

题目编号：XH-202627

基于 Intern-S1 的数学智能体设计与推理创新 比赛方案

一、发榜单位

上海人工智能实验室

二、题目名称

基于 Intern-S1 的数学智能体设计与推理创新

三、题目介绍

本赛题旨在探索基于 Intern-S1 模型的 API 构建数学智能体在自主解题、推理解释与学习启发方面的能力。

选手需设计一个具备以下能力的智能体系统：

- （一）能理解自然语言组成的数学问题；
- （二）能自主规划思路、求解并输出结果；
- （三）能以启发式表达方式解释推理过程；
- （四）能在多类型数学问题中表现稳健。

（五）比赛聚焦智能体推理链设计、表达清晰度与教育启发性，而非模型训练或算法优化，降低算力门槛，鼓励创新型实现。亦旨在推动基于大模型的数学研究智能体探索，促进 AI 在高难度数学问题自主求解与理论发现中的应用，为智能体助力数学科研究突破提供新的可能性。

四、参赛对象

学生赛道：2026 年 6 月 1 日以前正式注册的国内全日制非成人教育的普通高等学校在校专科生、本科生、硕士和博士研究生（不含在职研究生），以及全日制职业教育本科、高职高专在校学生，可通过学生赛道申报作品参赛。

高校青年教师在指导学生参赛的同时不得以参赛人员身份参加同一选题比赛。发榜单位及同发榜单位有相关隶属关系单位的青年不得参加本单位选题比赛。

参赛对象可以团队或个人形式参赛，每个团队不超过 10 人，每件作品可由不超过 3 名指导教师进行指导。可以跨专业、跨学校、跨单位、跨地域组队，但同一团队所有成员均应符合本赛道相关年龄、身份要求。每件作品只可由 1 所高等院校或科研院所作为参赛主体提交申报。

五、答题要求

大赛分初赛和决赛两个阶段

（一）初赛：单智能体解题与解释能力展示

选手需基于 Intern-S1 模型的 API 构建一个能自动求解多类型数学问题的智能体系统。

系统应具备：

- 推理求解；
- 过程解释与学习启发；
- 结构化输出（JSON 结果）；

题目来源：组委会提供 112 道覆盖偏微分方程、复分析、

拓扑学、运筹学等 18 个子领域的数学问题数据集，选手需使用构建的智能体完成数据集内所有问题的推理。

注：由于大模型发展迅速为了保证题目的时效性和评测的有效性，初赛结束后可能会根据所有参赛智能体初赛问题的通过率，视情况增加更实时、更高难度问题。

（二）总决赛：多智能体协作与场景创新

在初赛方案基础上，扩展为多智能体协作系统，具备如下特征：

- 展示一个可交互 Demo（Web / Notebook）。

可选挑战方向：

- 支持超长程推理（如实现复杂多轮推理等）；
- 能自动识别题型并调用不同 solver agent；
- 引入过程校验的智能体，实现推理过程的自主调控；

六、作品评选标准

本赛题采用客观评测与主观评测相结合的方式综合评分，总分 100 分。

（一）评分总表

类别	评审维度	权重	评 分 方 式	说明
客观评测	答案正确性	60%	线 性 计 分	根据统一评测集上的正确率按比例折算得分

主观评测	推理策略与系统设计	10%	分档评审	重点考察模块设计、推理链组织、智能体协作逻辑
主观评测	展示质量	20%	分档评审	重点考察 Demo、视频、报告与材料完整性
主观评测	创新性与可扩展性	10%	分档评审	重点考察机制创新、可迁移性、可持续拓展潜力

（二）分项评分细则

1. 客观评测：答案正确性（60分）

以组委会提供的统一数学问题评测集为准，按参赛系统在全部有效题目上的正确率（通过题数 / 有效评测题数）进行线性计分，满分 60 分。

计分方式如下：

答案正确性得分 = 正确率 × 60 分

其中：

- 正确率按实际通过题数占有效评测题数的比例计算；
- 若因接口异常、格式错误、未按要求输出结构化结果(JSON)等原因导致题目无法判分的，视为未通过；
- 如初赛结束后组委会根据整体通过率情况追加更高难度题目，则最终客观评测成绩以组委会最终公布的正式评测集结

