PROPOSAL 01

从方程结构到先验闭合：面向研究级 PDE 分析的 AI 方法

Closure-Guided AI for Research-Level PDE Analysis

核心主张：用可定位的闭合障碍改变下一步数学操作，比通用的“生成—审查—重试”更能推进原假设下的研究目标。
独立贡献：PDE 专用的研究状态、候选操作与障碍反馈；不是把检索、证明和验证三个代理重新组合。
状态：方法设计与实验计划；未报告尚未执行的实验或未完成的新定理。

摘要

研究级偏微分方程分析不仅要求生成正确的局部推导，还要求多个估计在导数、耗散、参数和时间尺度上协同支撑最终定理。局部正确而整体无法闭合的推导，因而构成了不同于单步证明错误的研究障碍。本文提出一种面向非线性耗散型 PDE 的闭合驱动 AI 研究方法。系统首先根据存在性、唯一性、初值连续依赖性或长期行为目标固定证明义务，再利用方程的耦合与耗散结构选择变量、分解、乘子及能量。候选估计被组织为保留数学对象和参数依赖的结构化状态；在可处理的组合子问题中，符号约束与可复核证据定位缺失控制，其余解析障碍则以明确的未完成义务返回。该反馈用于选择下一轮证明操作，而不是仅要求模型重复检查全文。系统始终区分原目标下的修复与通过加强条件获得的新命题。我们以流体与动理学中的稳定性和衰减任务为场景，设计预算匹配的强基线、操作级消融和跨母问题测试，分别评价原目标完成、错误结论接受和有效研究成本。Boltzmann 接触波案例提供长程研究的核心压力测试，但其尚未闭合的解析内容不作为已经取得的成果。该研究考察 PDE 结构能否成为证明搜索的可执行约束，使 AI 从生成分散的估计推进到构造可支持完整研究结论的论证。

研究边界

首篇聚焦能量方法占重要地位的流体、双曲—抛物系统及角截断动理学问题，不声称统一解决全部 PDE。存在性、唯一性和连续依赖是目标入口；长期稳定性与衰减是另外的任务，不与连续依赖混同。只有已明确建立的基础结果才能作为输入。最终评审依据是数学进展，而不是状态图更整齐。

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1. 相关工作与真正需要建立的新意

必须与能够开展研究的系统比较，而不是以竞赛题较简单来论证创新。

最接近工作	已有内容	本文需要独立验证的增量
LIPS，ICLR 2025	用不等式领域操作组织神经—符号搜索。[1]	PDE 中无限维范数、导数及时间耗散的兼容性如何控制搜索。
QED，2026 预印本	规划、证明与验证分离；报告包含流体 PDE 的研究项目。[2]	不是再做多代理，而是让具体闭合障碍触发不同解析操作。
RMA，2026 预印本	研究问题分析、文献与结构化记忆、迭代证明。[3]	在同样工具与记忆条件下，闭合反馈是否仍有可重复增益。
AI Co-Mathematician	交互式、状态化的开放数学研究工作台。[4]	专业先验是否减少错误研究方向及无效证明投入。

三个可证伪假设

H1：操作选择。遇到耗散缺口、导数损失、参数失配和不可积项时，结构化反馈提高下一步有效操作的比例。有效性由操作后获得的已核验控制定义，而不由模型解释是否自洽定义。

H2：研究完成。在同样总预算下，完整方法提高原假设下的目标完成率；只降低错误接受率而未提高有效推进，不足以支撑“更能做研究”。

H3：专业迁移。在未参与开发的母命题和至少一个保留方程子类上仍有收益。只在 HXY 的历史轨迹上有效，最多支持个案分析，不支持可复用方法。

创新应落在何处

核心不是某条代数不等式的自动验证，也不是发明全新的 PDE 技巧。拟建立的增量是**“已核验估计—闭合障碍—候选数学操作”之间的可执行接口**：反馈必须告诉搜索器哪一项控制不足、哪类操作能改变这种不足，并在操作实施后重新检验。

