

第 3 部分

理论知识复习题

一、单选题

序号	试题	选项 A	选项 B	选项 C	选项 D
1	人工智能训练师包含数据标注员和()两类人员。	人工智能算法测试员	软件开发人员	需求分析人员	系统测试人员
2	人工智能训练师需要具备哪种能力?	优秀的团队协作能力	高超的算法创新能力	准确的数据分析能力	良好的审美眼光
3	人工智能训练师的沟通和表达能力有助于什么?	加强与行业人员的联系	提高技术实施效率	提升数据分析准确性	减少团队协作难度
4	人工智能是通过什么来呈现人类智能的技术?	计算机程序	机械装置	化学合成物	电子设备
5	人工智能的研究内容主要包括什么?	计算机实现智能的原理	网络安全技术	建筑设计原理	动物行为研究方法
6	人工智能科学是一门研究、开发和应用智能体的()。	跨学科领域	计算机科学	脑科学	软件工程
7	智能体能够做什么?	感知环境、推理、学习和采取行动	控制机器人的运动	生成艺术作品	发布社交媒体帖子
8	人工智能科学属于哪个学科?	工程学	化学	历史学	音乐学
9	训练好的深度学习模型需要进行()。	评估验证	使用	模型加载	系统调用
10	在人工智能训练的模型设计阶段,根据人工智能主流算法框架和()选择最优的、与任务匹配的模型。	任务目标	项目预算	数据来源	项目截止时间

11	在评估人工智能系统时，下列哪个因素是最重要的评估指标？ a) 系统的准确性	系统的准确性	系统的响应速度	系统的外观设计	系统的品牌知名度
12	评估人工智能系统时，下列哪个因素可以衡量系统的可靠性和鲁棒性？	系统的安全性	系统的用户友好性	系统的社会影响	系统的数据处理能力
13	当模型在验证集上不能达到理想的评估指标时，()。	优化模型设计	重新进行模型训练	直接进行模型测试	继续进行模型验证
14	()是机器学习过程中的中间输出结果。	模型	输入数据	输出数据	参数
15	线性回归中，输入变量是()。	自变量	因变量	x	y
16	支持向量机是一种常用的()学习算法。	监督	无监督	半监督	对比
17	神经网络通常由三层组成：输入层、()和输出层。	隐藏层	辅助层	传输层	预测层
18	主成分分析中，主成分的排序是基于什么原则？	方差解释比例	相关系数大小	数据的原始顺序	样本的分类标签
19	卷积神经网络包含()、池化层和全连接层。	卷积层	输入层	输出层	中间层
20	RNN 由一个()组成，该层逐步处理序列数据，并将之前的信息传递到下一时刻。	循环层	输入层	隐藏层	输出层
21	在循环神经网络中，哪个组件用于传递信息并捕捉序列中的时间依赖关系？	隐藏层	输入层	输出层	激活函数
22	在知识图谱中，用于表示实体间关系的边通常具有什么类型？	有向边	无向边	加权边	线性边

23	类脑计算的主要目标是：	模拟人类的大脑功能 b) 创建一个全新的智能	创建一个全新的智能生物	提高计算机的运算速度	优化网络安全
24	在模式识别中，特征选择的目标是：	降低维度	增加噪声	增强数据	数据预处理
25	虚拟现实技术可以提供以下哪种体验？	沉浸式体验	增强现实	环境感知	人脸识别
26	虚拟现实是一种模拟的体验技术，下列哪个选项描述了虚拟现实的主要特点？	现实感强，用户可以与虚拟环境进行实时互动	只提供视觉上的模拟，没有其他感官参与	只适用于娱乐领域，没有其他实际应用	需要外部设备支持，无法独立运行
27	增强现实技术通常使用以下哪种设备或技术来实现虚拟信息的叠加？	头戴式显示器	3D 打印机	虚拟现实手柄	红外线传感器
28	scikit-learn (sklearn) 是一个用于机器学习的开源库，它基于以下哪个编程语言开发？	Python	Java	C++	R
29	scikit-learn 中的 fetch 类数据集是指：	加载现有数据集	根据公式生成数据集	随机构造数据集	读取文件中的数据集
30	以下哪个是 scikit-learn 中的 TruncatedSVD 的常见参数？	n_components	alpha	n_neighbors	threshold
31	下列哪个工具是用于增强现实 (AR) 开发的工具？	Unity	PyTorch	TensorFlow	Keras
32	由 Google 提供的深度学习平台具有广泛的用户，这个平台是：	TensorFlow	PyTorch	Keras	Caffe
33	PyTorch 是一个用于构建和训练机器学习和深度学习模型的：	框架	编程语言	算法	开发工具

34	数据分析中，为什么要先从业务角度切入，对业务流程进行梳理？	为了找到业务重点并提出关键的业务增长方案	为了收集大量的业务数据	为了规划公司的业务目标	为了计算企业的资产负债情况
35	为什么需要梳理业务流程？	为了制定业务数据的采集、处理、审核流程	为了详细了解公司的人员组成	为了修改公司的业务目标	为了便于管理资金
36	在人工智能业务流程中，什么是数据标注的作用？	准备数据用于模型训练	分析数据的特征和模式	测试和验证模型的性能	部署和监控模型的结果
37	处理单个业务数据时需要兼顾哪些方面？	业务流程和数据特征	系统硬件配置和网络带宽	处理时间和数据存储空间	系统安全性和数据来源可靠性
38	在处理单个业务数据时，以下哪个步骤通常首先进行？	数据清洗和预处理	特征提取和转换	数据可视化和探索性分析	模型训练和预测
39	对业务数据进行整体处理时需要根据什么进行处理？	业务规则	数据来源可靠性	系统硬件性能	数据处理时间
40	什么是数据标准化？	将数据处理为标准的数据类型和单位	将数据存储于云端服务中	对数据进行加密保护	将数据转化为一定的分布形式
41	业务数据处理中，制定输入输出关系是指？	确定输入数据与输出数据间的对应关系	确定硬件配置与网络环境间的协调关系	设定业务规则中的输入和输出参数	确定数据传输和存储的关系

42	特定业务字段获取是指？	从业务数据中获取符合要求的特定字段	将业务数据标准化为标准的数值范围	去除噪声和 无用信息	确定业务数据处理的目标和流程
43	业务数据采集和处理后，需要进行哪项工作？	审核	上传	下载	编辑
44	审核工作可以由以下哪些方式完成？	人工、软件、平台	人工、物联网、软件	软件、平台、机器学习	机器学习、人工智能、物联网
45	在获取业务数据后，查看数据是否与业务需求和目标相关联的流程是？	数据与业务关联度审核	数据质量审核	数据数量审核	数据指标分析
46	业务数据审核内容中，哪一项是指审核数据是否符合逻辑关系，比如数据之间的时间和数值关系？	逻辑性审核	数据质量审核	数据数量审核	经验审核
47	对文本类数据，通常需要进行哪些处理？	分词、情感分析、词性标注等	数值计算、单位转换、加密解密等	时间计算、日期格式转换等	逻辑运算、条件判断等
48	用于语义分析问题的文本类业务数据需要进行怎样的预处理？	需要进行标注分词、句法判别等操作。	需要进行数据存储操作。	需要进行文本统计分析操作。	不需要进行预处理。
49	视觉类数据以什么方式存储？	点阵图像	音频	数字	文字
50	何为视觉畸变？	任何视觉上与真实情况不同的变形或失真	图像的亮度、饱和度	图像的模糊程度	视觉检测的错误率

51	压缩视觉数据可以达到什么目的？	减少存储空间和传输带宽	增加存储空间和传输带宽	提高数据的质量	使数据能够在不同设备和系统上通用
52	图像描述主要使用什么方式进行？	文本描述	数字编码	图形符号	听觉信号
53	哪种审核是指用户对图像或视频的主观感受？	主观审核	客观审核	客户审核	内容审核
54	在审核中，数据质量如何定义？	合法性和规范性	趣味性和内容合法性	音频清晰度和图片大小	图像色彩鲜艳程度
55	SSIM 相比 PSNR 优势在哪里？	更精确评判图像质量	更易计算	更适合处理大量图像数据	优势不是很明显
56	VMAF 是如何评价视频质量的？	通过引入深度学习机制	通过逐像素计算比对	通过人工审核	通过播放用户反馈数据
57	VMAF 是用于评估视频质量的指标，它基于以下哪个原理进行评估？	人眼感知模型	编码效率	数据压缩算法	时空分析
58	音频可以包含哪些声音信号？	仅包含人耳可感知的声音信号	包含所有声音信号，无论人耳是否可感知	只包含音乐、鸣笛声等娱乐性的声音信号	只包含机床声等工业性的声音信号
59	在语音信号处理中，下列哪种畸变是由于信号传输过程中的时延引起的？	回声	噪声	多径效应	混响
60	语音类数据的处理目标是什么？	对声音波形和音频数值进行分析	将语音转录为文本	将语音转为图像	将语音转为视频
61	语音类业务数据主观审核主要包括以下哪些方面？	根据收听者的个人理解对语音类业务数据进行评估	静音、叠音检测	审查涉嫌内容是否合规	审查语音类业务数据是否符合业务指标

62	如何获取点云数据？	使用激光仪器扫描物体表面并记录激光的反射时间和位置信息	使用相机对物体进行拍摄	使用声波仪器扫描物体表面并记录声波的反射时间和位置信息	拍摄天空的云朵分布
63	使用相机获取点云数据时，下列哪种方式适用于 AR\VR？	深度相机	两个相机从不同位置获取物体的图像	激光雷达拍摄	在很近距离拍摄
64	人工智能在餐饮业中可以通过以下哪个技术实现个性化菜单推荐？	数据分析	地理信息系统	外卖配送管理	音频处理技术
65	人工智能在智能搜索中可以利用以下哪个技术？	机器学习	3D 打印	增强现实	虚拟现实
66	自动数据处理相比人工处理的优势是什么？	提高系统处理效率	提高数据存储能力	提高数据传输速度	提高数据加密技术
67	最优化决策中，可以利用人工智能计算来实现什么？	最大化业务指标	最小化系统节点	最少的费用	最低风险决策
68	业务构建流程的目的是什么？	描述系统功能	分析市场需求	优化业务流程	制定战略目标
69	业务流程优化的主要目标是：	提高效率	增加成本	扩大规模	提高风险
70	远程智能诊疗是指通过什么方式对患者进行疾病诊断及治疗的过程？	患者的视频	患者的电子病历	患者的面诊	患者的电话咨询
71	业务流程优化中的流程分析包括：	理解步骤、活动和决策点之间的关系	降低成本	提高效率	技术支持和自动化
72	在复杂综合业务流程优化中，下列哪项是确定的优化目标？	提高效率	增加复杂性	增加成本	降低质量

73	以下关于数据标注描述不正确的是哪一项？	数据标注规则不能改变	数据标注规则需要一致	数据标注规则需要不断完善	数据标注规则需要符合项目需求
74	数据清洗的前提和基础是以下哪个步骤？	数据分析	定义数据清洗的策略和规则	验证	干净数据回流
75	Smartbi 采用什么样的计算架构？	分布式计算架构	单机计算架构	多线程计算架构	云计算架构
76	属性标注是对事物属性进行标签，它包括以下哪些标注类型？	文本类别标注	指向关系标注	词性标注	实体标注
77	在文本类标注质量规范中，中文分词的质量标准是什么？	标注好的分词必须与词典中的词语一致，不存在歧义	标注好的分词必须包含所有可能的分词结果	标注好的分词必须考虑文本的上下文	标注好的分词必须包含所有专有名词
78	某公司需要对产品订单进行标注，提取订单中商品名称、购买数量、商品价格、订单日期，该标注属于以下哪种标注？	实体标注	关系标注	情感标注	属性标注
79	在视觉类数据清洗规范中，为什么需要去除已损坏图像或视频？	无法进行模型训练	占用了大量的存储空间	数据价值低	所有选项都正确
80	矩形框标注是一种什么样的标注方式？	对目标对象进行目标检测框标注的简单处理方式	将图像分成各具特性的区域并提取出所需部分	用一个或多个标签标注目标的属性	所有选项都正确
81	区域标注比矩形框标注更加准确的原因是什么？	区域标注可以标注出物体的形状更加准确	区域标注可以标注出物体的大小更加准确	区域标注可以标注出物体的颜色更加准确	所有选项都正确

82	对于人眼无法分辨的模糊目标，应该如何处理？	应该舍弃，无需标注	应该尽可能地标注，即使无法判断类别也应该标注	只需要标注清晰的目标即可	根据具体情况决定是否标注
83	标注框的大小和位置对目标检测任务影响，以下描述不正确的是？	标注框大小和位置对目标检测任务的效果没有影响	标注框大小和位置的不合适会导致目标检测任务的漏检和误检	标注框大小和位置的不合适会导致目标检测任务的准确率和召回率降低	标注框大小和位置的合适与否对目标检测任务的效果具有重要影响
84	语音切割的主要目的是什么？	便于对每个语音单元进行特征提取和模型训练	提高语音的质量	使语音更符合业务需求	去除语音中的噪声和失真
85	英文应该怎样转写？	根据情况转写成英文或汉字	转写成汉字	转写成英字	不转写
86	以下哪项不属于执行算法测试的操作？	测试数据采集	算法相关细节	测试数据处理	badcase 样例说明及分析
87	通过改变测试集对算法进行测试的目的是什么？	查看算法是否达到期望输出	提高算法的运行速度	减少算法的内存占用	提高算法的精度
88	准确率的计算公式是什么？	正确预测的样本数/总样本数	真正正类的样本数/全部预测为正类的样本数	能被算法正确预测为正类的样本数/真正正类的样本数	精确率和召回率的加权调和平均
89	在混淆矩阵中，TP 代表什么？	真正类	假正类	真负类	假负类
90	人工智能算法测试报告的测试目的是什么？	评估算法的性能和可靠性	展示算法的功能和特点	提供算法的使用方法	介绍算法的背景和发展历程

