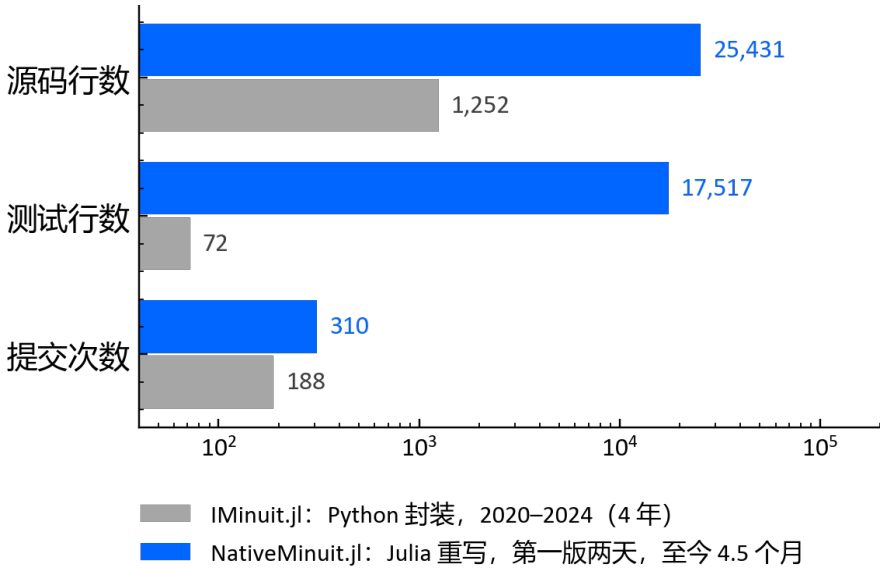


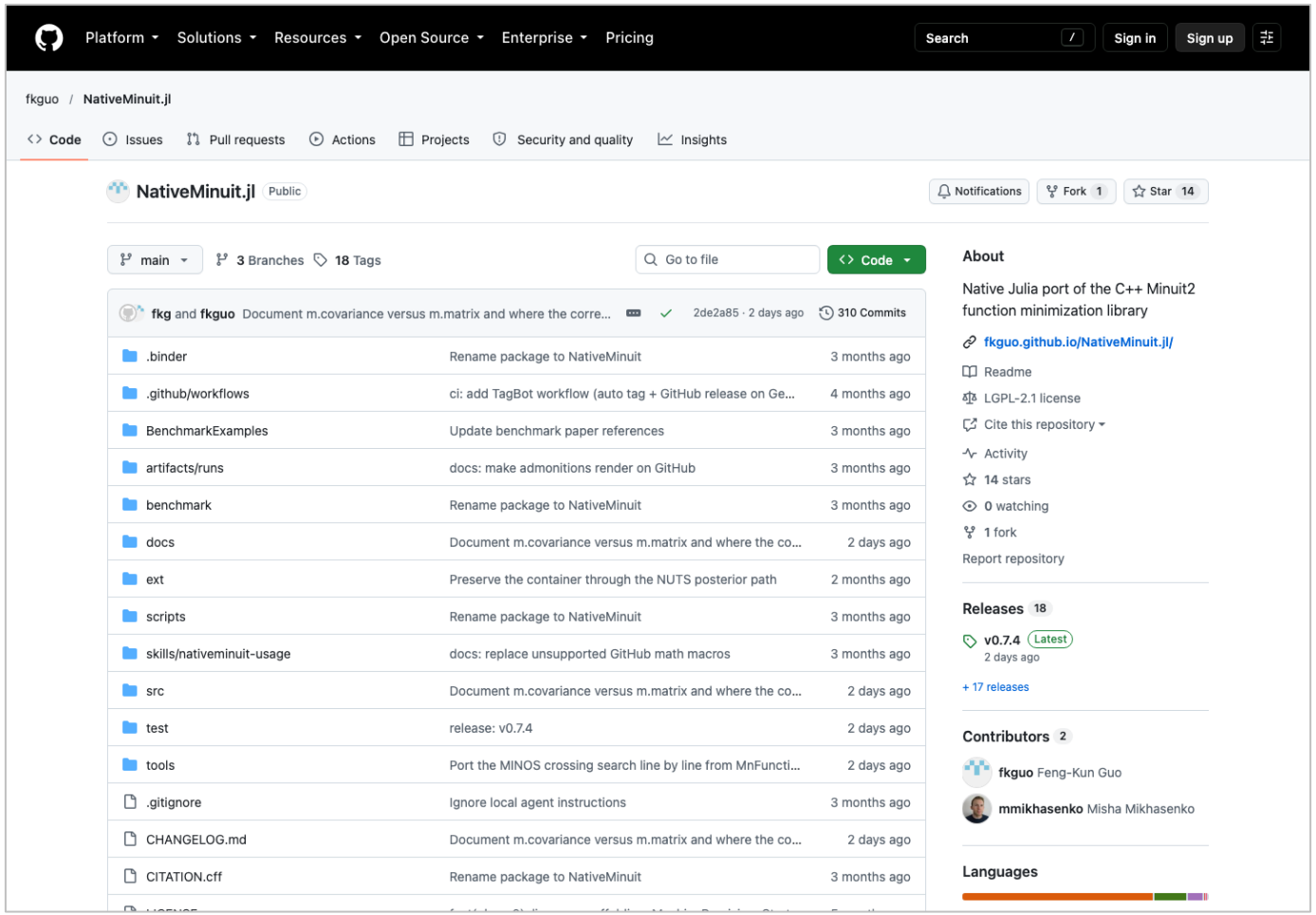


NativeMinuit.jl

- 此前用 IMinuit.jl 封装 Python iminuit。2026 年开始将 CERN Minuit2 移植到 Julia
- 第一版两天完成。之后把 MIGRAD、HESSE、MINOS 和等高线逐项与 C++ 版比较
- 已进入 Julia General。目前版本为v0.8.0

F. James, M. Roos, Comput. Phys. Commun. 10, 343 (1975)





NativeMinuit.jl 与 C++ Minuit2 的性能比较

性能超过 C++ 版本

- 数值梯度 MIGRAD，使用相同的输入、初值和停止设置
- Apple M3 单线程，10 对独立进程。计时包括求解和拟合结果的输出

- 预热后的拟合耗时（0.8.0 / 6.24.0）

目标函数	Julia (μs)	C++ (μs)	Julia/C++ 中位数	Julia/C++ 实测范围
四参数二次函数	0.93	4.2	0.22	0.21–0.24
Rosenbrock（2 参数）	8.3	48	0.17	0.17–0.19
Rosenbrock（10 参数）	54	200	0.27	0.24–0.30
高斯似然（2 参数，100 点）	5.8	16	0.35	0.31–0.44
高斯似然（10 参数，1000 点）	31	38	0.82	0.80–0.84

github.com/fkguo/NativeMinuit.jl#performance

- AI 辅助的程序改进

目标函数

按函数类型编译，减少运行时分派

工作数组

复用工作向量和逆 Hessian 的存储，减少分配与复制

矩阵更新

合并矩阵遍历，省去重复的梯度收敛检查

解析推导的独立核对

- agents 会写 sympy 等代码进行解析推导，从而提高正确性（并非仅靠训练集中学到的知识）

对象	AI 做了什么	发现	怎么核验的
超子 0νββ 单圈计算 Z.-Y. Zhao et al., JHEP 09 (2026) 185	从拉氏量重建顶角与振幅	一个重整化系数对不上， 学生推导有误： 维数正规化某个中间步骤 将 d 设为了 4	与学生的计算逐步比对
