

第 4 部分

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.1.1

试题名称：客服会话业务数据质量检测

考核时间：30 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

智能客服机器人能通过识别理解用户意图，将客户的需求进行合理分类、传递，从知识库中寻求答案，或者直接回答客户问题，或者辅助人工客服智能应答。本项目中，你作为一名人工智能训练师，将对预处理后的客服会话数据进行审核，并制定相应的采集和处理规范。

- (1) 请对文件(dataForReview.csv)中的预处理后会话数据进行审核，指出其中存在的问题。（4分）
- (2) 请基于标注后用于模型训练测试的需求，为会话数据制定采集规范。（8分）
- (3) 请基于标注后用于模型训练测试的需求，为会话数据制定处理规范。（8分）

3、技能要求

- (1) 能够对预处理后业务数据进行审核
- (2) 能够结合人工智能技术要求，梳理业务数据采集规范
- (3) 能够结合人工智能技术要求，梳理业务数据处理规范

4、质量指标

无

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.1.2

试题名称：客服会话数据处理方法优化

考核时间：30 min

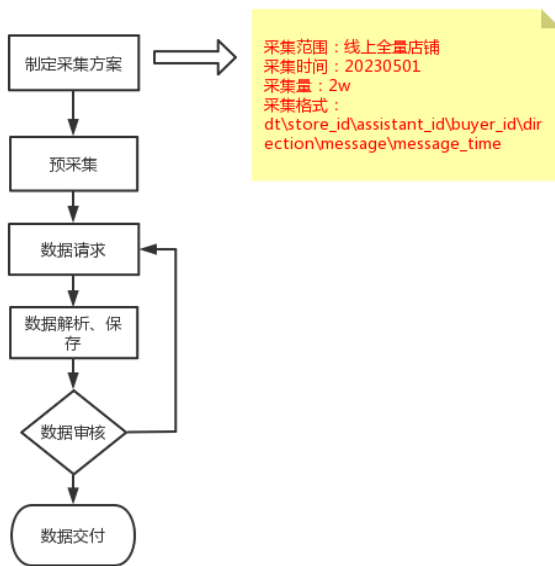
1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

智能客服机器人能通过识别理解用户意图，将客户的需求进行合理分类、传递，从知识库中寻求答案，或者直接回答客户问题，或者辅助人工客服智能应答。本项目中，你作为一名人工智能训练师，将对客服会话数据的采集处理流程提出优化建议。

- (1) 请阅读如下的会话数据采集流程，对采集流程提出优化建议。（5分）



(2) 请阅读如下的会话数据处理流程，对处理流程提出优化建议。

会话数据处理流程：

第一步，将数据按 store_id 进行归类；

第二步，对 message 进行 unique 处理，得到 unique_count；

第三步，将采集的会话数据进行撞库处理，得到模型预测识别结果；

第四步，汇总时筛选出 direction 为 buyer_to_assistant 的数据。（5分）

3、技能要求

(1) 能够对业务数据采集流程提出优化建议

(2) 能够对业务数据处理流程提出优化建议

4、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|---|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 5分 | 正确给出类似上述优化需求中的一条：+2分 正确给出类似上述优化需求中的两条及以上：+5分 若给出一个及以上不合理的优化需求：-2分 （优化需求合理即可） 加总得分在 0 到 5 分之间 | | | | |

| | | | | | | |
|-----|----|---|--|--|--|--|
| (2) | 5分 | 正确给出类似上述优化需求中的一条：+2分 正确给出类似上述优化需求中的两条及以上：+5分 若给出一个及以上不合理的优化需求：-2分 (优化需求合理即可) 加总得分在0到5分之间 | | | | |
|-----|----|---|--|--|--|--|

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.2.1

试题名称：客服会话数据归类和定义

考核时间：30 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

智能客服机器人能通过识别理解用户意图，将客户的需求进行合理分类、传递，从知识库中寻求答案，或者直接回答客户问题，或者辅助人工客服智能应答。本项目中，你作为一名人工智能训练师，将对已清洗完的数据进行分析，并根据分析结果进行数据归类和数据定义。

(1) 打开文件夹中的程序文件 2.1_clustering.ipynb，按下列要求及代码中的注释补全代码，并以原文件名保存在文件夹中。代码实现对已清洗完并获得相应 embedding 的会话数据进行 K-means 聚类分析。（10 分）

1. 导入 sikit-learn 包里的聚类方法
2. 读取 bert_embedding.csv 文件，查看数据
3. 使用 K-means 聚类算法，将数据聚为 3 类。

(2) 打开文件夹中的程序文件 2.1_clustering.ipynb，按下列要求及代码中的注释补全代码，找出最佳聚类分组，进行数据分类，提交分类后的数据。（10 分）

1. 计算 CH 系数，确认最佳 K 值。
2. 参考 dataSample.csv 文件，保存最佳聚类结果到 cluster_result.csv

(3) 请根据上述结果，对数据进行定义，以指导后续的数据标注；选取其中一簇判断聚类结果是否合理，给出理由。(10分)

聚类结果定义示例：

请求改价

定义：买家请求卖家改价，或者询问能否改价，不包括咨询卖家是否已经改好价格

例句：那你给我改价格吧

3、技能要求

- (1) 能够运用工具，对杂乱数据进行分析，输出内在关联及特征
- (2) 能够根据数据内在关联和特征进行数据归类
- (3) 能够根据数据内在关联和特征进行数据定义

4、质量指标

- (1) 代码书写规范
- (2) 代码正常运行

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|-----|---------------|--------|----|
| (1) | 10分 | 正确补全代码，每空格 2分 | | |
| (2) | 10分 | 正确补全代码，每空格 2分 | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|-----|---|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (3) | 10分 | 每个簇给出合适名称及定义各 2 分。 聚类结果判断为合理或不合理均正确，给出的合理理由即可：+2 分 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.2.2

试题名称：客服会话标注数据审核

考核时间：60 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

智能客服机器人能通过识别理解用户意图，将客户的需求进行合理分类、传递，从知识库中寻求答案，或者直接回答客户问题，或者辅助人工客服智能应答。本项目中，你作为一名人工智能训练师，对已标注的数据进行审核，并进行错误纠正及数据筛选。

- (1) 文件“标注结果_待审核.xlsx”包含两个 label 的标注结果，请对这份标注数据进行准确性和完整性审核，输出审核报告。（10 分）
- (2) 请列出审核过程中发现的三类错误并进行纠正，结果保存至“标注结果_已审核.xlsx”文件。（5 分）
- (3) 请根据审核结果，对数据进行筛选，选出适合用于模型训练测试的数据。结果保存至“标注结果_train.xlsx”文件。（5 分）

3、技能要求

- (1) 能够完成对标注数据准确性和完整性审核，输出审核报告
- (2) 能够对审核过程中发现的错误进行纠正
- (3) 能够根据审核结果完成数据筛选

4、质量指标

- (1) 错误描述或纠正正确
- (2) label 书写正确
- (3) 数据保存格式正确

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|------|---|--------|----|
| (1) | 10 分 | 答对一点得 5 分，答错不扣分。 | | |
| (2) | 5 分 | 5 条未标注数据填写正确的 label: +2 分 3 条标签有误的数据更改为正确的 label: +3 分 | | |
| (3) | 5 分 | 删除错误的系统消息、客服回复数据，每条 1 分。 | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.3.1

试题名称：客户体验管理系统维护

考核时间：15 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

客户体验管理（Customer Experience Management, CEM）智能系统是一种集成的解决方案，可帮助企业跟踪、分析和管理客户与公司的所有交互，从而提高客户体验和满意度。乐言科技的 CEM 产品是一款面向电商客服主管及质检团队的产品。本项目中，作为一名人工智能训练师，你将进行该智能系统的知识与数据维护，并为其单一智能功能找到合适的应用场景。请参考该智能系统的操作手册（CEM 操作手册.pdf）进行以下试题的作答。

- (1) 请对质检规则进行维护，添加一条自定义质检规则，检测客服是否有仅发送单个标点符号。请用文字描述如何操作来添加这条规则。（3 分）
- (2) 客服培训案例的维护需要会话数据，请用文字描述添加为客服培训案例的会话数据的判断标准。（3 分）
- (3) 请为该智能系统的客服对话质检功能找到合适的应用场景，并用文字描述应用场景。（4 分）

3、技能要求

- (1) 能够维护智能系统所需知识
- (2) 能够维护智能系统所需数据

(3) 能够为单一智能产品找到合适应用场景

4、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|--|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 3分 | 正确说明选择检测对象：+1分 说明质检关键词添加整句且正确描述至少一个标点符号：+2分 | | | | |
| (2) | 3分 | 提供一个合理标准：+1分 提供两个及以上合理标准：+3分 提供一个及以上不合理标准：-1分 加总得分在0到3分之间 | | | | |
| (3) | 4分 | 提供至少一个合理应用场景：+3分 提供两个及以上合理应用场景：+4分 提供一个及以上不合理应用场景：-2分 加总得分在0到4分之间 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：1.3.2

试题名称：客户体验管理系统优化

考核时间：15 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

客户体验管理（Customer Experience Management, CEM）智能系统是一种集成的解决方案，可帮助企业跟踪、分析和管理工作与客户的所有交互，从而提高客户体验和满意度。乐言科技的 CEM 产品是一款面向电商客服主管及质检团队的产品。本项目中，作为一名人工智能训练师，你将对其输出的数据进行分析，根据数据分析结论提出优化需求。请参考该智能系统的操作手册（CEM 操作手册.pdf）进行以下试题的作答。