91	海洋生特种类识别, 需要选择以下哪种模型?	图像分类	图像分割	物体检测	目标追踪
92	在 EasyDL 中, 哪种数据标注方式适用于数据集较小的情况, 可以提高标注准确度?	人工标注	自动标注	半自动标注	模型标注
93	以下哪个方法不能够提升模型训练效果?	减少样本数量	对标注结果进行修正	补充样本数量	优化数据集
94	智能产品最后是以硬件形式呈现, 需对产品进行_____。	工业设计、结构设计、硬件设计	物理设计、逻辑设计、硬件设计	结构设计、硬件设计、软件设计	工业设计、硬件设计、软件设计
95	以下代码的运行结果是_____。 import numpy as np a=np.arange(1,10) .reshape(3,3) print(a.size,a.shape)	9 (3, 3)	10 (3, 3)	(1, 10) (3, 3)	(1, 10) 9
96	在数据预处理时将数据集中某一列数值特征的值缩放到 0-1 区间内, 属于_____。	数据归一化	数据标准化	数据正则化	数据初始化
97	Tensorflow 中 Keras 序列模型的 fit 方法, 其 epochs 参数指定的是_____。	训练周期数	每批次的数据数目	训练样本数	测试样本数
98	针对 Tensorflow 中 Keras 序列模型, 用于模型评估的 evaluate 函数, 其默认返回值通常是_____。	损失值和准确率	预测值和实际值	准确率和精确率	精确率和召回率

99	<p>设有以下代码：</p> <pre>import tensorflow as tf model = tf.keras.Sequential() model.add(tf.keras.layers.Conv2D(16, kernel_size=(2, 2), activation='relu', input_shape=(28, 28, 1)))</pre> <p>则添加的卷积层中隐藏层的神经元个数为_____。</p>	16	$2 \times 2 = 4$	$28 \times 28 \times 1 = 784$	无法确定
100	<p>可以将实数范围的自变量映射到 (0, 1) 区间以实现二分类的激活函数是_____。</p>	sigmoid	relu	softmax	celu
101	<p>针对 Tensorflow 中 Keras 序列模型 model, model.compile 函数中用于指定损失函数的参数为_____。</p>	loss	metrics	optimizer	rossentropy
102	<p>人工智能分类算法测试常见的评价指标不包括_____。</p>	运行时间	准确率	精确率	召回率
103	<p>以下不支持 AI 在线开发功能的平台是_____。</p>	苹果手机“OpenAI ChatGPT”应用	百度飞浆 EasyDL	华为云	阿里云 AI 平台
104	<p>AI 模型优化的一般步骤不包括_____。</p>	模型创建	模型评价	数据分析	模型调整
105	<p>卷积神经网络模型优化的措施不包括_____。</p>	增加训练时间	增加卷积网络层数	增加池化层和 Dropout 层	改变损失函数

106	因为图像数据是一种深层次的结构化数据，所以深层次的数据，____，更能够更充分和准确地把这些层级信息表达出来。	卷积神经网络	循环神经网络	长短期记忆网络	感知机
107	目前中文票据 OCR 识别中的文本检测算法常用的是____和 EAST。	CTPN	LSTM	CRNN	Faster-RCNN
108	人工与智能的视觉交互方式不包括____。	情感识别	手势识别	表情识别	姿态识别
109	使用 AI 翻译软件翻译英文文章，属于人与智能的____交互方式。	文字	语音	视觉	触屏
110	用户在淘宝 APP 中输入问题，淘宝客服机器人自动回复用户提问，这属于人与智能的____交互方式。	文字	语音	视觉	触屏
111	触屏交互需要满足的条件不包括____。	系统支持人脸识别	触摸式屏幕	单指或多指直接或近距离接触该屏幕	触摸接触时移动（或不移动，具体取决于时间阈值）
112	智能客服机器人是____的一种典型的应用。	人工与智能交互	人脸识别	模式识别	数据挖掘
113	语音识别模块涉及的语言模型一般包括____和 RNN。	N-gram	LSTM	CTPN	CRNN
114	语音识别模块涉及声学模型及主流语音识别框架一般包括____和 ASRT。	HMM-GMM	RNN	CTPN	CRNN
115	智能客服机器人的工作原理一般不包括____。	训练 AI 模型	知识库构建	语义理解	问答匹配

116	智能客服机器人的工作原理主要包括四个部分： _____、语义理解、问答匹配和机器人深度学习。	知识库构建	知识库选择	知识点匹配	问答匹配
117	当智能客服机器人收到用户提出的问题后，通过自然语言处理技术和算法模型理解用户表达的含义，然后在 _____中找出与该问题匹配的答案并将其发送给用户。	知识库	语料库	信息库	关系数据库
118	智能客服机器人系统的主要构成模块中，_____仅在其与用户的交互方式为语音时才需要。	语音识别和语音合成	语音降噪和语音增强	语言模型和声学模型	语音识别框架和声学模型
119	属于主流的语音识别框架的是 _____。	HMM-GMM	CNN	RNN	LSTM
120	某厂家生产的翻译笔，按住语音键说话即可实现中英互译，翻译结果会显示在屏幕上。这款产品使用的技术中必定包含_____。	语音识别	增强现实	图像识别	数据挖掘
121	智能客服机器人要理解用户文字输入的问题，先要进行 _____操作，以便构成客户提问向量，从而获取客户提问标签。	分词	文本分析	同类词合并	词频统计
122	智能客服机器人中客户的提问和问题库问题，都属于非结构化数据，通常可用 _____算法将非结构化的客户提问和问题库问题用结构化的向量表示出来。	TF-IDF	HMM-GMM	N-gram	CNN

123	智能客服机器人的问答匹配阶段，会使用_____技术在知识库中寻找相应问题的答案。如计算用户的问题与知识库中保存的问答记录之间的相似性，选择相似性最高的答案进行回答。	检索、匹配	计算、匹配	检索、计算	比较、计算
124	培训讲义应该详略得当，对于员工难以理解的功能应该如何？	详细描述	省略	简单带过	绕过不提
125	数据处理过程中可能出现哪些问题？	数据缺失和数据格式标准化问题	数据缺失和数据标准化问题	数据缺失和数据收集问题	数据标准化和数据时效性问题
126	在结果评估中，采用哪些方法来评估所采集的数据？	实时指标和分析报表	机器学习模型	无监督学习模型	有监督学习模型
127	人工智能培训师在知识与技能方面，除了应该有深入的人工智能知识、具备熟练的技术操作能力之外，还需要有具备（）。	企业领导能力	行业领域知识	法律知识	人事管理能力
128	人工智能训练师通常需要与谁进行对接？	高级管理人员	客户和业务利益相关者	竞争对手	外部媒体
129	为什么人工智能训练师需要持续学习和更新知识？	提高团队协作能力	适应快速变化的技术环境	减少工作压力	提升个人社交能力
130	人工智能训练流程通常包括数据准备、模型设计、确定评估指标、（）、模型测试和模型验证	数据预处理	模型训练	模型部署	数据清洗
131	评估人工智能系统的性能时，下列哪个指标可以衡量系统的效率和资源利用率？	系统的灵活性	系统的能耗	系统的知识表达能力	系统的声音识别准确率

132	在人工智能训练的模型验证阶段,在()上通过评估指标验证模型的效果。	训练集	验证集	测试集	新的数据集
133	线性回归通过拟合一个()模型来预测连续的输出变量。	非线性	线性	三维	曲面
134	逻辑回归是一种用于解决什么类型的问题?	回归问题	分类问题	聚类问题	降维问题
135	支持向量机的目标是找到一个最优的()。	解	分隔超平面	方案	函数
136	主成分分析(PCA)是一种常见的数据()方法,它可以将高维数据转化为低维数据,同时保留数据的主要特征。	分析	降维	统计	可视化
137	以下哪个是深度学习的主要特征?	基于逻辑推理的模型	使用多层神经网络进行学习	依赖于传统机器学习算法	只能处理结构化数据
138	深度学习是基于()的学习研究。	机器学习	深度神经网络	支持向量机	逻辑回归
139	循环神经网络(RNN)在处理序列数据时最适合解决哪种问题?	图像分类	文本生成	物体检测	强化学习
140	长短时记忆网络的基本结构包括三个门控,分别是输入门、()、输出门。	概率门	遗忘门	反馈门	反向传播门
141	Transformer模型是一种采用()机制的深度学习模型。	长短时记忆	自注意力	门控循环	双向循环
142	在模式识别中,聚类是什么算法?	一种监督学习算法	一种无监督学习算法	一种增强学习算法	一种半监督学习算法
143	语音合成又称为:	语音识别	文字转语音	语音分割	文字识别
144	增强现实主要应用于什么场合?	用于实际工作场景	在真实世界上叠加虚拟内容	沉浸式体验虚拟场景	扩展视觉、听觉等多个感知能力

145	scikit-learn 中的 make 类数据集是指:	加载现有数据集	根据公式生成数据集	随机构造数据集	读取文件中的数据
146	在 scikit-learn 中, 以下哪个方法用于主成分分析降维?	SelectKBest	PCA	SelectFromModel	RFE
147	在 scikit-learn 中, 以下哪个方法用于构建决策树模型?	KNeighborsClassifier	DecisionTreeClassifier	LogisticRegression	SVC
148	PyTorch 最初由哪个公司开发?	谷歌	微软	Meta (原脸书)	亚马逊
149	以下哪个是业务流程五要素?	角色、活动、费用、产出物、规则	角色、活动、协作、产出物、规则	角色、系统、协作、产出物、规则	角色、活动、协作、产出物、管理
150	信息流主要包括哪些过程?	市场推广、库存管理、生产监控等	市场调研、需求预测、供应商寻源、采购招标等	售后服务、客户维护、质量监控等	原料采购、成品仓储、销售发运等
151	业务流程构建的哪个步骤是确定问题和目标的关键阶段?	数据收集和准备	模型选择和设计	算法训练和优化	解决方案部署和监控
152	在单个业务数据的处理过程中, 特征提取的主要目的是:	去除数据中的噪声和异常值	将原始数据转换为可供模型使用的特征表示	分析数据的分布和相关性	计算数据的统计指标和摘要信息
153	在数据标准化过程中, 以下哪种方法可以将数据转换为均值为 0, 标准差为 1 的分布?	最大最小值缩放	Z-score 标准化	对数转换	PCA 降维
154	业务数据审核内容中, 哪一项审核数据的数量是否符合预期?	数据质量审核	数据数量审核	数据指标分析	经验审核
155	日常的文本类数据通常包括哪些内容?	整篇文档、句子、图形	单词、句子、文本段落、整篇文档	数字、句子、文本段落、整篇文档	单词、句子、数字、整篇文档

156	在文本类数据中，哪一种类型的数据复杂性最高？	数值型数据	文本型数据	日期时间型数据	逻辑型数据
157	日常业务中，文本类数据的数据量通常是什么情况？	数据量比较小	数据量较大	数据量相对恒定	数据量忽略不计
158	文本类业务数据通常需要进行何种预处理？	存储	清洗	人工标注	统计分析
159	畸变可能导致什么问题？	计算机性能下降	视觉感知质量的变化	数据的安全性	存储空间变大
160	大量的图像信息每天会被生产和传输，这会导致什么问题？	未能及时存储会导致数据丢失	传输和存储的代价非常巨大	根本不可能传输这么多的数据	可以直接用于学习和研究
161	视觉数据在压缩后，如何进行传输和存储？	在被处理成数值格式后	在经过编码后	在经过解码后	在与其他数据合成后
162	什么是视觉类数据审核？	对文本数据的审核	对图像或视频的审核	对声音或音频的审核	对程序代码的审核
163	PSNR 和 SSIM 用于评价什么类型数据？	音频数据	视频数据	文字数据	程序代码
164	全球来看，人类最主要的沟通方式是什么？	视频沟通	语音沟通	文字沟通	动作沟通
165	与语音类业务数据相关的产品有哪些？	人脸识别和智慧城市	语音合成和声控系统	智能电视和智能手环	人类基因编辑和太空探索
166	具有音频畸变的语音类数据该如何处理？	静音处理	预处理算法	放大音量	无需处理
167	音频数据在处理前，要进行的离散化步骤是？	将波形分解为音频信号	将音频信号离散成数字信号	对音频质量进行评级	对声音波形进行抽样

168	什么是语音分析技术？	将文字转化为语音的技术	将非结构化的语音信息转化为结构化索引的技术	将图像转化为语音的技术	将数字信号转化为语音的技术
169	除了激光扫描外，还可以使用什么技术生成点云数据？	相机拍摄	三维建模	声波扫描	天气预报
170	颜色信息是如何赋予点云中对应的点的？	通过航拍获取颜色信息	通过相机获取彩色影像，然后将对应位置的像素颜色信息赋予点云中对应的点	通过激光扫描获取颜色信息	无法为点云赋予颜色信息
171	点云滤波的目的是什么？	提取点云的关键点	去除点云数据中的噪声	确定点云的几何和纹理等特征点	将不同视角下的坐标系变换到统一视角下
172	在点云数据处理中，下列哪种算法适用于点云配准任务？	K-means 聚类算法	最小二乘法	支持向量机	朴素贝叶斯分类器
173	点云数据审核是否包含色彩信息是为了什么？	确认点云中的噪声	确定点云的几何和纹理等特征点	了解数据采集的环境条件	确认点云中的噪声
174	人工智能对传统行业产生的影响是：	相互孤立的	深远的	无关紧要的	不确定的
175	人工智能为管理提供最优化决策支持的目的是什么？	提高系统稳定性	实现系统最优性能	减少系统错误	提高系统可扩展性
176	业务流程优化的第一步是什么？	设计业务优化流程	业务调研	诊断报告	技术评估
177	当外部经营环境处于剧烈波动时，企业更倾向于采用哪种方法进行业务流程优化？	智能系统改造法	全新设计法	冗余消除法	功能整合法