果为准。

2. 主观评测：推理策略与系统设计（10分）

重点考察参赛方案在数学问题求解中的系统架构完整性、推理链设计合理性、模块协同效率与可解释性组织能力。

档位	达标要求	分值区间	区间内浮动依据
A 档(优秀)	系统架构完整，具备清晰的任务拆解、推理规划、过程校验或多模块协同机制；推理链设计与数学问题特性高度匹配	8–10 分	根据模块清晰度、策略合理性、稳定性、可复现性、工程完成度浮动
B 档(良好)	具备较完整的解题流程设计，有基本的推理规划与解释组织；模块分工较清晰，但深度或稳定性一般	5–7 分	根据流程闭环程度、结构清晰度、策略适配度浮动
C 档(基础)	主要依赖单轮调用或简单串联，系统化设计较弱，缺少显式规划、校验或协作机制	0–4 分	根据是否形成基本可运行流程、是否具备最小可解释结构浮动

3. 主观评测：展示质量（20分）

重点考察作品提交材料与展示效果，包括 **Demo** 可用性、视频表达、技术报告完整度、结果呈现规范性。

档位	达标要求	分值区间	区间内浮动依据
A 档（优秀）	Demo 可稳定运行，交互流程完整；视频表达清晰；技术报告结构完整、逻辑严谨；结果文件规范、可复现性强	16–20 分	根据 Demo 稳定性、交互体验、视频质量、文档规范性、复现便利性浮动
B 档（良好）	Demo 基本可用或有完整演示；视频与报告较完整；材料基本齐全，但在表达、细节或复现性上存在不足	10–15 分	根据材料完整度、说明清晰度、结果呈现质量浮动
C 档（基础）	Demo 不完整或仅静态展示；视频/报告信息不足；材料缺漏较多，影响评审理解	0–9 分	根据最低提交完整度、核心信息可理解程度浮动

4. 主观评测：创新性与可扩展性（10 分）

重点考察作品是否在数学智能体机制设计、推理范式、协

作方式、验证机制、应用场景延展等方面具有创新价值，并具备进一步推广潜力。

档位	达标要求	分值区间	区间内浮动依据
A 档（优秀）	具有明确且有效的创新机制（如多智能体协作、过程校验、题型路由、长程推理控制、外部工具结合等），且具备较强迁移与扩展潜力	8–10 分	根据创新独特性、实际效果、可迁移性、推广潜力浮动
B 档（良好）	有一定改进或组合创新，能够体现一定的场景适配与扩展思路，但创新深度或验证力度一般	5–7 分	根据创新完整度、实际落地程度浮动
C 档（基础）	主要基于常规调用或通用范式，创新性较弱，可扩展性表述有限	0–4 分	根据是否具备基础改进、是否有明确扩展方向浮动

（三）补充说明

1. 同分处理原则

若综合总分相同，优先比较“答案正确性”得分；若仍相同，再依次比较“推理策略与系统设计”“创新性与可扩展性”得分。

2. 评测规范性要求

参赛系统须按要求输出结构化结果（JSON）及必要日志信息。因格式不符合要求导致无法解析、无法复核或无法纳入统一评测的，相关题目可判为未通过。

3. 作品导向说明

本赛题重点鼓励围绕数学问题求解过程的智能体设计创新，关注问题理解、推理规划、过程校验、解释生成与多智能体协作等能力，不鼓励仅通过提示词堆叠或固定模板完成单轮作答。

4. 违规处理说明

严禁人工逐题干预作答、赛后补填结果、伪造日志或使用未经允许的外部闭源服务代答。一经查实，取消参赛及评审资格。

七、作品提交时间

2026 年 5 月至 9 月上旬，各高校组织学生参赛，安排专业人员给予指导，为参赛团队提供支持保障。

2026 年 9 月 5 日前，各参赛团队通过邮箱的方式提交作品，具体要求详见作品提交方式。

2026 年 9 月 20 日前，由发榜单位完成初审，确定入围终审擂台赛的晋级作品和团队。

2026 年 10 月，发榜单位安排专门团队提供帮助和指导，各

晋级团队完善作品，冲刺攻关参加终审擂台赛。

2026 年 11 月，组织终审擂台赛，角逐“擂主”。

八、参赛报名及作品提交方式

1. 报名方式

(1) 参赛选手登录“挑战杯”官网 www.tiaozhanbei.net，在“揭榜挂帅”擂台赛报名入口注册账号，登录大赛申报系统在线填写报名信息。报名信息提交后，下载打印系统生成的报名表。

