



江汉大学
JIANGNAN UNIVERSITY



从大模型到智能体；

AI4Science与AI4Education的范式转变与实践探索

桑鸿乾

江汉大学精细爆破全国重点实验室 / 智算中心 副研究员

江汉大学青年教师专题报告 | 2026



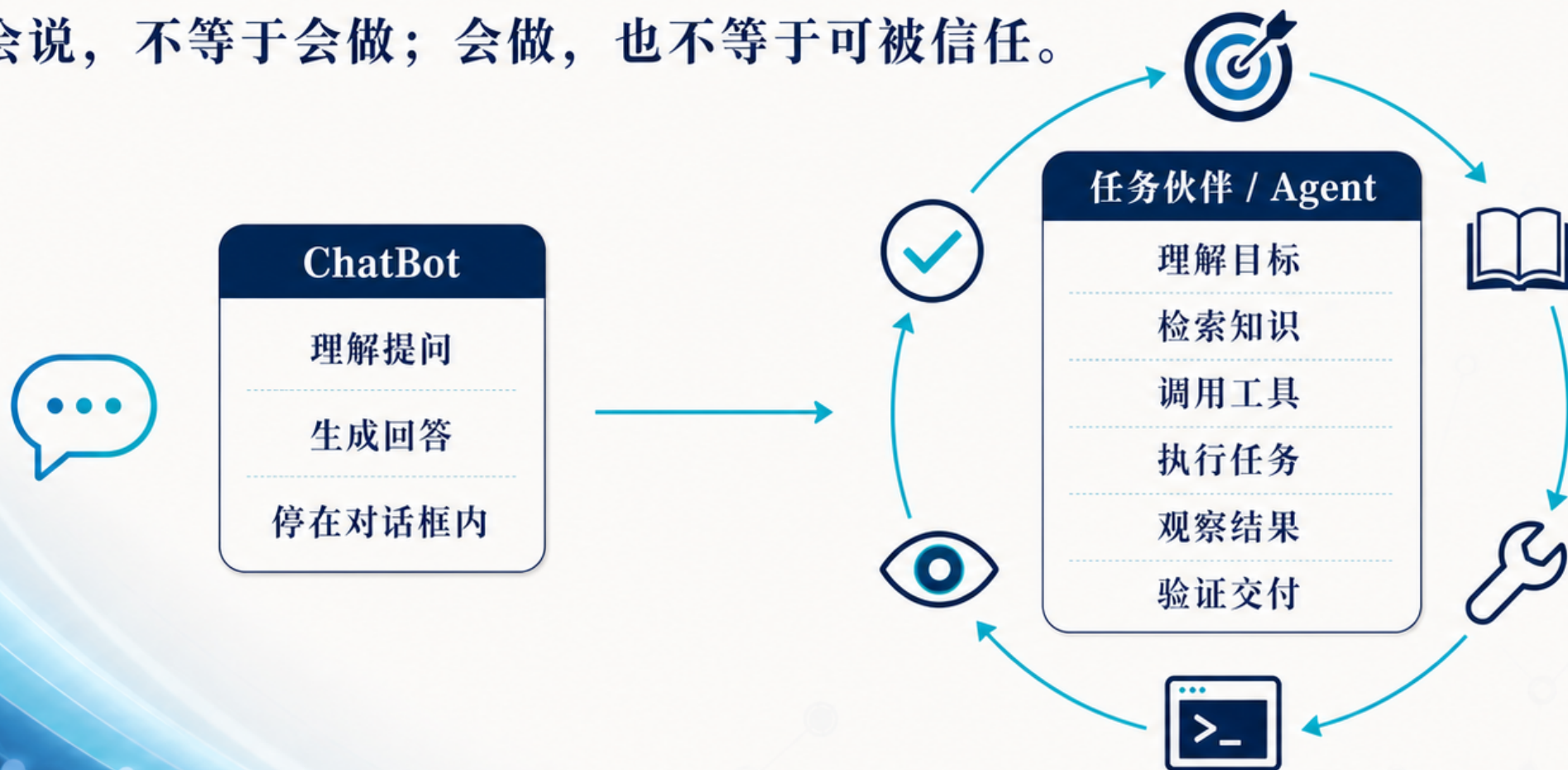


目录

- **01 大模型：生成、概率与幻觉**
- **02 智能体：工具、规划与协作**
- **03 AI4Science：科研流程重构**
- **04 AI4Education：教学伙伴与政策**
- **05 行动清单：可信使用与实践**

它只是 ChatBot，还是任务伙伴？

会说，不等于会做；会做，也不等于可被信任。



范式转变：从生成到任务闭环



模型负责理解与生成



外部系统提供知识、行动与反馈



可信交付来自证据、验证与人类责任



大语言模型先把语言切成 Token

模型并不直接“看到句子”，而是看到一串可计算的 token。



Token 不是简单的“字”或“词”；中文、公式、代码的切分方式也不同。

上下文窗口





自回归的核心：预测下一个 Token 的条件概率

$$P(x_t \mid x_1, x_2, \dots, x_{t-1})$$

给定前面的上下文，模型为“下一个 token”计算一个概率分布。

① 读取已有上下文

江 汉 大 学 很 美 ...