178	简单业务流程优化的第一步是什么？	培训和管理	确定优化目标	流程分析和重设计	技术支持和自动化
179	智能汽车的智能化程度快速提升主要得益于以下哪项技术？	4G 技术	5G 技术	3D 打印技术	太阳能技术
180	数据审核的目的是什么？	找寻数据集中存在的不符合规范的“脏”数据, 并进行处理	对标注人员标注的所有数据进行核对和校验	避免错误重复发生, 并出具数据质检报告	对质检合格数据进行抽检
181	每个标注对象都需要一个单独的标注框, 这是哪个原则？	贴边原则	独立原则	不越界原则	确定共同特征
182	标注框超出边界不会导致什么问题？	程序运行错误	标注数据丢失	标注数据不准确	
183	文本类数据清洗包括以下哪些操作？	统一数据规格、删除重复项、纠正错误数据、过滤图片数据	统一数据规格、删除重复项、纠正错误数据、过滤不符合项目需求的数据	统一数据规格、删除无用项、纠正错误数据、过滤不符合项目需求的数据	统一数据规格、删除无用项、纠正错误数据、过滤图片数据
184	在文本类数据清洗中, 数据清洗中的非空检查主要是检查什么？	数据的唯一性	数据的完整性	数据的准确性	数据的格式
185	在文本类数据清洗中, 对于文本非空检查的描述正确的是哪个？	所有数据字段都需要进行非空检查	只有一些数据字段需要进行非空检查	所有数据都不需要进行非空检查	对于空值字段可以直接删除
186	在文本类数据清洗中, 哪种数据清洗规则可以确保数据的主键唯一性？	非空检查	主键重复检查	错误值清洗	数据格式检查

187	数据清洗中的缺失值清洗策略不包括哪一项?	重要性高, 缺失率低: 通过计算进行填充	重要性高, 缺失率高: 删除该记录	重要性低, 缺失率低: 不作处理或通过计算填充	重要性低, 缺失率高: 删除该记录
188	文本类数据清洗后的数据质量评估指标中, 哪个指标要求数据不能存在错误或异常?	完整性	准确性	一致性	唯一性
189	在进行文本属性标注中, 应该根据哪个视角理解文本?	标注人员的个人喜好	项目需求	数据量大小	情感分类数量
190	关键点标注一般用于哪些场景?	自动驾驶下的车辆、行人等	人体面部轮廓、五官定位、身体部位、情绪识别、运动追踪等	对商品种类进行标注	所有选项都正确
191	对于视觉类数据标注质量规范, 以下描述哪个是不正确的?	标注像素点越接近标注物的边缘像素, 标注质量就越高	标注像素点越接近标注物的边缘像素, 标注质量就越低	视觉类数据标注的质量好坏主要取决于像素点的判定准确性。	视觉类数据标注的质量的好坏与标注时间无关性。
192	目标检测任务需要标注出哪些信息?	目标物体的大小	目标物体的位置和类别	目标物体的颜色和形状	目标物体的运动轨迹
193	在设置目标框时, 标注像素点越接近标注物的边缘像素, 标注质量会怎样?	标注质量越差	标注质量越好	不会影响标注质量	取决于标注物的大小

194	常见的语音类数据处理规范包括哪些方面?	语音处理类型、语音处理规则、数据输入格式要求、语音类数据质量要求	语音处理类型、语音处理规则、数据输出格式要求、语音类数据质量要求	语音处理类型、图像处理规则、数据输出格式要求、文本类数据质量要求	语音处理类型、文本处理规则、数据输出格式要求、语音类数据质量要求
195	语音切割需要考虑哪些方面?	语音信号长度、语音信号频率、语音信号时域特性	语音信号的语境、语法、语义	语音信号的音量、音调、语速	语音信号的语种、方言、发音特点
196	语音类数据质量要求包括以下哪些方面?	语音的音量、音调、语速等	语音的清晰度、无噪音等	语音的语境、语法、语义等	语音的语种、方言、发音特点等
197	在切割长篇语音时,不同说话人的语音应该切割为怎样的语音段?	合并到同一个语音段中	切割为不同的语音段		
198	为保证语音的完整性,切割语音段时应该保留多长的静音段?	0.1至0.2秒	0.2至0.3秒	0.3至0.4秒	0.4至0.5秒
199	在切割语音段时,如何避免突发噪音?	整句切割	选取没有突发噪音的语音段进行切割	将语音段切割到前一个语音段的末尾	将语音段切割到后一个语音段的开头
200	如果说话人发音时一字一顿,每个停顿时间超过1秒,会被判定为怎样的语音?	有效语音	无效语音	质量不佳语音	可疑语音
201	语音内容转写的目的是什么?	提高语音质量	便于后续的语音处理任务	减少语音处理时间	方便语音存档
202	增量式测试每次加入多少新数据?	一半	一定量	全部	

203	精确率的计算公式是什么？	正确预测的样本数/总样本数	真正正类的样本数/全部预测为正类的样本数	能被算法正确预测为正类的样本数/真正正类的样本数	精确率和召回率的加权调和平均
204	F1-score 值的取值范围是什么？	-1 到 1 之间	0 到 1 之间	0 到 100 之间	1 到 10 之间
205	什么情况下会出现假负类？	样本的真实值为负类，算法预测为负类	样本的真实值为正类，算法预测为负类	样本的真实值为负类，算法预测为正类	样本的真实值为正类，算法预测为正类
206	人工智能算法功能测试是用来测试算法的什么方面？	算法的性能	算法的正确性	算法的可扩展性	算法的可靠性
207	算法测试报告中的算法性能测试和结果分析可以包括哪些内容？	算法的输入和输出、测试用例和测试结果、算法的正确性和鲁棒性	算法的运行时间、内存占用、并行化能力	算法的背景和发展历程、算法的特点和优势、算法的使用方法	数据来源、数据规模、数据预处理方法、测试任务和测试时间
208	在 EasyDL 中，用户需标注数据集的百分比达到多少即可训练模型？	0.2	0.3	0.4	0.5
209	在 EasyDL 中，哪种数据标注方式适用于数据集较大的情况，可以提高标注效率？	人工标注	自动标注	半自动标注	模型标注
210	关于模型评估，以下哪个描述是正确的？	查看模型评估结果时，精确率和召回率都必须达到最优值。	更希望减少误识别的时候，更需要关注精确率指标	更希望减少误识别的时候，更需要关注召回率指标	更希望减少漏识别，更需要关注精确率指标

211	成功的 AI 产品往往是因为_____。	采用了最先进的技术	抓住了用户的痛点，能切实为企业或用户提供创新价值	价格非常便宜	请最红的明星代言
212	下列一般能作为 OCR 软件识别对象的文件是_____。	后缀为 MP3 的文件	后缀为 JPG 的文件	后缀为 XLSX 的文件	后缀为 EXE 的文件
213	当需要改变数据样本集的形状（或维度）时，可使用 python 的_____方法。	resize	reshape	change_size	change_shape
214	假设已利用“import tensorflow as tf”语句导入了 tensorflow 模块，则采用线性堆叠网络层来构建神经网络模型的语句为_____。	model = tf.Sequential()	model = tf.keras.Sequential()	model = tf.Linear_model()	model = tf.keras.Linear_model()
215	假设已创建了 Keras 的序列模型并添加了相应的层次，可以使用模型的_____方法对模型进行训练。	train	fit	train_model	fit_model
216	Tensorflow 中 Keras 序列模型的 fit 方法，其功能是_____。	函数拟合	模型训练	模型编译	模型评估
217	卷积神经网络中，如果输入是 4 x 4 的矩阵，采用窗口大小为 2 x 2 的最大池化层处理后，输出尺寸为_____。	1 x 1	2 x 2	3 x 3	4 x 4
218	针对 Tensorflow (tf 对象) 中 Keras 序列模型，全连接层用_____函数来定义。	tf.keras.Sequential()	tf.keras.layers.Dense()	tf.keras.layers.Dropout()	tf.keras.layers.Flatten()
219	函数解析式 $f(x) = \max(0, x)$ 对应的激活函数是_____。	sigmoid	relu	softmax	max

220	在机器学习中，softmax 激活函数常被用来进行_____。	二分类	多分类	聚类	降维
221	机器学习中的_____, 指的是模型在训练集上表现的很好, 但是在交叉验证集或测试集上表现一般, 也就是说模型对未知样本的预测表现一般。	欠拟合	过拟合	适度拟合	全拟合
222	为了解决模型训练中的_____问题, 可以将原始数据集再划分出除训练集和测试集以外的第三个数据集: 验证集, 用来微调模型超参数。	欠拟合	过拟合	适度拟合	全拟合
223	人工智能分类算法测试中, 全部预测为正类的样本中, 实际为正类的比例称为_____。	准确率	精确率	召回率	F1-score 值
224	针对 Tensorflow 中 Keras 序列模型 model, model.predict_classes 方法的参数为_____。	训练集的特征集	测试集的特征集	训练集的标签集	测试集的标签集
225	以下场合中不一定需要进行模型优化的是_____。	模型的评价指标不符合设计要求	模型复杂度很高, 训练时间长	模型在训练集上表现很好, 但是在测试集或者新数据上表现很差	模型在训练集和测试集上表现都不好
226	长短期记忆网络 LSTM 擅长的领域不包括_____。	中文 OCR 识别	人脸识别	语音识别	机器翻译
227	目前国标码收录的汉字数量为_____。	三千多个	六千多个	一万多个	二万多个
228	以下不属于人工智能框架和平台的是_____。	TensorFlow	NLP	PyTorch	PaddlePaddle

229	假设已正确加载 paddlepaddle 中文票据 OCR 识别预训练模型 ocr, 则 ocr.recognize_text() 方法返回值的类型是_____。	字典	列表	字符串	OCR 对象
230	使用苹果 Siri 手机语音助手, 属于人与智能的_____交互方式。	文字	语音	视觉	触屏
231	用户口头发布指令, 让智能音箱在某个时间开始播放某一首歌曲, 这种交互方式为_____。	文字交互	语音交互	自然语言交互	脑机接口
232	ChatGPT 属于_____程序。	人脸识别	聊天机器人	图像识别	工业控制
233	ChatGPT 是_____领域的 AI 工具。	计算机视觉	自然语言处理	模式识别	增强现实
234	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, 从中_____寻求答案并回答客户。	数据库	知识库	AI 模型	大数据
235	当智能客服机器人正确识别出用户提出的问题后, 下一步操作是从构建的_____中选择最匹配的问题和答案。	关系数据库	知识库	语料库	信息库
236	采用知识库的智能客服机器人, 其知识库通常由_____层架构组成。	2	3	4	5
237	属于常见语言模型的有_____。	HMM-GMM	N-gram	CNN	LSTM
238	导航仪能模拟人声播报道路信息和规划线路。这里面采用的技术是_____。 语音识别	语音识别	语音合成	图像识别	数据挖掘

239	用户想设计一个能及时回复学生校园招聘信息的智能客服机器人，最主要的工作是_____。	训练 AI 模型	建立专属知识库	建立领域知识库	建立通用知识库
240	智能客服机器人在数据采集阶段采集的信息不包括_____。	客户基本信息	客户隐私信息	客户提问信息	常见客户历史提问及答复
241	培训讲义应该遵循哪些原则？	内容完美、操作易懂、语言多样	有针对性、可操作性强、语言逻辑性、方法形式多样性	结构紧凑、内容详尽、逻辑清晰	创意突出、内容丰富、形式多变
242	培训讲义编写中，什么是确定培训对象需求的基础	培训目标	分析受众	编写核心内容	准备概述
243	哪一步骤是编写培训讲义的“灵魂”	分析受众	构思培训内容	编写目录和关键信息	将内容制定细节
244	演示法要求讲师做好以下哪个准备工作？	让每个受训者试一试	准备好所有用具并搁置整齐	对每个受训者的试做给予立即的反馈	向所有受训人员公布议程表
245	视听法的优点不包括以下哪中？	容易增加受训人员感性知识	视听教材可反复使用	适应受训人员个体差异	容易控制学习的进度
246	什么是数据标注的作用？	标注让人们可以更好地理解数据	标注让机器可以更好地理解数据	标注可以增强数据的可读性和可操作性	所有选项都是正确的
247	在数据标注流程中，如何帮助指导对象设定标注目标？	介绍标注工具和标注技巧	与指导对象协商标注目标，定义标注质量和依据标签要求	提供标注培训	所有选项都不是正确的
248	人工智能训练师包含()和人工智能算法测试员两类人员。	软件开发人员	需求分析人员	数据标注员	系统测试人员

249	人工智能训练师的创新能力主要用于什么方面?	开展市场营销活动	优化生产流程	开发新的人工智能产品和应用	提升员工福利待遇
250	确定评估指标,是为了对模型进行()。	训练	部署	验证	优化
251	在人工智能训练的模型测试阶段,在()上通过评估指标验证模型的效果。	训练集	验证集	测试集	新的数据集
252	()是实现人工智能的核心。	计算机	编程语言	算法	人工智能训练师
253	逻辑回归属于()算法。	生成	判断	二元分类	预测
254	决策树的叶节点包含()。	样本的全集	特征属性测试	决策的结果	判断的过程
255	支持向量机在决策边界上使用哪些数据点进行决策?	所有数据点	随机选择的数据点	支持向量	所有正确分类的数据点
256	KNN 算法的关键在训练集中找到最接近一个目标数据点的()个最近邻居	1	2	K	3
257	KNN 算法对于回归问题,预测目标数据点的值为 K 个最近邻居的()。	最大值	最小值	平均值	中位数
258	主成分分析 (PCA) 用于什么目的?	特征选择	数据可视化	数据降维	数据预处理
259	卷积层是卷积神经网络的核心组件,它可以学习数据的()。	时间特征	类别特征	空间特征	行为特征
260	循环神经网络主要用于处理()数据。	分类	空间	序列	特征
261	Transformer 模型是 2017 年由()提出的模型。	OpenAI	微软	Google	苹果公司
262	类脑计算的一个重要特征是:	只能处理数值数据	必须依赖云计算	具有自主学习和适应能力	无法处理自然语言

