

第 4 部分

操作技能复习题

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：1.1.1

试题名称：客服会话业务数据采集

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、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，将对客服会话数据进行采集。

(1) 打开文件夹中的程序文件“1.1.1_data_collection.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。程序实现将 txt 文件中的会话数据采集出来并梳理成标准格式，并提交梳理好的数据。（5 分）

- ① 导入 pandas，打开空格分隔的“1.1.1_rawdata.txt”文件。
- ② 删除包含缺失值的行。
- ③ 参考“dataSample01.csv”，另存为“dataCollected01.csv”文件。

(2) 打开文件夹中的程序文件“1.1.2_data_collection.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。程序实现从 SQLite

数据库中采集会话数据并梳理成标准格式，请提交代码及梳理好的数据。（10分）

- ① 导入 sqlite3, pandas 库，连接到现有数据库 exam.db。
- ② 开启游标功能，创建游标对象。
- ③ 使用 execute() 方法，执行 sql 语句，按以下要求取出数据：日期为 20230501；指定 dt, store_id, assistant_id, buyer_id, direction, message, message_time 字段；message 不为空。
- ④ 参考 dataSample01.csv 文件，将数据保存为 “dataCollected02.csv”。
- ⑤ 关闭数据库。

数据库表 schema:

字段	数据类型	字段说明
uuid	TEXT	message 唯一标识符
dt	TEXT	日期
store_id	TEXT	店铺 ID
assistant_id	TEXT	客服 id
buyer_id	TEXT	买家 id
direction	TEXT	消息方向，BUYER_TO_ASSISTANT 为买家咨询问题，ASSISTANT_TO_BUYER 为客服回复内容
message	TEXT	消息内容
message_time	TEXT	消息发生时间
message_type	TEXT	消息类型

3、技能要求

- (1) 能够利用设备、工具等完成原始业务数据采集
- (2) 能够完成数据库内业务数据采集

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)	5 分	正确补全代码，每个空格 1 分。		
(2)	10 分	正确补全代码，每个空格 2 分。		

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：1.1.2

试题名称：客服会话业务数据处理

考核时间：30 min

1、场地设备要求

- (2) 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

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，将进行客服会话数据的采集。

(1) 打开文件夹中的程序文件“1.2.1_data_split.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。程序实现将梳理好的数据合并为一份并按照店铺进行归类保存并提交。（5分）

① 导入 pandas，读取 answer 文件夹中的 dataCollected01.csv 和 dataCollected02.csv。

② 上下拼接两个 DataFrame，合并为一个 DataFrame 并保存。

③ 按照店铺拆分合并后的 DataFrame，以<店铺 ID>.csv 格式命名并保存。

(3) 请参考“dataSample02.csv”文件，对 message 为买家咨询问题的数据进行汇总排序，并提交文件及汇总报告 dataSorted.xlsx。（10分）

3、技能要求

- (1) 能够根据数据处理要求完成业务数据整理归类
- (2) 能够根据数据处理要求完成业务数据汇总

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)	5分	正确补全代码，每个空格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。

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	
(2)	10分	按照 store_id, message 进行分类汇总，总结某一店铺 topN 咨询内容：+8 分 汇总时限定 direction 为 buyer_to_assistant：+2 分 加总得分在 0 到 10 分之间				

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：1.2.2

试题名称：情感识别标注后数据分类与统计

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、工作任务

对话中的情感识别（Emotion Recognition in Conversation, ERC）应用广泛，可辅助智能客服机器人识别顾客情感进行相应的动作策略选择，可辅助企业洞察顾客的满意度，可质检客服对话中的违规行为进行相应的规范培训处理，可用于教育培训理解学生的学习状态与需求辅助提升教学效果，可帮助心理医生和研究人员了解患者情感状态帮助诊断治疗精神疾病。本项目中，你作为一名人工智能训练师，将进行对话数据的清洗与标注，为对话中的情感识别建模提供训练及测试验证数据。

此次标注需标注粗粒度情感标签和细粒度情感标签，具体的标签说明如下：

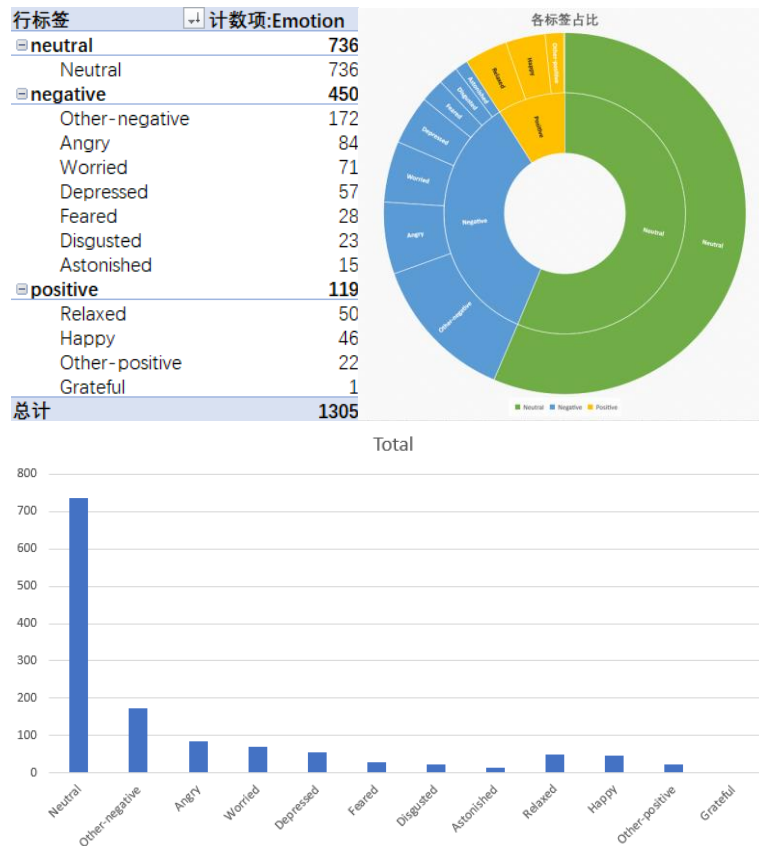
粗粒度情感标签：

标签名	说明	示例
Positive	正面情感	今天的晚餐很好吃。
Neutral	中性情感	我不确定是否应该接受这份工作。
Negative	负面情感	他的那番话让我非常不舒服。

