



第二届北京市大学生“人工智能+”创新大赛
THE 2ND BEIJING COLLEGE STUDENTS' AI+INNOVATION COMPETITION

智领新质 共创未来

第二届北京市大学生“人工智能+”创新大赛

赛题规则

赛题

“智训引擎”——工业制造场景下的
模型训练与超参数优化

软通计算机有限公司、华为技术有限公司

一、 项目背景

随着人工智能在智能制造中的规模化落地，设备故障诊断、剩余寿命预测、质量检测等工业模型已进入产线长期运行阶段。与实验室静态数据不同，工业数据随时间持续变化，这会使模型的输入分布及输入-输出关系发生漂移，导致模型逐渐“失准”，产生误报、漏报等隐患与质检失准等问题。因此，“一次训练、长期有效”的传统建模方式难以满足工业现场需求。与此同时，大模型智能体已具备编写代码、调度训练、读取实验记录并据此决策的能力，“由 AI 驱动模型持续优化”正走向工程实践。

本赛题旨在通过构建“工业模型自进化系统”，推动模型自我迭代与受控递归自改进（RSI, Recursive Self-Improvement）技术在真实工业场景中的落地应用。参赛者需基于组委会提供的官方数据包，在漂移数据流上自动维持并提升模型性能，自主判断并论证技术路线。参赛者可通过持续微调，或设计由工具（Harness）驱动的智能体系统，实现“模型层自迭代+策略层自改进”的两层自进化机制。

二、 赛题考察点及技术创新

本赛题重点考察选手在漂移监测与归因、技术路线判断与论证、自动化系统工程、实验管理与可观测性、受控自改进与安全等方面的综合能力。在有限算力与标注预算的条件下，参赛团队需通过合理的更新决策机制、规范的对照实验与成本收益分析，提升系统的长期稳定性与迭代效率。赛题支持参赛者自主设计基于大模型的智能体 Harness（工具与技能组合）或传统微调路线，鼓励在扩展数据验证与技术创新方面进行探索，推动持续学习与智能体工程技术的发展，为实际工业模型运维体系提供具备商业与工程价值的前瞻性方案。

三、 参赛要求

本次大赛面向北京地区普通高等学校在校本科生，不限制专业和年级。每支参赛队伍由 3—5 名成员组成，每名学生仅可参与一支队伍，每支队伍指导教师不多于 2 人。

四、 赛程安排与赛题规则

本赛题按高校预选推荐、企业分赛场选拔、北京市线下决赛三阶段实施，遵循大赛总体时间安排。

高校预选推荐阶段由各高校自行组织校内选拔，推荐队伍进入企业选拔阶段；

企业分赛场选拔阶段分为两个阶段：**第一阶段为线上提交与客观评测**，参赛队伍在专属评测平台提交系统镜像，由平台自动计算客观分数；**第二阶段为线下评审**，企业选取线上评分最高的 20 支队伍参加线下演示与答辩，综合评定推荐队伍晋级总决赛；

北京市线下总决赛安排由大赛组委会另行通知。

具体要求如下：

（一）校内选拔题：

校赛阶段由各高校自行组织，各高校可结合发布的赛题任务组织校内选拔，每所高校可推荐不超过 5 支队伍参加企业赛。

（二）企业选拔题：系统开发与统一评测

在参赛队伍进入企业赛阶段后，技术委员会将为参赛队伍提供 PyTRiO 模型训推服务使用额度。官方数据包包含训练集与验证集（含标签，公开发布）、评测脚本以及用于正式评测的隐藏测试流（不公开），其中公开集（训练集与验证集）将与专属评测平台同步开放。实验记录推荐使用 SwanLab 云端实验记录平台，该平台已上线并免费开放，参赛队伍在校选赛阶段即可自行注册使用。

1. 参赛队伍需完成以下核心任务：

任务 1：工业模型自进化系统开发。 基于官方数据包公开部分，完成系统的开发、训练与调优。系统需具备自动完成“漂移监测—诊断归因—实验优化—评估上线—回滚”闭环的能力，鼓励实现“模型层自迭代+策略层自改进”的两层自进化（受控 RSI）机制。