263	量子系统之间存在一种特殊的关联关系，改变一个系统的状态会立即影响另一个系统的状态，这种关系被称为：	量子隧道效应	量子态	量子纠缠	量子比特
264	自然语言处理（NLP）是指计算机对人类自然语言进行处理和理解的技术领域。以下哪个是NLP的主要任务之一？	图像识别	语音合成	文本挖掘	网络安全
265	以下哪个是计算机视觉的主要任务之一？	智能客服	视频压缩	目标检测	数据清洗
266	以下哪种生物特征常用于人类身份识别？	发型	体态	虹膜	脑电图
267	下列哪个选项最准确地描述了增强现实的特点？	通过虚拟环境完全替代现实环境	提供纯粹的虚拟体验，与现实世界没有关联	在现实世界中叠加虚拟信息，保持对现实环境的感知	只适用于娱乐领域，没有其他实际应用
268	以下哪个不是sklearn提供的机器学习算法和工具？	Kmeans	决策树	TenorFlow	数据降维
269	在scikit-learn中，用于处理缺失值的预处理方法是：	StandardScaler	MinMaxScaler	Imputer	OneHotEncoder
270	在scikit-learn中，用于将连续特征转换为离散特征的预处理方法是：	StandardScaler	MinMaxScaler	Binarizer	LabelEncoder
271	scikit-learn中的cross_val_score方法的作用是：	计算召回率	计算精确率	交叉验证评估模型性能	增加验证集
272	下列哪个工具是最适合用于自然语言处理（NLP）任务的开发工具？	TensorFlow	PyTorch	NLTK	OpenCV
273	以下哪个深度学习平台是专门用于移动设备和嵌入式系统的？	TensorFlow	PyTorch	Core ML	Azure

274	Keras 是一个用于构建和训练机器学习和深度学习模型的:	编程语言	开发工具	框架	算法库
275	PaddlePaddle 是由哪家公司开发的开源深度学习平台?	谷歌	微软	百度	腾讯
276	PaddlePaddle 最初被开发用于支持哪个领域的研究和应用?	自然语言处理	计算机视觉	机器翻译	语音识别
277	在业务数据采集集中, 需要关注哪些指标来推断结果?	数据采集频率及采集周期	数据分析的方法及工具	业务数据指标表现情况	业务数据的总量及类型
278	产品页面的操作信息不包括以下哪些内容?	操作类型	操作对象	报错信息	操作对象描述
279	以下哪个场合不太可能是业务数据的产生场合?	电子商务交易	社交媒体互动	系统错误调试	智能传感器监测
280	在数据分析和建模过程中, 数据规范化的主要目的是:	提高数据的安全性	降低数据存储成本	提高数据的一致性和可比性	加快数据处理速度
281	在数据处理过程中, 数据去噪的主要目的是:	去除无关的特征变量	减少数据的维度	去除数据中的异常值和噪声	增加数据的样本量
282	业务数据审核内容中, 哪一项是指通过指标分析, 审核数据是否存在异常情况?	数据质量审核	数据数量审核	数据指标分析	经验审核
283	文本类业务数据采集与分析的重要目的是什么?	用于判断文本数据的收益情况。	用于计算产品销售情况	用于提供营销方案和改进产品。	只用于抓取和存储数据。
284	文本类业务数据在采集过程中, 通常:	文本类业务数据数量比较少, 采集难度较低。	文本类业务数据是结构化数据, 采集比较容易。	文本类业务数据分布广泛、来源多样, 采集比较方便。	文本类业务数据难以采集。

285	为什么需要对文本进行特征提取？	方便保存文本数据。	方便人工处理文本数据。	方便智能产品处理和计算文本数据。	不需要进行特征提取。
286	特征提取之后需要进行什么操作？	进行文本存储操作。	进行文本清洗操作。	进行文本统计分析和计算操作。	特征提取后不需要进行其他操作。
287	文本类业务数据主要审核哪些方面的数量？	商品数量	用户数量	业务数据数量	软件质量
288	对文本类业务数据，需要审核的质量是？	服务质量	产品质量	数据质量	平台质量
289	现实中的视觉数据采集主要依靠哪些设备？	录音机	麦克风	相机、摄像头	扬声器
290	什么是数据的预处理？	对数据进行图像化展示	对数据进行抽象化处理	对数据进行清洗和过滤，以符合使用要求	对数据进行自动化分类和标注
291	图像分割是用来做什么的？	进行图像增强	找到图像中的关键物体	按像素分离出不同的区域	对图像进行分类
292	视觉类数据审核标准是按照什么来评价的？	内容是否有趣味性	是否符合市场需求	是否适宜观看	是否反映时事热点
293	视觉类业务数据主要审核哪几个方面？	数量、质量、是否有趣味性	数量、是否符合市场需求、声音质量	质量、数量、是否符合业务指标	质量、声音质量、是否符合社会热点
294	审核视觉类业务数据时，质量是否符合业务指标的标准由谁来制定？	用户自行决定	客户投票决定	项目经理制定或依据业务经验制定	计算得出
295	在视频业务数据中，下列哪种指标用于衡量视频的播放流畅性？	视频质量	视频分辨率	视频卡顿率	视频帧率
296	PSNR 是评价什么的数值评估指标？	图像大小	图像分辨率	图像质量	图像颜色

297	什么是 SSIM?	图像压缩算法	图像处理技术	图像相似度评价指标	条纹识别算法
298	VMAF 的全称是什么?	Video Mode Analysis Function	Video Multimethod Assessment Function	Visual Multimethod Assessment Fusion	Visual Mode Analysis Fusion
299	在 VMAF 中, 以下哪个因素不会对视频质量评估产生影响?	画面清晰度	色彩准确性	视频帧率	压缩比特率
300	声音类数据包括什么?	语言和文学	语音和音频	视频和音频	文字和模拟音乐
301	常见的音频文件存储的是什么?	逻辑信息	文本	语音波形	图像
302	语音类业务数据的广泛存在的基础是什么?	科技不断进步	声音优美	采集便捷	文化传承需要
303	在语音信号处理中, 下列哪种畸变是由于说话人与麦克风之间的距离变远导致的?	回声	噪声	多径效应	混响
304	什么是语音识别?	将语音中的文字信息转换为语音信息	将语音中的文字信息转换为数字信号	将语音中的信号转换为文字信息	将信号中的噪音提取出来
305	语音类业务数据审核主要:	检查语音声音的大小	模仿收听者的个人理解评估是否保留该语音	依据声音质量对音频进行审核	确定语音的节奏是否正确
306	点云数据可以用来做什么?	获取物体的重量数据	获取物体表面的纹理信息	建立三维模型	预测天气情况
307	星载激光雷达获取点云数据适用于哪种范围测绘?	街景测量	文物三维扫描建模	城市级别	室内测绘
308	点云配准的经典应用是什么?	点云分类	点云分割	场景重建	特征提取

309	在点云数据分割中,下列哪种方法常用于基于几何特征的分割任务?	基于颜色聚类算法	高斯混合模型	法线估计	RANSAC 算法
310	基于人工智能的智能搜索系统可以帮助提高以下哪方面效果?	搜索结果数量	搜索范围	搜索准确性	员工福利
311	智能交互通常服务于智能系统的哪个方面?	数据分析	系统维护	人机交互界面	电子设备控制
312	智能控制模块通过什么实现对机器人、智能设备等的控制?	人工智能算法	传感器技术	通信协议	机械结构
313	下列哪个工具适用于确定业务流程中的瓶颈和改进机会?	散点图	流程图	控制图	折线图
314	业务流程优化的最后一步是什么?	识别瓶颈和问题	实施改进措施	监测和评估	简化流程
315	智能化汽车的基础是什么?	汽油车技术	车联网	电动化	共享化
316	以下哪个工具可以帮助可视化整个业务流程?	数据分析工具	控制图	流程图	帕累托图
317	以下哪个环节不属于对数据集的处理?	数据采集	数据清洗	数据说明	数据标注
318	以下哪个环节是数据处理基本流程中的最后一个环节?	数据审核	数据质检	数据验收	数据采集
319	下列关于数据标注员对数据进行标注的描述,错误的是哪个?	需要参照数据标注规则	遇到数据标注规则模糊的问题需要及时询问	按照自己的想法猜测标注	数据标注员需要总结经验并及时汇报遇到的问题
320	以下哪项不是数据质检的任务?	质检人员发现数据标注问题并通知相关人员进行修正	出具数据质检报告	计算标注人员的标注准确率,输出审核报告	对于经过数据标注及数据审核处理后的数据随机抽检

321	数据处理规范主要是指什么？	数据采集过程中的规范	数据存储过程中的规范	数据清洗和数据标注过程中的规范	数据分析过程中的规范
322	VBA 是什么类型的宏语言？	C 语言宏语言	Java 宏语言	VisualBasic 宏语言	Python 宏语言
323	哪种数据清洗规则可以处理文本字段为空的情况？	主键重复	错误值清洗	非空检查	数据格式检查
324	以下哪一项不属于文本类数据清洗中需要去重的情况？	完全相同的两条数据	同一事物的不同表达方式	两个不同的数据类型	重复的数据行
325	关系标注是对复句的句法关联和语义关联作出重要标注的一种任务，它不包括以下哪些标注类型？	指向关系标注	平行语料标注	实体标注	修饰关系标注
326	在文本类标注质量规范中，多音字标注的质量标准是什么？	标注好的多音字必须包含所有可能的读音	标注好的多音字必须考虑文本的上下文	标注好的多音字必须使用专业性工具（如字典）来标注一个字的全部读音	标注好的多音字必须包含所有常用的读音
327	实体标注通常指具有特定意义或指代性强的名称词，不包括以下哪种类型？	地名	专有名词	动词	人名
328	在视觉类数据标注中，什么是属性标注？	对目标对象进行目标检测框标注的简单处理方式	将图像分成各具特性的区域并提取出所需部分	用一个或多个标签标注目标的属性	所有选项都不正确

329	对于被遮挡的目标，标注框应该如何选择？	只需要标注图像中完整的目标即可	不需要标注被遮挡的目标	应该标注被遮挡的目标，矩形框只需要包含看得见的部分	应该标注被遮挡的目标，矩形框应该包含整个目标
330	倾斜目标框标注与非倾斜目标框标注的区别是什么？	倾斜目标框和非倾斜目标框的标注规范完全一致	倾斜目标框和非倾斜目标框的标注规范完全不同	倾斜目标框可以更加贴合目标物体的轮廓，标注信息中多了旋转角度信息	非倾斜目标框可以更加贴合目标物体的轮廓，标注信息中多了旋转角度信息
331	以下哪个不属于语音标注任务特点的衡量维度？	应用场景	语音时长	语音质量	口音
332	在语音类数据处理中，为了保证语音质量，需要对语段进行正确准确的截取。这是因为语段截取的正确性与以下哪个方面有关？	语音的音量、音调、语速等	语音的清晰度、无噪音等	语音的语境、语法、语义等	语音的语种、方言、发音特点等
333	对于长篇语音的切割，应该以什么为切割单位，以保证语义连贯性？	单词	音节	句子	段落
334	下列哪种情况可能导致一段语音被判定为无效语音？	说话人的声音太大	语音中含有噪声	说话人的声音极小而无法听清语音内容	说话人说话速度过慢
335	在转写语音中出现感叹词时应该怎么处理？	按照习惯转写	不转写	使用标准拼写格式	转写成汉字
336	在语音转写中，陈述句应该使用什么符号？	?	!	。	
337	在人工智能算法测试中，以下哪项不是改变测试集时常用的操作？	输入与训练时一样的数据	输入与训练时完全不同的数据	输入测试时的随机数据	输入训练时的边界值

338	召回率的计算公式是什么？	正确预测的样本数/总样本数	真正正类的样本数/全部预测为正类的样本数	能被算法正确预测为正类的样本数/真正正类的样本数	精确率和召回率的加权调和平均
339	召回率的另一个名称是什么？	查准率	F1-score	查全率	准确率
340	在混淆矩阵中，TN 代表什么？	真正类	假正类	真负类	假负类
341	EasyDL 是哪家公司推出的云端 AI 开发平台？	腾讯	阿里巴巴	百度	新浪
342	工业零件划痕自动识别，需要选择以下哪种模型？	图像分类	图像分割	物体检测	目标追踪
343	关于模型评估，以下哪个描述是不正确的？	查看模型评估结果时，需要思考在当前业务场景，更关注精确率与召回率哪个指标。	更希望减少误识别的时候，更需要关注精确率指标	更希望减少误识别的时候，更需要关注召回率指标	更希望减少漏识别，更需要关注召回率指标
344	智能产品会在开发早期对其_____进行细致的考量。	产品属性	价格范围	商业模式	盈利模式
345	基于深度学习的算法模型通常将 OCR 识别分为_____两个阶段。	图像载入、图像分析	图像扫描、图像分割	文本检测、文本识别	文本扫描、文本分割

346	<p>关于以下代码段的说法，正确的有_____。</p> <pre>import numpy as np Z = np.arange(10).reshape(2, 5) Z = (Z - Z.min()) / (Z.max() - Z.min()) print(Z)</pre>	Z 是一个有 10 个元素的一维数组	Z 是一个 5 行 2 列的二维数组	代码段实现了对数组 Z 的归一化操作	代码段实现了对数组 Z 的正则化操作
347	<p>关于以下代码段的说法，正确的有_____。</p> <pre>import numpy as np Z = np.arange(9).reshape(3, 3) Z = (Z - Z.min()) / (Z.max() - Z.min()) print(Z)</pre>	Z 是一个有 9 个元素的一维数组	Z 的元素是从 1 到 9 的整数	代码段将数组 Z 的所有元素归一化为 0~1 范围的浮点数	代码段将数组 Z 的所有元素归一化为 0~9 之间的整数
348	<p>设有以下代码： <pre>import tensorflow as tf model = tf.keras.Sequential() model.add(tf.keras.layers.Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(28, 28, 1)))</pre> </p> <p>则添加的卷积层中输入层的神经元个数为_____。</p>	32	$3 \times 3 = 9$	$28 \times 28 \times 1 = 784$	无法确定
349	<p>针对 Tensorflow 中 Keras 序列模型，可用模型的_____函数查看模型性能的评价指标，如准确率。</p>	accuracy	precision	evaluate	predict