② 计算候选概率



③ 选择一个 token 并回填上下文

江 汉 大 学 很 美 丽 ...



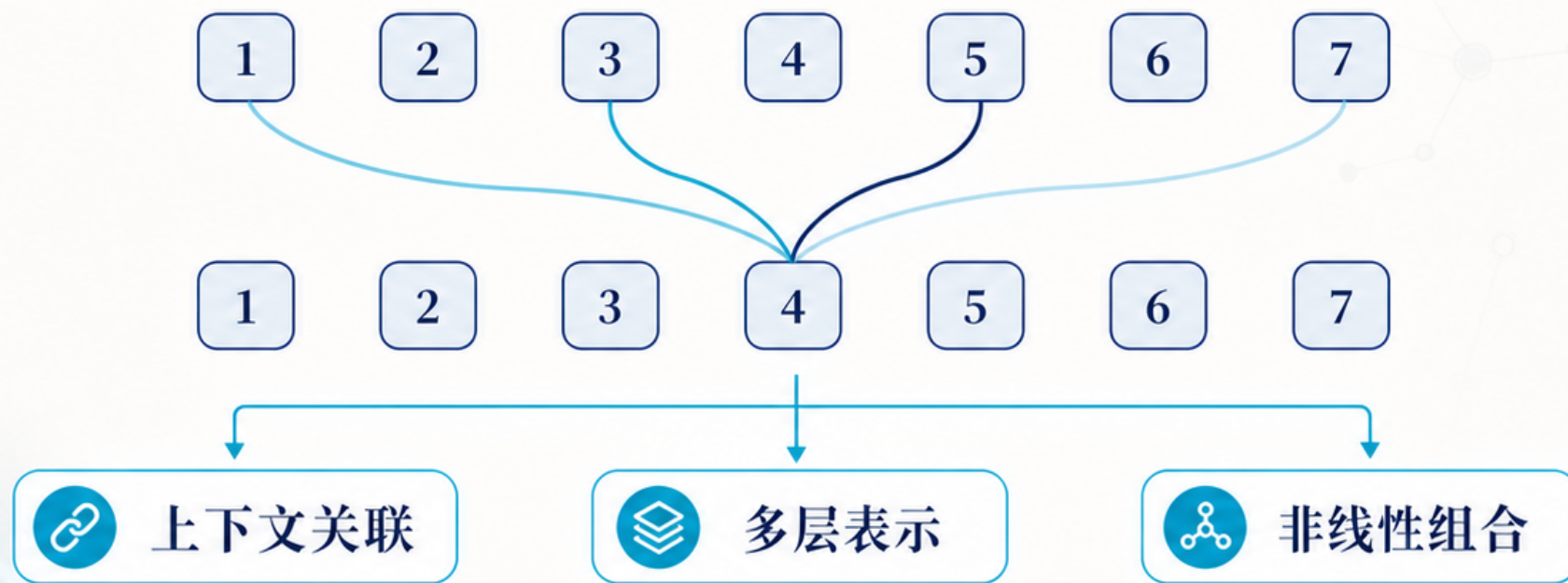
概率高 $\neq$ 事实真



06 | Transformer：为什么它能利用长上下文？

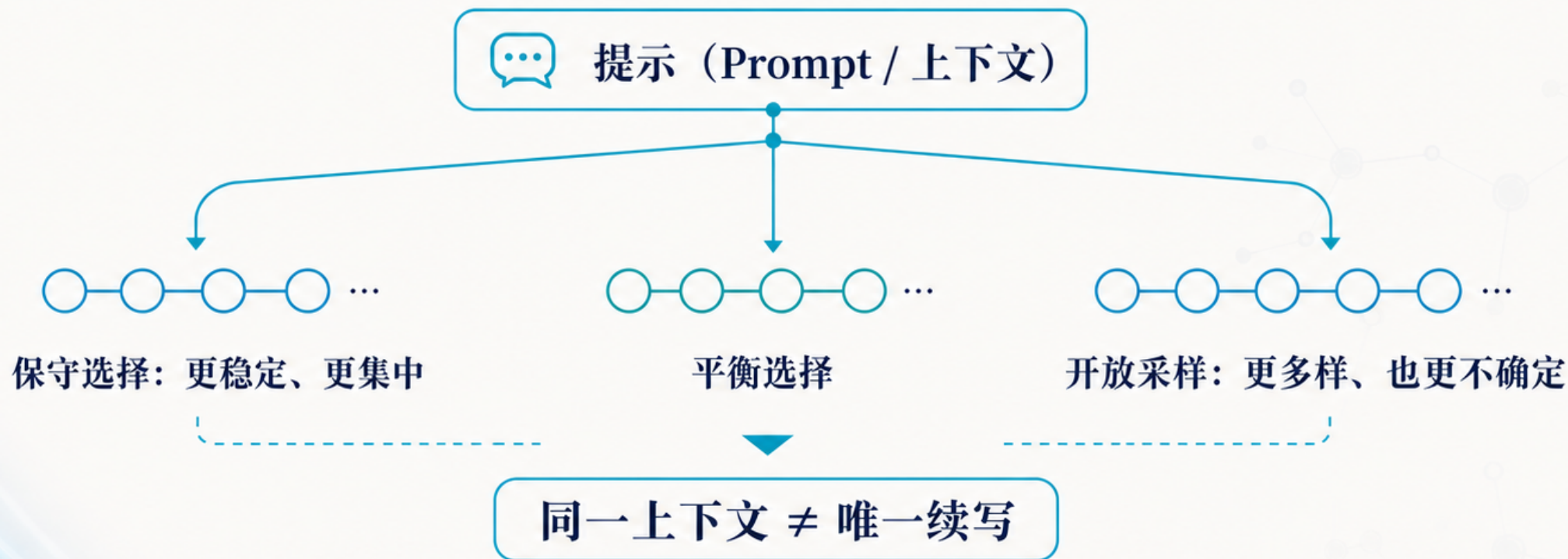


Attention 的直觉：当前 token 会动态“关注”上下文中更相关的位置。



这是一种计算关联机制，不等于人类意义上的“理解”或“注意”。

从概率分布到一段回答：生成中的选择



生成多样性、事实正确性与任务可靠性不是同一个指标。



幻觉：流畅文本为什么可能不真实？

两者并不天然等价。

语言上顺畅



证据上成立



来源

数据

计算

实验



虚构引用



错误关系



过期信息



错误计算/推理

流畅不是证据；需要外部事实、数据或计算结果来校验。

把概率输出变成可用结果：四种约束



【 目标不是“绝不出错”，而是让错误可发现、可追溯、可纠正。 】



More than a ChatBot 的必要条件

大模型 =
概率性生成
推理核心



智能体 = LLM + 目标 + 知识 + 工具 +
Observation + 规划 + 约束

LLM 怎样真的和外部世界交互？



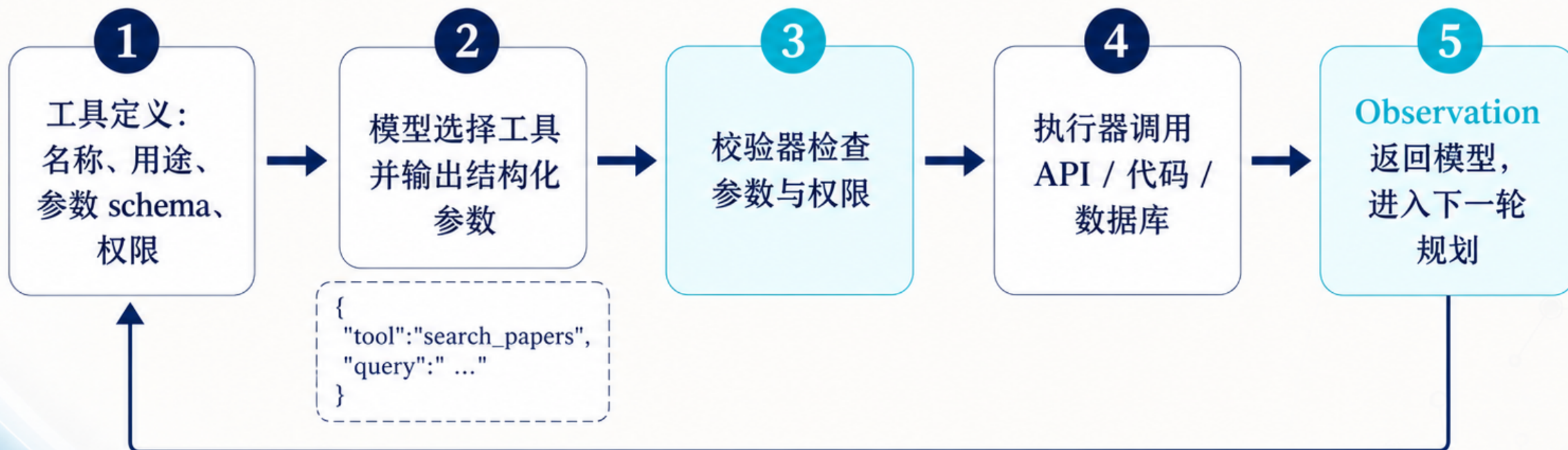
智能体的六个部件



智能体不是一个更大的模型，而是一个受约束的任务系统。