弦断裂导致的 string state 与介子对的耦合 Y. Zhang et al., arXiv:2610.19066	检查文献，从格点计算使用的算符，投影到物理基； 找到 2020 年的论文，其中更正了 2005年提出计算框架的格点计算文中的错误	与文献所用值差因子 2， 2019, 2024 两份 string breaking 格点计算原始文献有误，随后的所有 BOEFT 论文都用错了	✓ 四个独立agents重推一致； ✓ 我自己手推一致 ☐ 反馈给格点论文的作者

BOχEFT

Y. Zhang et al., arXiv:2610.19066

- 我们提出 BOχEFT 理论框架，做了大量前期推导和计算（从几年前开始的项目）
- 利用agents 快速推进：
  - 扩展理论框架（两体、三体compositeness）、cross check 与新计算
  - 约三天复现大量 BOEFT 结果，理解其中的参数依赖性
  - BOEFT, BOχEFT 代码整理进 `HadronDynamics.jl`

BOEFT 文献复现：首次提交后的墙钟时间（2026-10-07 14:14 → 10-10 11:30，483 次提交）

- +0 h 首次提交：新建仓库，读四篇 BOEFT 论文的 LaTeX 源码，调用 ThreeBodyToolkit 的求解器
- **+7.0 h** 三通道粲浅束缚态 E_B = 99.05 keV（论文 96 keV；步长/箱长收敛 0.22 keV）
- +20 h 四通道自旋分辨 1^{++} 浅态：步长减半能量变化 6×10^{-8} GeV
- +29.8 h 论文未做的参数曲线：束缚能对 Λ_spin（5 点）与 d（10 点）的依赖
- **+44.9 h** Table XIV 六个 1^{++} 态全部复现：质量差 −1.5 … +1.0 MeV；深束缚态半径/组分一致
- +45.3 h Table XVIII 介子组分：(3.29, 95.97, 0.36, 0.38)% vs 论文 (3.2, 96.1, 0.1, 0.6)%
- +62 h 自旋平均 S/P/D 波各表已算出，但空间精度未验证；两个近阈浅态的半径差异未解释