350	针对 Tensorflow (tf 对象) 中 Keras 序列模型, 处理图像数据的卷积层的构造函数通常为_____。	tf.keras.layers.Conv()	tf.keras.layers.Conv1D()	tf.keras.layers.Conv2D()	tf.keras.layers.Conv3D()
351	设有以下代码: import tensorflow as tf model = tf.keras.Sequential() model.add(tf.keras.layers.Conv2D(16, kernel_size=(2, 2), activation='relu', input_shape=(28, 28, 1))) 则添加的卷积层中输入层的神经元个数为_____。	16	2 x 2	28 x 28 x 1	无法确定
352	针对 Tensorflow (tf 对象) 中 Keras 序列模型 model, 关于语句 model.add(tf.keras.layers.Dropout(0.25)) 不正确的说法是_____。	语句用于添加一个 dropout 层	该语句会以 0.25 的概率随机关闭上一层的一小部分神经元	Dropout 层能够删除模型中的冗余数据, 从而减少计算工作量	Dropout 层能够降低网络对单个神经元的依赖, 以防止任何特定的神经元变得过于强大
353	激活函数_____可将实数范围的自变量值映射到 0~1 之间的输出值, 并且输出值的总和是 1, 其输出解释为“概率”。	sigmoid	relu	softmax	celu

354	当过分强调模型与训练集的符合程度时，模型可能会对未知的测试集中新样本的预测能力降低，导致模型的泛化能力下降，即造成虽然训练误差相对较低但测试误差高的现象，这种称为_____现象。	适度拟合	欠拟合	过拟合	不饱和拟合
355	人工智能分类算法测试中，实际为正的样本中被预测为正样本的比例称为_____。	准确率	精确率	召回率	F1-score 值
356	针对 Tensorflow 中 Keras 序列模型 model, model.fit 方法的参数 validation_split 可指定_____。	训练样本占全部样本的比例	测试样本占全部样本的比例	训练样本用于验证数据集的比例	测试样本用于验证数据集的比例
357	针对 Tensorflow 中 Keras 序列模型 model, model.fit 方法的参数 validation_split=0.2 表示_____。	全部样本中训练样本的比例为 20%	全部样本中测试样本的比例为 20%	训练样本用于验证数据集的比例为 20%	测试样本用于验证数据集的比例为 20%
358	长短期记忆网络的英文缩写是_____。	CNN	RNN	LSTM	NLP
359	目前国标码中收录的一级字库最常用汉字的数量为_____。	八百多个	一千多个	三千多个	六千多个
360	目前主流的中文票据 OCR 识别的流程一般包括_____和文本识别。	文本切割	文本抽取	文本检测	文本查找
361	目前中文票据 OCR 识别中的文本识别算法常用的是_____。	CTPN	EAST	CRNN	DETR
362	Python 开发环境中安装 paddlepaddle 开发包的命令是_____。	install paddlepaddle	uninstall paddlepaddle	pip install paddlepaddle	pip uninstall paddlepaddle

363	Transformer 模型中的_____是其最重要的特点,它能够对序列中的每个位置进行关注,并同时考虑序列中其他位置的信息,从而更好地捕捉序列之间的依赖关系。	编码器	解码器	自注意力机制	预训练网络
364	在使用手机时,如看到屏幕界面右边还留有一部分图片时,会尝试手指左滑查看;当看完手机当前页面时,会尝试用手指划动来翻页。这种交互方式称为_____。	文字交互	语音交互	触屏交互	脑机接口
365	使用轻点、按住、轻扫、滚动和缩放等几个简单的手势动作来控制 iPhone 及其 App,这种交互方式称为_____。	文字交互	语音交互	触屏交互	脑机接口
366	在支持多点触控的屏幕用手指触摸或移动作为输入,属于人机交互的_____方式。	文字交互	语音交互	触屏交互	脑机接口
367	车辆的智能驾驶系统实时监测司机头部姿态、眼睑开闭程度、视线方向以及眨眼频率等,由此来判断司机是否属于疲劳驾驶,这属于人与智能的_____交互方式。	文字识别	语音识别	姿态识别	触屏
368	能够准确对球员发球动作进行回应的球类陪练机器人中通常采用的人与智能的交互方式为_____。	文字识别	语音识别	姿态识别	表情识别
369	ChatGPT 属于_____。	泛问答系统	任务或目标型的对话系统	开放领域的闲聊系统	工业控制系统

370	能够处理语音的智能客服机器人所包含的系统模块一般不包括_____。	语音识别	自然语言理解	搜索引擎	回答语音合成
371	智能客服机器人的工作原理主要包括四个部分：知识库构建、_____、问答匹配和机器人深度学习。	用户输入	AI 模型训练	语义理解	数据挖掘
372	智能客服机器人的通用知识库一般由_____提供。	用户	行业协会	客服机器人的生产厂家	人工智能训练师
373	智能客服机器人的通用知识库中通常不包含_____。	通用对话库	通用本体库	领域场景库	抽象语义库
374	智能客服机器人在回复用户时，通过对比客户提问标签和数据库中问题标签，为用户推荐_____最大的回复信息。	重要性	优先级	相关性	迫切性
375	培训讲义内容的深度与广度应该符合员工什么需求？	知识需求	兴趣需求	实际需求	学术需求
376	哪一步骤可以提高培训对象的学习效果	验证培训内容	加入媒体材料	加入活动	审阅和完善讲义
377	哪些步骤可以确保讲义具有可读性和有效性？	构思培训内容和确定培训对象需求	安排活动和加入媒体材料	编辑讲义和审阅完善讲义	验证培训内容和将内容制定细节
378	常用的培训方法不包括以下哪个？	讲授法	演示法	咨询法	研讨法
379	下列哪个不是研讨法的缺点	讨论课题选择得好坏将直接影响培训的效果	参训人员自身的水平也会影响培训的效果	不利于参训人员系统地掌握知识和技能	容易受到人为因素的影响，缺乏客观性
380	与解决问题能力的培训相适应的培训方法是以下哪个培	讲授法	研讨法	案例分析法	角色扮演法

	训方法?				
381	在指导数据采集和处理时,首先需要了解的是什么?	数据存储技术	数据分析方法	数据采集和处理的步骤	数据应用的行业
382	在指导数据存储时,需要根据什么来确定最合适的方式?	数据分析方法	数据更新频率	数据类型、规模和存储性能	数据应用场景
383	在数据标注流程中,如何确定标注质量?	通过让指导对象自己评估	通过标注员对其它人的标注结果进行检查	检查标注的过程和结果,并提出改进意见,确保标注结果质量	所有选项都是正确的
384	在数据标注流程中,应如何进行标注工具的选择?	直接选择最流行的标注工具	不考虑标注任务难度,选择最简单的标注工具	根据标注目标和实际情况选择合适的标注工具	所有选项都不是正确的
385	人工智能主要的应用领域不包括()。	自然语言处理	机器人自动控制	语音识别与合成	数据库系统
386	在人工智能训练的模型设计阶段,根据人工智能主流算法框架和任务目标选择最优的、()的模型。	训练时间最短的	训练规模最小的	训练成本最低的	与任务匹配的
387	在评估人工智能系统时,下列哪个因素可以衡量系统的可解释性和透明度?	系统的算法复杂度	系统的可视化界面	系统的自主学习能力	系统的决策可追溯性
388	()是智能算法实现的关键。	代码	算力	程序员	知识表示
389	以下哪个过程不是机器学习的关键步骤?	数据收集和预处理	特征选择和提取	模型构建和训练	结果解释和可视化
390	()与“预测”是机器学习的两个过程。	学习	数据预处理	模型优化	训练
391	逻辑回归的输出是什么?	连续值	二元标签	多元标签	概率值
392	主成分分析中主成分的个数通常()原始数据的维度。	等于	大于	不小于	小于

393	多层感知机的基本结构不包括。	输入层	隐藏层	输出层	时间层
394	长短时记忆网络的基本结构包括三个门控，分别是输入门、遗忘门、输出门和一个贯穿整个结构的（）。	全连接网络	特征网络	状态	神经元状态
395	LSTM（长短期记忆网络）相对于普通循环神经网络的主要改进是什么？	更少的计算资源消耗	更高的模型容量	更快的训练速度	更好地处理长期依赖关系
396	知识图谱是一种用于表示和组织知识的数据结构，其基本元素是什么？	节点	边	属性	所有选项都是正确的
397	以下哪个是智能语音的主要应用之一？	文字识别	图像识别	视频处理	语音识别
398	在计算机视觉中，什么是目标检测？	标记图像中的关键点	估计图像的深度	分割图像中的对象	检测和定位图像中的特定对象
399	虚拟现实技术主要通过以下哪种方式创建逼真的体验？	刺激用户的听觉和嗅觉感官	利用激光投影创建立体影像	运用超高分辨率的显示屏幕	利用虚拟现实头戴设备提供全景视觉
400	以下哪个是 SelectFromModel 的常见参数？	n_neighbors	distance	alpha	threshold
401	在 scikit-learn 中，以下哪个方法用于计算分类模型的准确率？	roc_auc_score	mean_squared_error	precision_score	accuracy_score
402	以下哪个深度学习框架被设计用于构建和训练循环神经网络（RNN）？	TensorFlow	PyTorch	Keras	Theano
403	MindSpore 是由哪家公司开发的开源深度学习框架？	谷歌	微软	百度	华为
404	MindSpore 的设计目标是支持哪种类型的深度学习算法？	监督学习	无监督学习	强化学习	All of the above

405	产出物是指业务流程中的什么？	产品	科研成果	排出的废气 废水	活动的输出物
406	采集后的业务数据需要进行什么处理？	传输至其他系统	直接用于业务决策	加密保存以确保数据安全	进行处理以兼顾业务流程和数据特征
407	什么是数据规范化？	将数据转化为一定的分布形式	将数据传输到其他系统	将数据存储于云端服务中	将数据转化为标准的数值范围
408	数据规范化的一个主要好处是：	减少数据重复性	提高数据收集的速度	增加数据的数量	简化数据处理过程
409	文本类业务数据采集的来源有哪些？	只能来自网络。	只能来自软件平台。	只能来自文章。	可以来自网络、软件平台、文章，来源比较广泛。
410	为满足业务指标需求，对文本类数据进行哪些方面审核？	仅审核业务数据的质量	仅审核业务数据的数量	仅审核业务数据是否满足业务指标需求	以上都要审核
411	视觉类数据在存储上的特点是？	以数字方式存储	是纯文字数据	可以直接听到	以点阵方式存储
412	视觉类数据审核包括哪两大类？	主观审核和客户审核	内部审核和外部审核	质量审核和内容审核	主观审核和客观审核
413	哪个领域 PSNR 是目前应用较广的领域？	统计领域	经济领域	安防领域	视频/图像处理领域
414	现实中语音类数据主要靠什么设备进行采集？	PS4 主机	智能手机	音箱	声卡、录音机等常见设备
415	什么设备可以用于语音类数据的采集？	手机	平板	仪器仪表	以上都可以
416	点云数据中每一个点包含什么信息？	二维坐标和角度信息	天空的云朵分布	三角形的位置和颜色信息	三维坐标和颜色信息

417	点云分割的目的是什么？	提取点云的关键点	去除点云数据中的噪声	确定点云的几何和纹理等特征点	将点数据分块，便于单独处理
418	下列哪个不属于智能交互业务的应用？	智能客服	机器人	社交软件	智能空调
419	业务调研的目标对象是什么？	业务流程图	诊断报告	调研报告	人工智能业务系统
420	业务优化流程中的最后一步是什么？	业务调研	提出业务流程优化思路	诊断报告	新业务流程图
421	业务流程优化中的培训和教育的目的是什么？	降低成本	提高效率	简化流程	确保员工理解优化后的流程和具备必要技能
422	中国制定的汽车发展新四化战略方向不包括以下哪项？	汽车电动化	智能化	网联化	飞行化
423	在复杂综合业务流程分析中，发现改进机会的目的是什么？	识别瓶颈和问题	分析数据和问题	增加广告费用	提高效率和效果
424	数据清洗过程中需要进行哪些步骤？	数据去重	数据转换	数据过滤	所有选项都正确
425	以下对数据标注描述正确的是哪一项？	所有的数据标注都可以用脚本语言自动完成	所有的数据标注都必须通过人工完成	OCR 手写转录完全可以通过识别工具自动完成	数据标注是个重复性很强的工作
426	在进行视觉类数据标注时，使用 CVAT 不能实现以下哪个功能？	对图像进行多边形标注	视频标注	导出 VOC 格式的数据集	3D 图形标注
427	数据审核完成后的信息准确率要达到多少以上？	0.85	0.9	0.95	0.999

428	数据清洗和数据标注规范文档对于人工智能项目的重要性体现在哪里？	有助于算法团队进行模型训练	有助于标注团队进行数据标注	有助于保证数据质量和标注质量	以上选项都正确
429	Excel 数据处理支持以下哪种格式的数据？	文本数据	数值数据	日期时间	以上选项都正确
430	制定数据标注规范时需要注意什么？	对标注数据进行观察	找出标注对象的共同特征	定义标注的方法和流程	以上选项都正确
431	在进行文本标注时，使用 doccano 可以实现以下哪些功能？	文本分类	序列标注	序列到序列的标注	以上选项都正确
432	Labelme 能够对图像进行哪些形式的标注？	多边形	矩形	3D 建模软件	以上选项都正确
433	文本类数据清洗规范需要定义以下哪些要求？	文本数据清洗方法和规则	数据输出格式要求	文本类数据清洗质量要求	以上选项都正确
434	文本类数据清洗中哪些错误数据可能导致数据分析结果失真？	空白数据	拼写错误数据	过时数据	以上选项都正确
435	以下哪项不属于文本类数据清洗方法的是？	统一数据规格	过滤不符合项目需求的数据	删除重复项	控制数据样本数量
436	文本类数据清洗中的数据格式检查可以涉及哪些方面？	数值格式	日期时间格式	货币格式	以上选项都正确
437	在文本类数据清洗中，以下哪些方法可以用于处理缺失值？	计算估值进行填充	从其他渠道补充	删除该字段	以上选项都正确
438	文本类数据清洗后的数据质量评估指标中，哪个指标要求数据不能存在重复记录？	完整性	准确性	一致性	唯一性
439	序列标注包括哪些标注类型？	实体标注	词性标注	句法分析标注	属性标注
440	以下哪些属于文本类数据的关系标注？	指向关系标注	平行语料标注	修饰关系标注	以上选项都正确