- (1) 请对评价分析结果（评价分析.xlsx）进行数据分析，输出分析报告。（4分）
- (2) 请根据数据分析的结论，对评价分析提出优化需求。（6分）

3、技能要求

- (1) 能够利用分析工具进行数据分析，输出分析报告
- (2) 能够根据数据分析结论对智能产品的单一功能提出优化需求

4、质量指标

无

主观评分表/过程评分表

| 序号 | 配 分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|--------|---|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 4 分 | 正确给出标签统计：+2 分 正确给出占比最高的标签说明：+2 分 错误不扣分 加总得分在 0 到 4 分之间 | | | | |
| (2) | 6 分 | 正确给出类似参考答案优化需求中的一条：+3 分 正确给出类似参考答案优化需求中的两条及以上： +6 分 错误不扣分 加总得分在 0 到 6 分之间 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.1.1

试题名称：人类性别年龄智能分类系统图像数据质量检测

考核时间：30 min

一、场地设备要求

1. 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
2. 流程图管理软件 1 个，可绘制流程图

二、工作任务

人类性别年龄智能分类系统是一种基于人工智能技术的系统，旨在通过分析人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。该系统的应用领域广泛，包括社交媒体分析、人群统计、广告定向等。为持续的提高这个系统的准确率，需要持续提供人工标注的源数据图像进行深度学习，在这个过程中，高质量、高效率的提供大量测试图像数据极为重要。

根据文件夹「测试集」下的所有预处理后的图像数据，回答下列问题：

1. 请审核提供的图像数据集，指出其中存在的问题。（4 分）
2. 请根据图像数据采集标准，提出适合当前模型背景的图像数据采集规范要求。（8 分）
3. 请根据图像数据清洗标准，提出适合当前模型背景的图像数据处理规范要求。（8 分）

三、技能要求

1. 能够对预处理后业务数据进行审核

2. 能够结合人工智能技术要求，梳理业务数据采集规范
3. 能够结合人工智能技术要求，梳理业务数据处理规范

四、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|---|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 4分 | 正确指出上述存在问题中的一个：+1分 正确指出上述存在问题中的三个及以上：+4分 答错不扣分 | | | | |
| (2) | 8分 | 提供一个合理的采集规范：+2分 提供四个及以上合理的采集规范：+8分 提供一个及以上不合理的采集规范：-2分 加总得分在0到8分之间 | | | | |
| (3) | 8分 | 提供一个合理的处理规范：+2分 提供三个及以上合理的处理规范：+8分 提供一个及以上不合理的处理规范：-2分 加总得分在0到8分之间 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.1.2

试题名称：人类性别年龄智能分类系统图像数据集流程优化

考核时间：30 min

一、场地设备要求

1. 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
2. 流程图管理软件 1 个，可绘制流程图

二、工作任务

人类性别年龄智能分类系统是一种基于人工智能技术的系统，旨在通过分析人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。该系统的应用领域广泛，包括社交媒体分析、人群统计、广告定向等。为持续的提高这个系统的准确率，需要持续提供人工标注的源数据图像进行深度学习，在这个过程中，高质量、高效率的提供大量测试图像数据极为重要。

请阅读如下的会话数据采集流程，对采集流程提出优化建议。（10 分）

1. 数据采集：确定数据集的来源，确保数据采集过程符合法规和伦理标准。
2. 数据预处理：对原始数据进行处理，包括去除重复数据、处理缺失值、图像增强等。
3. 标注数据：对数据集中的图像进行人工标注，包括性别和年龄段。
4. 数据平衡处理：确保各类别样本数量相对均衡，避免数据集偏差。
5. 数据清洗和筛选：排除低质量或错误标注的样本。

三、技能要求

4. 能够对业务数据采集流程提出优化建议
5. 能够对业务数据处理流程提出优化建议

四、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|-----|--|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 10分 | 正确给出类似上述新增步骤中的一条：+3分 正确给出类似上述新增步骤中的两条及以上：+5分 正确给出类似上述优化步骤中的一条：+1分 正确给出类似上述优化步骤中的两条：+3分 正确给出类似上述优化步骤中的三条及以上：+5分 若给出一个及以上不合理的优化需求：-1分 （优化需求合理即可）加总得分在0到10分之间 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.2.1

试题名称：基于人类照片的性别年龄数据归类和定义

考核时间：60min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Labelme）

2、工作任务

人类性别年龄智能分类系统是一种集成的一种基于人工智能技术的系统，旨在根据人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。在本项目中，作为一名人工智能训练师，将会对已经清洗完成的数据使用 Labelme 进行标注，并对标注后的数据进行提取归档

- (1) 使用 labelme 对工程下 pic 文件夹中的图片进行标注，并对每张图片生成标注的结果的 json 文件保存在相同的目录下（15 分）

- ① 使用方形圈出照片中人物的脸
- ② 对图片进行标注，标注的格式为：性别,年龄段
- ③ 性别枚举值：[Male,Female]
- ④ 年龄段枚举值：[(0-2), (4-6), (8-12), (15-20), (25-32), (38-43), (48-53), (60-100)]
- ⑤ 标注样例：Female, (4-6)

- (2) 打开文件夹中的程序文件 4.2.1.ipynb，按照下列要求及代码中的注释补充代码，对上题每张图片标注的结果进行提取，将提取到的结果进行归档合并（15 分）

- ① 遍历 pic 文件夹中所有的 json 文件，提取标注内容

② 参考 Sample.csv 的格式，将最终的归档结果提取到 result.csv 中

3、技能要求

- (1) 能够运用工具，对杂乱数据进行分析，输出内在关联及特征
- (1) 能够运用工具，对标注后的数据进行提取归档

4、质量指标

- (1) 代码书写规范
- (2) 代码正常运行

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|------|--|--------|----|
| 01 | 30 分 | 1、正确补全代码，每空 4 分，共 20 分 2、成功生成 result.csv 5 分 3、result.csv 的 gender 字段和答案保持一致 5 分 | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.2.2

试题名称：基于人类照片的性别年龄标注数据审核

考核时间：20min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）

2、工作任务

人类性别年龄智能分类系统是一种集成的一种基于人工智能技术的系统，旨在根据人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。在本项目中，作为一名人工智能训练师，对已标注的数据进行审核，并进行错误纠正及数据筛选

- (1) 文件《标注结果_待审核.xlsx》为对目录下的 15 张图片的标注结果，其中 label_age 为对年龄段的标注，label_gender 为对性别的标注，请从数据的完整性和准确性两方面进行审核，输出报告（10 分）

- (2) 请列出审核过程中发现的错误，并进行纠正，将结果保存至《标注结果_已审核.xlsx》（5 分）

- (3) 请根据审核结果，对数据进行筛选，选出适合模型训练的测试数据，结果保存至《标注结果_train.xlsx》（5 分）

3、技能要求

- (1) 能够完成对标注数据准确性和完整性审核，输出审核报告
- (2) 能够对审核过程中发现的错误进行纠正
- (3) 能够根据审核结果完成数据筛选

4、质量指标

- (1) 错误描述或纠正正确
- (2) label 书写正确
- (3) 数据保存格式正确

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|------|---------------------------------------|--------|----|
| 01 | 10 分 | 答对一点得 5 分，共 10 分 | | |
| 02 | 5 分 | 对标注错误的数据修改正确得 2 分 对未标注的数据进行标注得 3 分 | | |
| 03 | 5 分 | 删除错误图片数据，模糊的图片 1 张得 2 分，2 张得 5 分 | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.3.1

试题名称：基于人类照片的性别年龄智能分类系统维护

考核时间：20 min

1、场地设备要求

- (1) 人工智能训练师主机 1 台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：labelme）
- (2) 流程图管理软件 1 个，可绘制流程图

2、工作任务

人类性别年龄智能分类系统是一种集成的一种基于人工智能技术的系统，旨在根据人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。在本项目中，作为一名人工智能训练师，你将进行该智能系统的知识与数据维护，并为其单一智能功能找到合适的应用场景。请对如下问题回答：

- (1) 要进行人类性别年龄智能分类系统的维护，需要哪些知识的储备。（3分）
- (2) 如何维护该系统所需数据。（3分）
- (3) 请为该智能系统找到合适的应用场景，并用文字描述应用场景。（4分）

3、技能要求

- (1) 能够维护智能系统所需知识
- (2) 能够维护智能系统所需数据
- (3) 能够为单一智能产品找到合适应用场景

4、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|--|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 3分 | 正确说明至少1个知识点：+1分 每多说明一项符合要求的知识点+1分，最多+2分 提供了一个不相关知识点：-1分 每多说明一项不符合要求的知识点+1分，最多-2分 加总得分在0到3分之间 | | | | |
| (2) | 3分 | 正确说明至少一个数据维护方法：+1分 每多说明一项符合要求的维护方法+1分，最多+2分 提供了一个不相关数据维护方法：-1分 每多说明一项不符合要求的维护方法+1分，最多-2分 加总得分在0到3分之间 | | | | |
| (3) | 4分 | 提供至少一个合理应用场景：+3分 提供两个及以上合理应用场景：+4分 提供一个及以上不合理应用场景：-2分 加总得分在0到4分之间 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：2.3.2

试题名称：基于人类照片的性别年龄智能分类系统优化

考核时间：20min

1、场地设备要求

- (1) 人工智能训练师主机1台（CPU：建议英特尔 i5 以上，操作系统：Windows 10；开发工具：Python 编译工具，深度学习训练环境：支持主流深度学习框架训练，标注工具：Microsoft Excel）
- (2) 流程图管理软件1个，可绘制流程图

