第七届“华为杯”区域职业院校新一代信息技术应用技能大赛人工智能技术应用赛项规程

一、赛项名称

赛项名称：人工智能技术应用

赛项组别：中职组、高职组（含职教本科、留学生）

二、竞赛目的

进一步强化职业院校本专业学生职业技能训练和职业能力的综合运用，促进校企合作、产教融合，完善“岗课赛证”教学模式，培育工匠精神，推动职业院校“双师型”师资队伍建设，大力培养适应经济与社会发展的高素质劳动者和技术技能型人才，进一步提升职业院校人工智能技术应用等相关专业学生能力素质与企业用人标准的吻合度，为在新形势下全面提高信息技术类专业教学质量、为扩大就业创业、运用人工智能新技术新模式赋能传统产业转型升级、培育经济发展新动能做出新贡献。

三、竞赛内容

本赛项围绕教育、工业、医疗等实际场景，考核选手完成从数据读取、数据清洗、特征工程、模型构建、结果落地到预警报告生成的全流程人工智能应用开发能力，全面检验数据工程、传统机器学习建模、轻量量化大模型离线应用等综合技能：

1. 数据读取与探索性分析：读取数据集，完成基础信息与缺失值统计、成绩统计描述、分数段分布统计与可视化。

2. 数据清洗与预警标签构建：完成缺失值填充、异常值处理，构建预警标签并统计标签分布。

3. 特征工程与数据集划分：完成特征筛选、独热编码、衍生特征构造、分层抽样划分与特征标准化。

4. 预警分类模型构建与评估：使用逻辑回归等构建双模型，计算评估指标，进行模型适配分析与模型保存。

5. 模型优化与风险画像分析：完成样本不平衡处理、网格搜索调优、特征重要性分析与高风险信息输出。

6. 大模型辅助个性化预警报告生成：加载轻量量化大模型，提取风险等特征，设计结构化提示词。

四、竞赛规则

（一）选手报名

各职业院校按照大赛组委会规定的报名要求，进行报名参赛。本赛项采用团队赛方式，每支参赛队由 2 名全日制在籍学生组成，参赛选手不得跨校组队；中职组参赛对象为中等职业学校在校生及五年制高职一至三年级学生，高职组参赛对象为三年制高职一至三年级学生及五年制高职四至五年级学生、职业本科在校学生和留学生。每个院校中职、高职组每赛项分别限报 2 支队伍；职教本科和留学生单设赛道，每赛项分别限报 1 支队伍。每队设指导教师 2 名，指导教师须为本校专职教师，竞赛期间不允许指导教师进入赛场进行指导。

（二）熟悉场地

熟悉场地时间以组委会统一安排为准，届时将安排参赛队熟悉比赛场地，召开领队会议，宣布竞赛纪律和有关事宜。

（三）赛场规范

赛前准备：选手抽签加密入场，参赛队就位并领取比赛任务，完成比赛设备、线缆和工具检查等准备工作。

正式比赛：参赛选手需按题目要求完成各道小题的全部任务。比赛全程离线进行，禁止任何联网操作；操作顺序和分工，由参赛队自行商定。

（四）成绩评定与结果公布

成绩评定和结果公布由裁判组、监督组和仲裁组共同负责。

裁判组实行“裁判长负责制”，设裁判长 1 名，全面负责赛项的裁判分工、裁判评分审核、处理比赛中出现的争议问题等工作。

裁判员根据比赛需要分为检录裁判、加密裁判、现场裁判和评分裁判。

（1）检录裁判：负责对参赛队伍（选手）进行点名登记、身份核对等工作；

（2）加密裁判：负责组织参赛队伍（选手）抽签，对参赛队信息、抽签代码等进行加密；

（3）现场裁判：按规定做好赛场记录，维护赛场纪律，评定参赛队的过程得分；评分裁判：负责按评分细则评定成绩。

监督组对裁判组的工作进行全程监督，并对竞赛成绩抽检复核。

仲裁组负责接受由参赛队领队提出的对裁判结果的申诉，组织复议并及时反馈复议结果。

最终成绩经裁判组、监督组和仲裁组审核无误后正式公布。

五、竞赛环境

（一）竞赛场地安排

竞赛场地按组委会统一安排设置，具体详见组委会发布的比赛地点及位置图。

（二）技能竞赛环境要求

竞赛工位内设有操作平台，每个工位为选手提供 PC 机，配备 220V 电源（带漏电保护装置）；竞赛工位标明工位号和参赛设备号，并配备竞赛平台和技术工作要求的软、硬件。环境标准要求保证赛场采光、照明通风良好、温度湿度适宜；留有出入和消防通道。比赛全程断开互联网，禁用无线网卡等外部通信模块，禁止接入任何外接存储设备。

