**安徽工程大学2022年大数据与人工智能应用竞赛执行规则**

**一、竞赛内容**

**1 .大数据现场赛**

**竞赛内容**分为大数据平台部署、大数据平台参数优化、大数据预处理、大数据分析、大数据可视化以及综合应用，具体包括：

大数据平台部署：参赛选手根据题目要求完成大数据平台的部署，包括但不限于Hadoop、HDFS、HBase、Hive、Spark、Zookeeper、SQOOP、flink，flume，storm等，并能够根据要求对部署的平台参数进行调整以及测试平台可用性，对多个平台进行联调，联测，实现数据在不同的平台中进行流转。

大数据预处理：参赛选手根据题目要求对给定的数据（数据可能不完整、不一致、重复、含噪音等）进行预处理，使得数据准确、完整、一致、可用，并能够对数据格式进行转换。

大数据分析：参赛选手根据题目要求在已搭建好的大数据平台上进行数据的分析。

大数据可视化：参赛选手根据题目要求对给定数据进行可视化展示，并能对可视化结果进行解释说明。

综合应用：对给定的数据集，参赛选手综合应用大数据预处理、大数据分析与挖掘等技术，求解一个给定的问题。

操作系统：Centos

编程语言：Python、Java

Python语言开发环境采用PyCharm，Java语言开发环境采用Eclipse。

**2.人工智能现场赛**

**竞赛内容**分为人工智能基础环境搭建部署、样本数据预处理、传统机器学习算法设计及应用、深度学习算法设计及应用、人工智能技术综合应用，具体包括：

人工智能基础环境搭建部署：参赛选手根据题目要求完成指定版本的人工智能相关开发环境的部署，包括但不限于Anaconda、深度学习库百度飞桨、旷视天元、PyTorch、并能够根据要求对部署的框架导入需要的相关库，包含但不限于sklearn，opencv。

样本数据预处理：对给定的样本数据，检测和去除数据集中的噪声数据和无关数据，处理遗漏数据，去除空白数据域，标准化、归一化、二元化数据以及特征提取等处理。

传统机器学习算法设计及应用：对于主流机器学习算法应用。包含但不限于逻辑回归、支持向量机、决策树、Kmeans等，根据题目或场景，通过选用机器学习算法训练模型，实现分类、聚类解决问题。

深度学习算法设计及应用：对于神经网络，自然语言处理等应用。包含但不限于卷积神经网络，递归神经网络等。根据题目或场景，通过选用深度学习相关算法，实现如图像、语言，文本分析相关结构化/非结构化数据的识别检测等。

人工智能技术综合应用：给定现实场景和数据集，综合运用人工智能相关技术解决给定问题。

操作系统：Centos

编程语言：Python、Java

Python语言开发环境采用PyCharm，Java语言开发环境采用Eclipse。

1. **大数据与人工智能作品赛**

参赛作品须围绕大数据与人工智能核心技术，探索有具体落地场景的技术应用创意方案，如大数据与人工智能技术在工业、农业、医疗、文化、教育、金融、交通、公共安全、日常生活、公益等行业领域的应用探索。

竞赛采用开放命题，参赛作品须使用大数据与人工智能相关技术。参赛者应充分发挥创新能力，自由探索应用场景并自行获取相关数据，最终提交具有原创性并能够进行可视化应用展示的参赛作品。

**二、竞赛规则**

**1.大数据与人工智能现场赛**

（1）报名资格和要求

比赛以队为参赛基础，每支参赛队由1-3名参赛学生和1-2名指导教师组成。

（2）赛前准备

赛前开展比赛规则进行进一步解读。各参赛队伍师生依据比赛规则和要求进行比赛准备。

（3）比赛期间

参赛队必须在指定的位置就坐和操作指定的计算机。

参赛队员可以携带诸如书、手册、程序清单等参考资料。

参赛队员不能携带任何可用计算机处理的软件或数据(不允许任何私人携带的U盘、磁盘或计算器)。

参赛队员不能携带任何类型的通讯工具，包括无线电接收器、移动电话。

在竞赛中，参赛队员根据题目要求完成相关工作，参赛队员不得和竞赛组委会指定工作人员以外的人员交谈，现场工作人员只能回答和处理与平台系统相关的问题。

参赛队员出现妨碍比赛正常进行的行为，诸如擅自移动赛场中的设备，未经授权修改比赛软硬件，干扰他人比赛等，取消其参赛资格。

竞赛的预定时间为4小时，但当竞赛进行一定时间后，竞赛专家组可以因为出现不可预见的事件而调整比赛时间长度，一旦比赛时间长度发生改变，将会以及时并且统一的方式通告所有参赛队员。