2、工作任务

人类性别年龄智能分类系统是一种基于人工智能技术的集成系统，旨在根据人的外貌特征和其他相关信息，自动识别和分类人的性别和年龄段。作为一名人工智能训练师，你被指派优化这个系统。请对以下问题进行回答：

- (1) 根据你的经验，简述可能会影响该系统性能的因素。（2分）
- (2) 针对上述你提到的影响因素，你会如何优化这个系统？请给出具体的优化策略和步骤。（3分）
- (3) 请为该系统设计一个实验，用来验证你的优化策略是否有效。需要包括实验设计、实验步骤、预期结果以及如何评估实验结果。（5分）

3、技能要求

- (1) 能够利用分析工具进行数据分析，输出分析报告
- (2) 能够根据数据分析结论对智能产品的单一功能提出优化需求

4、质量指标

无

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|--|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 2分 | 正确说明至少1个性能影响因素：+2分 未提供或者提供了1个与性能不相关的因素：不得分 | | | | |
| (2) | 3分 | 针对至少1个性能影响因素提供了有效的优化策略及步骤：+2分 未提供或者提供了无关的优化策略及步骤：不得分 | | | | |
| (3) | 5分 | 完整性得分（2分）： 包含实验设计、实验步骤、预期结果以及如何评估实验结果：+2分 实现内容不完整：酌情给分，得分不超过1分 合理性得分（3分）： 实验描述细节丰富，合情合理：+3分 实验描述细节缺失：酌情给分，得分不超过2分 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：3.1.1

试题名称：文本纠错-业务数据质量检测

考核时间：20 min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|---|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16G 操作系统：Windows10 开发工具：Python 编译工具 深度学习训练环境：支持主流深度学习框架训练 配件：鼠标、键盘及显示器等配件 | 台 | 1 | |

2、工作任务

项目背景：

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是支撑其良好表现和能力的基础。

你将作为一名人工智能数据分析师对需校对的文本分析审核：

根据业务特征和业务需求，能够利用设备、工具等完成原始业务数据和数据库内业务数据的采集，能够根据数据处理要求完成业务数据整理归类和汇总。

请回答以下问题：

- (1) 业务数据在采集过程中，需要遵循的规范和采集方法有哪些？
- (2) 业务数据读取：利用 pandas 读取 data 下的 xlsx 文件，并统计数据总条数。读取内容和统计数量截图提交至“提交资料/题目 1/试题_数据采集和处理.docx”文件中，能显示 wrong_text 和 correct_text 即可。



3、技能要求

- (1) 能够对预处理后业务数据进行审核
- (2) 能够结合人工智能技术要求，梳理业务数据采集规范
- (3) 能够结合人工智能技术要求，梳理业务数据处理规范

4、质量指标

- (1) 知道基本的采集规范和采集方法。
- (2) 完成要求的业务数据采集和处理。

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|-----|--|--|----|
| (2) | 5 分 | 打印显示读取到的文本，得分 2 分； 打印显示纠错数据数量，得分 3 分。 | 打印显示读取到的文本，即得分 2 分：  打印显示纠错数据数量，即得分 3 分：  | |
| | | | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|--------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| | | | | | | |

| | | | | | | |
|-----|----|---|--|--|--|--|
| (1) | 5分 | 合理的采集规范即可，每个+1分，采集规范得分最高3分； 合理的采集方法即可，每个+1分，采集方法得分最高2分 | | | | |
| | | | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：3.1.2

试题名称：文本纠错-业务数据质量检测

考核时间：40 min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|---|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16G 操作系统：Windows10 开发工具：Python 编译工具 深度学习训练环境：支持主流深度学习框架训练 配件：鼠标、键盘及显示器等配件 | 台 | 1 | |

2、工作任务

项目背景：

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是支撑其良好表现和能力的基础。

你将作为一名人工智能数据分析师对需校对的文本分析审核：

根据业务特征和业务需求，能够利用设备、工具等完成原始业务数据和数据库内业务数据的采集，能够根据数据处理要求完成业务数据整理归类和汇总。

请回答以下问题：

(1)业务数据清理：剔除数据当中的脏数据、乱数据（如：非中文，有缺省值，文本长度小于4的文本数据）。清洗后截图提交至“提交资料/题目1/试题_数据采集和处理.docx”文件中，显示清理后的 wrong_text 和 correct_text 即可。

(2)在原数据中添加一列表示修改时间（modify_date），该列值均为：“2022-03-25”。成功后，再把该列类型转换为时间类型。添加与转换完成后截图提交至“提交资料/题目1/试题_数据采集和处理.docx”文件中，截图中能显示修改时间列的前后类型变化、前/后任一时候的修改时间即可。

3、技能要求

- (1) 能够对业务数据采集流程提出优化建议
- (2) 能够对业务数据处理流程提出优化建议

4、质量指标

- (1) 根据要求完成数据采集。
- (2) 根据要求完成数据处理。

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|-----|--|--------|----|
| (1) | 10分 | 清理掉缺省值数据，得3分； 清理掉包含非中文数据，得4分； 清理掉文本长度小于4的数据，得3分。 | | |
| (2) | 10分 | 在原数据中添加一列表示修改时间（modify_date），该列值均为：“2022-03-25”，得6分； 把该列（modify_date）类型转换为时间类型，得4分。 | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：3.2.1

试题名称：文本智能校对-数据归类和定义

考核时间： min

1、 场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备 |
|----|-----------|---|----|----|---|
| 1 | 人工智能训练师主机 | CPU: 建议英特尔 i5 以上 内存: 建议不少于 16G 操作系统: Windows10 开发工具: Python 编译工具 深度学习训练环境: 支持主流深度学习框架训练 配件: 鼠标、键盘及显示器等配件 | 台 | 1 | |

1、 工作任务

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对

于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是支撑其良好表现和能力的基础。

你将作为一名人文本智能校对系统数据分析师对需校对的业务数据进行分析 and 标注，完成以下要求。

- (1) 定义三种纠错类型：多字、少字、错字
- (2) 对每行数据对进行错误标记，并把数据标注结果写到错误标记列（label）列中。

3、技能要求

- (1) 能够运用工具，对杂乱数据进行分析，输出内在关联及特征
- (2) 能够根据数据内在关联和特征进行数据归类
- (3) 能够根据数据内在关联和特征进行数据定义

4、质量指标

- (1) 能够正确标注各种错误类型数据得 10 分。

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|------|---|--------|----|
| (1) | 10 分 | 正确标记错字数据，4 分； 正确标记多字数据，3 分； 正确标记少字数据，3 分。 | | |

人工智能训练师（X 级）操作技能考核

试题单

准考证号：

试题代码：3.2.2

试题名称：文本智能校对-标注数据审核

考核时间： min

1、 场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备 |
|----|-----------|---|----|----|---|
| 1 | 人工智能训练师主机 | CPU: 建议英特尔 i5 以上 内存: 建议不少于 16G 操作系统: Windows10 开发工具: Python 编译工具 深度学习训练环境: 支持主流深度学习框架训练 配件: 鼠标、键盘及显示器等配件 | 台 | 1 | |

2、 工作任务

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对

于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是支撑其良好表现和能力的基础。

你将作为一名人文本智能校对系统数据分析师对需校对的业务数据进行分析 and 标注，完成以下要求。

- (1) 对错误类型进行类型数量统计：统计每个错误类型有多少条。
- (2) 对错误类型统计结果使用柱状图表示出来(列名用拼音的: duo_zi, shao_zi, cuo_zi 即可)。
- (3) 根据 create_date 对数据进行降序排序

3、技能要求

- (1) 能够完成对标注数据准确性和完整性审核，输出审核报告
- (2) 能够对审核过程中发现的错误进行纠正
- (3) 能够根据审核结果完成数据筛选

4、质量指标

- (1) 能够正确统计各种错误类型得 15 分。
- (2) 能够对统计出来对错误类型结果用柱状图表示出来 15 分。
- (3) 能够根据 create_date 字段对数据进行降序排序得 10 分

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|------|--|--------|----|
| (1) | 15 分 | 正确统计多字数量 5 分； 正确统计少字数量 5 分； 正确统计错字数量 5 分 | | |
| (2) | 15 分 | 展示出一个柱状图得 3 分； 展示出得柱状图有 duo_zi, shao_zi, cuo_zi 三列得 3 分； 柱状图中正确展示多字 (duo_zi) 数量得 3 分， 柱状图中正确展示少字 (shao_zi) 数量得 3 分， 柱状图中正确展示错字 (cuo_zi) 数量得 3 分。 | | |
| (3) | 10 分 | 正确对日期进行降序排序，得 10 分。 | | |

工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：3.3.1

试题名称：文本纠错-智能系统维护

考核时间：15 min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|---|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16G 操作系统：Windows10 开发工具：Python 编译工具 深度学习训练环境：支持主流深度学习框架训练 配件：鼠标、键盘及显示器等配件 | 台 | 1 | |

2、工作任务

项目背景：

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是

支撑其良好表现和能力的基础。

你将作为一名人工智能数据分析师对需校对的文本分析审核：

能够进行智能系统的开启，能够简单使用智能系统，能够记录智能系统功能应用和应用数据情况。

请回答以下问题：

- (1) 描述各个时间点各个错误类型的统计情况。

3、技能要求

- (1) 能够维护智能系统所需知识
- (2) 能够维护智能系统所需数据
- (3) 能够为单一智能产品找到合适应用场景
- (4) 能够将智能系统在真实应用场景中部署实施

4、质量指标

- (1) 根据要求记录应用数据情况。

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|-----|----------------------|--|----|
| (1) | 10分 | 正确展示要求记录应用数据情况，得10分。 | 按要求完成数据统计： 2021-03-17:4条,2021-03-16:3条, 2021-03-15:3条。 | |
| | | | | |
| | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：3.3.2