细粒度情感标签：

标签名	说明	示例
Happy	高兴的：表示开心、愉悦或快乐的情感状态	我很高兴能够和你聊天。
Grateful	感激的：表示感激或感谢的情感状态	我非常感激你的帮助。
Relaxed	放松的：表示轻松、平静或舒适的情感状态。	我感觉很放松，可以好好休息一下。
Other-positive	其他积极情感：表示积极的情感状态，但无法明确归类到以上三个情感标签中	当我完成这个任务时，我感到很满足。
Neutral	中性的：表示没有特定情感的状态	这件事情还没有确定下来。
Angry	愤怒的：表示愤怒、生气或激愤的情感状态	他一直在干扰我的工作，让我感到非常愤怒。
Sad	悲伤的：表示悲伤、难过、不开心的情感状态	我感到非常难过，因为我失去了我的宠物。
Feared	害怕的：表示害怕、恐惧、惊吓的情感状态	走在黑暗的小路上让我感到很害怕。
Depressed	沮丧的：表示沮丧、消沉或压抑的情感状态	最近我总是感到精神抑郁，不知道该怎么办。
Disgusted	厌恶的：表示厌恶、反感或感到心里不舒服、恶心的情感状态	我感到非常反感，因为他的行为很不礼貌。
Astonished	震惊的：表示惊讶、震惊、或吃惊的情感状态	我感到很震惊，因为他的话太过出人意料。
Worried	担忧的：表示担忧、忧虑或焦虑的情感状态	我担心自己不能如期完成任务，会给团队带来麻烦。
Other-negative	其他消极情感：表示消极的情感状态，但无法明确归类到以上七个负面情感标签中	我感到很疑惑，因为我不知道该怎么做。

(1) 请对标注后的数据(“dataLabelled.xlsx”)进行分类统计，使用 excel 数据透视表功能，参照下图生成透视图，包括各类标签占比(旭日图)及数量(柱状图)统计。(20分)



3、技能要求

- (1) 能够利用分类工具，对标注后数据进行分类
- (2) 能够利用统计工具，对标注后数据进行统计

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)	20分	生成格式正确的透视表: +4分 生成格式正确的旭日图: +8分 生成格式正确的柱状图: +8分 加总得分在0到20分之间				

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：1.3.1

试题名称：客户体验管理系统基础操作

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、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基本运维，记录系统的应用和数据情况。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）请浏览文件“CEM 系统功能.pdf”，描述如何开启“撤回消息”扣分及设置扣分规则为-1 分。（2 分）

- （2）请浏览文件“CEM 会话列表.pdf”中的每一页会话截图，模拟在 CEM 产品中挑选一条会话进行复检，并挑选一条包含质检标签的客服回答，编辑其质检标签，以上步骤请用文字及截图（可从会话列表、系统功能及操作手册中截图）描述。（3 分）

3、技能要求

- （1）能够进行智能系统开启
- （2）能够简单使用智能系统

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	正确描述操作流程: +2 错误描述不扣分 得分为2分或0分				
(2)	3分	准确描述操作步骤: +2分 正确提供相关截图: +1分 错误描述不扣分 加总得分在0到3分之间				

人工智能训练师（五级）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、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基本运维，记录系统的应用和数据情况。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) 请浏览文件“CEM 系统功能.pdf”中的每一页截图，记录该系统功能应用情况。（2分）

- (2) 请浏览文件“CEM 应用数据.pdf”中的每一页截图，记录该系统应用数据情况。（3分）

3、技能要求

- (1) 能够记录智能系统功能应用情况
- (2) 能够记录智能系统应用数据情况

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	正确描述至少三项功能：+2 分 错误描述不扣分 得分为 2 分或 0 分				
(2)	3 分	正确描述至少两项规则开启使用数量：+2 分 正确描述至少一项规则扣分分值：+1 分 错误描述不扣分 加总得分在 0 到 3 分之间				

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：2.1.1

试题名称：文本校对系统-业务数据采集

考核时间：30 min

1、场地设备要求

(1)

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

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，具体考核要求：

(1) 读取 data 下的 txt 文件。读取后截图提交至“提交资料/试题.docx”文件中。

(2) 读取 data 下的 xlsx 文件。读取后截图提交至“提交资料/试题.docx”文件中。

3、技能要求

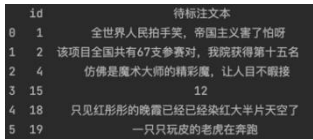
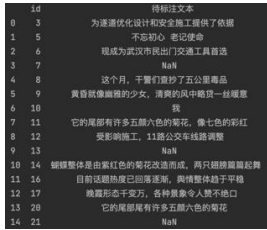
(1) 能够利用设备、工具等完成原始业务数据和数据库内业务数据的采集。

4、质量指标

(1) 能够利用工具顺利读取 data 下的 txt 文件

(2) 能够利用工具顺利读取 data 下的 xlsx 文件

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	5分	能够顺利读取 data 下的 txt 文件	真实数据读取结果如下： 	
02	5分	能够顺利读取工具读取 data 下的 xlsx 文件	真实数据读取结果如下： 	

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（五级）操作技能考核

试题单

准考证号：

试题代码：2.1.2

试题名称：文本校对系统-业务数据处理

考核时间：30 min

1、场地设备要求

(1)

