大学生 AI 创新创业组

[AI+加速卡模型适配赛道] 参赛指南

一、赛项基本信息

赛项支持单位：太初（杭州）集成电路有限公司
赛项名称：AI+加速卡模型适配挑战赛

二、赛项介绍

2.1 赛项背景

随着人工智能技术的快速发展，各类大模型、小模型在教育、办公、科研等场景中广泛应用。本次模型适配赛项旨在汇聚高校开发者的智慧，鼓励参赛者将指定的前沿模型与国产异构计算平台进行适配，形成可运行、可评估的推理/训练方案，加速国产 AI 生态建设。

2.2 参赛对象

高等院校及中高职院校在校师生、企业单位、科研院所。
选手以 1-5 人一队形式进行竞赛，需指定一名队长（主持人）。

2.3 参赛形式

作品赛

2.4 奖项设置

除了大赛组委会统一的奖项设置之外，会有以下额外奖励（不限队伍数量）；
针对完成赛题的团队(模型跑通且精度达标即可,不论成绩),奖励价值 1000 元的 token/团队,以 API 的形式赠与(https://tecoapi.thuwayteczz.com/ 全国产算力 API 平台)；

2.5 时间安排

报名/代码提交截止时间：即日起—2026 年 10 月 15 日
培训时间：2026 年 7 月下旬或 8 月上旬（具体时间视报名情况确定）
初赛评审时间：2026 年 10 月 16 日—10 月 31 日
决赛时间：2026 年 11 月初，线上线下待定
颁奖阶段：2026 年 11 月 12 日

三、赛题说明

3.1 赛题设计原则

- **真实场景导向：**选取当前学术界/工业界热门且尚未在目标平台上完成适配的模型，具有实际应用价值。
- **梯度难度：**提供不同参数量、不同类型的模型，适应不同能力水平的参赛队伍。
- **国产算力先行：**强化参赛者对国产 AI 加速卡（太初）生态的理解，推动自主可控技术栈的普及。

3.2 赛题考核范围

本赛道提供 5 个不同难度的模型，供参赛队伍选择进行适配。每个模型，开发者需要优化/开发 1~2 个算子（重点），并接入框架，从而完成模型的适配。
鼓励选择多个模型，若同时完成多个模型的适配，可在性能测试中获得额外加分。

赛题序号	类别	模型名称	难度系数
1	大模型	gemma-4-12B-it	1.0
2	大模型	InternVL3_5-8B	0.7
3	大模型	Hy-MT2-1.8B	0.5
4	大模型	MiniCPM5-1B	0.5
5	小模型	Deformable-DETR	0.9

注：难度系数用于得分归一化，最终解释权归组委会。

3.3 初赛阶段

3.3.1 作品创作要求

参赛队伍需从模型列表中任选至少 1 个模型完成适配。鼓励选择多个模型，若同时完成多个模型的适配，可在性能测试中获得额外加分。
核心要求：
1. 模型必须能在提供的环境上正常运行。
2. 跑通组委会提供的标准测例，精度达标（测例及精度要求联系组委会获取）。
3. 提供完整的适配代码及 readme。

3.3.2 初赛整体流程

环境准备

进入 Docker 容器后加载环境变量、拉取仓库（环境由组委会统一提供，每个队伍独享一张太初 AI 加速卡）。

模型迁移与适配

解析模型网络结构，区分大小模型制定适配方案：

- 大模型（TecoLLM 框架）：算子开发 / 接入 → 模型组网 → 模型注册 → 推理验证 → 精度验证 → 性能评测
- 小模型：缺失算子分析 → 算子开发 → 模型接入 → 迁移调优