试题名称：文本纠错-智能系统优化

考核时间：15 min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|---|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16G 操作系统：Windows10 开发工具：Python 编译工具 深度学习训练环境：支持主流深度学习框架训练 配件：鼠标、键盘及显示器等配件 | 台 | 1 | |

2、工作任务

项目背景：

近年来，随着社交网络和新闻传媒的蓬勃发展，人们获取信息的途径愈加多样化和便利化。但与此同时，由于信息来源的不确定性和信息传播的快速性，新闻报道中出现语法错误、拼写错误、事实型错误等问题就屡见不鲜。这不仅会影响信息的传递和理解，还会影响新闻报道的专业性和可信度，甚至引发一些不必要的误解和纠纷。因此，进行文本纠错的任务，对于新闻传媒来说是至关重要的。只有准确、规范、易懂的新闻报道，才能更好地满足公众需求，维护社会秩序和稳定。

基于人工智能技术的文本纠错任务的主要一环在于数据的采集、清洗和标注。在实现文本纠错任务时，准确可靠的数据是保证模型效果的关键。数据采集的过程需要处理大量来源不一、格式不统一的文本信息，并对其进行去重、筛选等操作，从中提取出具有代表性和可信度的数据。而数据清洗则涉及到对原始数据的繁琐处理，去掉冗余信息等。最后，数据标注是文本纠错任务中难度较大的问题之一，涉及到对错字、错别字、同音字等错误识别和标注的处理。总之，对于人工智能技术的文本纠错任务来说，精准的数据采集、清洗和标注是支撑其良好表现和能力的基础。

你将作为一名人工智能数据分析师对需校对的文本分析审核：

能够进行智能系统的开启，能够简单使用智能系统，能够记录智能系统功能应用和应用数据情况。

请回答以下问题：

(1) 已知模型效果依赖数据标注情况与数量，如果预计纠错模型可以获得更好的效果，请提出 3 条建议。

3、技能要求

- (1) 能够利用分析工具进行数据分析。
- (2) 能够根据数据分析结论对智能产品的单一功能提出优化需求。

4、质量指标

- (1) 根据要求记录应用数据情况。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|-----|---|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| (1) | 10分 | 给出合理建议 1 条总分得 3 分； 给出合理建议 2 条总分得 6 分； 给出合理建议 3 条总分得 10 分。 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.1.1

试题名称：自动驾驶-数据采集和处理-业务数据质量检测

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python ；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。

早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。

当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会

等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。数据采集和处理：根据业务特征与业务需求，能够对预处理后业务数据进行审核，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 原始数据采集的质量决定了数据预处理的难度和工作量，影响着生产过程的效率和效果，最终影响产品的质量，“采集文件”文件夹中是已经采集好的数据，需要对其质量进行审核，请根据数据质量审核需要遵循的六个核心纬度中的两个纬度，对此文件夹中的数据进行处理，并写出所参考的纬度。内容截图提交至“提交资料/1.1/试题_自动驾驶数据采集和处理”文件中。（4分）

(2) 人工智能行业对数据采集的质量要求比较高，数据服务团队的项目负责人需要深刻理解采集标准和规范，请完善自动驾驶数据采集员的岗位职责表格。内容提交至“提交资料/1.1/试题_自动驾驶数据采集和处理”文件中。（4分）

表 1-1 自动驾驶数据采集员的岗位职责（编号和采集顺序无关）

| 职责编号 | 职责描述 |
|------|-------------------------------|
| 1 | 知识: |
| 2 | 技术: |
| 3 | |
| 4 | 熟悉数据采集系统各个模块的工作原理，及时预判和解决设备问题 |
| 5 | 整理每日工作日志，定期输出工作总结和报告 |
| 6 | |

(3) 为了更好的处理数据，需要先对数据进行一些整合处理，请按照汽车类型、采集场景、数据类型三种方式任选其一对数据进行整合归类。内容截图提交至“提交资料/1.1/试题_自动驾驶数据采集和处理”文件中。（8分）

4、质量指标

| 题目号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|----|----|------------------------------------|---------|----|
| 1 | 01 | 2分 | 数据修改是否正确，需要修改内容为：“找到无效数据”和“删除多余数据” | 两个操作各1分 | |
| | 02 | 2分 | 修改对应的原则是否正确 | 两个原则各1分 | |
| 2 | 01 | 2分 | 职责1、2处合理即可 | 参考答案结果 | |
| | 02 | 6分 | 3、6任意填写下述的一项即可 | 参考答案结果 | |
| 3 | 01 | 8分 | 按照要求分类，汽车类型、采集场景、数据类型三种方式任选其一即可 | 参考答案结果 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.1.2

试题名称：自动驾驶-数据采集和处理-数据处理方法优化

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python ；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。

早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。

当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将

其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。数据采集和处理：根据业务特征与业务需求，能够对预处理后业务数据进行审核，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 根据自动驾驶功能需求，实现全部自动驾驶功能则至少需要安装 6 个以上的车载摄像头，安装位置为前视、环视、后视、侧视以及内置。请将表格中的内容补充完整。内容提交至“提交资料/1.2/试题_自动驾驶数据采集和处理”文件中。(5分)

将以下表格内容补充完整：

| 传感器 | 感知范围 | 性能 |
|-------|-----------|---|
| 环视摄像头 | 50° /150m | 主要应用于中远距离场景，能识别清晰的车道线、交通标识、障碍物、行人，但对光照、天气等条件很敏感，而且需要复杂的算法支持，对处理器的要求也比较高。 |
| 超声波雷达 | | 毫米波雷达可有效提取景深及速度信息，识别障碍物，有一定的穿透雾、烟和灰尘的能力，但在环境障碍物复杂的情况下，由于毫米波依靠声波定位，声波出现漫反射，导致漏检率和误差率比较高。 24GHz 雷达主用于中短测距，77GHz 雷达用于长测距。 |
| 激光雷达 | | |

(2) 关于数据整合中的优化，查看文件夹“数据整合优化”中的图像，判断出它是非结构化数据整合类型下的哪一种方式，且写出在哪些自动识别领域广泛使用了这种整合方式。并将文件夹“数据整合优化”中有重叠区域的两张图片移动到一起，并截图保存到文档中。内容提交至“提交资料/1.2/试题_自动驾驶数据采集和处理”文件中。(5分)

4、质量指标

客观评分表/结果评分表

| 题目号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|----|----|----------------------|------------|----|
| 1 | 01 | 5分 | 按照答案进行评分 | 填错一空扣除0.5分 | |
| 2 | 01 | 2分 | 答出“图像拼接”，并写出至少一个应用领域 | 参考答案结果 | |

| | | | | | |
|--|----|-----|------------|------------------------|--|
| | 02 | 3 分 | 按照要求进行图像拼接 | 参考答案结果, 拼错 一张扣除 1 分 | |
|--|----|-----|------------|------------------------|--|

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.2.1

试题名称：“自动驾驶”-数据标注-数据归类和定义

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。数据标注：根据业务需求，完成对标注数据准确性和完整性审核，利用 Python 完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。根据业务目标，基于工具完成天气分析，并可以对数据进行归类和汇总，满足分析和处理的要求。

打开项目中的数据分析（试题）文件，考生在规定的时间内完成以下任务：

(1) 读取 imgs 文件夹下所有的图片，筛选出格式为 jpg 和 png 的图片文件，并去除所有分辨率低于 500×500 像素及文件大小小于 200K 的图片，筛选后的图片名字和分辨率信息输出至 image_info.json 文件中，格式为[{"img": image name, "shape": shape}], 文件内容提交至“提交资料/2.1/试题_自动驾驶数据标注”。

(15 分)

(2) 得到筛选后所有图片的长宽信息，其中图片 i 的宽为 W 像素，高为 H 像素，通过公式 $x_i = \lfloor \frac{W}{16} \rfloor$, $y_i = \lfloor \frac{H}{16} \rfloor$ 求得 x_i, y_i ，把 x, y 相同的图片置为一组，以二维列表的方式输出，输出格式为[[{"img": "image name", "shape": shape, 'heightDivisor': y1, 'widthDivisor': x1}], {"img": "image name", "shape": shape, 'heightDivisor': y2, 'widthDivisor': x2}]] 内容提交至“提交资料/2.1/试题_自动驾驶数据标注”。(15 分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|------|---------------|--------|----|
| 1 | 01 | 15 分 | 编写代码，完成筛选 | 同答案操作 | |
| 2 | 01 | 15 分 | 编写代码，完成输出二维列表 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.2.2

试题名称：“自动驾驶”-数据标注-标注数据审核

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。数据标注：根据业务需求，完成对标注数据准确性和完整性审核，利用 Python 完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。根据业务目标，基于工具完成天气分析，并可以对数据进行归类和汇总，满足分析和处理的要求。

打开项目中的数据分析（试题）文件，考生在规定的时间内完成以下任务：

(1) 把所有筛选过的图片 padding 为 1024 像素×1024 像素的图片并保存至 padding 文件夹下，文件夹保存至“提交资料/2.2/试题_自动驾驶数据标注”。(6分)

(2) 对筛选后的图片做归一化，其中 mean=(0.485, 0.456, 0.406)，std=(0.229, 0.224, 0.225)，并将归一化后的图片保存至 normalization 文件夹下，文件夹保存至“提交资料/2.2/试题_自动驾驶数据标注”。(6分)