工具调用如何实现：从自然语言到结构化 Action



不是模型直接操作数据库或电脑

来源：Yao et al., ReAct, 2022, <https://arxiv.org/abs/2210.03629>; Schick et al., Toolformer, 2023, <https://arxiv.org/abs/2302.04761>.



例子：查文献并生成证据表

“给出某研究问题的证据表”，而不是“帮我写一段综述”。

01

检索

02

获取元数据/全文

03

摘取证据

04

冲突比对

05

生成可复核表格

论文	核心主张	证据位置	证据等级	局限
Paper A	—	—	—	—
Paper B	—	—	—	—
Paper C	—	—	—	—

高质量交付物 应能回链到论文、证据位置和局限。



代码、浏览器与文件：行动能力也是风险能力



能力越接近真实世界，权限、日志和人工确认越重要。

来源：Yao et al., ReAct, 2022, <https://arxiv.org/abs/2210.03629>; UNESCO, *Guidance for Generative AI in Education and Research*, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000386693>。



知识增强、记忆与 RAG：让模型先查再答



检索到出处 ≠ 出处支持主张；
知识库必须版本化。

规划、反思与协作：为什么不只是一轮工具调用？



多智能体是可选机制：可分工，
也会增加协调成本和错误传播。

OpenClaw：个人智能体热潮的一个人口

个人智能体的核心不是一个聊天窗口，而是一套 **channels + skills + tools** 的任务网络。



 GitHub 热度是生态信号，不是安全或教学适用性的证明。

中美自动化工具生态：按任务与边界比较

比较公开定位与 workflow，不做能力排行榜。

工具	主要任务	交互入口	外部工具/环境	教师的低风险试点
 Codex	代码理解、生成、审查与调试	CLI/IDE/App	代码库与工具链	课程代码示例、资料整理
 Claude Code	终端式代码工作流	CLI	代码库与命令环境	小型教学项目
 Trae	AI辅助软件开发	IDE	开发环境	学生项目脚手架
 WorkBuddy	办公任务与多模态交付	桌面/办公入口	腾讯办公生态	教务/资料流程草案
 OpenClaw	个人智能体与技能生态	多渠道	技能/插件	受控的个人知识任务

来源：OpenAI Codex, <https://developers.openai.com/codex/>；Anthropic Claude Code, <https://code.claude.com/docs/en/overview>；
Trae, <https://www.trae.ai/>；WorkBuddy, <https://www.codebuddy.cn/docs/workbuddy/Overview>；
OpenClaw, <https://github.com/openclaw/openclaw>。