N. Brambilla et al., PRD 114, 054049 (2026)
arXiv:2604.12603, Table XIV

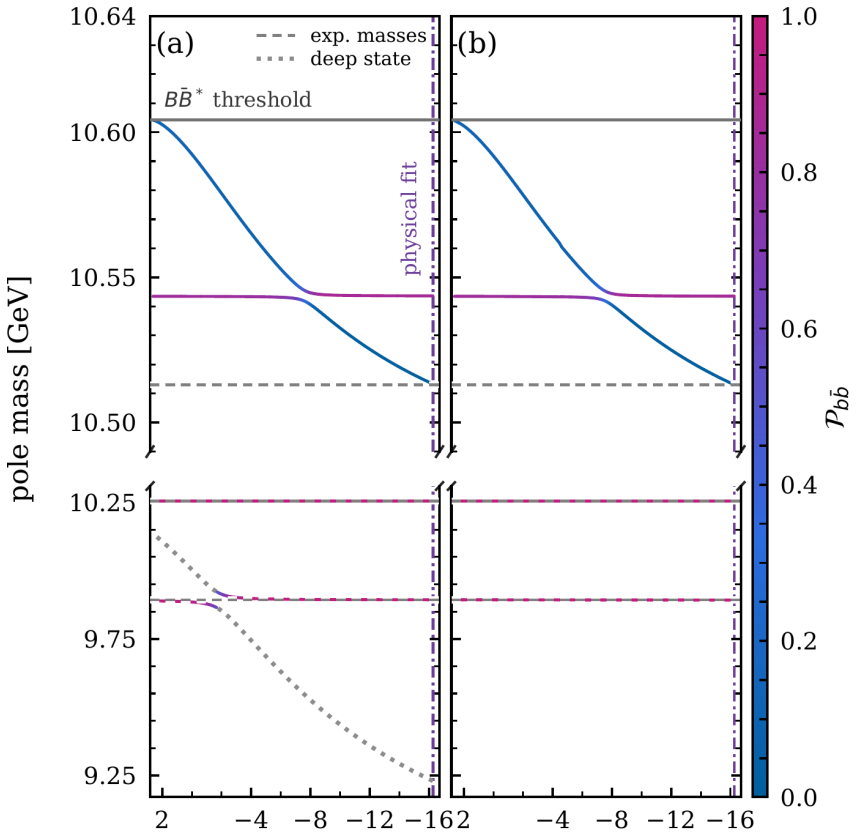
1^{++} 态	复现 / MeV	论文 / MeV
粲 1P	3536.56	3536.7
$\chi_{c1}(3872)$	3874.51	3875.4
底 1P	9881.80	9882.9
底 2P	10234.50	10235.0
底 3P	10516.22	10515.2
X_b	10594.40	10595.9

X_b 的耦合道计算

Y. Zhang et al., arXiv:2610.19066

- 与学生计算的结果有差别：统一归一化约定和输入
- 发现 singular potential 导致的深束缚伪束缚态问题 （图上的回避交叉）引出伪束缚态问题
  - 智能体快速找到核力中采用的处理方案 V.I. Kukulin and V. N. Pomerantsev (1978); E. Epelbaum et al. (2002); A. Nogga et al. (2005)， 扫描截断并做投影比较

谁做了什么	人	AI
归一化约定不一致	对约定做出选择	逐项比对两套代码
弦断裂耦合系数	判断要重推	四路不看答案重推，结果一致
能级图上的伪态	看图发现；想起手征核力文献	截断扫描、投影去除



能级随接触项 C（GeV⁻²）的变化：(a) 含深的伪束缚态（灰色点线）；(b) 投影去除后，其余能级的变化在 1 MeV 以内

三体振幅中的运动学错误

- 独立计算组内三体振幅，发现原实现把 π 的能量写成了 K 的能量
- 检查能量与质量关系，修正后重新拟合并检查极点
- 重拟合后检查新解的稳定性和附加极点对谱形的影响