(3) 读取 label.jpg 图片和 label.json 文件，将 label.json 内的标注信息可视化至 label.png 图片中，图片文件夹保存至“提交资料/2.2/试题_自动驾驶数据标注”。(8分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|---------|--------|----|
| 1 | 01 | 6分 | 完成筛选 | 同答案操作 | |
| 2 | 01 | 6分 | 完成归一化操作 | 同答案操作 | |
| 3 | 01 | 8分 | 完成可视化操作 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.3.1

试题名称：“自动驾驶”-智能系统运维-智能系统维护

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(1) 查看当前 MySQL 的最大连接数和已经连接数。内容提交至“提交资料/3.1/试题_自动驾驶数据智能系统运维.docx”。(10分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-----|---------------------------|--------|----|
| 1 | 01 | 10分 | 成功查看当前 MySQL 的最大连接数和已经连接数 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：4.3.2

试题名称：“自动驾驶”-智能系统运维-智能系统优化

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）自动驾驶是一项颠覆性的技术，它允许车辆在不需要人类干预的情况下自动驾驶。这项技术已经在汽车、飞机、船只和无人机等领域得到广泛应用。自动驾驶技术的发展至今已有数十年，但是直到最近才取得了较大的进展。早期的自动驾驶技术主要是依赖于传感器和计算机系统来实现的。传感器可以感知车辆周围环境，计算机系统则将传感器信息解析为车辆的控制指令。但当时的计算机能力和传感器技术还无法满足实现自动驾驶的要求。随着计算机技术和传感器技术的不断发展，自动驾驶技术也逐渐成熟。例如，现在广泛使用的激光雷达传感器就可以检测并识别车辆周围的物体。同时，计算机算法的提高也让自动驾驶系统变得更加智能化和有效。当前，全球各大汽车厂商和科技公司都在积极开发自动驾驶技术，并试图将其商业化、普及化。不过，自动驾驶技术在面临着许多的技术、法律、金融和社会等方面的挑战，其商用化进程还需要进一步推进。

(2) 你将作为一名人工智能数据分析师对“自动驾驶”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(1) 打开 GT. json 和 Predict. json，计算各个标签的 IOU。内容提交至“提交资料/3.2/试题_自动驾驶数据智能系统运维”。(4分)

(2) 根据标签 IOU 结果，提出优化意见。内容提交至“提交资料/3.2/试题_自动驾驶数据智能系统运维.docx”。(6分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|------------------|--------|----|
| 1 | 01 | 4分 | 编写代码，计算各个标签的 IOU | 同答案操作 | |

主观评分表/过程评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|----|-------------|-------|---|---|----|
| | | | | 1 | 2 | 3 | |
| 2 | S1 | 2分 | 写出1个合理的优化建议 | | | | |
| | S2 | 4分 | 写出2个合理的优化建议 | | | | |
| | S3 | 6分 | 写出3个合理的优化建议 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.1.1

试题名称：玩具市场-数据采集和处理-业务数据质量检测

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发展具有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。数据采集和处理：根据业务特征与业务需求，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 请打开“采集文件”文件夹，其中是本次采集的全部数据，请根据数据质量审核遵循的六个核心纬度中的两个纬度，对已采集的数据进行处理，并写出所参考的纬度。内容截图提交至“提交资料/1.1/试题_玩具数据采集和处理”文件中。(4分)

(2) 采集规范通常涉及到的内容有：采集范围、规范性引用文件、术语和定义、基本原则、采集要求、人员要求、场所要求、设备要求、注意事项等。根据此次采集的图片内容，梳理对应的数据采集规范，补充填写试题中的表格。内容提交至“提交资料/1.1/试题_玩具数据采集和处理”文件中。(4分)

表 1-1 所示为玩具市场分析数据的采集规范

| 文档内容 | 描述 |
|--------|--|
| 1.范围 | 本文件规定了玩具市场采集的技术要求 |
| 2.采集要求 | 2.1 采集对象：玩具 2.2 采集目的：对比分析玩具的产地价格等和其销量之间的关系 2.3 采集内容： 2.4 文件格式： 文件命名： |
| 3.设备要求 | 3.1 安装 python, 安装导入 beautifulsoup 库 3.2 相机拍摄 |

(3) 数据常规处理包括整理、归类、汇总等，依据数据的分类分级原则，数据常规处理应该遵循的规范有多种。请先对图片文件按照材质再进行归类，并按照销量进行排序汇总。内容截图提交至“提交资料/1.1/试题_玩具数据采集和处理”文件中。(8分)

4、质量指标

客观评分表/结果评分表

| 题目号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|----|----|-------------------------------------|---------|----|
| 1 | 01 | 2分 | 数据修改是否正确，需要修改内容为：“修改文件名”和“删除重复的图片” | 两个操作各1分 | |
| | 02 | 2分 | 修改对应的原则是否正确 | 两个原则各1分 | |
| 2 | 01 | 6分 | 考生须填写出5个大类给5分，写出大类别下的小分类可给全部分数（同答案） | 参考答案结果 | |
| | 02 | 1分 | 填写正确的文件格式 | 参考答案结果 | |
| | 03 | 1分 | 填写正确的文件名命名 | 参考答案结果 | |
| 3 | 01 | 4分 | 按照材质正确归类 | 参考答案结果 | |
| | 02 | 4分 | 按照题目要求对数据进行排序 | 参考答案结果 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.1.2

试题名称：玩具市场-数据采集和处理-数据处理方法优化

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python ；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发

展具有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。数据采集和处理：根据业务特征与业务需求，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 在采集玩具的流程中，关于采集工具和采集环境都有可优化的地方，请明确写出可优化的方案。方案存储在新建的 word 文档中，文档命名为“关于玩具数据采集的优化建议”。内容提交至“提交资料/1.2/试题_玩具数据采集和处理”文件中。（5分）

(2) 仔细阅读上面第三题，思考在“汇总”之前是否还能添加一项工作进行数据处理的优化。写出可以添加的那一项数据处理的工作，并写出原因和优化方案。新建一个 word 文档，文档命名为“关于玩具数据处理的优化建议”。内容提交至“提交资料/1.2/试题_玩具数据采集和处理”文件中。（5分）

4、质量指标

主观评分表/过程评分表

| 题目号 | 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|----|-------------------------------------|-------|---|---|----|
| | | | | 1 | 2 | 3 | |
| 1 | S1 | 5分 | 分别写出“采集工具”和“采集玩具”可以优化的地方，合理即可，答案不唯一 | | | | |
| 2 | S1 | 5分 | 考生提出合理的建议即可，答案不唯一 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.2.1

试题名称：“玩具市场”-数据标注-数据归类和定义

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发

展具有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。数据标注：根据业务需求，完成对标注数据准确性和完整性审核，利用相关软件完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。

打开“数据标注”项目中的数据标注（试题）文件，考生在规定时间内完成以下任务：

(1) 标注项目需要开发一个基于 Flask 的 demo 系统，请在 app 文件夹构建一个 flask 项目，在访问网页时输出”hello world”，并将结果截图内容提交至“提交资料/2.1/试题_玩具数据标注”。（15分）

(2) demo 系统需要容器化的方式进行部署，请拉取 python:3.6 镜像，并使用拉取的镜像新建一个名字为 python-demo 的容器。文件内容提交至“提交资料/2.1/试题_自动驾驶数据标注”。（15分）

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-----|---------------------------|--------|----|
| 1 | 01 | 15分 | 编写代码，成功创建网页 | 同答案操作 | |
| 2 | 01 | 15分 | 完成新建一个名字为 python-demo 的容器 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.2.2

试题名称：“玩具市场”-数据标注-标注数据审核

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（2）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发

展具有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。数据标注：根据业务需求，完成对标注数据准确性和完整性审核，利用相关软件完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。

打开“数据标注”项目中的数据标注（试题）文件，考生在规定时间内完成以下任务：

(1) 请根据“app”文件中的代码创建出 flask 项目，并通过 docker build 封装成名字为 flask-test 的镜像，文件内容提交至“提交资料/2.2/试题_玩具市场数据标注”。(6分)

(2) 使用(1)中新建 flask-python 镜像新建一个名字为 flask-demo 的容器，让 demo 运行起来并能被同个局域网的其他设备通过浏览器访问，并把容器的详情信息输出到 describe.txt 文件中，文件内容提交至“提交资料/2.2/试题_玩具市场数据标注”。(6分)

(3) 输出 flask-demo 的 log 信息，把容器的详情信息输出到 flask.log 文件中，并查看端口占用情况，文件内容提交至“提交资料/2.2/试题_玩具市场数据标注”。(8分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|--------|--------|----|
| 1 | 01 | 6分 | 完成封装 | 同答案操作 | |
| 2 | 01 | 6分 | 完成输出操作 | 同答案操作 | |
| 3 | 01 | 8分 | 完成输出操作 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.3.1

试题名称：“玩具市场”-智能系统运维-智能系统维护

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发展具

有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(1) 用 python 遍历 document 文件夹，计算当前任务文件夹的总大小和文件数量。内容提交至“提交资料/3.1/试题_玩具数据智能系统运维”中。(5分)

(2) 获得计算机的 IP 地址，并通过 IP 地址登录 MYSQL，查看当前连接到 MYSQL 服务器的用户。内容提交至“提交资料/3.1/试题_玩具数据智能系统运维”中。

(5分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|----------------------|--------|----|
| 1 | 01 | 5分 | 计算当前任务文件夹的总大小和文件数量 | 同答案操作 | |
| 2 | 01 | 5分 | 查看当前连接到 MYSQL 服务器的用户 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：5.3.2

试题名称：“玩具市场”-智能系统运维-智能系统优化

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）玩具数据分析可以帮助玩具公司更好地了解消费者需求和市场趋势，从而制定更有效的市场策略和产品开发计划。具体来说，玩具数据分析能够帮助实现以下目标：