提交代码

提交代码到指定算子和模型仓库，由 CI 自动执行评测，全部通过方可合入。

3.3.3 作品提交要求

- **代码托管方式：**
  - 本赛事采用组织+私有仓库模式进行，在报名参赛后需提供队名及队长 github 账号名称，将为每个参赛队伍创建带队伍名称的私有仓库，并添加队长权限(需要队长通过邮箱接受邀请)，后续由队长负责添加队员私有仓库权限，各队伍在自己的私有仓库下进行开发。
    - 赛题 1~4 提交至模型推理仓库。
    - 赛题 5 提交至模型训练仓库。
    - 开发或优化的算子提交自定义算子仓库。
  - 参赛队伍在各自私有仓库进行开发。
  - 赛事工作人员每天定时拉取代码进行测试打分,比赛排名定时公布在 github 社区。
  - 待比赛结束后由赛事方收集所有改动，选取最优方案提交至上游官方仓库。参加比赛即视为接受该规则。
- **提交规范：**
  - 参赛作品必须为原创，不得侵犯第三方知识产权。若使用他人开源代码，必须在 README 中注明出处、许可协议及修改部分。不得使用 GPL、LGPL、Mozilla 等强约束协议代码。
  - 代码注释应当充分，关键函数、适配接口、复杂算法需有清晰注释。
  - 作品应能正常运行并达到预期结果。

3.3.4 评审评分标准

将基于各模型精度与性能结果计算的总积分，从高到低排序，确定晋级总决赛名单，选手积分步骤如下：

1. 计算单个模型得分：
   a. 大模型：

i. 精度得分：通过得 50 分，达标后可进行性能测试。
ii. 性能得分：性能越好，得分越高。根据各测例排名，加权后得到 0~50 分。
iii. 单个模型得分：（精度得分+性能得分）× 难度系数。

b. 小模型：
i. 精度基础得分：通过得 50 分，达标后可进行性能测试。
ii. 优化得分：进一步优化精度和提升性能，依据横向排名加权计算，得分区间 0～50 分。
iii. 单个模型得分：（精度基础得分+优化得分）× 难度系数。

2. 计算多个模型累加积分：
a. 取单个模型得分的最高值作为基础分。
b. 其他模型得分分别乘以衰减系数：第二高模型 × 0.8，第三高模型 × 0.5，第四高模型 × 0.25、第五高模型 × 0.125
c. 累加后得到总积分。示例：完成 A、B、C 三个模型，单个模型得分分别为 90、50、40，则总积分 = 90 + 50 × 0.8 + 40 × 0.5 =150

3.4 全国总决赛阶段

3.4.1 答辩要求

- 展示环节（15 分钟）：PPT 讲解优化思路、技术难点、性能提升效果，可配合播放视频、图片佐证。
- 问答环节（5 分钟）：评委针对技术细节、创新性、国产平台适配经验进行提问。

3.4.2 评审评分标准

最终总分 = **模型适配得分 × 80% + 路演答辩得分 × 20%**

- **模型适配得分**：即初赛总积分。
- **路演答辩得分**（满分 100 分）：由 3-5 名专家现场打分，取平均分。评分维度：

  - 优化思路清晰度与合理性（30 分）
  - 技术难度与创新性（30 分）
  - 现场问答（20 分）
  - 文档及代码质量（20 分）

3.5 竞赛环境说明

比赛将为所有参赛队伍提供配备太初 AI 加速卡的云服务器。
参赛队伍报名后请联系工作人员获取（QQ 号：309491611；手机号：15311426217），将分配 VPN 及服务器 IP 地址、账号密码等信息，参赛者登录即可。具体的 IP、账号、密码因为涉及到参赛者的独占资源，不在参赛指南里公开。

3.6 作品原创性声明及要求

- 参赛队伍同意将其为比赛所开发的适配代码（包括修改、优化部分）以 Apache 2.0 或 MIT 协议开源，并授权组委会在比赛结束后将优秀代码合并到官方仓库。
- 参赛队伍保留其代码的署名权，合入时会在提交记录中注明作者信息和参赛队伍名称。
- 若代码中包含第三方代码（如模型原仓库的修改），需明确标注原始许可证，并确保兼容性。

四、报名方式

本赛道提交的作品需要基于组委会提供的云服务器进行开发，因此，区别于其它“提交作品即报名成功”赛道，**本赛道完成“作品信息”的预填写并保存，即视为已报名。**报名系统后续的步骤，可在完成赛题开发任务后，再进行开展。