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

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，具体考核要求：

(1) 将读取的 txt 文件和 xlsx 文件内容进行合并。合并后截图提交至“提交资料/试题.docx”文件中。

(2) 对合并后的内容，只保留“待标注文本”列。截图提交至“提交资料/试题.docx”文件中。

3、技能要求

(1) 能够根据数据处理要求完成业务数据整理归类和汇总。

4、质量指标

(1) 能够顺利合并读取的 txt 文件和 xlsx 文件内容。

(2) 能够顺利筛选数据，仅保留所需的“待标注文本”列。

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	能够顺利合并读取的 txt 文件和 xlsx 文件内容	真实数据合并后结果如下： 	
02	10 分	能够顺利筛选数据，仅保留所需的“待标注文本”列	真实数据筛选后结果如下： 	

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（五级）操作技能考核

试题单

准考证号：

试题代码：2.2.1

试题名称：文本校对系统-原始数据清洗与标注

考核时间：70 min

1、场地设备要求

(1)

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

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，具体考核要求：

(1) 查看业务数据，判断数据中是否存在缺失值和文本长度小于 3 的数据，将这两类数据去除。截图提交至“提交资料/试题.docx”文件中。

(2) 查看业务数据，添加“正确文本”列，并对“待标注文本”列的数据进行标注，将每个待标注文本对应正确文本写入“正确文本”列。截图提交至“提交资料/试题.docx”文件中。

(3) 查看业务数据，添加“错误类型”列，并对“待标注文本”列的数据进行标注，将每个待标注文本对应错误类型写入“错误类型”列（错误类型仅从[错字错误、多字少字错误、乱序错误]中进行选择）。截图提交至“提交资料/试题.docx”文件中。

3、技能要求

(1) 能够根据标注规范和要求，完成对文本、视觉、语音数据的清洗和标注。

4、质量指标

(1) 能够顺利去除数据中的缺失值和文本长度小于 3 的数据。

(2) 能够将错误句正确标注为正确句。

(3) 能够正确标注出错误句的错误类型。

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	能够顺利去除数据中的缺失值和文本长度小于 3 的数据	去除缺失值和文本长度小于 3 的后的数据结果如下：	

			<p style="text-align: center;">待标注文本</p> <p>0 全世界人民拍手笑，帝国主义害了怕呀</p> <p>1 该项目全国共有67支参赛队，我院获得第十五名</p> <p>2 仿佛是魔术大师的精彩魔，让人目不暇接</p> <p>4 只见红彤彤的晚霞已经染红大半片天空了</p> <p>5 一只只顽皮的老虎在奔跑</p> <p>0 为隧道优化设计和安全施工提供了依据</p> <p>1 不忘初心 牢记使命</p> <p>2 现成为武汉市民出门交通工具首选</p> <p>4 这个月，干警们查抄了五公里毒品</p> <p>5 黄昏就像幽雅的少女，清爽的风中略带一丝暖意</p> <p>7 它的尾部有许多五颜六色的菊花，像七色的彩虹</p> <p>8 受影响施工，11路公交车线路调整</p> <p>10 蝴蝶整体是由紫红色的菊花改造而成，两只翅膀翩翩起舞</p> <p>11 目前话题热度已回落逐渐，舆情整体趋于平稳</p> <p>12 晚霞形态千变万，各种景象令人赞不绝口</p> <p>13 它的尾部尾有许多五颜六色的菊花</p>
02	20分	能够将错误句正确标注为正确句	<p>标注后的结果如下：</p> 
03	20分	能够正确标注出错误句的错误类型	<p>标注后的结果如下：</p> 

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（五级）操作技能考核

试题单

准考证号：

试题代码：2.2.2

试题名称：文本校对系统-标注后数据分类与统计

考核时间：20 min

1、场地设备要求

(1)

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

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，具体考核要求：

(1) 将相同错误类型的数据进行分组。截图提交至“提交资料/试题.docx”文件中。

3、技能要求

(1) 能够利用分类和统计工具，对标注后的数据进行分类和统计。

4、质量指标

(1) 能够将相同错误类型的数据分组。

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	能够将相同错误类型的数据分组	<p>分组后的结果参考答案如下：</p> 	

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（五级）操作技能考核

试题单

准考证号：

试题代码：2.3.1

试题名称：文本校对系统-智能系统基础操作

考核时间： 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

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) 打开智能标注系统并成功登录。

(2) 在智能标注系统当中查找最近一个月时间字词错误，语法错误，语序错误，三种错误类型的数据分布情况，并导出相关统计结果，统计结果包括数据标注量，标注时间，各个错误类型。

(3) 在智能标注系统中点击数据管理->数据标注，筛选出未标注数据，对人工智能模型标注的文本校对数据进行正误判断，如果模型标注正确，则人工标注该数据为正报，反之则标为误报。

3、技能要求

(1) 能够进行智能系统开启

(2) 能够简单使用智能系统

4、质量指标

(1) 能够正常访问并登录智能标注系统得 2 分。

(2) 能够在智能标注系统查找业务数据并且成功导出数据得 4 分。

(3) 能够在智能标注系统完成数据标注工作，并且正确标注数据得 4 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
01	2	能够正常访问并登录智能标注系统		
02	4	能够在智能标注系统查找业务数据并且成功导出数据		
03	4	能够在智能标注系统完成数据标注工作，并且正确标注数据		

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						

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：2.3.1

试题名称：文本校对系统-智能系统维护

考核时间： 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

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) 打开智能标注系统并以管理员身份成功登录，将帮助用户名为 user2 的用户密码修改为 mima123。

(2) 测试智能标注系统各个功能是否正常运转，测试内容包括：查看数据，筛选数据，标注数据，导出数据。

(3) 查看已经被人工标注过的数据，并复核，对误报数据进行驳回（取消接受）。

3、技能要求

(1) 能够记录智能系统功能应用情况

(2) 能够纪律智能系统应用数据情况

4、质量指标

(1) 能够正常登录智能标注系统，并能够修改密码，维护账号安全得 2 分。

(2) 能够在智能标注系统对系统功能进行测试并总结得 4 分。

(3) 能够在智能标注系统对人工误标的数据进行驳回得 4 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
01	2	能够正常登录智能标注系统，并能够修改密码，维护账号安全得 3 分。	2	2
02	4	能够在智能标注系统对系统功能进行测试并总结	4	4
03	4	能够在智能标注系统对各个普通用户的操作进行统计和导出	4	4