确定市场趋势：通过对市场数据的统计分析，了解消费者的兴趣、需求以及行为模式，及时发现市场趋势并响应。通过分析消费者的购买数据，发现促销活动的效果、产品热销时间、促销力度等，从而调整销售策略，提高销售额。

优化产品开发：通过对消费者的反馈数据进行分析，了解用户对产品的需求和评价，从而调整产品设计，优化产品体验，提升市场竞争力。

降低风险：玩具行业周期较短，产品更新换代较快。通过对市场数据的分析，及时了解产品失利原因，找到企业的短板并及早调整。

因此，玩具数据分析对于玩具公司在市场竞争中立于不败之地和实现持续发

展具有重要意义。

(2) 你将作为一名人工智能数据分析师对“玩具市场”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(3) 根据当前数据经历，请给出优化建议，内容提交至“提交资料/3.2/试题_玩具数据智能系统运维.docx”。(10分)

4、质量指标

主观评分表/过程评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|----|---|-------|---|---|----|
| | | | | 1 | 2 | 3 | |
| 1 | S1 | 6分 | 缺少“具体市场侧分析”和“对应的风险预警提醒数据”。以及其他合理建议均可,不少于三条。 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.1.1

试题名称：城市天气-数据采集和处理-业务数据质量检测

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python ；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：

1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。

2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。

3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分

析审核报告。数据采集和处理：根据业务特征与业务需求，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

（1）国际数据管理协会英国分会关于数据质量提出了六个核心纬度，请根据其中的两个纬度审核修改已采集的数据，并写出修改所参考的纬度。内容截图提交至“提交资料/1.1/试题_天气数据采集和处理”文件中。（4分）

（2）根据此次采集的内容，请梳理对应的数据采集规范，补充填写试题中的表格。内容提交至“提交资料/1.1/试题_天气数据采集和处理”文件中。（4分）

（3）面对复杂的数据我们需要进行规范的数据处理，根据数据的属性或特征，按照一定的原则和方法进行整理、归类、汇总。请整理下述天气信息数据表，对数据进行格式的规范并按照时间将数据进行归类排序，再新建 sheet 根据时间将数据整合为不同的表格，表格名命名为时间段。内容截图提交至“提交资料/1.1/试题_天气数据采集和处理”文件中。（8分）

4、质量指标

客观评分表/结果评分表

| 题目号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|-----|----|----|------------------------------------|---------|----|
| 1 | 01 | 2分 | 数据修改是否正确，需要修改内容为：“删除降水量一列”和“删除℃符号” | 两个操作各1分 | |
| | 01 | 2分 | 修改对应的原则是否正确 | 两个原则各1分 | |
| 2 | 01 | 1分 | 出现“天气”两个字即可 | 参考答案结果 | |
| | 02 | 6分 | 每种内容各对应一分 | 参考答案结果 | |
| | 03 | 1分 | 出现“整数类型（整型）”或“文本类型”均可 | 参考答案结果 | |
| 3 | 01 | 4分 | 将风力风向分为两列 | 参考答案结果 | |
| | 02 | 4分 | 按照题目要求对数据进行归类排序 | 参考答案结果 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.1.2

试题名称：城市天气-数据采集和处理-数据处理方法优化

考核时间：30min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言:Python ；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：

1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。

2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。

3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

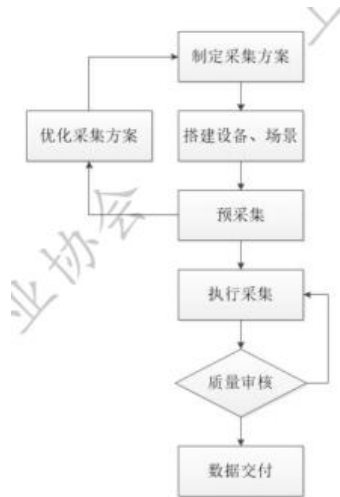
（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分

析审核报告。数据采集和处理：根据业务特征与业务需求，梳理业务数据采集和处理规范，对业务采集流程和处理流程提出优化建议。

3、技能要求

打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 参考以下“数据采集流程-基本框架图”。



制作一个关于天气数据采集的详细流程图；将该数据导出成 jpg 图片格式，命名为“关于天气数据采集-详细流程图”。内容提交至“提交资料/1.2/试题_天气数据采集和处理”文件中。（5分）

(2) 根据(1)中的流程图制作一个关于天气数据处理的基本框架图；将该数据导出成 jpg 图片格式，命名为“关于天气数据处理-基本框架图”。建立一个新的 word 文档，命名为“对处理流程的优化建议”；将该图片“关于天气数据处理-基本框架图”复制到此文档中；并写出自己对处理流程的优化建议。内容提交至“提交资料/1.2/试题_天气数据采集和处理”文件中。（5分）

4、质量指标

主观评分表/过程评分表

| 题目号 | 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|-----|----|----|-------------|-------|---|---|----|
| | | | | 1 | 2 | 3 | |
| 1 | S1 | 2分 | 包含题目中的全部步骤 | | | | |
| | S2 | 1分 | 包含“1”步骤 | | | | |
| | S3 | 2分 | 包含“2”步骤 | | | | |
| 2 | S1 | 2分 | 合理的流程即可 | | | | |
| | S2 | 3分 | 提出合理的两点优化即可 | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.2.1

试题名称：“城市天气”-数据标注-数据归类和定义

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分析审核报告。数据标注：完成对标注数据准确性和完整性审核，利用各项软件完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。

根据业务目标，基于工具完成天气分析，并可以对数据进行归类和汇总，满足分析和处理的要求。

打开项目中的数据分析（试题）文件，考生在规定的时间内完成以下任务：

（1）数据文件收集了一段时间内的天气信息，包括天气摘要、气温、表面温度、湿度、风速等信息，请读取文件，删除掉有缺失值、重复值和异常值的行，并绘制相关性矩阵。内容提交至“提交资料/2.1/试题_天气数据标注.docx”文件中。

（15分）

（2）请以年为单位利用折线图可视化 2006 年的以下数据：温度和湿度之间的 Pearson 相关系数、温度和风速之间的 Spearman 相关系数、温度和气压之间的 Kendall 相关系数。内容提交至“提交资料/2.1/试题_天气数据标注.docx”文件中。（15分）

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-----|-----------------------------------|--------|----|
| 1 | 01 | 10分 | 编写代码，成功删除掉有缺失值、重复值和异常值的行 | 同答案操作 | |
| | 02 | 15分 | 编写代码，成功删除掉有缺失值、重复值和异常值的行，并绘制相关性矩阵 | 同答案操作 | |
| 2 | 01 | 15分 | 完成折线图 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.2.2

试题名称：“城市天气”-数据标注-标注数据审核

考核时间：45min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分析审核报告。数据标注：完成对标注数据准确性和完整性审核，利用各项软件完成数据内在关联和特征进行数据归类。

3、技能要求

经过数据预处理，对数据按月进行分析和对用户个体消费数据进行分析，得出分析报告。

根据业务目标，基于工具完成天气分析，并可以对数据进行归类和汇总，满足分析和处理的要求。

打开项目中的数据分析（试题）文件，考生在规定的时间内完成以下任务：

（1）绘制一个行为天气摘要、列为年份的表格，统计每年每种天气摘要各有多少天。内容提交至“提交资料/2.2/试题_天气数据标注.docx”文件中。（6分）

（2）绘制从2006年1月1日起至2007年1月1日为止的以星期为单位的温度箱线图。内容提交至“提交资料/2.2/试题_天气数据标注.docx”文件中。（6分）

（3）绘制2007年天气摘要为Partly Cloudy且湿度大于20%的日子的温度曲线图。内容提交至“提交资料/2.2/试题_天气数据标注.docx”文件中。（8分）

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|---------|--------|----|
| 1 | 01 | 6分 | 完成绘制表格 | 同答案操作 | |
| 2 | 01 | 6分 | 完成绘制箱线图 | 同答案操作 | |
| 3 | 01 | 8分 | 完成绘制曲线图 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.3.1

试题名称：“城市天气”-智能系统运维-智能系统维护

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(1) 在 MYSQL 新建一个名为” label” 的数据库。相关报告内容提交至 “提交资料/3.1/试题_天气数据智能系统运维.docx”。(5分)

(2) 创建一个名为”user1”的用户并赋予数据库”label”的可读权限。内容提交至 “提交资料/3.1/试题_天气数据智能系统运维.docx”。(5分)

4、质量指标

客观评分表/结果评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|----|--------------|--------|----|
| 1 | 01 | 5分 | 完成创建，并提交相应代码 | 同答案操作 | |
| 2 | 01 | 5分 | 完成创建，并提交相应代码 | 同答案操作 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：6.3.2

试题名称：“城市天气”-智能系统运维-智能系统优化

考核时间：15min

1、场地设备要求

（1）CPU：建议英特尔 i5 以上；

内存：建议不少于 16GB；

操作系统：Windows10；

开发语言：Python；

开发工具：Python 编译工具；

深度学习训练环境：支持主流深度学习框架训练；

标注工具：图像文本标注工具；

（2）配件：鼠标、键盘、麦克风及显示器等配件；

2、工作任务

（1）天气数据分析可以通过对气象数据的收集、处理和分析来获得天气变化的趋势和规律。天气数据分析可以用于识别天气模式、预测气候变化、评估气象风险等方面，对于决策制定、气象预警、环境管理等方面具有重要意义。主要通过以下三个步骤进行：1. 数据收集。这一步骤是收集和整理各地的气象数据，并将其推送到云端进行存储。2. 数据处理。这一步骤是使用大数据技术对数据进行处理和分析，比如可以使用数据挖掘技术对数据进行可视化处理，以便更好地理解和分析数据。3. 数据应用。基于数据处理的结果，可以对天气进行预测和分析，从而可以进行监测和预警，为人们的生活和对物质资源的管理提供参考意见。