任务 2：实验记录与提交。 全部训练、微调、调整与评估实验须按《实验记录规范》在云端实验记录平台完整记录。企业赛第一阶段线上提交截止时间前（以赛题组通知为准），提交模型配置与 Harness 的系统镜像，其中驱动 Harness 的大模型须选用选手通过 PyTRiO 平台训练得到的模型或 PyTRiO 的 Qwen3.8-27B 模型。

赛题组将在指定设备上，以统一、离线的运行环境和资源限制，在隐藏测试数据集上对各队伍镜像进行正式评测。系统须无人工干预地自主运行，自动维持并提升工业模型在测试流上的性能。

2. 提交资料：

模型配置与 Harness 的系统镜像。

完整的云端实验记录链接或查询依据。

系统运行说明文档。

3. 评分标准：

组委会依据正式评测结果生成排行榜，排行榜前 20 支队伍晋级现场答辩。

分数构成：作品评分由客观分和主观分两部分组成。客观分由评测平台按步骤自动计算，主要考察单位增益消耗的算力（GPU 时长或训练额度）、标注配额与大模型 Token。主观评分由评审专家独立打分，包括方法设计与技术路线、实验规范与分析深度、安全可控与工业落地、报告与答辩等维度。

重新运行：如系统因自身原因运行失败，可申请重新运行一次，取两次运行中的较高成绩。

基础门槛：未满足以下基础完成门槛的队伍，成绩不进入排名：①系统在封闭评测中无人工干预地完成全部阶段并按时输出结果；②全部训练、微调、调整与评估实验均按《实验记录规范》留存完整记录；③封闭评测综合性能不低于组委会静态基线（部署后不再更新的模型）。

五、 设备要求及资源支持

（一）设备要求

自备设备：参赛队伍需自备比赛所需的电脑等设备，需能正常连接竞赛指定的云端实验记录平台，并拥有代码运行环境（可使用 VSCode 等编辑器编写代码）。

开发环境：要求 Python 3.10 及以上；深度学习框架、实验跟踪工具与智能体框架均不限。官方数据包规模较小并提供预提取特征，基础任务使用普通 CPU 即可完成。

评测环境：企业赛正式评测由组委会在指定设备上以统一的离线环境运行，不受参赛队伍本地环境影响。

（二）资源支持

比赛期间，命题企业统一提供以下资源支持：

支持内容	提供方式与范围
官方数据包	上传至指定平台固定目录下，包含训练集、验证集、评测脚本及隐藏测试流。
算力支持	企业赛阶段，用于驱动 Agent 的大模型推理与微调算力，由组委会提供 PyTRiO 模型训推服务的使用额度。
平台与规范	提供评测平台 SDK、《实验记录规范》及云端实验记录平台。
实验记录平台	提供 SwanLab 云端实验记录平台（已上线，免费开放）及《实验记录规范》，校选赛阶段即可自行注册使用。

平台使用安排：SwanLab 云端实验记录平台为已上线系统，免费开放，参赛队伍

在校选赛阶段即可自行注册使用。PyTRiO 模型训推服务使用额度仅在企业赛阶段提供，校选赛阶段不提供额度。专属评测平台（客观分数自动评测）正在开发中，预计 11 月初上线，官方数据包公开集（训练集与验证集）将随平台同步开放；相关账号与额度最迟于企业赛开赛前一周发放到位。

（四）安全及风险提示

数据与账号安全：参赛队伍须妥善保管云端实验记录平台的账号，不得泄露或转借。禁止使用违规手段攻击评测平台或获取隐藏测试集。

模型与代码安全：提交的智能体系统需具备受控机制，防止无限制的恶意迭代。不得在系统中植入恶意代码或进行未经授权的外部网络访问。

知识产权与合规：参赛作品须为原创，引用开源框架或预训练模型须遵守相应的开源许可证。

六、 附加说明

本规则未尽事宜，由赛题技术委员会与大赛组委会协商确定。

本规则如与大赛组委会的其它规定不一致时，以大赛组委会规定为准。

“智训引擎”——工业制造场景下的模型训练与超参数优化

赛题答疑群