（三）医疗服务及要求

配备防疫和急救人员与设施。

（四）裁判员工作场所及要求

现场裁判工作场地与要求详见（一）竞赛场地安排；评分裁判需要 2 间各 30 平米的房间，每间需要提供两张不小于 3*2 的方桌，每张方桌配置 6 把靠椅，配备 220V 电源（带漏电保护装置），提供打印机。

（五）赛场保密场所及要求

保密室需要提供可容纳 160*50 张 A4 纸的保险柜，两台不低于 20 张/分钟的激光打印机及用于装订试卷的必要设备设施。

六、技术规范

（一）教学标准

序号	标准单位	中文标准名称
1	教育部	高等职业教育专科专业简介：51 电子与信息大类-5102 计算机类-510209 人工智能技术应用
2	教育部	高等职业教育专科专业简介：51 电子与信息大类-5102 计算机类-510205 大数据技术

（二）行业标准

序号	标准号	中文标准名称
1	GB/T 41867-2022	信息技术 人工智能 术语
2	GB/T 42755-2023	人工智能 面向机器学习的数据标注规程
3	GB/T 41864-2022	信息技术 计算机视觉 术语
4	GB/T 35295-2017	信息技术 大数据 术语
5	GB/T 11457-2006	信息技术 软件工程术语
6	2-02-10-09	人工智能工程技术人员国家职业技术技能标准
7	2-02-10-12	数据安全工程技术人员国家职业技术技能标准
8	GB/T 45288.1-2025	人工智能 大模型

七、技术平台

（一）竞赛设备

序号	设备类型	设备规格及型号
1	通用设备	每组选手配备 2 台普通计算机（CPU 主频不低于 2.5GHz；内存不低于 8GB；硬盘不低于 256GB）。

（二）竞赛软件

类别	名称	版本
操作系统	Windows 10/11	Windows 10 专业版（64 位）及以上
开发工具	PyCharm	社区版 Version 2022.3 及以上
软件环境	Python	3.9.13 及以上
	pandas	1.5.3 及以上
	numpy	1.23.5 及以上
	scikit-learn	1.2.2 及以上
	lightgbm	3.3.5 及以上
	matplotlib	3.7.1 及以上
	joblib	1.2.0 及以上
	llama-cpp-python	0.2.77 及以上
	llama.cpp	build 12100
浏览器	Chrome	版本 131.0.6778.205（64 位）及以上
预置资源	数据集	Csv 格式
文档编辑	WPS	主流 WPS 版本

类别	名称	版本
大模型	轻量量化大模型	qwen2-0_5b-instruct-q4_k_m.gguf

以上软件环境均支持全离线安装部署，无联网授权依赖。

（三）裁判工作需要的办公用品及设备、场所等要求及清单

赛点提供两间评分裁判工作室；竞赛区域需设置现场裁判工作区域，需配备计算机和打印机各一台。裁判用的办公用品主要包括：黑色和红色水笔、订书机、计算器、A4 打印纸等。

八、成绩评定

（一）评分方法

本赛项各道小题由评分裁判依据组委会发布的评分表逐项评分，重点考察选手的题目完成正确性与代码质量。代码无法运行或关键结果缺失的，按评分细则相应扣分。

（二）成绩复核与解密

监督、仲裁组将对赛项总成绩排名前 30%的所有参赛队伍（选手）的成绩进行复核；对其余成绩进行抽检复核，抽检覆盖率不得低于 15%。如发现成绩错误以书面方式及时告知裁判长，由裁判长更正成绩并签字确认。复核、抽检错误率超过 5%的，裁判组将对所有成绩进行复核。成绩复核、确认无误后进行成绩排名。