**2 .大数据与人工智能作品赛**

（1）报名资格和要求

比赛以队为参赛基础，每支参赛队由1-3名参赛学生和1-2名指导教师组成。

（2）赛前准备

赛前开展比赛规则进行进一步解读。各参赛队伍师生依据比赛规则和要求进行比赛准备。

（3）比赛期间

作品赛采取现场答辩的形式，根据答辩需要，允许答辩队伍通过计算机、手机、实物等方式辅助展示作品。

以下情况将视为违规，竞赛组织委员会有权取消参赛队伍的参赛资格：

（a）在参赛过程中出现违反相关法律、法规的行为；

（b）作品参加过其他校级及以上比赛并获奖；

（c）作品涉嫌抄袭，侵犯他人知识产权等；

（d）作品涉及不健康、淫秽、色情或毁谤第三方等内容；

（e）参赛期间发现或被举报认定存在其他违法、违规行为。

**三、竞赛成绩评定**

**1. 大数据现场赛**

评分标准如下：

|  |  |  |
| --- | --- | --- |
| **竞赛内容** | **竞赛考核要求** | **成绩占比** |
| 大数据  平台部署 | 根据要求完成大数据平台的部署，并能够对部署的平台参数进行调整以及测试平台的可用性，对多个平台进行联调，联测，实现数据在不同的平台中进行流转。 | 15% |
| 大数据预处理 | 对给定的数据（数据可能不完整、不一致、重复、含噪音等）进行预处理，使得数据准确、完整、一致、可用，并能够对数据格式进行转换。 | 30% |
| 大数据分析 | 在已搭建好的大数据平台上进行数据分析 | 20% |
| 大数据可视化 | 对给定数据进行可视化展示，并能对可视化结果进行解释说明 | 15% |
| 综合题 | 对给定的数据集，参赛选手综合应用大数据预处理、大数据分析与挖掘等技术，求解一个给定的问题 | 20% |

1. **人工智能现场赛**

评分标准如下：

|  |  |  |
| --- | --- | --- |
| **竞赛内容** | **竞赛考核要求** | **成绩占比** |
| 人工智能基础环境搭建部署 | 参赛选手根据题目要求完成指定版本的人工智能相关开发环境的部署，包括但不限于Anaconda、深度学习库百度飞桨、旷视天元、PyTorch，并能够根据要求对部署的框架导入需要的相关库，包含但不限于sklearn，opencv。 | 15% |
| 样本数据  预处理 | 对给定的样本数据，检测和去除数据集中的噪声数据和无关数据，处理遗漏数据，去除空白数据域，标准化、归一化、二元化数据以及特征提取等处理。 | 30% |
| 传统机器学习算法设计  及应用 | 对于主流机器学习算法应用。包含但不限于逻辑回归、支持向量机、决策树、Kmeans等，根据题目或场景，通过选用机器学习算法训练模型，实现分类、聚类解决问题。 | 20% |
| 深度学习算法设计及应用 | 对于神经网络，自然语言处理等应用。包含但不限于卷积神经网络，递归神经网络等。根据题目或场景，通过选用深度学习相关算法，实现如图像、语言，文本分析相关构化/非结构化数据的识别检测等。 | 15% |
| 人工智能技术综合应用 | 给定现实场景和数据集，综合运用人工智能相关技术解决给定问题 | 20% |

1. **大数据与人工智能作品赛**

|  |  |  |
| --- | --- | --- |
| **评价内容** | **说明** | **成绩占比** |
| 选题定位 | 1、创意与独创性  2、落地转化可行性 | 20% |
| 社会价值 | 1、用户需求贴合度  2、效率提升的明确表现  3、市场价值及推广性 | 30% |
| 技术能力 | 1、技术综合能力  2、任务处理效果 | 30% |
| 现场答辩 | 根据现场表现情况打分 | 20% |