“自动化”不是“放手”：安全与评价框架

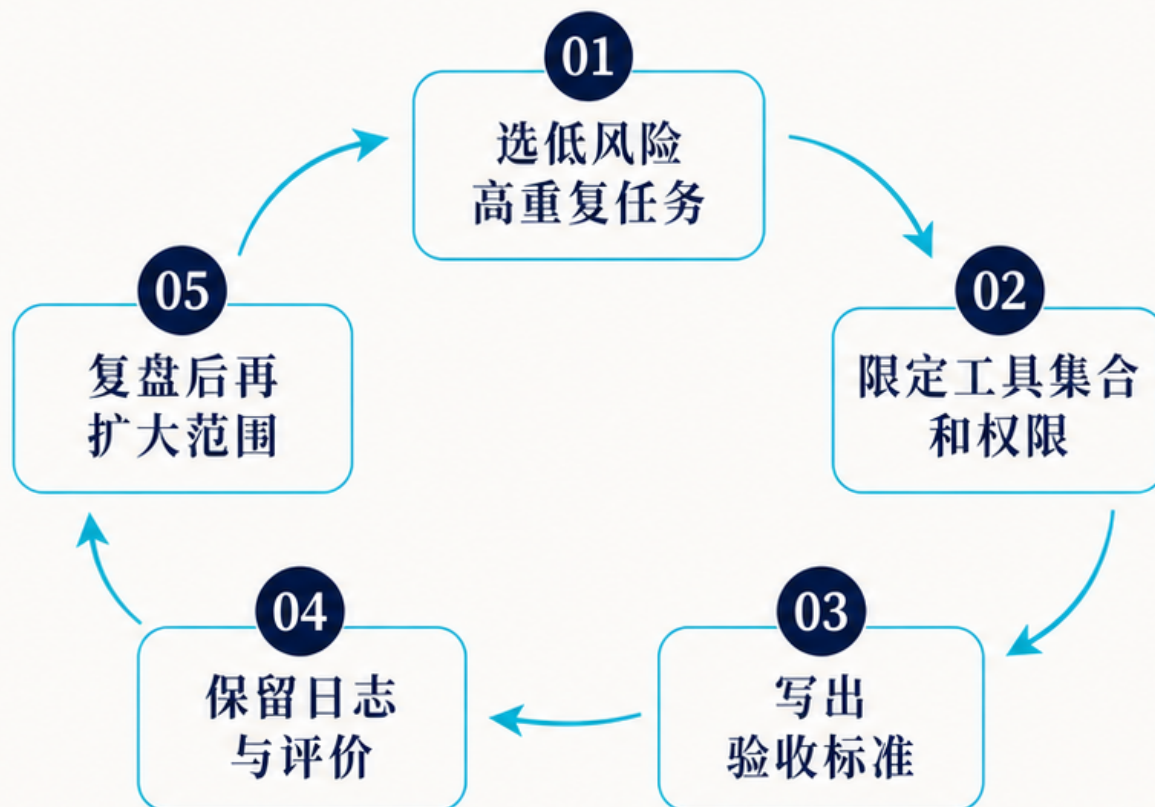


任务完成率 · 事实/引用错误率 · 人工修改量 · 完成时间 · 用户满意度 · 可复现率



自动化必须伴随边界、日志与人工责任。

起步原则：一个任务、一组工具、一次复盘



先做可验证的小闭环，不从自动评分、自动发布或真实设备控制起步。

来源：UNESCO, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000386693>; Gao et al., 2025, <https://arxiv.org/abs/2502.05151>。

AI4Science: 瓶颈往往在流程，而非单一步骤



🔗 智能体连接原本断裂的步骤：从文献问题到计算计划，从结果到实验候选。

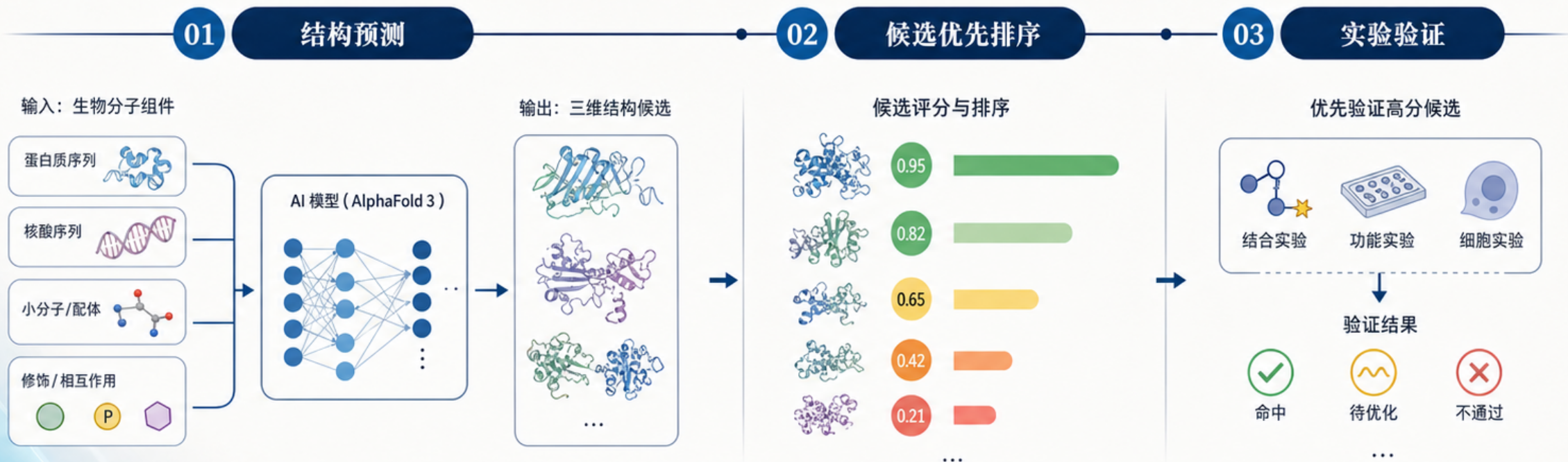
👤 人类研究者负责问题价值、方法选择、数据质量、结论解释和学术责任。

科研智能体的最小闭环：Plan-Tool-Observation-Verify



来源：Yao et al., ReAct, 2022, <https://arxiv.org/abs/2210.03629>; Boiko et al., Autonomous chemical research with large language models, 2023, <https://doi.org/10.1038/s41586-023-06792-0>.