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						

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.1.1

试题名称：“古诗学习”-数据采集和处理-业务数据采集

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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，能够对数据库内的业务数据进行数据采集，梳理业务数据采集和处理规范，整理并归纳数据，做好数据分类。

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) 实时采集数据的过程中需要先安装 Pyaudio 并导入库。请完成这个操作，将操作后的内容截图提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。(5分)

(2) 安装完成后，使用 PyAudio 打开麦克风并设置参数，参数分别是：采样深度为 16、声道数为 2、采样率为 16000、模式为输入、采样点的缓存数量为 2048，再使用 PyAudio 库的 read () 方法从麦克风读取语音数据，请完成这个操作并将操作后的内容截图提交内容提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。(10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	3分	完成Pyaudio安装操作	同答案操作	
	02	2分	导入Pyaudio库	同答案操作	
2	01	6分	参数设置对应正确	参数赋值对象严格按照答案对应，顺序无要求	
	02	4分	可以正确使用read函数读取	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.1.2

试题名称：古诗学习-数据采集和处理-业务数据处理

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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，能够对数据库内的业务数据进行数据采集，梳理业务数据采集和处理规范，整理并归纳数据，做好数据分类。

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）请打开未处理的采集文件，将其中的图片文件和音频文件先进行归类，再将音频分为有效音频和无效音频，截取文件路径图片和内容提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。（5分）

（2）打开提交资料/1.2/试题_数据采集和处理.excel表格，通过查看文件夹的属性，可获知此文件夹下有多少子文件夹和文件等信息，并汇总图片、有效音频和无效音频的数量。内容截图提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。（10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	将数据正确归类为两个文件夹：图片和音频	同答案截图，具体格式无要求	
	02	3分	将音频正确归类为两个文件夹：有效音频和无效音频	同答案截图，具体格式无要求	
2	01	10分	正确汇总和整理各文件夹信息	同答案，统计数据正确即可，格式无要求	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.2.1

试题名称：“古诗学习”-数据标注-原始数据清洗与标注

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：Praat 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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据标注工作。
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 Praat 完成数据的标注和处理。

3、技能要求

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

（1）在“语音采集数据”文件中对已采集的语音数据进行筛选，将有其他语言和卡顿的语音数据筛选掉，完成语音清洗，建立新的文件夹。文件夹名为“正式数据”。将筛选好的数据“正式数据”提交至“提交资料/2.1/试题_数据标注”文件夹中。（15分）

（2）请打开 Praat 软件，使用该软件对之前筛选好的语音数据进行裁剪和转写的标注工作，并保存标注文件，文件名命名为“原名.TextGrid”。并都存储到文件夹“导出文件”中。将文件夹“导出文件”提交至“提交资料/2.1/试题_数据标注”文件夹中。（25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	将有其他语言和卡顿的语音数据筛选掉，完成语音清洗，建立新的文件夹。	同答案操作	
2	01	10分	使用 Praat 软件对之前筛选好的语音数据进行裁剪和转写的标注工作，并保存标注文件。	同答案操作，完成并导出正确的9-13个数据文件	
	02	15分		同答案操作，完成并导出正确的20个数据文件	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.2.2

试题名称：“古诗学习”-数据标注-标注后数据分类与统计

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：Praat 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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据标注工作。
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 Praat 完成数据的标注和处理。

3、技能要求

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

（1）请打开 Excel 软件，新建表格，文件名命名为“分类与统计.xlsx”。将“导出文件.TextGrid”中的 txt 内容导入到该表格中，并按照语音内容主题进行分类。主要分成五言诗、七言诗、含“春”的诗句、含“秋”的诗句、含“风”的诗句、含“雪”的诗句、含“花”的诗句、含“山”的诗句、含“草”的诗句和涉及动物的诗句，这十类。（10分）

（2）使用 Excel 软件，根据刚刚分好类别的数据信息，创建饼状图进行统计分析。并将“分类与统计.xlsx”文件提交到“提交资料/2.2/试题_数据标注”文件夹中。（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	将“导出文件.TextGrid”中的 txt 内容导入到该表格中，并按照语音内容主题完成分类。	同答案操作	
2	01	2分	根据分类创建出了其他表格类型。	同答案操作	
	02	8分	根据分好类别的数据信息，创建饼状图完成统计分析。	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.3.1

试题名称：“古诗学习”-智能系统运维-智能系统基础操作

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 开发工具；

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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据标注工作。
智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

（1）打开 Python，记录当前版本信息，数据结果提交至“提交资料/3.1/试题_智能系统运维.docx”文件中。（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	要求符合考试版本情况	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：3.3.2

试题名称：“古诗学习”-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 开发工具；

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）信息爆发时代，如何快速有效地处理视频、音频、文字等信息显得尤为重要。无论是政企会议、公检法办案、还是教学培训、个人录音等场合均需要形成完整的文字记录材料，音频文件也需要转成文字，为解决各类场景下的音频转文字问题，语音标注应运而生。

某公司正在进行“古诗学习”机器人的研发，本款机器人主要针对古诗进行转写识别以应用于教育事业，帮助学龄儿童进行古诗学习。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据标注工作。

智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

（1）查看 windows 系统日志，记录 3 条日志信息。数据结果提交至“提交资料/3.2/试题_智能系统运维.docx”文件中。（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	要求符合考试版本情况	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.1.1

试题名称：保护动物-数据采集和处理-业务数据采集

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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，能够对数据库内的业务数据

