

题目编号：XH-202619

基于国产开源大模型的 AI Scientist 的研发与应用比赛方案

一、发榜单位

浙江阿里巴巴云计算有限公司

二、题目名称

基于国产开源大模型的 AI Scientist 的研发与应用

三、题目介绍

当前，人工智能正加速融入科学研究全过程（AI for Science）。传统科研模式高度依赖研究者的个人经验与文献积累，在面对海量数据、跨学科交叉和复杂系统时，往往效率低下且易陷入思维定式。大语言模型（LLMs）凭借强大的知识整合与逻辑推理能力，为科学发现提供了新路径——有望成为“AI Scientist”（人工智能科学家），自动提出新颖、可验证的科学假设，推动科研范式从“数据驱动”向“智能驱动”跃迁。

基于国产开源大模型（如 Qwen/千问）研发面向科学假设生成的 AI 系统，不仅关乎 AI 技术的自主可控，更是构建安全、开放、可持续的智能科研基础设施的关键一步。

由阿里云联合中国科学院国家天文台人工智能推进委员会、

他山学科交叉创新协会共同发起基于国产开源大模型的 **AI Scientist** 的研发与应用赛事，聚焦“可验证科学研究假设的自动生成”，鼓励青年学子结合具体学科领域（包括自然科学及社会科学），利用国产大模型开发出能理解科学问题、识别知识缺口并生成高质量、可验证假设的 **AI** 工具。我们鼓励青年学子基于专业深度的垂直研究，也鼓励参赛选手跨学科的创新融合，期待通过此次赛事推动科研智能化、促进国产大模型在垂直领域的落地、加速科研创新，并以青年创新力量服务国家科技自立自强战略。

赛题设置如下：

1. 核心任务：参赛团队需围绕特定学科领域，基于超级智能体（如 **OpenClaw**、**Hermes**）或编程设计多智能体系统（**Multi-Agent Systems**）架构，利用国产开源大模型——千问（**Qwen**）系列，开发一个具备问题理解、知识整合、关联发现与可验证假设生成能力的 **AI** 系统原型，实现从“数据/文献输入”到“可验证科学假设输出”的智能闭环。

2. 能力项参考：参赛者需构建一个智能化的“科研灵感流水线”，需要具备以下核心能力：

（一）**文献挖掘与事实提取：**结合选题的结构化信息，提取关键科学事实，避免断章取义。

（二）**逻辑驱动的假设生成：**利用归纳与演绎推理，基于已知事实生成初步假设。

（三）论证可行与多轮迭代：挖掘跨学科技术迁移潜力，确保引用真实可靠、假设具备可行路径，并进行多轮迭代以完善提案。

（四）智能体思辨与人在回路：构建可交互、具备教学意义的人机协作流程，对假设进行进一步辩论与迭代。

3. 科研问题集参考：参赛者可以基于自己的学科研究方向，参考公开发布的相关科学问题集进行选题。

（一）自然科学方向：

构建/开发适配具体自然科学领域的 **AI Scientist**，可链接该领域多维度、多模态的实测数据（如观测数据、实验数据、临床数据等）与文献知识、历史数据库资源，针对该领域的某核心科学问题（可涵盖异常现象识别、极端事件预测、内在机制挖掘、性能突破、路径优化、靶点发现等各类场景），通过智能体实现数据的智能清洗、分析与关联挖掘，精准识别关键信息、捕捉核心规律，进而自动生成具备科学性、可验证性的科学假设、内在驱动机制解释或工程优化策略，打破传统人工研究的局限，助力该领域的科学发现与技术创新。

（二）人文社科方向：

构建/开发适配某人文社科领域的 **AI Scientist**，链接该领域多维度、多模态的实测数据（如行为数据、交易数据等）与相关规则（法规、监管政策等），针对该领域的核