科学发现：AlphaFold 3 如何改变候选空间



AI 的科研价值是帮助研究者在巨大的候选空间中优先排序。



正确叙事不是替代科学家，而是改变假设生成、筛选和实验设计的效率。

文献理解：从“问模型”到“让证据可回链”



当证据不足或冲突时，保留不确定性，而不是平均成确定结论。



仿真与代码：LLM 编排，科学软件给出可检验输出



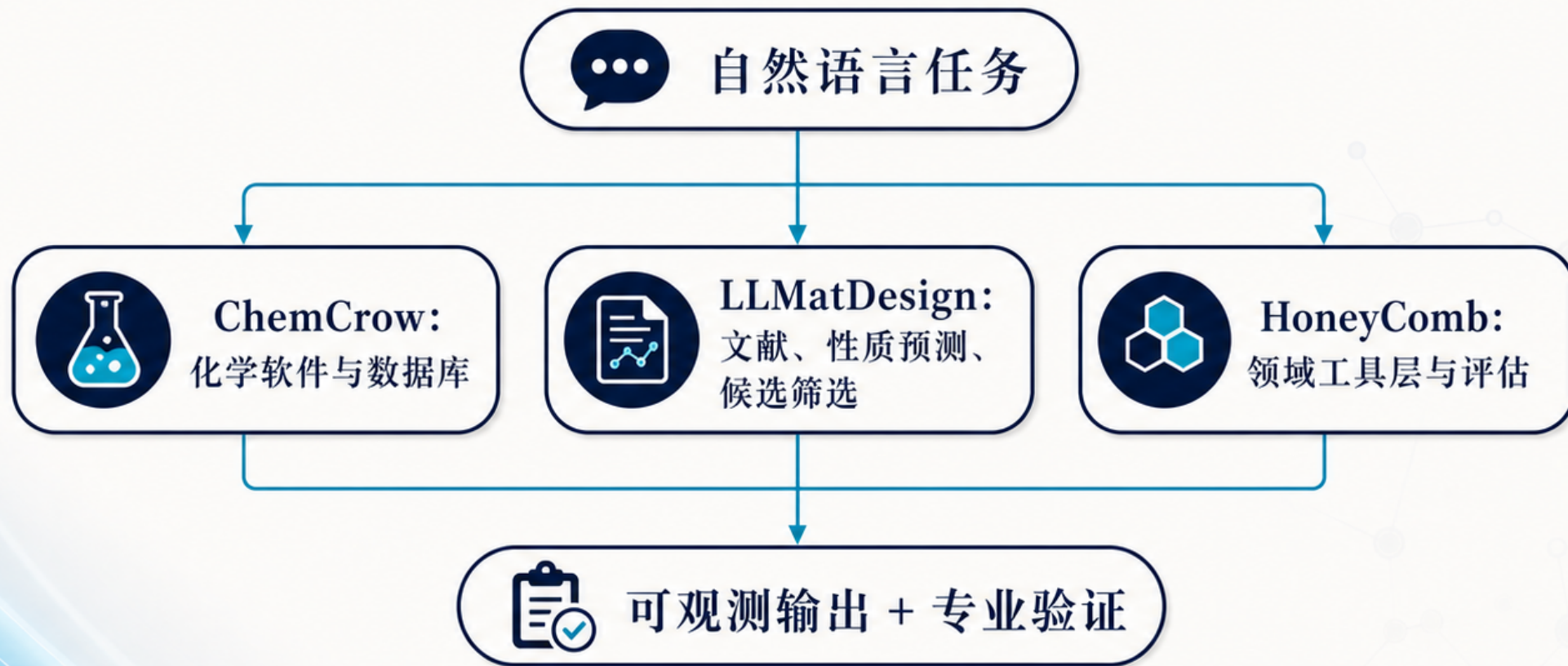
记录与可复现清单

- ✓ 代码版本
- ✓ 参数
- ✓ 随机种子
- ✓ 环境
- ✓ 输入数据
- ✓ 收敛信息



数值真相来自经过验证的求解器、数据和模型假设。

26 • 化学与材料：工具调用把语言任务变成领域动作



候选生成不等于合成成功，工具输出不等于材料性能兑现。

来源：Bran et al., ChemCrow, 2023, <https://arxiv.org/abs/2304.05376>; Chen et al., LLMatDesign, 2024, <https://arxiv.org/abs/2406.13163>。