进行数据采集，梳理业务数据采集和处理规范，整理并归纳数据，做好数据分类。

3、技能要求

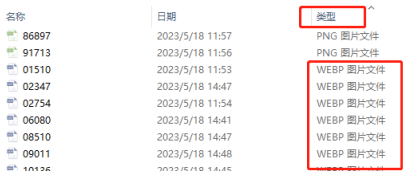
打开“数据采集和处理”项目中的数据采集和处理（试题）文件，考生在规定的时间内完成以下任务：

(1) 采集图片时候利用了多种设备和工具进行了采集，需要对数据进行预处理，请打开“未处理的采集文件”，筛选掉谷歌网页中采集的图片，将其另存在“未处理的采集文件/网页采集文件”，其余数据归类为“非网页采集文件”。请将路径内容提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。（5分）

(2) 由于数据不充足，要再用实时采集的方式采集一些图片，需要完成以下三个初始步骤：安装 cv2 库、导入 cv2 库、为采集图片指定文件夹（D:\\testData\\img.jpg）。请完成这个操作并将操作的内容截图提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。（10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	查看筛选后的图片类型是否包含 WEBP 图片类型 	筛选后的结果不可包含 WEBP 格式的图片文件	
2	01	3分	是否按照要求正确安装 cv2 库	同答案操作	
	02	3分	是否按照要求正确导入 cv2 库	同答案操作	
	03	4分	是否按照要求正确为采集的图片制定文件夹	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.1.2

试题名称：保护动物-数据采集和处理-业务数据处理

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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“古诗学习”识别机器人的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，能够对数据库内的业务数据进行数据采集，梳理业务数据采集和处理规范，整理并归纳数据，做好数据分类。

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）请将“未处理的采集文件/非网页采集文件”中的数据按照动物的类别进行归类：大熊猫、金丝猴、华南虎、白鳍豚、扬子鳄和其他动物。截取文件路径图片和内容提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。（5分）

（2）打开提交资料/1.2/试题_数据采集和处理.excel 表格，通过查看文件夹的属性，可获知此文件夹下有多少子文件夹和文件等信息，请汇总各个动物的数据数量和其数据格式。内容截图提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。（10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	归类的图片是否整理正确	归类错误一种动物错误扣除1分	
2	01	10分	汇总整理的数据结果是否正确	汇总错误一个数据扣除2分	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.2.1

试题名称：“保护动物”-数据标注-原始数据清洗与标注

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 开发工具、Labelme 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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“动物识别”数据清洗和标注工作。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 完成数据的清洗，使用，利用 Labelme 进行数据标注。

3、技能要求

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

（1）请打开 Python 工具，对“图片库”文件夹中的数据进行数据清洗，将“图像模糊不清”、“图像高度相似”、“图片重复”的数据清洗掉，完成图像清洗；并建立新的文件夹，文件夹名为“正式数据”。将“正式数据”文件夹提交至“提交资料/2.1/试题_图像标注”文件夹中。（15分）

（2）请打开 Labelme 工具，使用该工具对文件名为“64645”、“50827”、“08130”的图像进行语义分割的标注工作。

具体标注规则如下：

①标注对象：耳朵、眼睛、鼻子、嘴巴、头、上肢、下肢、躯干、尾巴；

②标注方法：多边形框；

③标注正确率：大于等于 95%；

④允许误差像素值：3px；

⑤导出格式：每张图片对应一个 VOC 文件；

⑥显示语义分割“PNGlabel 图片”，并截图保存；

每张图片导出的 VOC 文件命名和截图文件名称都为原数据原名，并存储到文件夹“导出文件”中。将文件夹“导出文件”提交至“提交资料/2.1/试题_图像标注”文件夹中。（25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	写出任意 2 段正确的代码	同答案操作	
	02	15分	写出所有正确的代码，并筛选出符合要求的图片	同答案操作	
2	01	10分	完成了任意 2 张标注，但没有正确导出	同答案操作	
	02	15分	完成了所有标注，但仅导出了一份正确的导出文件	同答案操作	
	03	25分	完成了所有标注，且导出的数据都符合要求	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.2.2

试题名称：“保护动物”-数据标注-标注后数据分类与统计

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 开发工具、Labelme 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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“动物识别”数据清洗和标注工作。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 完成数据的清洗，使用，利用 Labelme 进行数据标注。

3、技能要求

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

（1）请打开 Excel 工具，新建表格，文件名命名为“分类.xlsx”。将“正式数据”文件夹中的图像数据和导出数据的标签数量进行分类与统计。分类与统计需要包含以下这些信息：文件名、动物类别、是否进行标注、标签数量。将文件“分类.xlsx”提交至“提交资料/2.2/试题_图像标注”文件夹中。（10分）

（2）请使用 Python 工具，使用代码将文件“分类.xlsx”进行导入和按照“标签数量”进行排序。最后，将运行出来的结果进行截图保存，截图命名为“统计结果”，并提交到“提交资料/2.2/试题_图像标注”文件夹中。（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	分类与统计需要包含以下这些信息：文件名、动物类别、是否进行标注、标签数量	同答案操作	
2	01	10分	使用 Python 工具，使用代码将文件“分类.xlsx”进行导入和按照“标签数量”进行排序	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.3.1

试题名称：“保护动物”-智能系统运维-智能系统基础操作

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 开发工具；

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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“动物识别”数据采集系统运维工作。

智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

数据标注完成后，结合具体的运行情况，进行系统内信息记录。

打开“智能系统运维”项目中的智能系统运维（试题）文件，考生在规定时间内完成以下任务：

（1）请检查是否正常运行 MySQL 服务，如果服务没有启动请点击启动，并将结果提交至“提交资料/3.1/试题_智能系统运维.docx”文件中。（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	要求符合考试版本情况	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：4.3.2

试题名称：“保护动物”-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 开发工具、MySQL 数据库；