具体报名步骤如下：

1. 登录大赛官网，选择本赛道，单击“赛道报名”。将跳转进入报名系统登录页。



2. 点击"立即注册"，按要求完成注册，注册成功后将自动跳转登录页，根据注册信息完成登录。



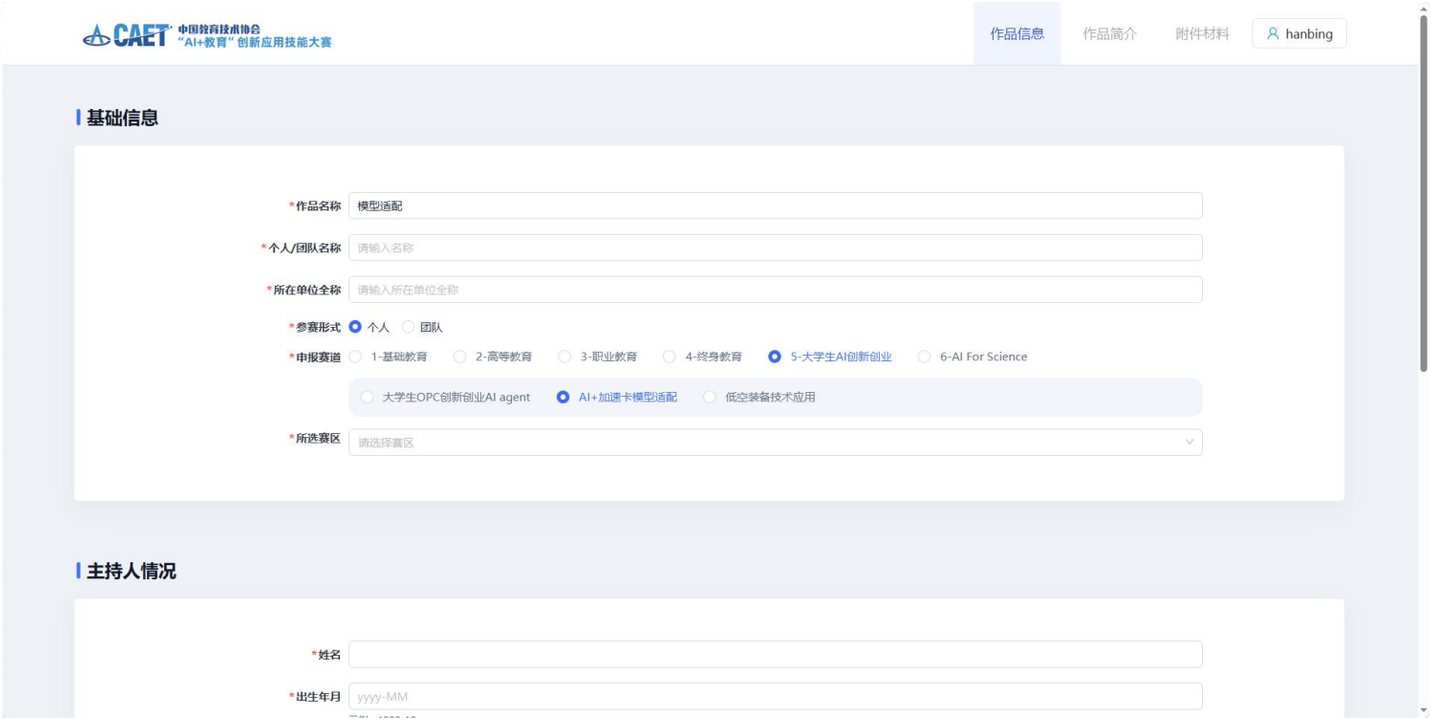


3. 在"报名管理"页签下，单击“添加报名信息与作品”按钮。



4. 在“作品信息”页面填写详细报名与作品信息。其中：

- 作品名称：填写模型适配即可。
- 参赛形式：若队伍中存在多名选手，或需要填写指导老师，请选择“团队”，并在下方“作品其他完成人情况”处添加队伍其他选手/指导老师信息。
- 申报赛道：选择“5-大学生 AI 创新创业”-“AI+加速卡模型适配”
- 所选赛区：按照实际单位省份填写即可，本赛道为全国统一评比，不区分区域。



5. “作品信息”下所有带*字段内容全部填写完成后，请点击页面下方"保存"按钮。**该步骤完成后，可联系赛道协办方（QQ：309491611）进入赛事交流群并申请开发环境。**

【提醒：点击“保存”按钮，可以在报名截止前反复修改或删除！】



6. 至此，您便可以进行赛题任务的开发任务了。在完成赛题开发任务后，可按照大赛统一的“报名申报流程指南”完成后续“填写作品简介”-“导出报名表”-“上传签章版报名表”-“上传附件材料”-“最终提交”流程。