自动化实验室：真正的“行动”发生在受控接口后



AI Scientist 与 AI co-scientist: 看见未来，也看清原型边界



江汉大学
JIANGNAN UNIVERSITY

可能性



想法生成



代码实验



报告草拟



反思评估

边界



新颖性



可靠性



可复现性



统计严谨性



学术价值



显著标记：预印本 / 研究原型；用证据等级而非宣传热度解释其位置。

来源：Google Research, 2025, <https://arxiv.org/abs/2502.18864>; Lu et al., 2024, <https://arxiv.org/abs/2408.06292>。

可信 AI4Science：四道闸门



通过四道闸门，输出才可能成为科研工作流的可靠组成部分。

AI4Education 的北极星：提升学习与教师判断



当系统帮助学生暴露过程、获得提示和反思，它才可能增强学习。

来源：UNESCO, Guidance for Generative AI in Education and Research, 2023,
<https://unesdoc.unesco.org/ark:/48223/pf0000386693>; UNESCO, AI competency framework for teachers, 2024,
<https://unesdoc.unesco.org/ark:/48223/pf0000391104>。

教育智能体最小闭环：目标—任务—证据—教师干预



真正的个性化不是给每个学生不同答案，
而是让教师看见不同理解路径与困难。

五类应用：预习、辅导、资源、评价、教师增能



最终课程设计主权属于教师。

eduJupyter: 把可执行环境转化为可观察的学习过程



任务



运行

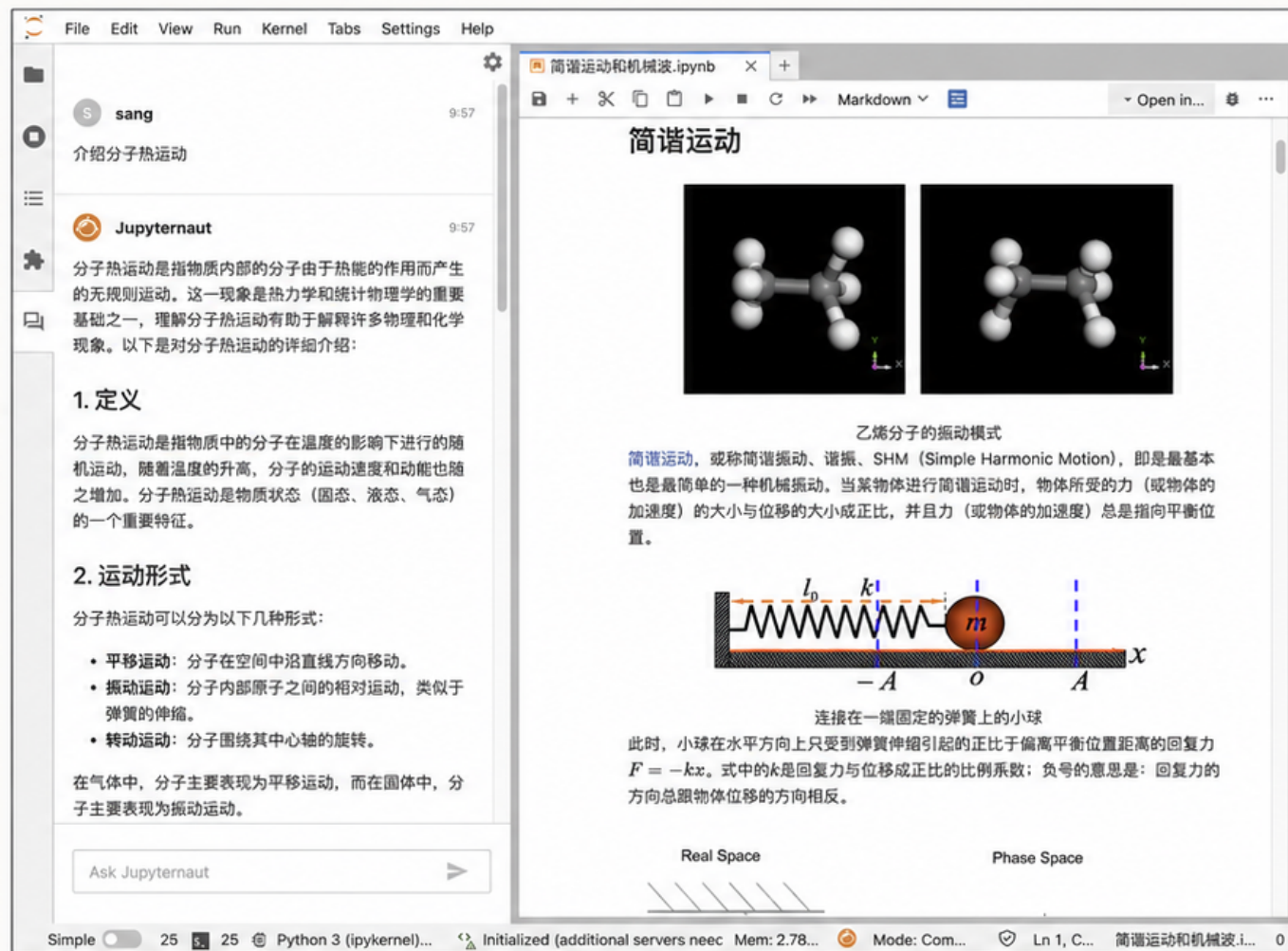


报错/修改



解释与教师反馈

形成性评价从
是否提交最终答案,
转向如何尝试、
如何修订、
能否解释结果。



The screenshot displays the eduJupyter web interface. On the left, a sidebar shows the file explorer with a notebook named '简谐运动和机械波.ipynb'. The main area shows the notebook content, which includes a title '简谐运动', two molecular vibration diagrams, a definition of simple harmonic motion, a diagram of a mass-spring system, and a discussion of real vs. phase space. The interface includes a top menu bar with options like File, Edit, View, Run, Kernel, Tabs, Settings, and Help. The bottom status bar shows the current kernel is 'Python 3 (ipykernel)' and the mode is 'Com...'.