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）随着人类的不断发展和扩张，野生动物的生存环境受到了严重的破坏和威胁。许多动物种群数量急剧减少，甚至濒临灭绝。为了保护这些珍贵的生物资源，保护动物的业务逐渐兴起。进行动物语义分割标注的重要原因是为了推动人工智能技术在动物保护和研究领域的应用。通过训练计算机视觉模型，可以实现自动化的动物监测和识别，从而更好地保护野生动物和研究动物行为和生态环境。总之，动物语义分割标注是一项非常重要的任务，它可以帮助计算机视觉模型更好地理解和分析动物图片，从而实现更精确的图像识别和分析，同时也可以推动人工智能技术在动物保护和研究领域的应用。

（2）你将作为一名数据标注员参与“动物识别”数据采集系统运维工作。智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

(1) 进行 MySQL 数据库的导出操作，将“数据采集和处理”文件夹导出，并将结果提交至“提交资料/3.2/试题_智能系统运维.docx”文件中。(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	要求导出数据文件正确	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.1.1

试题名称：政策分析-数据采集和处理-业务数据采集

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。

政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地理解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可

以帮助企业、个人等各个方面更好地了解政策信息，提高大数据分析的效率和准确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，及时发现政策变化和风险，为企业、个人等各个方面提供预警和应对措施，整理并归纳数据，做好数据分类。

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）.ipynb 文件，考生在规定的时间内完成以下任务：

(1) 需要采集中文文稿，请打开“材料/文稿库”文件，采集其中有关中文的文稿，提取每段独立的文稿新建为一个单独的 docx 文件，文件命名从 1 开始，按照字段出现的顺序命名，将其存在“中文文稿”文件夹。请将路径内容提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。（5 分）

(2) 由于数据不充足，需要再浏览通过视频手工采集一些，请打开“材料/视频文本采集”文件，通过手工采集后新建文档（文档名命名同上），并保存到“采集数据/中文文稿”文件中。路径内容截图提交至“提交资料/1.1/试题_数据采集和处理.docx”文件中。

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	3 分	查看采集的数据的数量是否正确	同答案结果	
	02	2 分	查看采集的数据的排序是否符合要求	同答案结果	
2	01	7 分	查看手工采集的数据的数量是否正确	同答案结果	
	02	3 分	查看采集的数据的排序是否符合要求	同答案结果	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.1.2

试题名称：政策分析-数据采集和处理-业务数据处理

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。

政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地理解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可以帮助企业、个人等各个方面更好地了解政策信息，提高大数据分析的效率和准

确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，及时发现政策变化和风险，为企业、个人等各个方面提供预警和应对措施，整理并归纳数据，做好数据分类。

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）.ipynb文件，考生在规定的时间内完成以下任务：

(1) 请将采集后的中文文稿文件中按照政府类和故事类别进行归类。截取文件路径图片和内容提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。

(2) 打开“提交资料/1.2/试题(4)_数据采集和处理.excel”表格，请分别汇总政府类文档的数量和故事类文稿的数量。内容截图提交至“提交资料/1.2/试题_数据采集和处理.docx”文件中。

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	3分	查看采集的数据的数量是否正确	同答案结果	
	02	2分	查看采集的数据的分类是否符合要求	同答案结果	
2	01	5分	查看政府类文档数据的数量是否正确	同答案结果	
	02	5分	查看故事类文档数据的数量是否正确	同答案结果	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.2.1

试题名称：“政策分析”-数据标注-原始数据清洗与标注

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 开发工具；

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地了解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可以帮助企业、个人等各个方面更好地了解政策信息，提高大数据分析的效率和准确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，及时发现政策变化和风险，为企业、个人等各个方面提供预警和应对措施，整理并归纳数据，做好数据分类。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 Doccano 完成数据的标注和处理，利用 SPSS 完成数据统计。

3、技能要求

经过数据预处理，对图片数据进行清洗和标注，得出分析报告。

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

(1) 请打开 Python 工具，使用 Re (Regular expression) 工具包，对“文本数据库”文件夹中的文本数据进行数据清洗，将文本中的特殊字符、链接标签、多余空格进行去除，完成文本数据清洗；并建立新的文件夹，文件夹名为“正式数据”，将完成数据清洗的文本存储到该文件夹中。将“正式数据”文件夹提交至“提交资料/2.1/试题_文本标注”文件夹中。(15分)

(2) 请打开 Doccano 工具，需对“正式数据”文件夹中的语料进行命名实体标注，其实体代码有政府名称 (NOG)、省名称 (PRN)、市名称 (CIN)、年 (YEAR)、月 (MON)、日 (DAY)、数字 (NUM) 等；将标注好的数据文件，导出成 jsonl 格式，文件命名为“原名.jsonl”，并存储到文件夹“导出文件”中。将文件夹“导出文件”提交至“提交资料/2.1/试题_文本标注”文件夹中。(25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	写出任意 2 段正确的代码	同答案操作	
	02	15分	写出所有正确的代码，完成数据清洗并保存	同答案操作	
2	01	10分	完成了任意 2 个文本的标注，但没有正确导出	同答案操作	
	02	15分	完成了所有标注，但没有正确导出	同答案操作	
	03	25分	完成了所有标注，且导出的数据都符合要求	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.2.2

试题名称：“政策分析”-数据标注-标注后数据分类与统计

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 开发工具；

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地了解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可以帮助企业、个人等各个方面更好地了解政策信息，

提高大数据分析的效率和准确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据采集和处理工作。数据采集和处理：根据业务特征与业务需求，及时发现政策变化和风险，为企业、个人等各个方面提供预警和应对措施，整理并归纳数据，做好数据分类。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 Doccano 完成数据的标注和处理，利用 SPSS 完成数据统计。

3、技能要求

经过数据预处理，对图片数据进行清洗和标注，得出分析报告。

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

(1) 通过 Excel 工具完成数据集，数据集需要 5 个文本数据的文件名、实体名称和他们出现的次数，将该表格存储并命名为“实体出现频率记录表.xlsx”。将该文件“实体出现频率记录表.xlsx”提交至“提交资料/2.2/试题_文本标注”文件夹中。（10 分）