441	为以下文本标题进行类别标注：注意！这样的房子千万不要买！	房价走势	房产资讯	房产政策相关	房产科普知识
442	实体标注需要注意哪些方面？	实体的数量是否完整	实体的类型是否正确	实体的边界是否统一	以上选项都正确
443	对于关系标注，以下描述哪项是不正确的？	控制实体数量	实体名称不能和关系类型、属性名称重复	关系类型名称不能和实体名称相同	关系类型名称应该包含语义
444	对于情感数据标注规范，以下描述哪项是不正确的？	理解文本时，需要选择一定的标注视角，如客服视角、客户视角等。	④结合上下文理解情绪。	需要结合人们的表达习惯	不需要区分文字表达和口语表达的情绪差别
445	在视觉类数据清洗规范中，哪些图像需要被去除？	损坏的图像	难以处理的图像	清晰度低的图像	所有选项都正确
446	以下哪个选项不是常见的视觉类数据标注类型？	关键点标注	属性标注	区域标注	关系标注
447	区域标注比矩形框标注更加准确，标注边缘可以是什么形状？	圆形	五边形	多边形	所有选项都正确
448	如果说话人第一遍读错了句子，停顿后又正确朗读了该句子，应该如何处理？	将两次朗读都标注	将两次朗读都不标注	只标注第一次朗读	只标注第二次朗读
449	以下哪种异常情况可能导致一段语音被判定为无效语音？	语音中含有短暂的静音	说话人的声音清晰可辨	语音音量较小	语音存在切音、吞音、丢帧、喷麦、重音等异常
450	语音内容转写时可能需要注意以下哪些信息？	说话人的年龄	说话人的身份	说话人的性别	所有选项都正确

451	人工智能算法测试不需要结合哪个方面做针对性的测试?	算法架构	算法应用场景	算法输入/输出数据要求	算法的开发时间
452	人工智能算法测试需要结合以下哪项进行?	人工智能架构	算法	应用场景	所有选项都正确
453	在编写算法测试方案时,需要确认哪些信息?	算法测试场景信息	各场景测试数据需求量	算法测试评价指标及计算公式	所有选项都正确
454	F1-score 值的计算公式是什么?	正确预测的样本数/总样本数	真正正类的样本数/全部预测为正类的样本数	能被算法正确预测为正类的样本数/真正正类的样本数	精确率和召回率的加权调和平均
455	在混淆矩阵中, FN 代表什么?	真正类	假正类	真负类	假负类
456	以下哪个参数表示正样本被算法错误识别,即漏报	TP	TN	FP	FN
457	在图像分类数据准备中,以下哪个说法是正确的?	如果不同分类的图片具有相似性,需要增加更多图片,尽量提升图片数据的丰富度	每个分类的图片需要覆盖实际场景里面的可能性	训练图片和实际场景要识别的图片拍摄环境一致	所有选项都正确
458	在 EasyDL 中创建数据集可以使用以下哪种方式?	本地导入	导入公开数据集	导入平台已有数据集	所有选项都正确
459	在 EasyDL 中,当数据集的标签数量过多,或多个数据集需共享标签时,可以使用以下哪个功能进行管理?	创建数据集	数据标注	智能标注	标签组
460	在 EasyDL 平台上,高性能算法相对于高精度算法具有以	模型体积更大	模型速度更慢	模型效果更好	模型所占用的资源更小

	下哪个特点?				
461	如果更关注模型准确性, 或硬件资源较为充足, 建议选择哪种算法?	均衡算法	随机算法	高性能算法	高精度算法
462	当模型训练结果不理想时, 需要对训练数据进行哪种检查?	检查是否存在训练数据过少的情况	检查不同标签的标注框数量是否均衡	检查测试模型的数据与训练数据的采集来源是否一致	所有选项都正确
463	小明同学将报纸中的一篇文章扫描成图像, 为识别图像中的文字并形成文本文档, 那么他电脑中必须安装的软件是_____。	杀毒软件	Photoshop 软件	FrontPage 软件	OCR 软件
464	早期 OCR 文字识别一般采用_____。该方法模型泛化能力较差, 不能灵活适应字体变化及背景干扰。	卷积神经网络	循环神经网络	Transformer	特征提取方法
465	以下代码的运行结果是_____。 <pre>import numpy as np a=np.array([1, 2, 3, 4, 5, 6, 7, 8, 9]).reshape(3, 3) print(a.shape)</pre>	6	9	(1, 9)	(3, 3)
466	机器学习的聚类分析会根据各样本的特征值将特征相似的样本聚成一类。但由于样本特征值的量纲, 量级都不尽相同, 量级大的特征值会压低甚至排除某些量级低的特征值的作用, 为使特征值的量级统一, 聚类前需进行的操作是_____。	归一化	标准化	正则化	上述三者之一均可

467	在创建 Tensorflow 中 Keras 序列模型后, 测试模型、查看模型的评价指标可用_____函数。	score	accuracy_score	predict	evaluate
468	针对 Tensorflow 中 Keras 序列模型, 用于模型评估的 evaluate 函数, 前两个参数一般是_____。	全部训练集, 全部测试集	全部特征集, 全部标签集	训练集的特征集, 训练集的标签集	测试集的特征集, 测试集的标签集
469	卷积神经网络中 Dropout 层的功能是_____。	通过卷积计算获取图像的高级特征	减少神经网络中的参数数量, 加快计算速度	删除冗余的数据	随机关闭上一层的一小部分神经元, 以防止任何特定的神经元变得过于强大
470	针对 Tensorflow (tf 对象) 中 Keras 序列模型 model, 关于语句 model.add(tf.keras.layers.Dropout(0.25)) 的正确说法是_____。	该语句会随机添加 25% 的训练数据到 dropout 层	该语句会随机删除 dropout 层 25% 的训练数据	该语句会以 0.25 的概率随机复制上一层的一小部分神经元	该语句会以 0.25 的概率随机关闭上一层的一小部分神经元
471	神经网络中常用的激活函数一般不包括_____。	Sigmoid 函数	Tanh 函数	ReLU 函数	Cos 函数
472	以下措施中不能对已有卷积神经网络模型起到局部优化作用的是_____。	增加网络层数	增加池化层或 Dropout 层	改变损失函数	增加迭代次数
473	对于不同类型的模型所采用的评价指标也不尽相同。机器学习中根据任务类型划分、可计算评价指标的模型类别一般不包括_____。	分类	回归	聚类	池化

474	针对 Tensorflow 中 Keras 序列模型 model, model.fit 方法的参数_____可指定训练样本用于验证数据集的比例。	split	data_split	train_test_split	validation_split
475	针对 Tensorflow 中 Keras 序列模型 model, model 的_____方法用于对数据进行预测, 并返回预测结果。	train	fit	predict	predict_classes
476	以下关于 Transformer 网络模型的说法中错误的是_____。	Transformer 最初应用于自然语言处理并取得较好的效果	Transformer 内部由编码器和 Decoder 构成, 支持更快的并行计算	GPT 模型(生成式的预训练模型)就是基于 Transformer	Transformer 只能用于图像识别
477	常见的深度学习网络模型不包括_____。	CNN	RNN	LSTM	NLP
478	假设已正确加载 paddlepaddle 中文票据 OCR 识别预训练模型 ocr, 则用于实施 OCR 识别的方法是_____。	ocr.predict()	ocr.predict_classes()	ocr.recognize()	ocr.recognize_text()
479	以下属于神经网络中的注意力机制的是_____。	模型训练需集中注意力	只关注训练集, 忽略测试集	只关注测试集, 忽略训练集	对于一个词根据其上下文的关系赋予不同的权重, 然后使用该权重向量来表示这个词

480	指在人或动物大脑与外部设备之间创建的直接连接以实现脑与设备的信息交换。常用于辅助、增强、修复人体的感觉-运动功能或提升人机交互能力。	文字交互	语音交互	视觉交互	脑机接口
481	通过在猴子大脑中植入芯片，让猴子能用意念玩游戏，这种人机交互方式属于_____。	文字交互	语音交互	视觉交互	脑机接口
482	交互式人工智能系统的应用领域一般不包括_____。泛问答系统	泛问答系统	任务或目标类型的对话系统	开放领域的闲聊系统	工业控制系统
483	ChatGPT 聊天机器人的模型是以_____网络为基础的。	CNN	RNN	LSTM	Transformer
484	智能客服机器人的工作原理主要包括四个部分：知识库构建、语义理解、_____和机器人深度学习。	用户问题收集	知识库选择	知识点匹配	问答匹配
485	假设要设计一个能够查询话费余额的智能客服机器人，知识库中可添加的关于查询话费余额的相似问法中，一般不包括_____关键字。	余额	话费	多少钱	套餐价格
486	智能客服机器人的语义理解部分，其技术要点除了过滤和纠错、词法分析、上下文处理外，还包括_____。	AI 模型构建	AI 模型训练	语音识别	知识点匹配
487	以_____为代表的深度学习技术出现后，智能客服机器人能够通过与用户互动和互联网数据挖掘自动进行学习，不断积累新知识，完善知识库。	监督学习	无监督学习	半监督	强化学习

488	智能客服机器人能够通过_____，不断积累新知识，完善知识库。	构建知识库	拓展知识库	反复试错	自主深度学习
489	智能客服机器人知识库的组成通常不包括_____。	专属知识库	领域知识库	通用知识库	图像知识库
490	培训讲义编写中，“血肉”是什么？	主题	提纲	媒体材料	素材
491	演示法在培训中发挥的作用不包括哪种	可利用多种感官	可加深对所学内容的印象	容易激发受训者的学习兴趣	可训练体能仪容和言谈举止
492	讲授法缺点中不包括以下哪种？	学习效果易受培训师讲授的水平影响	讲授内容具有强制性	只是培训师讲授，没有反馈	受训者之间不能讨论，不利于促进理解
493	以下哪种培训方法适宜知识类培训的直接传授培训方式。	案例分析法	研讨法	角色扮演法	讲授法
494	选择正确的培训方法不需要考虑以下哪个因素？	目标受众	培训时间	培训效果	战略目标
495	在确定最合适的数据采集和处理工具时，可以考虑什么技术？	暴力破解技术	系统开发技术	计算机网络技术	开放源代码技术

二、多选题

序号	试题	选项 A	选项 B	选项 C	选项 D
1	Transformer 模型的网络结构主要由两个部分组成：（）和（）。	编码器	解码器	输入	输出层
2	在模式识别中，以下哪些是常见的分类算法？	朴素贝叶斯分类器	K 近邻算法	小波变换	傅里叶变换
3	下列哪些是生物特征识别的常见技术？	指纹识别	声纹识别	体温识别	血压测量
4	对哪些因素进行分析处理，可以帮助提高文本类业务数据的质量？	字符编码	干扰数据	数据分析软件	数据存储设备

5	文本类数据可以应用于哪些领域?	文本统计分析	舆情分析	虚拟现实	网络优化
6	视觉类数据预处理有哪些常见方式?	归一化	白化	回归	梯度下降
7	静音、叠音检测包括哪些问题?	是否存在声音重叠现象	音频数据是否为无声	业务数据是否符合业务标准	语音数据的正确语序
8	人工智能对传统行业产生的影响包括以下哪些方面?	新技术的应用	业务流程的优化	新产品的开发	产品价格的下降
9	以下哪个是综合人工智能应用场景?	智慧园区	智慧交通	浏览器软件	电子游戏
10	下面哪些步骤是复杂综合业务流程分析的一部分?	绘制流程图	分析数据和问题	增加广告费用	提高产量
11	在切割语音段时, 如何避免突发噪音?	缩短语音前后的预留静音时间	选取没有突发噪音的语音段进行切割	将语音段切割到前一个语音段的末尾	将语音段切割到后一个语音段的开头
12	OCR 识别中文本检测阶段的工作包括_____。	文本定位, 正确标注出文本框的位置	图像预处理, 包括图像角度矫正和去噪	识别结果矫正	字符识别
13	文本识别算法 CRNN 中采用的神经网络模型包括_____。	CNN	RNN	NLP	Transformer
14	百度地图 APP 可以为用户提供智能推荐交通搭配、预测拥堵、调整出行时间及路线等服务, 其与用户的交互方式可包括_____。	文字交互	语音交互	视觉交互	脑机接口
15	标注质量监控应该从哪些方面进行检查?	标注过程的合理性和正确性	标注结果的可靠性和有效性	标注人员的经验和知识水平	标注工具的技术
16	以下哪些是人工智能训练师应具备的能力?	强大的编程技巧	专业的统计知识	创造性思维能力	良好的心理辅导能力
17	以下哪些学科对人工智能科学的研究具有重要影响?	神经科学	神经科学	仿生学	绘画学

18	通过知识表示使得机器具备了()、()和()的能力	演绎	推理	解决问题	数据采集
19	类脑计算的应用领域包括:	机器人技术	医学诊断	金融风险评估	太空探索
20	下列哪些是常见的自然语言处理技术?	词嵌入	文本分类	关系抽取	遗传算法
21	智能语音技术包括以下哪些主要领域?	语音识别	语音合成	语音增强	人脸识别
22	以下哪些是常见的人机交互技术?	视觉界面	手势识别	脑机接口	视频播放
23	sklearn 中, 评估模型的指标有:	F1 分数	均方误差	召回率	网络层数
24	以下哪些信息可以从混淆矩阵中获得?	Accuracy	Recall	Precision	MSE
25	Keras 提供以下哪些常见的深度学习模型的构建块?	Layers	Optimizers	Loss functions	Datasets
26	产品页面的时间信息包括以下哪些内容?	访问时间	停留时间	页面路径	报错信息
27	单个数据的处理包括哪些方面?	数据类型	数据溢出	数据值的规范化	硬件配置
28	以下哪些类型的数据属于结构化数据?	数值型	日期型	逻辑性	文本型
29	影响视觉类数据的体量大小的因素是?	存储类型	时长	尺寸	字幕多少
30	语音类业务数据的采集质量受什么影响?	环境噪声	辐射畸变	声音时长	节奏和旋律
31	智能交互中, 用户的输入形式通常包括以下哪些?	文字	语音	点击	天气预报
32	自动数据处理的优势包括以下哪些?	提高处理效率	提高准确率	降低成本	提高数据存储能力
33	最优化决策支持的应用场景包括以下哪些?	仓储优化	物流订单分配	交通规划	系统故障维护
34	智能控制模块的应用场景包括以下哪些?	工业机器人	无人驾驶	智能家居	人工智能开发
35	智慧医疗模块通常包含以下哪些智能模块?	图像识别	生理信号判别	数据挖掘	智慧交通