（2）你将作为一名人工智能数据分析师对“城市天气”数据进行分析，输出分析审核报告。智能系统运维：能够维护业务所需要的数据，利用分析工具进行数据分析，输出分析报告。

3、技能要求

(1)根据已有文件数据,给出本次数据的优化建议。内容提交至“提交资料/3.2/试题_天气数据智能系统运维.docx”。(10分)

4、质量指标

主观评分表/过程评分表

| 题号 | 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|-----|------------|-------|---|---|----|
| | | | | 1 | 2 | 3 | |
| 1 | S1 | 3分 | 写出1个合理优化建议 | | | | |
| | S2 | 6分 | 写出2个合理优化建议 | | | | |
| | S3 | 10分 | 写出3个合理优化建议 | | | | |

人工智能训练师（四级）操作技能考核 试题单

准考证号：

试题代码：7.1.1

试题名称：业务数据质量检测

考核时间：40min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|----|----|----|----|
| 1 | 人工智能训练师主机 | | 台 | 1 | |

2、工作任务

（1）业务背景

梳理业务数据采集规范是一个重要的步骤，因为它有助于确保数据的准确性和一致性。在现代商业环境中，大量的数据被收集和存储，包括客户信息、销售记录、库存水平等等。然而，这些数据可能来自于不同的来源，并且可能存在格式不统一、缺失值、错误等现象。因此，为了确保数据的可靠性和可用性，需要对数据采集过程进行规范化，以确保数据的准确性和一致性。

（2）业务目标

梳理业务数据采集规范的目标是确保所有数据收集活动都符合公司的政策和标准，从而提高数据的准确性、完整性和可靠性。

（3）考试内容

【1】在（考试）文件夹中查看法律法规数据（试题）.xlsx，对名称为“数据”的 sheet 页进行信息审核，若数据均符合要求，则不用进行任何操作；若其中有部分数据不符合要求，需要将不符合要求的数据进行删除，并将删除多余数据后的文档另存到桌面上，文件名称为**【修改后数据】**。

审核要求：爬取数据需要为法律法规类相关数据

【2】业务数据采集规范是指在业务运营过程中收集、存储、处理和利用的数据集合，包括数据收集、存储、处理、使用等方面的规定。

下面是一些可能包含在业务数据采集规范中的内容：

- 01) 数据源选择。
- 02) 数据格式要求。
- 03) 数据质量控制。
- 04) 数据安全保护。

- 05) 数据整合和清洗。
- 06) 数据可视化展示。
- 07) 数据更新频率。
- 08) 数据共享权限管理。

请针对以上几点，梳理出一份业务数据采集规范，并给出正确的答案。

3、技能要求

- (1) 能够对预处理后业务数据进行审核
- (2) 能够结合人工智能技术要求，梳理业务数据采集规范

4、质量指标

详情请见评分表

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|------------------|--|----|
| 01 | 4 | 答对即得 4 分 | 详情请见【修改后数据.xlsx】 | |
| 02 | 16 | 见主观评分表，答对一条得 2 分 | 1. 数据源选择：根据业务需求，选择合适的数据源进行采集。常见的数据源包括数据库、API、日志文件等。 2. 数据格式要求：明确数据的存储和传输格式，例如数据的字段名、数据类型、数据长度等。 3. 数据质量控制：在数据采集过程中，对数据的质量进行控制和筛选，确保数据的准确性和完整性。 4. 数据安全保护：考虑数据的安全性，防止数据泄露、篡改等问题。在采集数据时需要采取相应的措施，如加密传输、访问控制等。 5. 数据整合和清洗：在采集到数据后，需要进行整合和清洗工作，以确保数据的正确性和一致性。 6. 数据可视化展示：将采 | |

| | | | | |
|----|--|--|---|--|
| | | | 集到的数据进行可视化展示，方便用户查看和分析数据。 7. 数据更新频率：确定数据更新的频率，以保证数据的及时性和准确性。 8. 数据共享权限管理：对数据进行权限管理，确保不同人员只能访问自己需要的数据。 | |
| 03 | | | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|-------------------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| S1 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S2 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S3 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S4 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S5 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S6 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S7 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |
| S8 | 2 | 逻辑正确，确实写出了相关规范即得分 | | | | |

人工智能训练师（四级）操作技能考核 试题单

准考证号：

试题代码：7.1.2

试题名称：数据处理方法优化

考核时间：20min

1、场地设备要求

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|----|----|----|----|
| 1 | 人工智能训练师主机 | | 台 | 1 | |

2、工作任务

（1）业务背景

在现代商业环境中，企业需要收集和分析大量的数据来支持业务决策和运营。然而，业务数据的采集通常是一项耗时且繁琐的任务，可能会导致数据质量不佳、数据一致性不强等问题。我们可以通过优化业务数据采集流程来提高数据质量和效率

（2）考试内容

假设你正在开发一个用于预测股票价格的人工智能系统。为了收集足够的训练数据，你需要从多个来源获取市场信息、公司财务报告和其他相关数据。

请从以下几点提出自己对数据采集流程的建议

- 01) 数据采集
- 02) 自动化工具
- 03) 采集频率
- 04) 数据质量
- 05) 数据可视化

3、技能要求

能够对业务数据采集流程提出优化建议

4、质量指标

详情请见评分表

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-------------------|------------------------------------|----|
| 01 | 10 | 答对一条及得2分，全部答对得10分 | 1. 制定清晰的数据采集标准：在开始数据收集之前，确保已经确定了所有 | |

| | | | | |
|----|--|--|--|--|
| | | | <p>需要收集的数据类型和格式，以及每个数据源的可靠性和完整性。</p> <p>2. 利用自动化工具：使用自动化工具来自动化数据收集过程，例如爬虫或API调用器。这可以减少手动操作的时间和工作量，并提高数据的准确性和一致性。</p> <p>3. 确定最佳数据采集频率：根据你的数据需求和分析目标，确定最佳的数据采集频率。如果需要实时数据分析，可能需要更高的频率。同时，确保不超采样或漏采样任何关键数据。</p> <p>4. 确保数据质量：对于每种数据类型，都要进行必要的预处理和清洗，以确保数据的准确性和一致性。这包括去除错误数据、填充缺失数据和标准化数据格式等。</p> <p>5. 提高数据可视化和报告能力：通过使用数据可视化和报告技术来更清晰地展示数据，帮助管理层更好地理解业务状况。这可以通过使用数据仪表板和报表工具来实现。</p> | |
| 02 | | | | |
| 03 | | | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|---------------------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| S1 | 2 | 逻辑正确，确实写出了相关优化建议即得分 | | | | |
| S2 | 2 | 逻辑正确，确实写出了相关优化建议即得分 | | | | |

| | | | | | | |
|----|---|-------------------------|--|--|--|--|
| S3 | 2 | 逻辑正确，确实写出了 相关优化建议即得分 | | | | |
| S4 | 2 | 逻辑正确，确实写出了 相关优化建议即得分 | | | | |
| S5 | 2 | 逻辑正确，确实写出了 相关优化建议即得分 | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：7.2.1

试题名称：基于数据内在关联和特征对数据进行分类

考核时间：60min

1、场地设备要求

(1) 场地要求：指定考试场地

(2) 设备要求：

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|--|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16GB 操作系统：Windows10 | 台 | 1 | |

2、工作任务

(1) 业务背景

- 数据管理需求：**随着企业规模的不断扩大，数据量也变得越来越庞大。为了有效地管理和利用这些数据，需要对数据进行分类和归档。
- 数据可视化需求：**在数据可视化的过程中，可以将数据按照不同的维度进行分组展示，方便用户快速了解数据的整体趋势和关系。
- 业务分析需求：**在进行业务分析时，需要将数据按照一定的规则进行分类和分析，以便更好地理解业务情况并制定相应的决策。
- 数据标准化需求：**在数据标准化方面，需要对不同类型的数据进行分类，以便更好地实现数据的统一管理和处理。
- 安全管理需求：**在安全管理方面，需要对敏感数据进行分类和保护，防止未经授权的人员访问和泄露。

(2) 业务目标

数据归类是人工智能训练师日常使用的工作技能，你是一家科技公司的人工智能训练师，今日的工作任务是基于数据内在关联和特征对数据进行分类。

(3) 考试内容

假设我们有一个数据集，其中每个样本都有三个特征：发文单位、文件类型、发文时间。目标是对这些样本进行分类，可以将它们分为以下四个类别：

1. 证监会，属于部门规范性文件，2022 年之前发布的文件；
2. 深交所，属于行业规定，2020 年发布的文件；
3. 上交所，属于行业规定，2022 年发布的文件；
4. 银保监会，属于解释性公告及规范问答，2020 年之后发布的文件。

要求考生在给定的数据集中找到正确的类别，新建 4 个以发文单位为名的文件夹，把找到的正确类别的数据分别放到 4 个对应的文件夹里。

3、技能要求

- (1) 能够根据数据内在关联和特征进行数据归类
- (2) 能够根据数据内在关联和特征进行数据定义

4、质量指标

详见人工智能训练师（四级）操作技能考核试题评分表及答案

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-----------------|--------------|----|
| 01 | 8 | 对证监会类别文件进行完整归类 | 参考证监会类别分类文件 | |
| 02 | 7 | 对深交所类别文件进行完整归类 | 参考深交所类别分类文件 | |
| 03 | 8 | 对上交所类别文件进行完整归类 | 参考上交所类别分类文件 | |
| 04 | 7 | 对银保监会类别文件进行完整归类 | 参考银保监会类别分类文件 | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：7.2.2