T.-R. Hu et al., arXiv:2606.19040

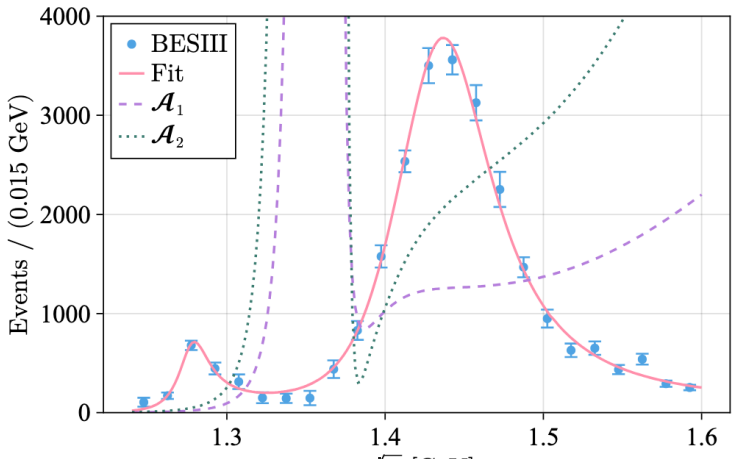
可疑之处
一个粒子的能量公式

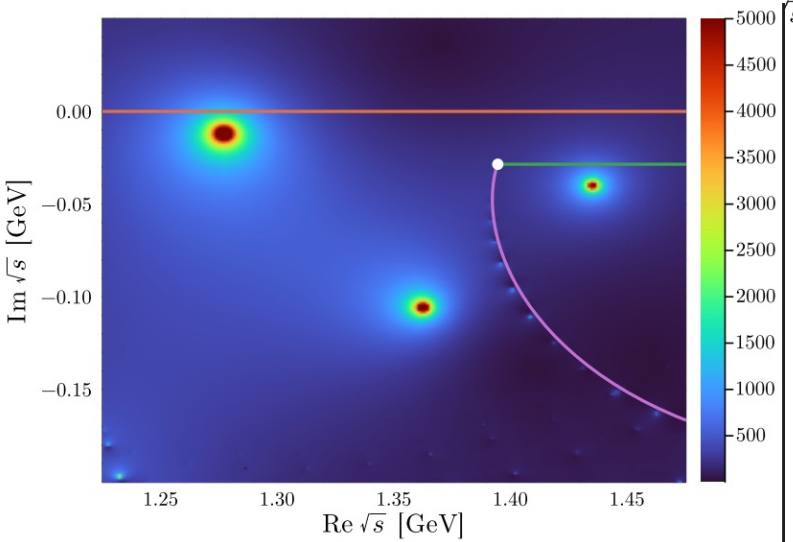
解析论证
分解动量对两质量交换对称，错的只能是能量

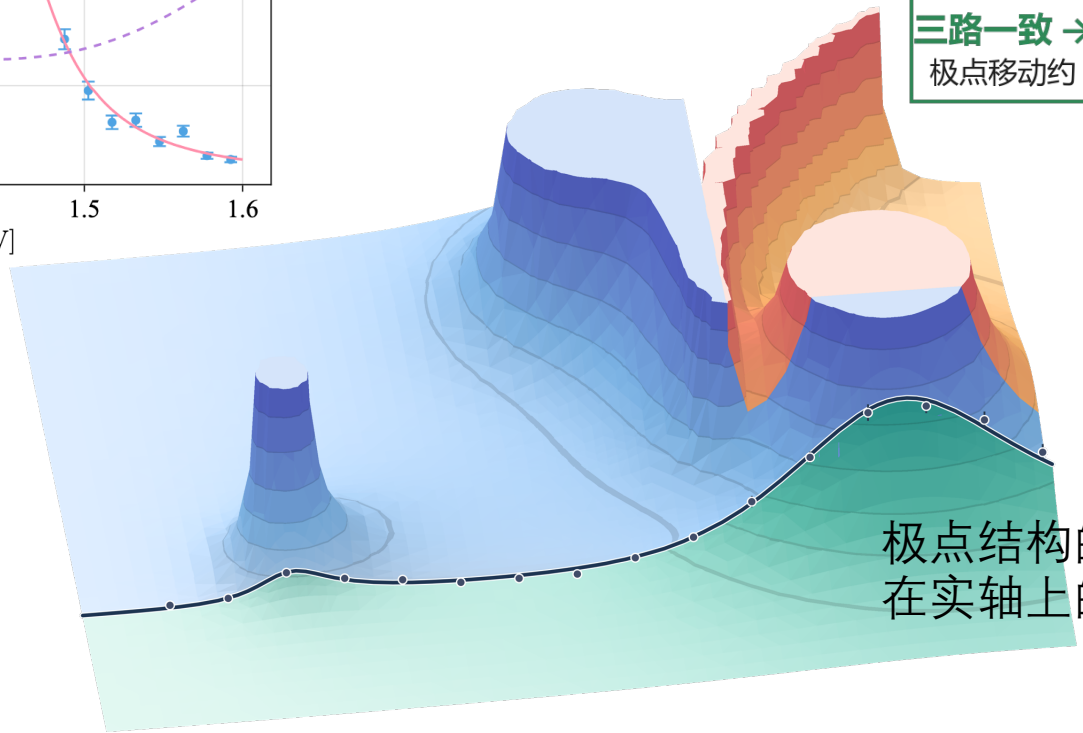
单点数值
不变质量变成了另一粒子的质量²
能量守恒差几十 MeV

另一模型家族
不看原代码
盲重推，得到同一修正式

三路一致 → 修正，重新拟合
极点移动约 2 MeV







极点结构的三维图及其在实轴上的投影

非微扰强相互作用的理论方法

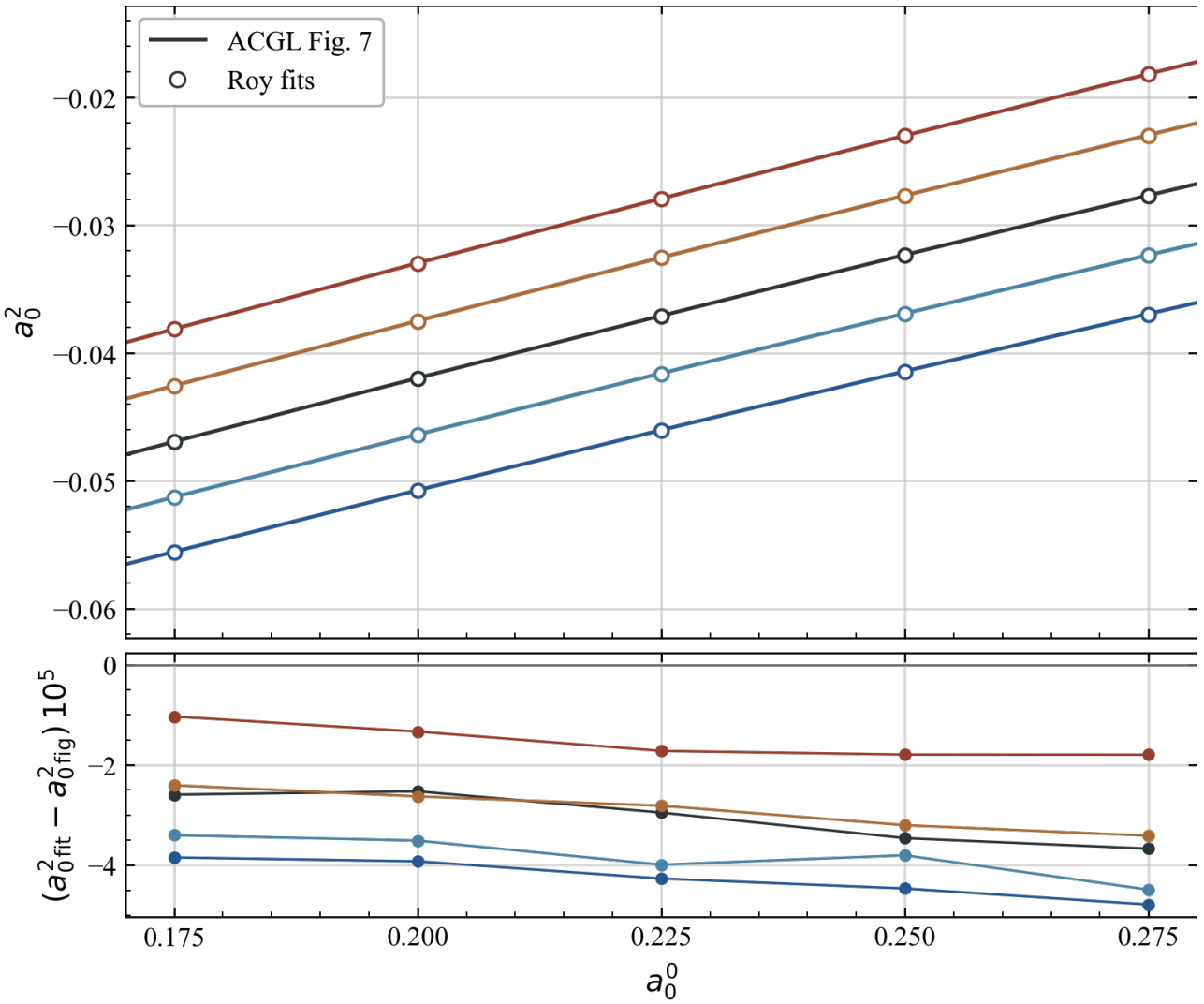
- 非微扰强相互作用动力学的 agent-ready 基础设施：
  - 强子谱学、强子相互作用、标准模型的精确检验、核结构与反应、…
- Julia 程序包附物理输入、文献对照和算例，可直接用于新课题

软件包	内容
`DispersiveToolkit.jl`	色散关系与分波输入。Roy / GKPY 方程、Roy-Steiner、Omnes、Khuri-Treiman、…
`ThreeBodyToolkit.jl`	通用少体求解。Lippmann-Schwinger, Faddeev, AGS 积分方程、束缚与共振态、解析延拓和极点搜索；complex scaling; …
`HadronDynamics.jl`	Cornell; Godfrey-Isgur; BOEFT; BOχEFT; …
`ChiralNuclearForces.jl`	连续空间的手征核力 (NNLO, N3LO, N4LO)，提供相互作用与相应的参数