心需求（可涵盖违规行为识别、风险预警、机制挖掘、假设生成、合规建议等各类场景），通过智能体实现数据的智能清洗、分析与关联挖掘，精准识别关键信息、捕捉核心规律或风险点，进而自动生成具备合理性、可验证性、合规性的行为假设、证据推理链、风险预警或合规操作建议，打破传统人工研判的局限，助力该领域的规范治理与行业智能化升级。

4. 生成结果规范：系统最终生成的《科学假设与研究计划》需包含以下标准化字段：

待研究问题（Problem Statement）：明确指出当前领域存在的具体局限性。

解决思路（Rationale）：基于逻辑推理的创新点阐述，展示推导链条。

必要的技术手段（Technical Details）：详细列出验证假设所需的具体技术栈（如具体统计与机器学习、深度学习方法等）。

数据集（Datasets）：使用来源合规真实的数据集。

Source：假设推演依据的历史数据。

Target：验证实验所需的拟采集数据特征。

标题（Paper Title）：符合学术出版规范的标题。

摘要（Paper Abstract）：包含背景、方法、预期结果的完整摘要。

方法论（**Methods**）：具体的实施步骤，包括模型架构或实验流程。

实验设计（**Experiments**）：包含基线对比（**Baselines**）及评估指标（**Metrics**）。

实验结果（**Results**）：通过公式推导或实际执行，在一定范围内验证该实验可行性。

参考论文（**References**）：系统生成假设所引用的真实文献列表（严禁虚构）。

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、科研院所等作为参赛主体提交申报。

五、答题要求

技术基础：基座模型，必须基于千问开源模型（Qwen-Max/Plus/Turbo 等）；开发平台，需通过阿里云百炼平台调用模型 API 并提供调用凭证/截图；允许微调，允许基于下游任务、领域数据的 SFT（监督微调）。鼓励演示，推荐搭建前端、制作演示视频（推荐基于秒悟、万镜一刻）

参赛作品提交形式与规范：参赛团队需提交以下材料：

技术方案文档（PDF≤20 页），包含研究问题与解决方法、AI Scientists 架构设计与讲解（基于 Qwen 的多智能体架构或超级智能体的 Skills）、真实案例（基于问题集、满足规范的生成结果）、源代码（包含智能体工作流程的核心代码、上下文工程设计等）等内容。

附加提交（可选）：可交互前端页面、10 分钟内的演示视频。

六、作品评选标准

本榜题初审和终审决赛的评分标准一致，赛事评分采用“专家评审+AI 辅助评审”相结合的方式，通过多个维度进行评审：

指标	分值	评判标准
科学价值	40	核心假设创新性与自洽性（0-20 分） 方案可落地验证性（0-20 分）

指标	分值	评判标准
		注：该项由科学专家结合领域专业度完成评价
技术深度	30	超级智能体或多智能体协作设计（0-15分） 基于多模态大模型对科学模态数据的处理成效（0-15分） 注：该项由技术专家结合领域专业度完成评价
应用潜力	30	实际场景问题支撑能力（0-10分） 论文/专利成果转化潜力（0-10分） 代码与结果可复现性（0-10分） 注：该项由科学及技术专家综合完成评价
总分	满分 100	—

七、作品提交时间

2026 年 5 月至 9 月上旬，各参赛团队选择榜单中的题目开展研发攻关，各高校、科研机构等组织协调机构应组织学生参赛，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，各参赛团队要向发榜单位完成作品提

交，具体要求详见本方案第八点第（二）款，并严格遵照发榜单位明确的提交规范执行。

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。

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

（一）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或所在单位公章。

（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。

（二）作品提交方式

请已在官网报名成功的团队，于9月5日前将盖章的参赛

申报表 pdf、作品所有相关材料提交至 https://survey.aliyun.com/apps/zhiliao/A4e_qqNGu。压缩包命名规则：学校-姓名-作品名-联系电话。作品材料打包压缩，并将压缩包上传至“夸克网盘”，并将夸克网盘分享链接和提取码以及夸克网盘文件截图（含上传时间），整理成单独附件文档。