分类模板只提供候选方向，不能预置关键乘子或隐藏标准答案。一般约束求解、依赖图及自然语言反思均作为组件；是否构成足够方法贡献，由操作级实验和完整研究结果共同决定。

预先承认的失败情形：若一个得到充分 PDE 指令、拥有同样工具和结构化笔记的通用系统即可达到相同结果，本方法的专用状态与反馈尚未证明必要性。

2. 工作流程与各模块的数学责任

实线为研究推进；虚线为失败或预算状态。每次回路都应留下明确的新估计或反例。

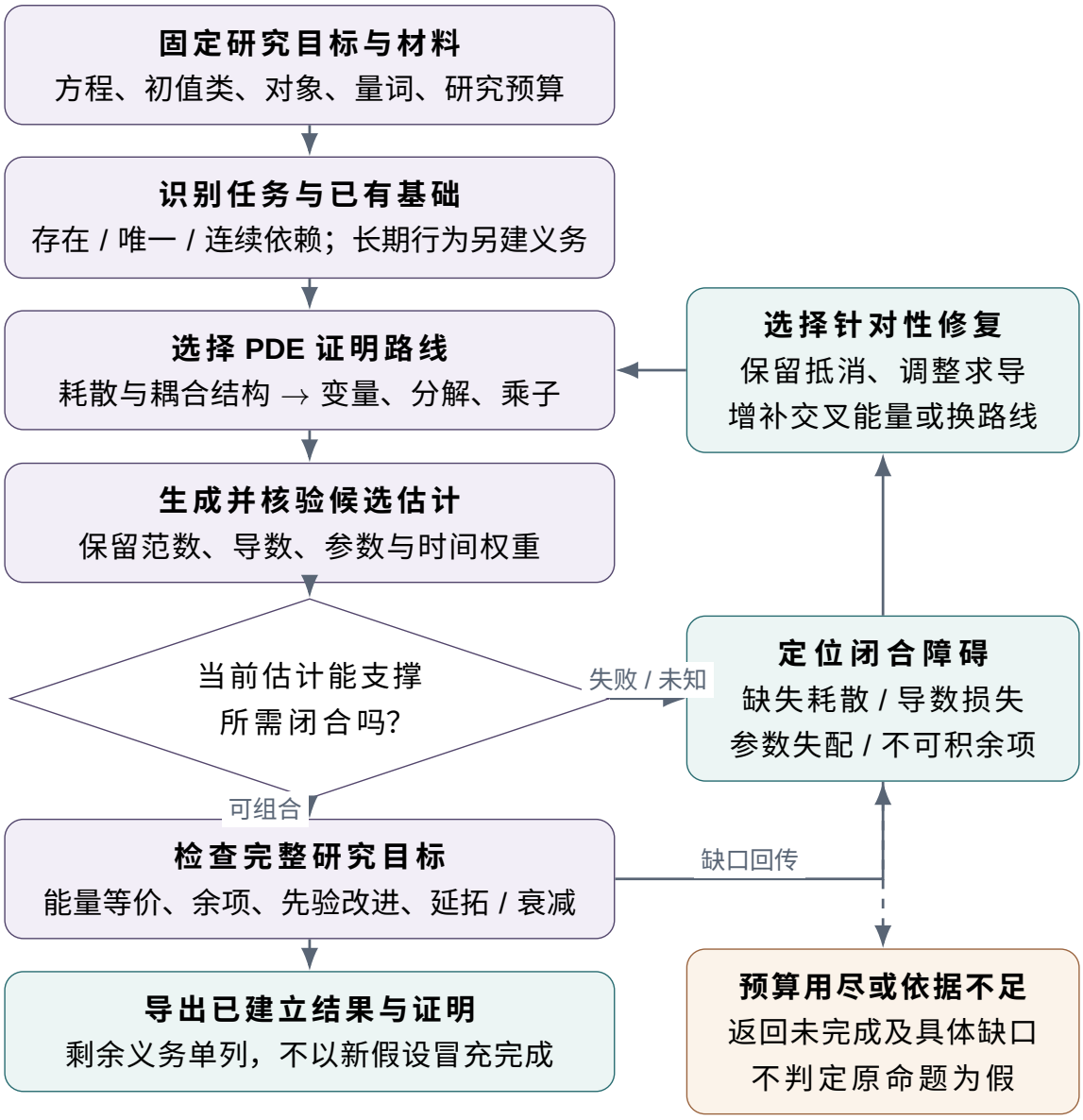

图 1：闭合检查不是一般 PDE 正确性判定器。可组合仅指现有估计通过指定组合测试；仍须检查能量等价、余项处理、局部解基础及延拓或衰减结论。

3. 最小可实现算法：从估计到障碍，再到操作

先实现一个可审查的专业控制器，不以大规模训练或通用自动证明为前提。

3.1 输入与状态

输入固定方程、区域和边界条件、初值类、比较对象、正则性、参数范围、量词与可用文献。估计记录只增加机器可读的元数据，不向 PDE 定理添加新条件。

每条估计保留：原始等式或不等式；各项对应的对象与范数；导数阶数；空间、速度及时间权重；系数的实际依赖；有效来源。**不同范数不因名字相似而合并；未证明的嵌入与插值必须另建义务。**