`LatticeEFT.jl`	格点有效场论，用于少体计算和轻核结构和反应
`NativeMinuit.jl`	已发布的通用拟合工具，负责极小化和不确定度分析

Roy 方程：ππ 散射长度的普适带

- 文献输入
  - ❑ Roy 方程的积分核、减除项和匹配条件
  - ❑ 图 5 的五组等张量 S 波相移
- Julia 程序
  - ❑ DispersiveToolkit.jl 提供色散积分与分波相移接口
  - ❑ NativeMinuit.jl 用于低能分波拟合
- 与文献图 7 比较
  - ❑ 曲线取自原图，25 个圆点由 Roy 方程求得
  - ❑ 下图放大计算值与原图的差
  - ❑ 等张量散射长度最大差值约 4.8×10^{-5}

B. Ananthanarayan et al., Phys. Rept. 353, 207 (2001), Figs. 5, 7



ChiralNuclearForces.jl：两核子势矩阵构造

- Reinert、Krebs、Epelbaum 手征核力，450 MeV 截断，84 个动量点
- Apple M3 单线程，构造同一个 168×168 势矩阵
- AI 辅助改写矩阵构造，逐项对照作者的 Fortran 结果

P. Reinert, H. Krebs, E. Epelbaum, EPJA 54, 86 (2018) [arXiv:1711.08821]

- 势准备后的矩阵构造耗时

阶次	Fortran (ms)	Julia (ms)	Fortran/Julia
N^4LO	317.5	1.79	177
N^4LO^+	303.2	1.96	155