（三）成绩公布

记分员将各参赛队竞赛成绩进行汇总制表，经裁判长、监督仲裁组签字后向全体参赛队进行公布。公布无异议后，经裁判长、监督仲裁组长审核签字后，在闭赛式上宣布。

（四）评分标准

1. 本赛项按组委会发布的“评分表”对每道小题评分，评分表按题明确各评分项、分值、关键截图与评分依据。

2. 选手须按题目要求提交源代码、结果文件及关键运行截图作为评分证据，对应关键截图缺失或无法证明该评分项的，该项记 0 分。

九、竞赛须知

（一）参赛队须知

1、参赛队名称统一使用规定的代表队名称；

2、参赛队员在报名获得审核确认后，原则上不再更换，如筹备过程中，选手因故不能参赛，所在学校需出具书面说明、按相关规定补充人员、经组委会同意后方可参赛；

3、参赛队按照大赛赛程安排，凭大赛组委会颁发的参赛证、有效身份证件参加比赛及相关活动；

4、各参赛队统一参加比赛前熟悉场地环境的活动；

5、各参赛队准时参加赛前领队会，领队会上举行抽签仪式、抽取场次号；

6、各参赛队要注意饮食卫生，防止食物中毒；

7、各参赛队要发扬良好道德风尚，听从指挥、服从裁判、不弄虚作假。

（二）指导老师须知

1、各指导老师要发扬良好道德风尚，听从指挥，服从裁判，不弄虚作假。指导老师经报名、审核后确定，一经确定不得更换。

2、对申诉的仲裁结果，领队和指导老师应带头服从和执行，还应说服选手服从和执行。

3、指导老师应认真研究和掌握本赛项比赛的技术规则和赛场要求，指导选手做好赛前的一切准备工作。

4、领队和指导老师应在赛后做好技术总结和工作总结。

（三）参赛选手须知

1、参赛选手应遵守比赛规则，尊重裁判和赛场工作人员，自觉遵守赛场秩序，服从裁判的管理。

2、参赛选手应佩戴参赛证，带齐身份证、注册的学生证。在赛场的着装，应符合职业要求。在赛场的表现，应体现自己良好的职业习惯和职业素养。

3、进入赛场前须将手机等通讯工具交赛场相关人员保管，不能带入赛场。未经检验的工具、电子储存器件和其他不允许带入赛场物品，一律不能进入赛场。

4、比赛过程中不准互相交谈，不得大声喧哗；不得有影响其他选手比赛的行为，不准有旁窥、夹带等作弊行为。

5、参赛选手在比赛的过程中，应遵守安全操作规程，文明的操作。通电调试设备时，应经现场裁判许可，在技术人员监护下进行。

6、比赛过程中需要去洗手间，应报告现场裁判，由裁判或赛场工作人员陪同离开赛场。

7、完成比赛任务后，需要在比赛结束前离开赛场，需向现场裁判示意，在赛场记录上填写离场时间并签工位号确认后，方可离开赛场，离开赛场后不可再次进入。未完成比赛任务，因病或其他原因需要终止比赛离开赛场，需经裁判长同意，在赛场记录表的相应栏目填写离场原因、离场时间并签工位号确认后，方可离开；离开后，不能再次进入赛场。

8、裁判长发出停止比赛的指令，选手（补时选手除外，等延时结束）应立即停止操作进入通道，在现场裁判的指挥下离开赛场。

9、遇突发事件，立即报告裁判和赛场工作人员，按赛场裁判和工作人员的指令行动。

（四）工作人员须知

1、工作人员必须服从赛项组委会统一指挥，佩戴工作人员标识，认真履行职责，做好服务赛场、服务选手的工作。

2、工作人员按照分工准时上岗，不得擅自离岗，应认真履行各自的工作职责，保证竞赛工作的顺利进行。

3、工作人员应在规定的区域内工作，未经许可，不得擅自进入竞赛场地。如需进场，需经过裁判长同意，核准证件，有裁判跟随入场。

4、如遇突发事件，须及时向裁判长报告，同时做好疏导工作，避免重大事故发生，确保竞赛圆满成功。

5、竞赛期间，工作人员不得干涉及个人工作职责之外的事宜，不得利用工作之便，弄虚作假、徇私舞弊。如有上述现象或因工作不负责任的情况，造成竞赛程序无法继续进行，由赛项组委会视情节轻重，给予通报批评或停止工作，并通知其所在单位做出相应处理。