3.2 可计算的组合子问题

在固定已证明估计族中，写出

$$\frac{d}{dt}E_i + \sum_j M_{ij} D_j \le R_i, \qquad D_j \ge 0.$$

M_{ij} 是同一耗散对象的净系数，可为负。寻找非负组合权重 w_i，使

$$\sum_i w_i M_{ij} \ge c_j > 0, \qquad \mathcal{E} = \sum_i w_i E_i \asymp E_{\text{target}}.$$

常系数有理数情形可用线性规划提出候选，并用精确运算复核；能量等价另行检查，尤其 E_i 含交叉项时不可省略。系数依赖退化参数时，必须证明整个规定参数区间中的不等式，有限网格只能用于探索。

不可行证据只针对**当前估计族、允许组合和指定目标控制**；它不否定 PDE 命题，也不排除另一种能量或估计路线。未知系数、未控制余项或无法认证的数值解均返回“未判定”，不伪装成证书。

3.3 障碍如何生成下一步任务

具体障碍	首轮候选操作	操作的验收
缺失某分量耗散	利用耦合构造交叉能量	新耗散明确出现，且能量等价成立
最高阶导数损失	调整分部积分、求导次序或变量	不超出原正则性且余项可处理
退化参数不可吸收	保留系数、改变配对或重新加权	常数选择满足原有一致性要求
时间项不可积	保留抵消、时间权或渐近主项分析	真正改善时间控制，而非加强初值

候选仅表示值得尝试，不保证成功。低成本代数检查用于筛选，复杂解析估计仍需完整证明。优先修复支撑多个目标控制的障碍；若连续若干次修复无新增有效控制，允许切换路线，而不是无限重复同一提示。

4. 一个完整的小例子：反馈必须改变数学操作

这是验证工作流设计的教学示例，不是新 PDE 成果，也不计入研究级任务完成数。

考虑实值、充分光滑且在无穷远衰减的解

$$v_t + u_x = 0, \qquad u_t + v_x = u_{xx}, \qquad x \in \mathbb{R}.$$

所有积分分部均在上述条件下合法。令

$$E_0 = \tfrac{1}{2}(\|v\|_2^2 + \|u\|_2^2), \quad E_1 = \tfrac{1}{2}(\|v_x\|_2^2 + \|u_x\|_2^2).$$

直接能量等式为

$$E_0' + \|u_x\|_2^2 = 0, \qquad E_1' + \|u_{xx}\|_2^2 = 0.$$

结构性障碍：这两个等式的任何正权组合都没有 $\|v_x\|_2^2$ 耗散。反馈不应是“继续求导”，而应定位为**无直接粘性的分量缺少耗散，需要利用双向耦合**。

考虑交叉项 $J = \int_{\mathbb{R}} uv_x\,dx$。由方程和积分分部，

$$\begin{aligned} J' &= \int (-v_x + u_{xx})v_x\,dx - \int uu_{xx}\,dx \\ &= -\|v_x\|_2^2 + \int u_{xx}v_x\,dx + \|u_x\|_2^2 . \end{aligned}$$

选取固定 $0 < \eta < 1$，令 $\mathcal{E} = E_0 + E_1 + \eta J$。因为

$$|J| \le \tfrac{1}{2}(\|u\|_2^2 + \|v_x\|_2^2) \le E_0 + E_1,$$

有 $(1-\eta)(E_0+E_1) \le \mathcal{E} \le (1+\eta)(E_0+E_1)$，且

$$\boxed{\mathcal{E}' + (1-\eta)\,\|u_x\|_2^2 + (1-\eta/2)\,\|u_{xx}\|_2^2 + \tfrac{\eta}{2}\,\|v_x\|_2^2 \le 0.}$$

最后一步使用 $|\int u_{xx}v_x| \le (\|u_{xx}\|_2^2 + \|v_x\|_2^2)/2$。

本例检验的能力：不是让系统复述交叉能量的名字，而是从缺失耗散定位所需机制，给出符号正确的交叉项，核实新余项的代价，并完成正定性及组合。主实验必须进一步进入非线性、变系数或退化参数问题；本例不支撑衰减率或全局非线性结论。