(2) 申报人在报名表对应位置加盖所在学校公章。

(3) 将盖章版报名表扫描件上传至报名系统，等待系统审核。请参赛选手注意查看审核状态，如审核不通过，需重新提交。

(4) 系统开放报名时间为 2026 年 5 月 30 日—6 月 30 日，逾期后系统将自动关闭报名功能。

2. 作品提交方式

提交具体作品至发榜单位邮箱：changshuai@pjlab.org.cn，务必一并提交 1 份报名系统中审核通过的参赛报名表（所有信息与系统中填报信息保持严格一致）。不需要额外在挑战杯官网提交参赛报名表及相关作品材料。

在初赛/总决赛截止时间点前，参赛团队应将提交所有要求的材料，包括技术方案文档（PDF）、日志与结果文件（JSON 格式）、展示视频（≤10 分钟，总决赛）、系统介绍 PPT（总决

赛)、源代码或 Notebook (可选, 建议提交或开源) 打包成.zip 格式压缩包。压缩包命名方式为: 申报人所在单位-申报人姓名-作品名称-联系电话(例如: XX 大学-张 XX-XX 方案-手机号)。

九、赛事保障

1. 资源支持: 参赛选手在完成报名后, 系统会自动提高报名账户 intern-s1 模型的流控策略;

后续可前往书生 API 控制台 (<https://studio.intern-ai.org.cn/console/dashboard>) 生成 API token 调用 intern-s1 模型。注意系统仅调整了报名账户的流控策略, 同一团队内的其他成员可使用此账户生成的 API key 共同使用 API。

2. 技术培训: 赛前赛中至少组织 2 场线上培训, 提供培训回放及答疑文档; 另外届时会根据实际情况, 决定是否组织线下的专场培训。

3. 专家指导: 建立线上答疑社群, 由上海人工智能实验室的研究员组成指导团队, 定期回复技术问题。

4. 交流平台: 在书生社区开设赛事专属社区板块, 支持参赛团队分享经验、交流问题, 促进技术共创。

5. 技术文档和课程:

(1) 数学智能体的官方基线: GitHub - InternLM/lagent: A lightweight framework for building LLM-based agents

(2) Intern-S1 模型介绍: GitHub - InternLM/Intern-S1: A

Scientific Multimodal Foundation Model

(3) 书生 API 技术文档：

<https://internlm.intern-ai.org.cn/api/document?lang=zh>

十、设奖情况及奖励措施

1. 设奖情况

分按参赛作品数量比例设奖，原则上评出“擂主”1名、特等奖5名，一等奖5名、二等奖5名、三等奖5名，获奖比例不超过参赛作品总数的30%。

2. 奖励措施

- 奖金奖励：擂主奖励10万元/个，特等奖奖励2万/个，一等奖奖励1万元/个，二等奖奖励0.5万元/个，三等奖奖励0.2万元/个。
- 发展支持：获奖团队有机会获得上海人工智能实验室的实习机会；所有获奖团队可获得赛事荣誉证书
- 曝光支持：获奖作品可在书生开源社区展示

3. 奖金发放方式

赛事终审结果公示无异议后（公示期约1个月），在30个工作日内通过银行转账一次性发放至团队负责人指定账户，税费由发榜单位承担。

十一、比赛专班联系方式

1. 专家指导团队

顾问专家：毛老师，联系电话：18607133138

顾问专家：顾老师，联系电话：15171489817

顾问专家：常老师，联系电话：13917557174

负责比赛期间技术指导保障。

2. 赛事服务团队

联络专员：常老师，联系电话：13917557174

联络专员：王老师，联系电话：15330019133

负责比赛期间组织服务及后期相关赛务协调联络。

3. 联系时间

比赛期间工作日（9:00-17:00）

附：上海人工智能实验室简介

上海人工智能实验室是我国人工智能领域新型科研机构，深耕战略性、原创性、前瞻性科学研究与技术攻关，突破 AI 重要基础理论和关键核心技术，打造“突破型、引领型、平台型”一体化综合研究基地，支撑我国 AI 产业跨越式发展，目标建成国际一流实验室与全球 AI 原创理论和技术策源地。

实验室聚焦通用人工智能（AGI）核心技术突破，构建了丰富的技术成果体系，包括覆盖语言、视觉、多模态等方向、支撑科研与产业应用的书生系列大模型，面向材料、药物等领域、推动 AI 与科学融合的 AI 科学计算共性平台，以及司南评测体系等产品矩阵。

其中，科学多模态大模型 Intern-S1 表现亮眼，在 2025 年全国中学生数学奥林匹克竞赛（CMO）中斩获卓越成绩，以 102 分位列大模型榜首且远超金牌线，其解题思路被阅卷专家评价为“新的解法”，充分彰显了实验室在复杂逻辑推理领域的前沿实力。