(2) 使用 SPSS 工具打开“实体出现频率记录表.xlsx”，进行实体名称的频数统计分析。最后，将完成的统计结果进行截图保存，命名为“频数统计结果”。将文件“频数统计结果”提交至“提交资料/2.2/试题_文本标注”文件夹中（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	完成数据集，并存放到相应位置	同答案操作	
2	01	10 分	制作出表格或者饼状图都可以	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.3.1

试题名称：“政策分析”-智能系统运维-智能系统基础操作

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 开发工具；

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地理解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可以帮助企业、个人等各个方面更好地了解政策信息，提高大数据分析的效率和准确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据维护工作。智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

数据清理完成后，结合具体的运行情况，

打开“智能系统运维”项目中的智能系统运维（试题）文件，考生在规定时间内完成以下任务：

(1) 打开 Pandas 工具包，并进行 `dropna()` 去空值操作，数据结果提交至“提交资料/3.1/试题_智能系统运维.docx”文件中。(5分)

3、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	要求代码执行无误	同答案操作	

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：5.3.2

试题名称：“政策分析”-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 开发工具；

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）政府公告是重要的政策信息来源。政府公告通常包含了政策、法规、规章等重要信息，对于企业、个人等各个方面都有着重要的影响。因此，通过 AI 学习政府公告，可以更好地了解政策信息，及时掌握政策变化，从而更好地应对政策风险和机遇。政府公告是大数据分析的重要数据源。政府公告中包含了大量的文本信息，这些信息可以被用于进行大数据分析，从而发现政策变化、社会热点等信息。通过 AI 学习政府公告，可以更好地处理和分析这些文本信息，从而提高大数据分析的效率和准确性。政府公告是智能决策的重要依据。政府公告中包含了大量的政策信息，这些信息可以被用于进行智能决策。通过 AI 学习政府公告，可以更好地理解政策信息，从而为智能决策提供更准确的依据。AI 学习政府公告具有重要的意义，可以帮助企业、个人等各个方面更好地了解政策信息，提高大数据分析的效率和准确性，为智能决策提供更准确的依据。

(2) 你将作为一名数据标注员参与“政府政策解析”方向的数据维护工作。智能系统运维：能够维护业务所需要的数据，并记录智能系统应用数据情况。

3、技能要求

(1) 查看 windows 系统日志，记录 5 条日志信息。数据结果提交至“提交资料/3.2/试题_智能系统运维.docx”文件中。(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	要求符合考试版本情况	同答案操作	

人工智能训练师（五级）操作技能考核 试题单

准考证号：

试题代码：人工智能训练师（五级）6.1.1

试题名称：法律法规数据业务数据采集

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	安装 OFFICE 或者 WPS 等办公软件	台	1	

2、工作任务

（1）业务背景

法律智能问答系统是一种基于自然语言处理技术的智能对话系统，可以通过人机交互的方式回答用户的法律问题。法律智能问答系统可以帮助公众快速准确地获得法律知识，也可以为律师等法律从业者提供便捷的法律咨询服务。

（2）业务目标

你将作为一名人工智能数据分析师，参与法律法规相关的数据清洗与数据分析工作，输出分析报告。

数据采集和处理：根据业务特征与业务需求，能够对业务数据按照要求进行数据清洗，并对关键信息进行数据分析。

（3）考试内容

根据业务目标，需要为法律智能问答系统提供文本数据集，并对数据集进行数据处理与分析，使其能满足法律智能问答系统的训练需求。

打开（考试）文件夹，考生在规定时间内完成以下任务：

（1）在采集数据的过程中，我们应该遵循哪些采集原则？并将内容提交至“提交资料/题目 1/试题_数据采集和处理.docx”文件中。【能够利用设备、工具等完成原始业务数据采集】

（2）业务数据在采集过程中，对数据库进行数据采集时，一般的采集方法有哪些，请至少写出两种？并将内容提交至“提交资料/题目 2/试题_数据采集和处理.docx”文件中。

3、技能要求

- (1) 能够利用设备、工具等完成原始业务数据采集
- (2) 能够完成数据库内业务数据采集

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	6	答对任意两条即可算对，答对 1 条得 3 分	<p>1) 明确采集目的及数据类型：采集之前，必须清楚地知道需要什么类型的数据以及采集的目的是什么。</p> <p>2) 选择可靠的来源：来源可以是官方机构发布的数据、权威的研究报告和调查数据，以及经过严格审查和筛选的第三方数据。</p> <p>3) 确保数据采集过程中合法：遵循 robots 协议，明确对方运行采集哪些，不允许采集哪些。对于涉及到个人隐私的数据，应该放弃采集。</p> <p>4) 确保数据安全性和保密性：采集到的数据必须安全的存储在可靠的平台上，防止数据丢失、泄漏或者被篡改。</p>	
02	4	至少写出两种，写出任意两种即可，答对一种得 2 分	<p>1. 使用 ETL 工具：ETL 工具可以将数据库中的数据提取出来，并进行清</p>	

			<p>洗和转换，最后加载到目标数据库或数据仓库中。常见的 ETL 工具有 Talend、Pentaho、Informatica 等。</p> <p>2. 写 SQL 语句：可以通过编写 SQL 语句来查询数据库中的数据，并将结果导出为 CSV 或 Excel 文件。</p> <p>3. 使用 API 接口：如果数据库中已经有相应的 API 接口，可以通过调用接口来获取数据，并进行采集和处理。</p> <p>4. 数据库复制：某些数据库系统支持数据复制功能，可以将数据库中的数据自动复制到另一个数据库中，再进行进一步的数据处理和分</p> <p>析。</p> <p>1. 数据库日志：部分数据库系统会记录操作日志，可以从中提取所需的</p>	
--	--	--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

			业务数据，并进行采集和处理。	
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						
S2						
S3						
S4						

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：人工智能训练师（五级）6.1.2

试题名称：法律法规数据业务数据处理

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	安装 OFFICE 或者 WPS 等办公软件	台	1	

2、工作任务

（1）业务背景

法律智能问答系统是一种基于自然语言处理技术的智能对话系统，可以通过人机交互的方式回答用户的法律问题。法律智能问答系统可以帮助公众快速准确地获得法律知识，也可以为律师等法律从业者提供便捷的法律咨询服务。