5. 任务集：研究级，但不能不可判定或只剩一篇案例

先导阶段只用于校准难度；母命题及其所有变体在开发与测试之间整体隔离。

任务层	输入与目标	成功标准
机制校准	已知结构的小系统；定向移除一个关键控制	恢复实际缺失控制，而非复述标准术语
研究检查点	给定文献和已有推导；补齐明确解析义务	原假设下完整证明预登记的目标
受控推广	改变背景、参数或可用正则性，目标预先核验	有效证明；不成立任务接受严格反驳但分开统计
前瞻研究	专家提供尚未解决、范围明确的研究任务	独立核验新证明，并单独审查新颖性

拟定规模，而非功效保证

先导：来自 3 个母问题的 12 个检查点，不进入正式测试。主实验：另取 12 个母问题，每个 2 个检查点，共 24 项，涵盖双曲—抛物能量、Navier–Stokes 波稳定、角截断 Boltzmann 分析。另设 2–4 项前瞻研究，单独报告而不与已知答案任务混算。数量在先导结束后、主测试运行前冻结。

主比较采用 4 种系统、2 个可复现模型家族、每任务 3 次运行，共 576 次中预算运行。其他预算档及消融在预先分层抽取的子集执行。专家投入可能是最大成本：所有宣称完成的产物都须核验；对失败与拒绝样本也作随机复查，检验是否过度保守。记录每次审查用时。

HXY 案例的准确位置

核心开发案例是 Huang–Xin–Yang 的接触波研究。[5] 已有内部方案将既有衰减结论重建与特定条件下更快衰减的探索分开，后者主体解析输入仍须独立建立。[6]

因此，HXY 用于展示初值空间、比较对象、上下界与参数量词之间的真实研究约束；**不能把条件性 $t^{-1/2}$ 论证预写为已经完成的新定理**。HXY 的变体也不能作为独立跨领域测试。

任务完整性与保密

每项任务交付自包含的目标、精确文献范围、合法输入和专家标准。已发表材料存在预训练污染可能，只能作为机制评测；前瞻任务须在运行前封存，并取得合作者使用与公开授权。私有题未公开的失败也要计入分母，至少公开不泄露数学内容的任务编号和状态。

6. 对照、消融与主指标

公平比较的单位是完整研究预算，不是单次模型调用次数。

6.1 必须包含的强基线

基线 A： 同一模型，完整 PDE 专业说明、检索、工具、结构化笔记及多轮反思。不能只给一句“证明这个定理”。

基线 B： QED 或同等实现的研究代理，使用相同模型与资源；如无法复现原系统，明确标注为同结构实现，不将其结果冒充原论文系统。[2]

基线 C： 相同候选数学操作与工具，但不提供结构化闭合反馈，以隔离知识库及额外工具的作用。

完整方法： 保留操作库、状态及障碍反馈。主比较同时给各方相同可用文献、人工提示额度及总成本上限。闭合检查器自身消耗计入预算。

6.2 机制实验

对同一冻结状态分派不同反馈，比较**下一轮候选操作产生的新有效控制**。分别移除目标诊断、耗散对象匹配、参数一致性检查及障碍回传。某模块无稳定贡献应如实报告。不要把全部模块同时删去后与完整方法对比，再声称每个模块都有用。

6.3 定量指标

以母问题 f 为等权单位，定义预算 B 内的原目标完成率

$$\widehat{C}_m(B) = \frac{1}{F}\sum_{f=1}^{F}\frac{1}{N_f R}\sum_{i=1}^{N_f}\sum_{r=1}^{R}\mathbf{1}\{\text{原假设下，目标由独立审查确认完成}\}.$$

并列报告： 已确认错误的结论接受率；尚无充分证据却宣称完成的比例；首次有效完成成本；所有运行总成本。诊断成功、严格反驳和条件性推进分别统计，不能替代证明完成。

原子进展指标： 只对预先登记的关键义务计数, 不能靠自动拆出大量小引理抬高进度。路径长度、token 数和模型自信均不是数学进展。

6.4 统计与不确定性

按母问题聚类报告差值与区间；重复运行并非独立研究问题。主要对比和最小有意义效应在先导后冻结。模型家族分层报告，不用总体均值掩盖单模型失效。对仅有 2–4 项前瞻任务的结果作案例级陈述，不给虚假的精确成功概率。