试题名称：人工智能训练师数据审核能力考核

考核时间：30min

1、场地设备要求

(1) 场地要求：指定考试场地

(2) 设备要求：

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|--|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16GB 操作系统：Windows10 | 台 | 1 | |

2、工作任务

(1) 业务背景

数据标注业务是指将大量的未标注数据进行标注的过程。在人工智能和机器学习领域中，数据是构建模型和算法的基础，而标注数据是为了训练机器学习算法。因此，数据标注业务是人工智能和机器学习应用的重要环节。

在许多应用场景中，人工智能和机器学习需要处理大量的数据。例如，在自然语言处理领域中，机器学习需要对文本进行分类、情感分析、关键词提取等，这需要对大量的文本数据进行标注。在计算机视觉领域中，机器学习需要对图像、视频等进行分类、目标检测、语义分割等处理，也需要对大量的图像和视频进行标注。在自动驾驶领域中，机器学习需要对路况、车辆、行人等进行识别和预测，也需要对大量的场景进行标注。

总之，数据标注业务是实现人工智能和机器学习应用的必要步骤，也是改善算法

和模型性能的重要手段。

(2) 业务目标

数据标注业务的目的是为机器学习和人工智能算法提供高质量、准确和可靠的数据集，以便这些算法能够更好地识别、分类、预测和推荐。标注任务的目标包括以下方面：

1. 提高数据的质量：通过数据标注，可以提高数据的质量，例如准确性、一致性和完整性，使得机器学习算法的准确率更高。
2. 提高算法的效果：数据标注可以为机器学习和人工智能算法提供高质量的训练数据，因此算法的效果会更好。
3. 提高业务效率：数据标注可以同时处理大量的数据，有效地提高了数据处理的效率，提高了业务的效率。
4. 节省成本：机器学习和人工智能算法需要大量高质量的数据，数据标注可以帮助节省因建立训练数据而产生的成本。
5. 保护数据安全：数据标注业务通常会签署保密协议，确保数据的安全和保密，避免敏感数据外泄。

(3) 考试内容

数据标注和审核是人工智能训练师日常使用的工作技能，你是一家科技公司的人工智能训练师，今日的工作任务是完成通用合同的审核。

通用合同的定义如下：

通用合同是一种一般适用于多个领域、多种业务活动的标准化合同。通用合同通常包含一些基本的法律条款和条款，如合同的目的、双方的义务和责任、价格和支付方式、违约责任等等。通用合同可以减少合同的谈判和签署的时间和费用，提高合同的效率和质量，简化企业之间的商务交易。它也可以降低风险，保障双方权益。

(1) 要求标注的字段信息如下：

| 序号 | 字段名 |
|----|------|
| 1 | 甲方名称 |

| | |
|---|----------|
| 2 | 乙方名称 |
| 3 | 合同金额（小写） |
| 4 | 合同金额（大写） |
| 5 | 币种 |
| 6 | 签订时间 |
| 7 | 合同份数 |

(2) 文档中多次出现的字段信息均需标注

注：币种仅标注合同大小写金额所涉及到的相应字段信息

请在规定时间内完成如下操作：

- (1) 请对标注数据的完整性和准确性进行审核，输出审核报告
- (2) 能够对审核过程中发现的错误进行纠正

3、技能要求

- (1) 考生能熟练掌握数据清洗并理解标注要求和规范；
- (2) 考生能按照试题要求规范进行标注。

4、质量指标

详见评分表 2-2.1

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|---------------------|---------------|----|
| 01 | 3 | 字段标注完整且正确。(注：漏标不得分) | 详情请见【字段参考答案表】 | |
| 02 | 3 | 字段标注完整且正确。(注：漏标不得分) | 详情请见【字段参考答案表】 | |
| 03 | 3 | 字段标注完整且正确。 | 详情请见【字段参考答案表】 | |
| 04 | 3 | 字段标注完整且正确。 | 详情请见【字段参考答案表】 | |
| 05 | 3 | 字段标注完整且正确。(注：漏标不得分) | 详情请见【字段参考答案表】 | |
| 06 | 3 | 字段标注完整且正确。 | 详情请见【字段参考答案表】 | |
| 07 | 2 | 字段标注完整且正确。 | 详情请见【字段参考答案表】 | |

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|--------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| S1 | | | | | | |
| S2 | | | | | | |
| S3 | | | | | | |
| S4 | | | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：7.3.1

试题名称：人工智能训练师培训能力考核

考核时间：10min

1、场地设备要求

(1) 场地要求：指定考试场地

(2) 设备要求：

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|--|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16GB 操作系统：Windows10 | 台 | 1 | |

2、工作任务

(1) 业务背景

随着人工智能技术在各领域的应用不断深入，人工智能训练师这一职业不断发展壮大，作为一名四级/中级人工智能训练师，需要熟练使用各种工具完成数据采集、标注、训练和分析。

(2) 业务目标

您将作为一名四级/中级人工智能训练师。

(3) 考试内容

在 word 中简单描述智能系统部署步骤。

3、技能要求

(1) 能够将智能系统在真实应用场景中部署实施。

4、质量指标

见试题评分表。

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | 得分 |
|----|----|-------------------|------------------------|----|
| 01 | 1 | 包含“确定系统需求和功能”步骤 | 回答内容相似即可得分，不要求答案表述一模一样 | |
| 02 | 1 | 包含“选择合适的硬件设备”步骤 | 回答内容相似即可得分，不要求答案表述一模一样 | |
| 03 | 1 | 包含“安装操作系统和软件环境”步骤 | 回答内容相似即可得分，不要求答案表述一模一样 | |
| 04 | 1 | 包含“配置系统参数”步骤 | 回答内容相似即可得分，不要求答案表述一模一样 | |
| 05 | 1 | 包含“进行数据备份和恢复”步骤 | 回答内容相似即可得分，不要求答案表述一模一样 | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|--------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| S1 | | | | | | |
| S2 | | | | | | |
| S3 | | | | | | |
| S4 | | | | | | |

人工智能训练师（四级）操作技能考核

试题单

准考证号：

试题代码：7.3.2

试题名称：人工智能训练师培训能力考核

考核时间：20min

1、场地设备要求

(1) 场地要求：指定考试场地

(2) 设备要求：

| 序号 | 名称 | 规格 | 单位 | 数量 | 备注 |
|----|-----------|--|----|----|----|
| 1 | 人工智能训练师主机 | CPU：建议英特尔 i5 以上 内存：建议不少于 16GB 操作系统：Windows10 | 台 | 1 | |

2、工作任务

(1) 业务背景

随着人工智能技术在各领域的应用不断深入，人工智能训练师这一职业不断发展壮大，作为一名四级/中级人工智能训练师，需要熟练使用各种工具完成数据采集、标注、训练和分析。

(2) 业务目标

您将作为一名四级/中级人工智能训练师，需要利用分析工具进行数据分析，输出分析报告。

(3) 考试内容

在桌面上有一个名为“系统日志.xlsx”的文件，该文件记录了网站的访问日志。请完成以下操作：

(1) 统计每种事件（event_id）发生的次数和占比。

(2) 统计每个按钮（button_id）被点击的次数和占比。

(3) 统计每个用户 (user_id) 进行搜索的次数和占比。

(4) 找到使用最频繁的版本号 (app_version)。

(5) 将以上四个要求的答案写入 EXCEL，并保存到桌面，命名为“答案.xlsx”。

3、技能要求

(1) 能够利用分析工具进行数据分析，输出分析报告。

4、质量指标

客观评分表/结果评分表

| 序号 | 配分 | 评分细则描述 | 规定或标称值 | | | | 得分 |
|----|----|--------|-----------------------|-----------------------|-----|-----|----|
| 01 | 1 | 答对得4分 | event_id=click_button | 60 | 60% | | |
| | | | event_id=search | 40 | 40% | | |
| 02 | 1 | 答对得5分 | user_id=12345 | 39 | 39% | | |
| | | | user_id=23456 | 31 | 31% | | |
| | | | user_id=34567 | 30 | 39% | | |
| 03 | 1 | 答对得5分 | user_id=12345 | 合计 | 39 | | |
| | | | | event_id=click_button | 25 | 46% | |
| | | | | event_id=search | 14 | 36% | |
| | | | user_id=23456 | 合计 | 31 | | |
| | | | | event_id=click_button | 17 | 55% | |
| | | | | event_id=search | 14 | 45% | |
| | | | user_id=34567 | 合计 | 30 | | |
| | | | | event_id=click_button | 18 | 60% | |
| | | | | event_id=search | 12 | 40% | |
| 04 | 1 | 答对得5分 | app_version=1.0.1 | | | | |

| | | | | | | | | | |
|----|---|-----------|-----------------|-------------------|----|-----|--|--|--|
| | | | | app_version=1.0.0 | 42 | 42% | | | |
| | | | | app_version=1.0.1 | 58 | 58% | | | |
| 05 | 1 | 答对得 1分 | 详情请见【系统日志.xlsx】 | | | | | | |

注：如有分解为子项目评分表的，请分列客观评分表，并注明子评分表名称。

主观评分表/过程评分表

| 序号 | 配分 | 评分细则描述 | 考评员评分 | | | 得分 |
|----|----|--------|-------|---|---|----|
| | | | 1 | 2 | 3 | |
| S1 | | | | | | |
| S2 | | | | | | |
| S3 | | | | | | |
| S4 | | | | | | |