提交具体作品时，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）。

以上材料无需在“挑战杯”官网提交。

九、赛事保障

阿里云为参赛团队提供全方位支持：

算力支持：为参赛者提供 300 元/每人算力补贴，为所有指导教师提供算力 5 折专属权益。

AI 大模型工具：为参赛者提供基于千问大模型及 AI 应用开发平台百炼平台的 AI 工具，并且配有专家资源进行使用教学。

文献与数据：推荐接入公共开放的科学文献库及标准化数据集。

技术辅导：提供阿里云百炼平台使用培训、多智能体开发课程及科研专家在线答疑。

实践证明：为成功完成作品并提交参赛的队伍提供参与 AI Scientist 科技实践证明。

阿里云可同时为共青团中央 2026 年度挑战杯所有赛事的参赛学生和教师提供算力资源支持：

所有赛事的大学生：阿里云为参赛大学生提供 300 元算力资源支持。

所有参赛的高校教师：阿里云为参赛/指导教师提供 5 折算力（最高 20 万）资源支持。

十、设奖情况及奖励措施

1. 设奖情况

根据评分规则，综合评定参赛队伍成绩。设擂主 1 个（从特等奖中产生），特等奖 5 个，一等奖 5 个，二等奖 10 个，三等奖 10 个。奖项不重复，奖金按队伍所获最高奖项授予。

最终授奖数量视作品申报数量和质量情况，报组委会同意后动态调整。

2. 奖励措施

特等奖（擂主）：1 名（从特等奖中决出），奖金 100,000 元，并向团队全部成员提供实习/校招内推机会。

特等奖：5 名，税后奖金 20,000 元/人，并向团队全部成员提供实习/校招内推机会。

一等奖：5 名，税后奖金 10,000 元/人，并向团队全部成员提供实习/校招内推机会。

二等奖：10 名，税后奖金 5,000 元/人，并向团队全部成员提供实习/校招内推机会。

三等奖：10 名，税后奖金 2,000 元/人，并向团队全部成员提供实习/校招内推机会。

3. 奖金发放方式

以上奖金以汇款方式兑现，获奖者需提供接收奖金的银行卡信息，奖金在赛事结束并经公司审批后 6 个月内发放，所有奖金均为税后金额。

十一、比赛专班联系方式

1. 赛事服务团队

联络人：左老师，联系方式：xiaoan.zj@alibaba-inc.com

钉钉答疑群，钉钉搜索群号：162255026342

2. 联系时间

比赛期间工作日（9:30-18:00）

附：浙江阿里巴巴云计算有限公司简介

作为全球领先的全栈人工智能服务商，阿里云坚持让计算成为公共服务，助力全球客户加速价值创新。自 2009 年成立伊始，坚持自主可控的技术路线，自研了中国唯一的公共云计算操作系统。目前为全球 200 多个国家和地区、92 个可用区的客户提供稳定性全球领先的产品技术。

阿里云是国内率先布局自研大模型的企业之一，其千问大模型代表中国大模型多次刷新世界纪录，登顶全球最强开源模型，千问大模型在大语言、视觉理解、图像生成、视频创作、语音交互等核心赛道实现全面突破。已经服务超过 100 万企业，千问开源模型数量已突破 400 个，涵盖从 0.5B 到 480B 的全尺寸范围，以及文本、视觉、语音等全模态，并支持 119 种语言及方言，成为全球 AI 开源社区采用率最高的模型家族。衍生模型超 20 万，稳居全球开源大模型榜首。此外，千问开源模型的累计下载量已超过 10 亿次，活跃开发者和企业用户数量庞大，形成了全球最大的开源模型生态之一。

阿里云拥有全栈自研能力，包括自研的飞天操作系统和含光 800 芯片，这些自研技术构建了坚实的技术壁垒。