7. 实施、风险与投稿门槛

一个可运行控制器、一个可信任务集、一组决定性实验，比三个宽泛的智能体名称更重要。

实施顺序

阶段 A：冻结语义。完成任务目标表与来源定位；实现范数、导数、参数和时间权重的结构化记录；用第四节例子测试符号及接口。

阶段 B：闭合子程序。实现固定估计族的精确组合检查、已知失配定位及“未知”出口；将反馈对接少量可执行 PDE 操作，而非建立无边界工具大全。

阶段 C：研究运行。先导校准后冻结系统、预算和测试集；保存显式证明、调用与修订产物，不依赖模型隐藏思维链。对人工帮助作逐次标注。

阶段 D：独立验证。外部专家核验完成任务，另一名研究者重放局部产物；审查错误样本与全部失败分母。按同一规则评判所有方法。

最可能的四个质疑及应对证据

质疑	应提交的直接证据
“只是更长的提示词”	同预算、同知识与工具、只有反馈机制不同的比较
“关键技巧还是专家给的”	封存输入；展示专家未提供关键操作的运行；人机协作另列
“闭合检查器不可靠”	明确可认证子集；精确证书；未判定出口；抽查输入提取
“只适配 HXY”	按母问题隔离的保留任务与跨方程子类结果

作为独立 ICLR 论文的最低证据包

必须有可运行机制、强基线与机制消融；在多个独立研究任务上得到可核验的改进；至少完整交付一项非教学级的研究义务链，且不靠隐藏新增条件。若主张数学发现，还必须拿出具体的新结果与独立新颖性判断。**“至少一项”是项目内部质量门槛，不是会议接收标准**。

若只能展示正确识别缺口，论文应收缩为诊断方法；若仅有一个手工打磨的成功案例，不宣称一般研究系统。不存在“PDE 很难所以审稿人必须认可”的论证。

主文的九页组织建议

第 1 页问题与结果；第 2 页相关工作及专业差异；第 3–4 页方法和一条完整反馈链；第 5 页任务与评估；第 6–7 页对照和消融；第 8 页真实研究案例；第 9 页局限与结论。核心数学证据在正文可定位，完整推导与复现实验放附录。

8. English abstract and references

英文摘要为研究方案版本；正式投稿须补入真实数学产出与关键实验结果。

English abstract

Research-level PDE analysis requires more than locally correct derivations: estimates must jointly control the derivatives, dissipative directions, parameters, and time scales required by the final theorem. We propose a closure-guided AI research method for nonlinear dissipative PDEs. Starting from a fixed research objective and its established foundations, the system uses the equation's coupling and dissipation structure to organize variable choices, decompositions, multipliers, and energy estimates. Candidate estimates retain their mathematical objects and parameter dependencies in an explicit research state. For tractable composition subproblems, checkable constraints identify missing control; unresolved analytic questions remain explicit obligations. These obstructions guide the next mathematical operation rather than merely triggering another unrestricted proof attempt. Goal-preserving repair is distinguished from changing the theorem's assumptions. We design budget-matched comparisons, operation-level interventions, and held-out parent problems in fluid and kinetic analysis to measure valid completion of the original objective, unsupported claims, and research cost. A Boltzmann contact-wave investigation supplies a demanding longitudinal case, without treating its unresolved analytic inputs as established results. The central question is whether PDE structure can become an operational constraint on proof search, enabling AI systems to construct arguments whose estimates support a complete research conclusion.

参考文献

[1] Li et al. *Proving Olympiad Inequalities by Synergizing LLMs and Symbolic Reasoning*. ICLR 2025. 正式论文集.

[2] An et al. *QED: An Open-Source Multi-Agent System for Generating Mathematical Proofs on Open Problems*. arXiv:2604.24021v4, 2026. 预印本；研究结果为作者报告.

[3] Zhao et al. *RMA: an Agentic System for Research-Level Mathematical Problems*. arXiv:2605.22875, 2026. 预印本.

[4] Zheng et al. *AI Co-Mathematician: Accelerating Mathematicians with Agentic AI*. arXiv:2605.06651, 2026. 预印本.

[5] Huang, Xin and Yang. *Contact discontinuity with general perturbations for gas motions*. Advances in Mathematics 219 (2008), 1246–1297. DOI: 10.1016/j.aim.2008.06.014. 本方案不另行宣称对该文全类结果作出改进。

[6] 项目工作材料：*HXY PDE* 可审计推理：摘要与研究方案，中文版，2026 年 9 月，§6–8。用于定位案例与未完成状态；不是已发表或已验收结果。