- 原作者 Fortran 用 -O3 编译，两边各构造 5 次，取中位数
- 两个阶次均逐项核对 28,224 个矩阵元

Julia 势准备约 42–44 ms。固定势参数时，多次构造可共用准备结果。

- 计算结构

原作者 Fortran

逐对动量调用分波程序，按互易性补齐矩阵

Julia：径向系数预先计算

每套势准备一次，后续计算共用径向系数

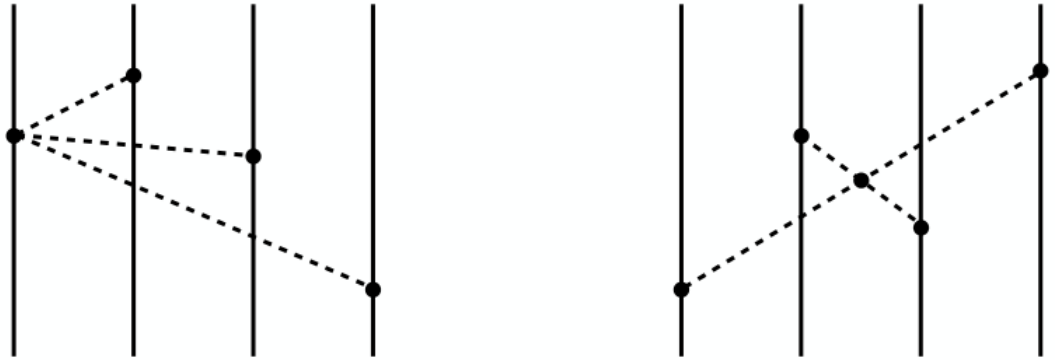
Julia：成批构造分波矩阵

在动量网格上预先计算球贝塞尔函数，再用矩阵乘法求各分波块

Gradient flow 与四核子力

- Krebs、 Epelbaum 用 gradient flow 正则化 π 场，计算图中四核子力的两项贡献
- 两种计算：直接实现式 (4.59)，另从手征拉氏量的两个顶点独立推导