36	智能系统改造法是通过 对现有业务流程进行 以下哪些活动来达 成优化目标?	冗余消除	功能整合	功能拆分	全新设计
37	业务流程优化中的重 设计阶段可能包括	去除冗余 步骤	简化流程	优化资源利 用	培训和教育
38	数据标注规范通常包 括哪些内容?	数据标注 的基本步 骤	数据标注的 方法	数据标注的质 量要求	数据标注的 价值
39	下列哪些属于文本类 数据清洗的内容?	统一数据 规格	删除重复项	纠正错误数据	过滤敏感数 据
40	关系标注中, 哪些内 容不能重复?	实体名称	关系类型名 称	属性名称	业务场景
41	情感数据标注规范 中, 应该注意哪些方 面?	根据项目 需求确定 标注维度	选定标注视 角	区分文字表达 和口语表达的 情绪差别	忽略文本外 部因素
42	倾斜目标框标注同非 倾斜目标框标注一致 的规范有哪些?	物体遮挡	模糊目标	标注框的位置	标注框的大 小
43	EasyDL 平台上可以选 择哪些算法?	高性能算 法	高精度算法	均衡算法	随机算法
44	交互式人工智能系统 的应用领域包括 _____。	泛问答系 统	任务或目标 型的对话系 统	开放领域的闲 聊系统	工业控制系 统
45	智能客服机器人的智 能语义理解阶段, 包 括的步骤一般有 _____。	自然语言 理解	语义分析	结果处理	模型优化
46	研讨法可以采取哪些 形式?	演讲	讨论	沙龙	视听
47	研讨法可带来以下哪 些好处?	有利于激 发学习兴 趣	能够主动提 取出疑问	可帮助提高观 察能力和解决 问题的能力	可训练体能 仪容和言谈 举止
48	在数据采集和处理 中, 以下哪些问题应 该考虑?	数据的有 效应用	数据收集 和记录方法	数据存储场所	数据消费信 息的方式

49	人工智能训练师是指使用智能训练软件，在人工智能产品实际使用过程中进行()、()、()、()及其他辅助作业的人员。	数据集管理	算法参数设置	人机交互设计	性能测试跟踪
50	人工智能主要的应用领域包括()、()、()、()等。	自然语言处理	机器人自动控制	语音识别与合成	计算机视觉
51	下列哪些是人工智能应用?	自动驾驶	人脸识别	机器翻译	智能客服机器人
52	人工神经网络用于解决()、()、()、()等问题。	分类	回归	聚类	识别
53	虚拟现实技术的应用领域包括以下哪些方面?	游戏和娱乐	医疗和康复	建筑和设计	教育和培训
54	以下哪些是开源深度学习框架?	Caffe	Torch	TensorFlow	Keras
55	平台和系统的信息包括以下哪些内容?	系统启动信息	报错信息	错误编号	操作内容
56	以下哪些场合可能会产生大量的业务数据?	金融交易	物流运输	医疗保健	娱乐和媒体
57	在整体处理业务数据时，需要处理哪些方面的内容?	获取特定字段	去除噪声	数据标准化	制定输入输出关系
58	通常的文本数据预处理流程包括哪些步骤?	数据清洗	分词	词性标注	去除停用词
59	点云数据在哪些方面具有特点?	数据量大	带有颜色信息	离散性	立体化
60	常见的三维点云关键点提取算法有哪些?	ISS3D	Harris3D	NARF	SIFT3D
61	智慧医疗的目标是实现以下哪些方面的改善?	提高医疗服务的效率	降低医疗成本	改善医患关系	提高医疗服务的安全性
62	智能汽车的业务涵盖以下哪些方面?	智能汽车软件研发	智能硬件的设计和制造	智能导航	汽车售后服务
63	数据审核员的主要工作内容包括以下哪些?	对标注人员标注的所有数据进行核对和校验	记录标注错误点，并及时反馈标注员	统计审核数量与错误数量	输出审核报告

64	数据清洗工具 Smartbi 支持以下哪些数据预处理功能?	内置排序	去重	映射	去空值
65	Labelme 具有哪些主要功能?	对图像进行多边形标注	支持视频标注	支持 3D 图形标注	可以进行二次开发和优化
66	在文本类数据清洗中, 以下哪些情况需要进行错误值清洗?	文本取值错误	文本格式错误	日期时间格式错误	E. 缺失值
67	文本数据格式检查包括哪些?	数值格式	日期时间格式	货币格式	特殊格式
68	常见的视觉类数据标注类型包括哪些?	关键点标注	矩形框标注	区域标注	属性标注
69	关键点标注可以应用于以下哪些应用?	人脸识别	车辆关键点	人体骨骼检测	手势确认
70	语音内容转写规范包含哪些方面?	噪音	感叹词	数字	标点符号
71	在编写测试数据方案时, 需要确认哪些信息?	数据采集场所	数据采集指标范围	数据采集设备	数据采集人员
72	人工智能算法测试的基本方法包括以下哪些方面?	改变测试集	人机交互测试	增量式测试	调整参数或修改配置文件
73	EasyDL 中可以使用以下哪些人工智能模型?	图像分类	图像分割	物体检测	目标追踪
74	EasyDL 的智能标注支持以下哪些功能?	根据所选模型自动标注	自动筛选出数据集中难例图片	手动确认预标注结果	一键标注难例图片
75	通过优化数据集及标注的方式来提升模型效果时, 数据集可能出现的情况有哪些	错标	漏标	样本不平衡	标签数据量过少
76	人工智能分类算法测试常见的评价指标包括_____。	准确率	精确率	召回率	F1-score 值
77	常见的深度学习网络模型包括_____。	CNN	RNN	LSTM	Transformer
78	人工与智能的交互方式有_____。	文字	语音	视觉	脑机接口

79	引进智能客服机器人能够_____。	降低运营成本	降低人力成本	分担人工客服咨询流量	提升客户的体验
80	能够处理语音的智能客服机器人通常包含的系统模块有_____。	语音识别	自然语言理解	对话管理	回答语音合成
81	智能客服机器人的工作原理主要包括_____。	知识库构建	语义理解	问答匹配	机器人深度学习
82	下列哪些是角色扮演法的缺点？	人为性因素对培训效果有较大的影响	强调个人	容易影响态度、不易影响行为	角色扮演的设计和实施具有较大的挑战
83	哪些因素需要考虑在选择正确的培训方法时？	培训目标	培训时间	培训费用	培训地点
84	在人工智能训练过程中，为什么需要团队协作能力？	提高沟通效率	保证项目进度	加强媒体宣传	改善训练模型质量
85	在知识图谱中，属性通常用于描述实体的哪些特征？	名称	类型	关系	值
86	计算机视觉涉及以下哪些主要任务？	图像分类	目标跟踪	视频压缩	图像修复
87	企业的业务流程包括以下哪个核心流程？	信息流	物流	网络流	资金流
88	文本类数据的处理通常需要使用哪些技术？	自然语言处理	数据挖掘	图像处理	机器学习
89	以下哪些类型的数据属于非结构化数据？	图像	视频	逻辑性	文本型
90	对于语音类数据的预处理，以下哪些是常用的预处理操作？	滤波	降噪	压缩	平滑
91	在视觉类数据清洗中，应该去除哪些图像？	损坏的图像	重复的图像	数据量大的图像	难以处理的图像
92	中文票据 OCR 识别相比英文识别，增加的困难包括_____。	汉字字符数量要多很多	存在着很多形近汉字的组合	汉字输入计算机比英文输入更困难些	汉字字形复杂度较高

93	数据标注目标的设定应该考虑哪些方面?	标注结果的准确性和有效性	可用的数据量和种类	标注工具的技术	标注人员的经验和知识水平
94	TensorFlow 支持以下哪些编程语言?	Python	HTML	C++	JavaScript
95	PaddlePaddle 支持以下哪些编程语言?	Python	R	C++	JavaScript
96	数据标准化的作用是:	转换数据,使其易于处理	将数据数量增加	将数据缩放到特定的范围	将数据转换为分类变量
97	下列哪些步骤是复杂综合业务流程优化的关键步骤?	确定优化目标	增加瓶颈和问题	流程分析和重设计	增加复杂性
98	对于文本类数据标注质量规范,以下哪项的描述是正确的?	情感标注的质量标准则要求对标注句子的情感分类级别正确	标注好的多音字必须包含所有可能的读音	标注好的语义必须标注清楚词语或句子的真实语义	标注好的分词必须考虑文本的上下文
99	对于视觉类数据标注质量规范,以下描述哪些是正确的?	标注像素点越接近标注物的边缘像素,标注质量就越高	标注像素点越接近标注物的边缘像素,标注质量就越低	视觉类数据标注的质量好坏主要取决于像素点的判定准确性。	视觉类数据标注的质量的好坏与标注时间有关
100	假设数组 Z 由以下代码产生,则语句_____可实现对数组 Z 的归一化操作。 import numpy as np Z = np.random.randint(0, 101, (3, 3))	$Z = Z/100$	$Z = Z/9$	$Z = (Z - Z.min()) / (Z.max() - Z.min())$	$Z = Z.Standard()$
101	针对 Tensorflow (tf 对象) 中 Keras 序列模型 model, 关于语句 model.add(tf.keras.layers.Dense(100, activation='softmax')) 的正确说法是_____。	添加 softmax 层, 其输出值采用 softmax 逻辑回归进行分类	添加池化层, 其输出值采用 softmax 逻辑回归进行分类	添加 softmax 层, 输出神经元个数为 100	添加池化层, 输出神经元个数为 100

102	以下属于 AI 云平台的是_____。	百度飞浆 EasyDL	百度网盘	华为云	天猫商城
103	决策树由 ()、 ()、 () 元素构成	根节点	超节点	内部节点	叶节点
104	以下哪些工具是用于数据可视化的开源工具?	Matplotlib	NLTK	seaborn	Plotly
105	PyTorch 提供以下哪些功能?	动态图计算	静态图计算	自动微分	分布式训练
106	语音类业务数据客观审核主要包括以下哪些方面?	语序审核	语音的个人理解评估	对审核语音类业务数据数量的审核	违规内容审核
107	数据清洗的内容包括以下哪些?	处理缺失值	处理重复数据	处理数据量不足	处理数据格式不一致
108	在切割语音段时, 应该注意哪些事项?	以说话人的整句话为单位进行切割	E. 选取含有突发噪音的语音段进行切割	尽可能保留 0.2 至 0.3 秒的静音段	将每个语音段的时间边界设置在语音波形图的最低点
109	以下人机交互方式中属于视觉交互的是_____。	人脸识别	语音识别	表情识别	姿态识别
110	培训讲义编写中, 分析受众包括哪些方面?	技能水平	岗位职责	学习风格	受众需求
111	在数据处理中, 以下哪些技巧可以用来处理数据?	数据转换	数据可视化	数据过滤和清洗	使用机器学习模型
112	视觉类数据包括哪些类型?	视频	文字	数值	图像
113	在文本类数据清洗中, 以下哪些情况可能导致数据不满足有效性指标?	数据对象取值不在预定范围内	数据存在重复记录	E. 数据存在不一致性	数据存在错误或异常
114	长短时记忆网络可以有效地解决传统循环神经网络中的 () 和 () 问题。	准确性	梯度消失	梯度爆炸	过拟合
115	以下哪些深度学习框架提供了动态图 (Dynamic Graph) 的功能?	TensorFlow	PyTorch	Keras	Theano

116	MindSpore 支持以下哪些编程语言？	SQL	Python	C++	JavaScript
117	目标检测任务的标注框应该包含哪些信息？	目标物体的名称	目标物体的位置	目标物体的类别	目标物体的颜色
118	scikit-learn 中的模型构建模块提供了以下哪些常见的分类算法？	KMeans	Random Forest	Logistic Regression	Support Vector Machines
119	AI 模型优化的步骤一般包括_____。	模型创建	模型评价	数据分析	模型调整
120	针对智能客服机器人与客户的交互过程，其给出准确答案的步骤主要包括_____。	AI 模型创建	数据采集	数据处理	答复推荐
121	相较其他类数据，文本类数据更为复杂，主要因为：	文本类数据通常是英文的	文本类数据通常包含较长的字符串	文本类数据通常没有标点符号	文本类数据通常包含较多的特殊符号和代码
122	针对 Tensorflow (tf 对象) 中 Keras 序列模型 model, 关于语句 model.add(tf.keras.layers.Dense(128, activation='relu')) 的正确说法是_____。	添加池化层, 每个神经元的激活函数采用 ReLU 函数	添加全连接层, 每个神经元的激活函数采用 ReLU 函数	添加池化层, 输出神经元个数为 128	添加全连接层, 输出神经元个数为 128
123	下列哪几类人员属于人工智能训练师？	软件开发人员	需求分析人员	数据标注员	人工智能算法测试员
124	在 scikit-learn 中, 以下哪些是基于树模型的特征选择方法？	SelectKBest	RFE	SelectFromModel	ExtraTreesClassifier
125	在业务数据采集中, 需要重点关注哪些因素？	数据采集员	企业效益	业务数据的数量	业务数据的类型
126	业务流程优化中的监测和评估阶段的目的是什么？	确定优化目标	识别瓶颈和问题	检查是否达到预期的优化目标	跟踪流程优化的效果
127	下列哪些情况语音为有效语音？	语音是用规定语言朗读的	整段语音只含有噪声或静音	说话人的声音太大	说话人说话速度较慢

128	混淆矩阵包括哪些维度?	置信度	样本数量	真实值	预测值
129	假设已利用“import tensorflow as tf”语句导入了 tensorflow 模块, 则能够成功构建 Keras 序列模型的语句为_____。	model = tf.Linear_model()	model = tf.keras.Linear_model()	model = tf.keras.Sequential()	model = tf.keras.models.Sequential()