（2）业务目标

你将作为一名人工智能数据分析师，参与法律法规相关的数据清洗与数据分析工作，输出分析报告。

数据采集和处理：根据业务特征与业务需求，能够对业务数据按照要求进行数据清洗，并对关键信息进行数据分析。

（3）考试内容

根据业务目标，需要为法律智能问答系统提供文本数据集，并对数据集进行数据处理与分析，使其能满足法律智能问答系统的训练需求。

打开（考试）文件夹，考生在规定时间内完成以下任务：

（1）在（考试）文件夹中查看法律法规数据（试题）.xlsx，1、对 content 列的数据进行去重；2、筛选出 publish_date 列 2000 年以后的数据，3、对 2000-01-01 以后发布的规定按照年度进行数量统计并形成表格。最后将筛选后所有数据及表格提交至“提交资料/题目 1/试题_ 数据采集和处理.xlsx”文件对应 sheet 页中。

（2）在（考试）文件夹中查看法律法规数据（试题）.xlsx，1、对 content 列的数据进行去重；2、筛选出 publish_date 列 2020 年-01-01 以后的数据；3、对 2020-01-01 以后发布的规定按照年度进行数量统计并绘制成折线图（其中图表标题为“法律法规 2020-2023 年度发布数量统计”，横坐标为“年度”，纵

坐标为“发布数量”，折线图上并显示发布数量)。最后将去重后的数据及折线图提交至“提交资料/题目2/试题_数据采集和处理.xlsx”文件对应 sheet 页中。

3、技能要求

- (1) 能够根据数据处理要求完成业务数据整理归类
- (2) 能够根据数据处理要求完成业务数据汇总

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	8	答对得 8 分	详情请见【题目 1 答案.xlsx】	
02	12	答对得 12 分	详情请见【题目 2 答案.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						

人工智能训练师（五级）操作技能考核

试题单

准考证号：

试题代码：人工智能训练师（五级）-6.2.2

试题名称：人工智能训练师数据分类统计能力考核

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

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

基础的数据分类统计能力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，今日的工作任务是对已完成标注文档的数据进行分类统计。

将标注完成文档中已标注的字段内容，按如下形式在 Excel 中进行记录统计。

字段A名称	字段B名称	字段C名称
字段内容	字段内容	字段内容
字段内容	字段内容	字段内容
字段内容	字段内容	字段内容

3、技能要求

- (1) 考生能够对已标注数据进行基础的分类统计。

4、质量指标

详见评分表 2-2.2

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	1.5	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
02	1.5	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
03	1	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
04	1	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
05	1.5	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
06	1	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	
07	1	字段记录格式正确,字段内容完整且正确(包括字段名称,字段内容在内,一个字段0.5分)	详情请见【字段分类统计参考答案表】	

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（五级）操作技能考核

试题单

准考证号：

试题代码：人工智能训练师（五级）6.2.1

试题名称：人工智能训练师数据清洗标注能力考核

考核时间：70 min

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

数据清洗和标注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文档标注。

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) 下载“考试”文件夹，并筛选正确的标注文档，删除无效文档；
- (2) 打开正确的标注文档，将需要标注的字段信息标红；
- (3) 标注完成后保存标注文档。

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	3.5	删除掉 3 个错误文档即得分（错删、漏删均不得分）	删除非通用合同文档，即文件 1、文件 2 和文件 4	
02	8	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	8	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】	
04	6	字段标注完整且正确。	详情请见【字段参考答案表】	
05	6	字段标注完整且正确。	详情请见【字段参考答案表】	
06	8	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】	
07	6	字段标注完整且正确。	详情请见【字段参考答案表】	

08	6	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（五级）操作技能考核

试题单

准考证号：

试题代码：人工智能训练师（五级）6.3.1

试题名称：人工智能训练师培训能力考核

考核时间：15min

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，作为一名五级/初级人工智能训练师，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

您将作为一名五级/初级人工智能训练师，有能力记录智能系统应用数据情况、记录智能系统功能数据情况等多项工作。

(3) 考试内容

- ① 打开智能系统的控制台或管理界面。
- ② 进入智能文本页面。
- ③ 新建一个文档抽取任务。
- ④ 新建一个表格解析任务。

⑤ 新建一个文字识别任务。

⑥ 进入日志界面，导出以上三个操作的日志记录，并保存为“操作日志.xlsx”。

3、技能要求

(1) 能够进行智能系统开启。

(2) 能够简单使用智能系统。

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	正确新建一个文档抽取任务		
02	1	正确新建一个表格解析任务		
03	1	正确新建一个文字识别任务		
04	1	正确导出日志文件		
05	1	正确命名日志文件		

注：因为系统不确定，暂无对应正确答案

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（五级）操作技能考核

试题单

准考证号：

试题代码：人工智能训练师（五级）6.3.2

试题名称：人工智能训练师培训能力考核

考核时间：15min

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，作为一名五级/初级人工智能训练师，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

您将作为一名五级/初级人工智能训练师，有能力记录智能系统应用数据情况、记录智能系统功能数据情况等多项工作。

(3) 考试内容

- ⑦ 打开智能系统的控制台或管理界面。
- ⑧ 找到管理日志的选项。
- ⑨ 筛选日志类型为“抽取”的日志记录，并将其导出。
- ⑩ 筛选近六十天内的日志记录，并将其导出。

11 筛选操作员为 admin 的日志记录，并将其导出。

12 将导出的三个文件，保存至“日志”文件夹。

3、技能要求

(3) 记录智能系统应用数据情况。

(4) 记录智能系统功能数据情况。

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	正确导出类型为“抽取”的日志记录		
02	1	正确导出六十天内的日志记录		
03	1	正确导出操作员为 admin 的日志记录		
04	2	正确保存导出的日志文件		

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						