H. Krebs, E. Epelbaum, arXiv:2312.13932v1, Fig. 1 and Eq. (4.59)

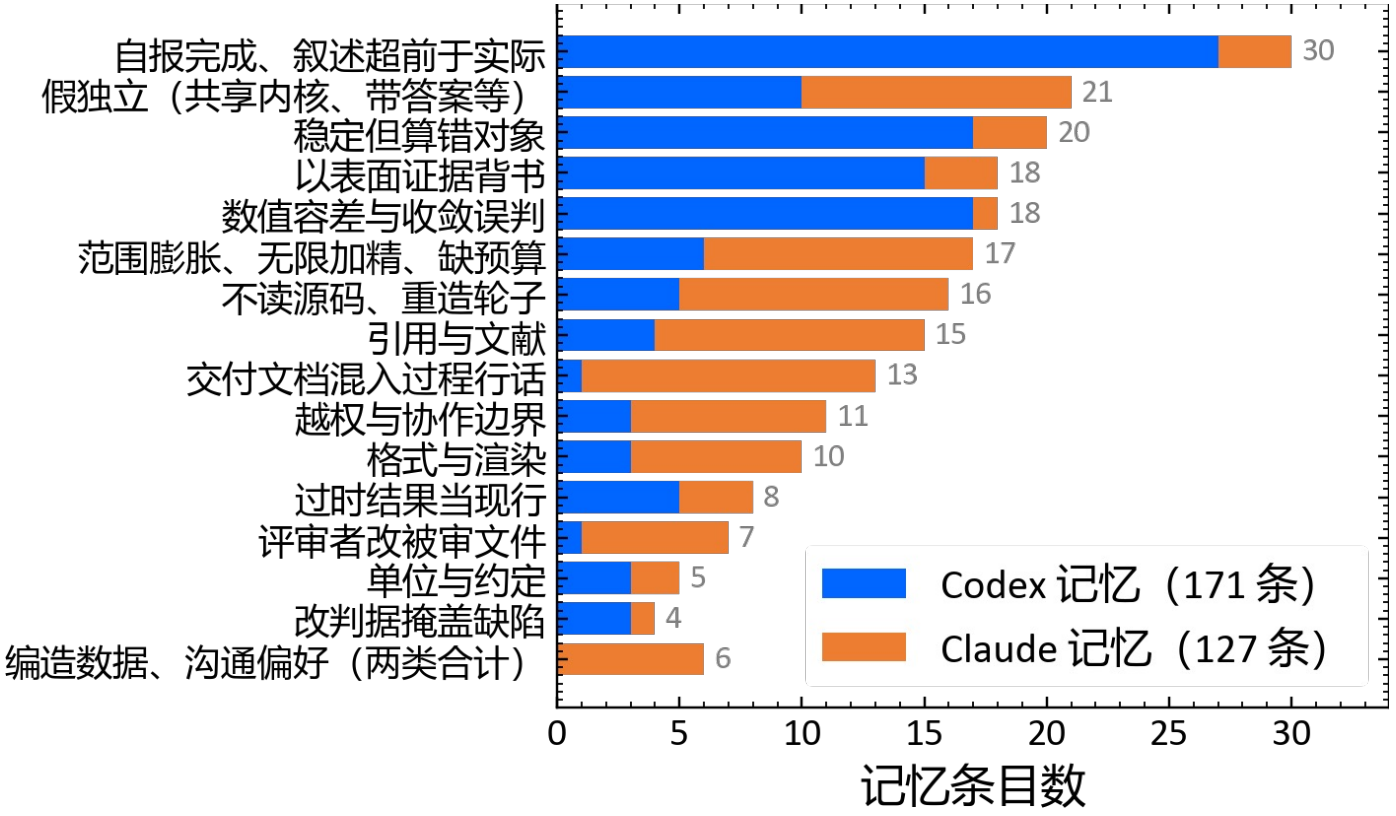


两幅四核子力树图（原文图 1）

- 流时为零，两种结果一致。有限流时，粒子置换求和后仍有差异
- 原作者式 (4.59) 的两个分子可能漏了与逆 π 传播子成正比的项
- ChiralNuclearForces.jl 保留两种表达式，可按同一组输入比较

AI 的失败模式：298 条记忆条目

- 两套智能体的记忆（截至 2026-10-10：Codex 171 条、Claude 127 条），按主要失败模式分类
- 多数是：拿表面的证据（链接存在、测试通过）代替物理检验；只在局部检验就当整体成立；说做了其实没做。也有能力不足的情形。学生也会这样



记忆文件原文：

"a worktree/task exists and is described as completed"

"all three included the same production … kernel. The agreement MASKED a 27–76% … discrepancy"

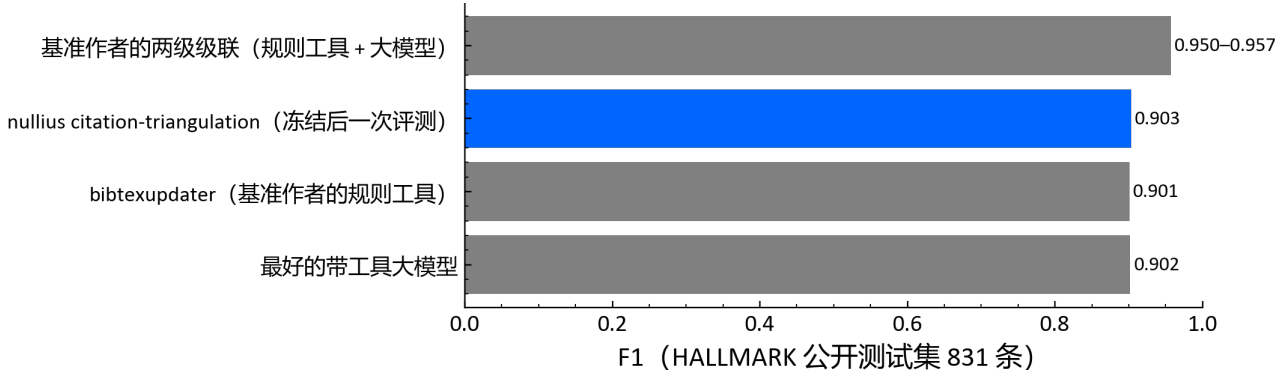
"conv=true only means det≈0 at the root the solver found, not that it is the physical pole"

"links do not prove derivation correctness"

"a 'recalled, DOI-probably-correct' arXiv ID … turned out to be an unrelated quantum-finance paper"

文献引用核对

- 智能体经 hep-mcp 查 INSPIRE：每条引用解析到 recid，作者、卷页逐项核对；重要文献读 LaTeX 源码，摘要只用来筛选
- citation-triangulation：每条引用从两个以上独立索引（INSPIRE、CrossRef、OpenAlex、DBLP）取元数据，离线比对，有不一致就不算通过
- claim-grounding：查文献内容是否支持这句话，不只查文献是否存在
- HALLMARK 公开基准（约 2,500 条标注引用、14 类伪造）：公开测试集 831 条冻结后只评一次，F1 0.90，与最好的带工具大模型相当，误报率 0.21 对 0.34；比对不调用大模型，基本不消耗 token；基准作者的两级级联（规则工具 + 大模型）为 0.950–0.957