（五）裁判员须知

1、裁判员执裁前应参加培训，了解比赛任务及其要求、考核的知识与技能，认真学习评分标准，理解评分表各评价内容和标准。不参加培训的裁判员，取消执裁资格。

2、裁判员执裁期间，统一佩戴裁判员标识，举止文明礼貌，接受参赛人员的监督。

3、遵守执裁纪律，履行裁判职责，执行竞赛规则，信守裁判承诺书的各项承诺。服从赛项专家组和裁判长的领导。按照分工开展工作，始终坚守工作岗位，不得擅自离岗。

4、裁判员有维护赛场秩序、执行赛场纪律的责任，也有保证参赛选手安全的责任。时刻注意参赛选手操作安全的问题，制止违反安全操作的行为，防止安全事故的出现。

5、裁判员不得有任何影响参赛选手比赛的行为，不得向参赛选手暗示或解答与竞赛有关的问题，不得指导、帮助选手完成比赛任务。

6、公平公正的对待每一位参赛选手，不能有亲近与疏远、热情与冷淡差别。

7、赛场中选手出现的所有问题如：违反赛场纪律、违反安全操作规程、提前离开赛场等，都应在赛场记录表上记录，并要求学生签工位号确认。

8、严格执行竞赛项目评分标准，做到公平、公正、真实、准确，杜绝随意打分；对评分表的理解和宽严尺度把握有分歧时，请示裁判长解决。严禁利用工作之便，弄虚作假、徇私舞弊。

9、竞赛期间，因裁判人员工作不负责任，造成竞赛程序无法继续进行或评判结果不真实的情况，由赛项组委会视情节轻重，给予通报批评或停止裁判资格，并通知其所在单位做出相应处理。

十、申诉与仲裁

1、各参赛队对不符合赛项规程规定的设备、工具、材料、计算机软硬件、竞赛执裁、赛场管理及工作人员的不规范行为等，可向赛项仲裁组提出申诉。

2、申诉主体为参赛队领队，申诉启动时，参赛队以该队领队签字同意的书面报告的形式递交赛项仲裁组。报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

3、提出申诉应在赛项比赛结束后 2 小时内提出，超过 2 小时不予受理。

4、赛项仲裁组在接到申诉报告后的 2 小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由领队向大赛仲裁工作组提出申诉。大赛仲裁工作组的仲裁结果为最终结果。

5、申诉方不得以任何理由拒绝接收仲裁结果；不得以任何理由采取过激行为扰乱赛场秩序。仲裁结果由申诉人签收，不能代收；如在约定时间和地点申诉人离开，视为自行放弃申诉。

6、申诉方可随时提出放弃申诉。

十一、其他

本技术文件的最终解释权归大赛组织委员会。

样题：学业预警 AI 系统构建（100 分）

（一）赛题背景

本样题围绕职业院校智慧校园核心场景“学业预警”展开，要求选手基于公开学生行为数据集，完成从数据读取、数据清洗、特征工程、模型构建、结果落地到个性化学业预警报告生成的全流程人工智能应用开发，最终输出一套可落地的学业风险识别系统。

（二）示例数据集说明

采用学生数据集，原始格式为分号分隔。核心字段包括：所在学校、性别、年龄、每周学习时长、过往挂科次数、缺勤次数、期中成绩、期末成绩等。

（三）竞赛任务

根据上述背景设计 5~8 个任务，以下为示例

数据读取与探索性分析

1. 读取数据集，在控制台输出总行数、总列数及所有字段的名称与数据类型。

2. 统计各字段缺失值数量并保存为 CSV 文件。

3. 对期末成绩进行统计描述，输出最大值、最小值、均值、中位数。

数据清洗与预警标签构建

1. 数值型字段缺失值用均值填充，类别型字段缺失值用众数填充，删除空样本。

2. 对缺勤次数大于指定值的记录判定为异常，替换为该字段中位数。

3. 新增预警标签：期末成绩<10 标记为 1（高风险），期末成绩≥10 标记为 0（低风险）。

特征工程与数据集划分

1. 剔除平时成绩字段，拆分特征集与标签集。

2. 对类别型特征执行独热编码，输出编码后特征总数。

3. 构造至少 2 个有效衍生特征，并说明计算逻辑与业务意义。

学业预警分类模型构建与评估

1. 使用逻辑回归等算法训练二分类模型。

2. 在测试集计算并输出准确率、精确率、召回率、等指标。

3. 说明结合学业预警场景（优先避免漏判）哪类模型更适配及理由。