

附件 2

人工智能竞技类竞赛规则

一、竞赛背景

近年来，以 ChatGPT 为代表的大模型技术应用飞速发展，标志着“强人工智能”时代的到来，一场“新工业革命”正促使新一代科技与产业变革加速演进，全球科技创新的广度深度也在持续攀升。大模型技术在加速千行百业数字化与智能化转型的同时，也给数智化创新人才培养带来了巨大挑战。

该赛项旨在促进人工智能大模型技术赋能跨专业融合教育，促进数智化创新人才培养，鼓励高校人工智能相关专业学生充分学习大模型技术，使用大模型技术解决专业问题，提升大模型技术思维能力，提高学生对新时代人工智能技术的认知和兴趣。

二、竞赛器材

比赛任务由一台包含基座大模型的桌面级教学实验箱支撑完成。教学实验箱需是一体化结构，需具备人工智能大模型训练、推理、展示能力、内置麦克风、音箱、视觉摄像头等软硬件设备，有主副两个显示屏，可以显示训练过程日志，硬件内存容量不超过 16GB，显存容量不超过 24GB。

三、竞赛规则

基于有限硬件资源及标准软件环境，使用人工智能大模型技

术，训练出具有一定学科属性的人工智能对话大模型，该模型经训练后能准确回答相关知识。最终考核参赛团队对大模型相关技术的掌握与应用能力。

（一）竞赛任务

比赛任务分为数据整理、大模型调试训练、大模型性能评测、页面对话展示、自然语音对话五部分，总分为 100 分。

1.数据整理（10 分）

比赛现场各团队通过抽签的方式抽取所要构建的学科大模型任务，之后参赛团队通过擅长的方式、工具进行数据的搜集、整理、生成。数据质量对最终大模型训练效果影响较大，需要对数据进行一定的处理，以最终达到高质量的模型评测效果。

参赛队伍通过 Excel 文件提交数据。数据量大于等于 50 条，可得 3 分。数据量大于等于 100 条，可得 6 分，不累积得分。没有重复数据、问答数据宽泛质量高，可再最高得 4 分。

2.大模型调试训练（6 分）

基于 ChatGLM2-6B 基座大模型进行任务竞技，提供训练基础代码，各参赛团队需要进行模型参数调整、代码调整、启动训练微调大模型、产出学科大模型文件。参赛队伍需要理解模型参数的意义及对应调整的效果影响，掌握大模型技术知识，对代码进行调整编写。

参赛队伍能够基于硬件资源，选择合适的微调算法，启动大模型训练可得 3 分，能够训练完成大模型，生成大模型文件，提

交模型文件可再得 3 分。

3.大模型性能评测（70 分）

训练完成后，参赛队伍需提交训练成功的大模型文件及学科数据文件（格式见附录），通过多维度学科数据集，完成对大模型性能效果的评测，选手每隔 20 分钟，可以重新提交模型效果评测。

竞赛所用的大模型评测平台基于“融合智创”多维度学科数据集，该数据集涵盖人文、社科、理工等近 80 个专业数据主题，具有完整可复现的学科大模型评测能力，支持大语言模型一站式评测，将为参赛过程中参赛队伍所产出的大语言模型提供统一评测支持。

大模型性能评测任务占总分的 70%，评测平台是百分制，所得分将乘以 70%后，再计入总分。

4.页面对话展示（10 分）

通过页面形式，在本地加载模型，用中文实现与大模型文字对话。比赛现场跟随抽签所得任务，会发放 3 个中文问答问题，需让大模型完成连续回答。3 个问题需要连续问答，输入第 1 个问题，大模型回答正确得 2 分；输入第 2 个问题，大模型回答正确得 4 分；输入第 3 个问题，大模型回答正确得 4 分。若问题回答错误，该任务评测便终止。

5.自然语音对话（4 分）

通过自然语音对话的方式，与大模型进行语音对话，并对回

复内容进行语音播报。能够进行中文语音输入提问，得 2 分，对回答的问题，能够进行中文语音播报，得 2 分。

（二）比赛时间

本次比赛采用多轮排位赛，参赛团队可在 120 分钟内进行多次调试训练，训练的大模型文件可在规定时间内多次提交，每次提交后对应需要进行五项任务的总体评分，比赛将以最后一次提交的结果作为最终成绩。

（三）排名规则

本次比赛的成绩为单轮任务得分的总和，得分高的团队排名在前，若出现同分情况，则较早提交大模型评测的团队排名在前。

附录：数据提交格式

序号	问题	答案