项目截图：报告人授权提供（已脱敏）

BlastGPT：把教育智思体放进教师协同流程



项目截图：报告人授权提供（已脱敏）

评价与诚信：从“抓 AI”到“看学习过程”



检测器可以是辅助信号，但不能作为唯一裁决。

政策窗口：从教育数字化到“人工智能+教育”

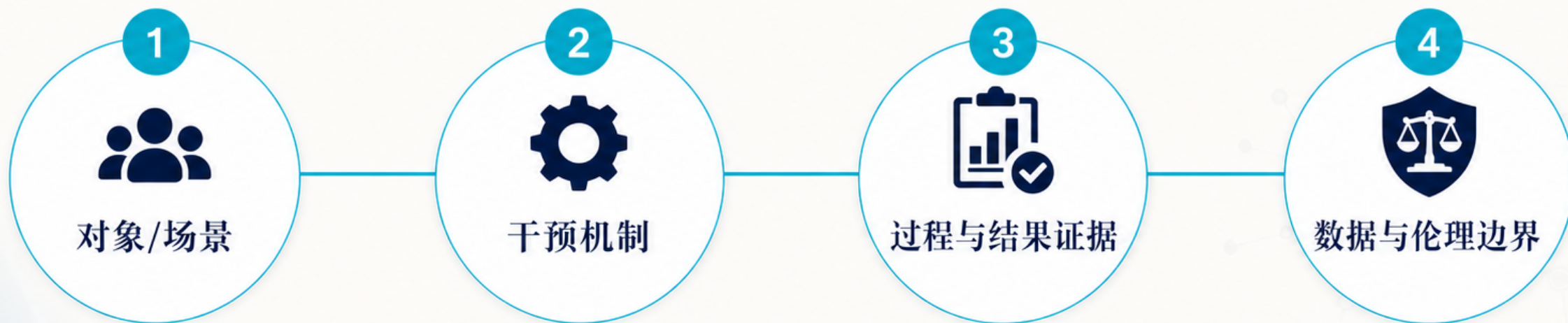


政策是研究背景和方向依据，不是教学效果的自动证明；
课题仍须回答对象、机制、证据和边界。

来源：国务院，2017，https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm；
教育部等五部门，2026，https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html。



把工具试用写成教研问题：四段式问题链



? 示例问题：eduJupyter 的运行轨迹与错误分类，能否提高计算类课程形成性反馈的及时性与针对性？



青年教师行动清单：本学期就能做的三件事

01

选一个小任务：高频、低风险、可验收

02

设一条边界：资料、工具、隐私、确认

03

留一组证据：输出、修改、错误、时间、反馈

先做可验证的小闭环，再扩大自动化程度。

来源：UNESCO, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000386693>;

教育部等五部门, 2026, https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html.



More than a ChatBot, 也必须 More than a Demo

可信人机协同流程



AI4Science 接入文献、计算、实验与可复现证据链。

AI4Education 接入课程目标、学习过程、教师判断与公平治理。

演示只能说明可能；证据、评价、权限与制度设计才能说明值得推广。

来源：Ji et al., 2023, <https://arxiv.org/abs/2309.05922>; Yao et al., 2022, <https://arxiv.org/abs/2210.03629>;
UNESCO, 2023, <https://unesdoc.unesco.org/ark:/48223/pf0000386693>。

你最希望智能体替你消除哪一段重复劳动？

它需要什么工具、什么证据、什么边界，才值得交给它？

从一个低风险任务开始，把实践做成可观察、可评价、可复盘的闭环。



江汉大学
JIANGHAN UNIVERSITY

谢谢！

桑鸿乾 | 江汉大学