P. Reizinger, W. Brendel, arXiv:2607.18360 (HALLMARK)
github.com/rpatrik96/hallmark (v1.0)

引文核对
保存数据库记录和比对结果
用摘要筛选论文，读正文核查所引结论
从原文查明是谁提出或证明

nullius 复核记录

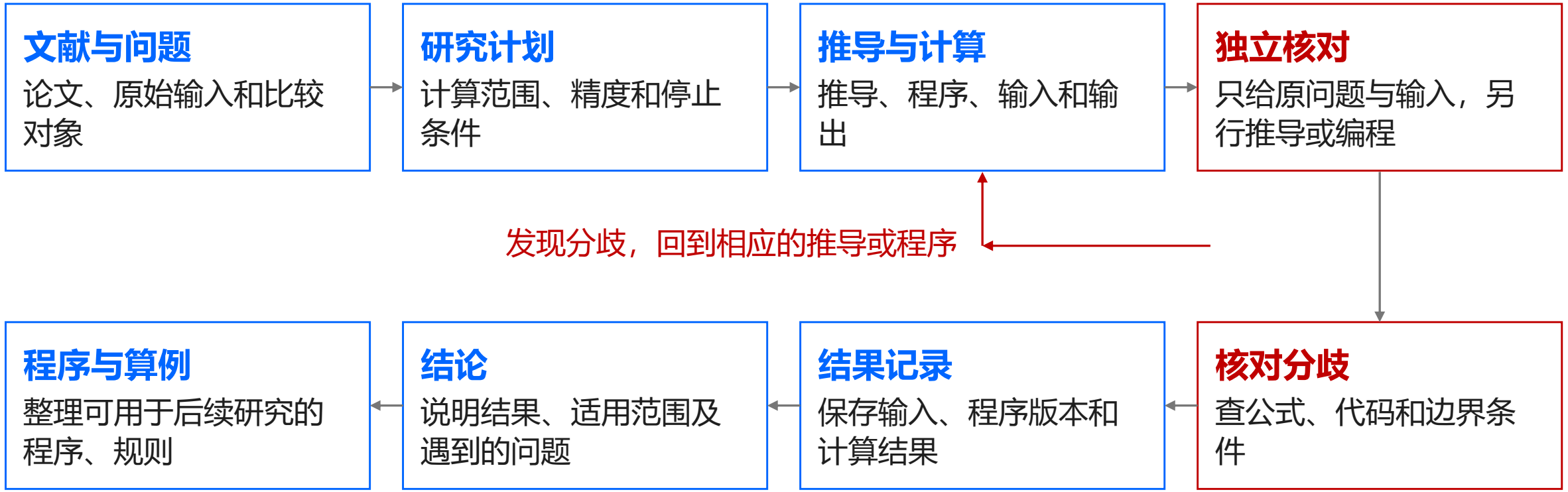
- 2026 年 4 月 28 日至 11 月 23 日，有报告格式检查记录的复核共 483 批，分派 561 次
- 348 份报告通过格式检查，46 份未通过。另有 109 次超时或无进展

记录的配置	分派	READY	NOT_READY	格式未通过	超时等	其他失败
GLM-5.2	203	114	21	23	37	8
Claude Opus（别名）	148	65	27	18	31	7
Kimi（默认）	50	20	12	1	5	12
Claude Sonnet（别名）	31	14	3	0	14	0
Claude Fable 5.1	29	24	2	3	0	0
Codex（合计）	12	1	7	0	3	1
其他配置	88	27	11	1	19	30
合计	561	265	83	46	109	58

同一批内重试只计一次。READY / NOT_READY 表示复核者认为可继续或仍需修改。各配置的任务与预算不同。

研究计划、计算与复核

- 给出代表性论文、物理问题、原始输入与比较对象
- 独立复核关键推导和数值。出现分歧时核对原始公式、数据或代码
- 保存计算结果和输入，把常用的做法写成规则、skill 或程序



选题与物理判断

- 针对具体问题，比较不同假设下的计算结果
- 选题、物理模型的选择和结果的解释

T. Tao, Mathematics in the age of AI, arXiv:2608.16753 (2026)
T. Bloom, Erdős problems in the OpenAI proofs (2026-10-08)

问题的类型	例子	AI 能做什么	人要做什么
定义清楚的问题	数学猜想、两体算符纠缠功率的极小值（已有前期工作）	求解、找反例、证特例、形式化验证	判断问题值不值得做；解出之后读懂、核验、讲清，放进已有的理论框架：讲不清的证明不算完成
本领域专家熟悉的问题	三体幺正框架下的 f_1 极点；X_b 与 BOEFT	把专家设想的路线做出来；多种方案并行，快速排除错路	知道问题卡在哪里、有哪些可能的方案；判断结果是否合理
完全跨领域的问题	受控核聚变：借助 AI 研究有效场论的处理方案	补背景、提出方案、做计算	只靠自己不确定其意义：问题是否重要、关键在哪里，要请该领域的专家

这是最好的时代还是最坏的时代?

- 提出重要问题：问题意识、物理积累与创造力决定研究方向
- 善用新的工具：AI 辅助推导、编程与计算，扩大探索范围
- 作出科学判断：读懂、检验并解释结果，形成新的物理认识
- 让成果成为知识：进入人类的理解、交流与实际应用



转化成人类知识的研究才是科学的一部分

科学家是研究的主体，AI 解放科学家的生产力，加速推进科学技术发展

- 如何进行 AI 时代的人才培养（提出问题、独立思考、善用 AI ）？

谢谢大家，欢迎交流

- nullius： github.com/fkguo/nullius
- zotero-inspire： github.com/fkguo/zotero-inspire
- NativeMinuit.jl： github.com/fkguo/NativeMinuit.jl

nullius

zotero-inspire

NativeMinuit.jl