三、判断题

序号	对/错	试题
1		人工智能训练师不需要具备团队协作能力, 独立完成任务就可以了。
2		人工智能科学不需要其他学科的支撑。
3		一般情况下, 用于训练模型的数据不需要任何处理就可以直接使用。
4		在人工智能训练的确定评估指标阶段, 根据任务类别, 选择定性的评估指标。
5		在人工智能训练的模型训练阶段, 需要在新的数据上达到理想的评估指标。
6		在人工智能训练的模型测试阶段, 在验证集上通过评估指标验证模型的效果。
7		机器学习包括人工智能。
8		线性回归中, 输出变量是自变量。
9		KNN 算法对于分类问题, 预测目标数据点的标签为 K 个最近邻居中出现最少的标签。
10		池化层是一种重采样技术, 可以增加数据的空间维度, 从而降低计算复杂度。
11		量子计算是一种能够解决所有经典计算问题的方法。
12		模式识别只适用于图像和声音等传感器数据的处理。
13		生物特征识别是一种身份验证技术, 因此可以不经许可获取用户生物特征。
14		人机交互是指用户向计算机单向传递信息。
15		在 scikit-learn 中, LabelEncoder 用于处理连续数值特征。
16		scikit-learn 中的特征选择模块是基于统计方法的, 不包括基于模型的特征选择技术。
17		scikit-learn 中的 LinearDiscriminantAnalysis (LDA) 方法是一种基于分类方法的降维方法。

18		scikit-learn 中的 KMeans 构建模块是用来构建回归算法的。
19		TensorFlow 和 PyTorch 都是静态图 (Static Graph) 的深度学习框架。
20		深度学习平台包含深度学习开发框架, 不包含模型部署和监控等功能。
21		PyTorch 仅支持 Python 编程语言。
22		PaddlePaddle 深度学习框架仅适用于中文文本处理。
23		MindSpore 是一个基于静态图计算的深度学习框架。
24		单据流是企业业务流程的核心流程之一。
25		在人工智能业务中, 所有的业务数据都是可直接用于模型训练和分析的。
26		业务数据可以直接进行使用, 不需要进行规范化
27		在单个业务数据的处理过程中, 数据可视化和探索性分析是可选步骤, 对最终的数据处理结果没有影响。
28		在文本类数据处理中, 分词和词性标注的主要目的是进行文本数据的加密和解密处理
29		对文本数据进行特征提取之后, 可以进行图像处理操作。
30		环境中出现静音时, 会导致语音数据质量的下降。
31		当音频类业务数据内出现静音时, 需要放大音量来提高声音质量。
32		点云数据能够体现物体的内部结构。
33		智能控制模块通过数据分析技术实现对设备的自动控制。
34		数据质检的核查比例一般为总数据量的 20%-25%。
35		数据标注规范文档只需要按照标注团队的需求制定就可以了。
36		Excel 不能进行数据清洗。
37		文本类数据清洗后, 所有不符合项目需求的数据都应该删除。
38		数据清洗中的非空检查主要针对数值型数据。
39		数据清洗中的错误值清洗只需要考虑文本类型的数据。
40		数据清洗中的缺失值清洗策略不需要考虑缺失值的重要性。

41		中文分词的目的是将一段中文文本划分成一个个单独的字。
42		关系标注主要是对单句话进行句法关联和语义关联的标注。
43		关系标注中，实体数量应该尽量多，以便提升关系标注质量。
44		关系标注中，关系类型名称应该尽量包含语义。
45		情感数据标注规范中，情绪只包含主观和客观两种。
46		相邻三辆汽车可以拉一个框进行标注。
47		属性标注对像素点的判定准确性要求极高。
48		目标检测任务中，标注框的大小和位置对于检测任务的效果没有任何影响。
49		语音清洗是对语音进行重新录制和编辑，以提高语音质量。
50		语音切割规范中提到，如果说话人第一遍读错了句子，停顿后又重复了该句子的朗读，则应该将两次朗读都标注。
51		语音是用规定之外的语言朗读的，不会被判定为无效语音。
52		语音内容转写的基本原则是“所看即所写”。
53		在明确算法测试需求时，只需要明确测试目的及需求。
54		badcase 样例说明及分析的作用是为了指导测试数据采集。
55		准确率是指全部预测为正类中的真正正类的比率。
56		F1-score 值越大，表示模型的性能越差。
57		在混淆矩阵中，FP 代表负样本被算法正确识别。
58		算法性能测试是用来测试算法的正确性。
59		电商网站上将商品描述分类为多个标签，需要选择单标签文本分类
60		EasyDL 不支持自动标注。
61		若智能产品最后是以硬件形式呈现，只需要增加产品的外观设计这项工作。
62		印刷体识别的 OCR 识别相对手写体的 OCR 识别要略微复杂些，因为印刷过程中字体很可能变得断裂或者墨水粘连，增加了 OCR 识别的难度。

63		语句 <code>X_train.reshape(X_train.shape[0], 28, 28, 1)</code> 能将原来训练样本集 <code>X_train</code> 中的每个样本的形状变换为三维张量。
64		针对 Tensorflow (tf 对象) 中 Keras 序列模型, 使用 <code>tf.keras.layers.Conv2D()</code> 构造卷积层, 其中 “activation” 参数用于指定损失函数。
65		卷积神经网络中的 Dropout 层能够删除模型中的冗余数据, 从而减少计算工作量。
66		ReLU 函数能将线性模型输出的实数域映射到 [0, 1] 表示概率分布的有效实数空间, 可用于多分类。
67		F1-score 值是准确率和召回率的加权调和平均。
68		在机器学习中, 模型预测一般是指用模型验证已有标签的测试数据集并计算性能评价指标。
69		AI 云平台中已训练好的内置模型, 只能直接使用, 无法根据任务类型调整参数。
70		循环神经网络 (RNN) 和长短期记忆网络 (LSTM) 比较适合图像处理。
71		Transformer 只能用于图像识别
72		在深度学习中, 卷积神经网络 CNN 最常用于自然语言处理。
73		票据 OCR 识别中的文本检测可直接使用经典的目标检测算法。
74		用于英文票据 OCR 识别的 AI 模型一般也能用于中文票据的识别。
75		CTPN 文本检测算法最后会将文本切割成单个字符。
76		目前中文票据 OCR 识别中使用的文本识别算法 CRNN, 要求识别前先进行单字符切割。
77		文本识别算法 CRNN 中使用 ReLU 损失函数, 将神经网络获取的标签特征分布通过一系列的计算操作转换为真实的预测值。
78		要使用百度飞浆提供的中文票据 OCR 识别接口, 只需要安装 <code>paddlepaddle</code> 开发包, 不用安装其他模块。
79		以手机为例, 人工与智能的交互方式有文字与语音两种。
80		人与智能的视觉交互是以视频识别为特征的交互方式。
81		泛问答系统通常包括自动问答、阅读理解等。属于多轮次的对话系统。

82		设计智能客服机器人，需向领域知识库添加问答条目，一般来说，标准问题需要扩展多个类似的句子，机器人才可以正常使用。
83		如果智能客服机器人与用户的交互方式还包括语音，则先用语音识别理解用户提问，在机器人回答生成后直接以文字形式回复用户，由用户自行阅读。
84		语音识别解决的主要问题就是如何将文字信息转化为可听的声音信息，也即让机器像人一样开口说话。
85		智能客服机器人能够通过以监督学习为代表的深度学习技术进行自主深度学习，不断积累新知识，完善知识库。
86		搭建智能客服机器人的专属知识库一般属于人工智能算法工程师的工作内容。
87		智能客服机器人的通用知识库通常由该客服机器人的用户根据业务内容自行搭建。
88		制定内容的步骤被认为是培训讲义编写的“骨架”。
89		角色扮演只是为了让受训者娱乐并学习如何表演。
90		角色扮演法的主要目的是让受训人员表演，提高娱乐性。
91		对数据进行结果可视化的目的是提高数据安全性。
92		数据集收集的目的是为了增加标注完成速度。
93		标注工具的作用是保证标注结果的质量。
94		人工智能训练师的沟通和表达能力对于与行业人员对接和了解应用需求是有帮助的。
95		人工智能训练师需要具备创新能力来应对不断发展的人工智能技术。
96		人工智能训练师应该具备持续学习和更新知识的能力，具备创新能力。
97		人工智能科学是一门综合学科。
98		人工智能科学是一门综合学科，涉及多个学科的研究。
99		深度学习是现阶段实现人工智能的最常用的方法。
100		在评估人工智能系统时，系统的安全性是一个重要的评估指标。
101		在人工智能训练的模型训练阶段，需要在训练集上达到理想的评估指标。

102		在人工智能训练的模型验证阶段，需要在验证集上达到理想的评估指标。
103		通过知识表示使得机器具备了演绎、推理和解决问题的能力。
104		机器学习“训练”产生“模型”，“模型”指导“预测”。
105		决策树算法采用树形结构，使用层层推理来实现最终的结果分类。
106		人工神经网络是一种模拟人脑神经元之间相互连接的计算模型。
107		每个神经网络层都包含多个神经元，可以通过反向传播算法进行优化。
108		Transformer 模型是一种采用自注意力机制的深度学习模型。
109		知识图谱的目标是准确地模拟现实世界中的所有事物和关系。
110		类脑计算是一种基于仿生学原理的计算方法。
111		智能语音可以实现人机对话，并进行语义理解和回应。
112		虚拟现实和增强现实可以共同创造一个融合虚拟与真实的混合现实体验。
113		增强现实技术可以将虚拟信息叠加在现实世界中，与用户的实际环境进行交互。
114		sklearn 中提供了常用的数据集和数据预处理方法。
115		TensorFlow 2.x 版本引入了 Eager Execution（即时执行）模式，可以动态执行模型的每一步操作，无需构建图。
116		业务流程的梳理是数据分析的基础和前提。
117		物流是商品在交易过程中所有权或控制权的转移
118		在人工智能业务流程构建中，模型训练的核心目标是通过数据进行学习，生成能够对新数据进行预测和推断的模型。
119		数据规范化可以消除数据中的冗余和不一致，提高数据的质量和可信度。
120		结构化数据具有明确定义的数据模式和固定的数据格式，便于存储、管理和分析。
121		个性化推荐系统的目标是提供与用户兴趣相关的产品或服务建议。
122		智能搜索业务可以通过自然语言处理技术来解析和理解搜索查询。
123		智能搜索系统可以利用机器学习算法来提供个性化的搜索结果。

124		自动数据处理能够通过人工智能模型和算力,挖掘出稳定且准确的分析结果。
125		最优化决策支持利用人工智能计算来实现系统的最优性能,以及得出达到最优业务指标的分配或决策。
126		综合人工智能系统中的智能控制模块用于实现设备控制。
127		智慧医疗利用物联网技术实现患者与医务人员、医疗机构、医疗设备之间的互联。
128		业务流程分析的目标是发现或确定流程中的瓶颈和改进机会,以提高效率和效果。
129		业务流程优化是一个长期持续的过程。
130		在复杂综合业务流程分析中,控制图和帕累托图是常用的分析工具。
131		改变流程执行方式是复杂综合业务流程优化的一个方法。
132		监测机制的建立有助于跟踪流程优化的效果。
133		标注框应该紧贴标注对象的边缘
134		数据质检应该按比例随机抽检。
135		在人工智能项目中,数据处理规范文档是必不可少的。
136		文本类数据清洗是对采集的文本类数据进行审查和校验的过程。
137		文本类数据清洗包括统一数据规格。
138		在文本类数据清洗中,非空检查主要是检查数据是否为空值。
139		乱码是一种错误值,需要进行清洗。
140		文本类数据清洗质量要求有完整性、准确性、一致性、唯一性和有效性。
141		新闻分类标注属于文本类数据标注。
142		知识图谱可以用来进行关系标注。
143		属性标注是根据事物的属性进行打标签。
144		情感数据标注规范中,应该区分文字表达和口语表达的情绪差别。
145		视觉类数据清洗规范中,需要去除重复和相似度高的图像。
146		视觉类数据清洗规范中,需要去除模糊图像或视频。
147		标注框不能在图片的边界上。
148		对行人随身携带物品(比如背包、挎包、手提包、行李箱、打伞等)进行标注属于属性标注。

149		标注像素点与标注物的边缘像素点的误差应该在 1 个像素以内。
150		倾斜目标框标注比非倾斜目标框更加贴合目标物体的轮廓。
151		外语相关的语音标注任务需要标注人员具备相应专业知识或者由专业人员完成
152		语音类数据质量不仅与语音本身的特性有关，还与采集设备、采集环境等因素有关。
153		语音类数据质量要求包括语音清晰、无噪音，语段截取正确准确、有逻辑性等。
154		语音存在影响语音清晰度的空旷音、混响等异常，会被判定为无效语音。
155		所有数字应根据实际发音转写为文本，不能写成阿拉伯数字。
156		在算法测试中，修改参数配置文件可能提高算法的精度。
157		在进行图像分类时，每个类别的图像数量应该尽可能保持一致。
158		在准备数据的时候，应该尽可能采集真实场景下的数据。
159		在数据导入到人工智能测试平台前需要进行数据清洗。
160		在 EasyDL 中，创建模型后可持续新增模型版本。
161		在模型训练时，应该根据硬件资源的充足程度来选择合适的算法类型。
162		训练后的模型在正式集成之前，需要评估模型效果是否可用。
163		在 OCR 识别中，卷积神经网络可以取代传统方式中特征提取器和分类器的角色。
164		在 numpy 中，shape 属性和 reshape() 函数的功能都是对于数组的形状进行操作。shape 属性可以了解数组的结构，reshape() 函数可以对数组的结构进行改变。
165		机器学习时，当数据集的特征间具有不同的值范围时，通常采用数据归一化的手段使特征具有相同的度量尺度。
166		Tensorflow 中 Keras 序列模型的 fit 方法，其 validation_split 则指定验证数据集占全部训练样本的比例
167		对于不同类型的 AI 模型，其所采用的评价指标也不尽相同。如分类和回归常用的评价指标就不相同。

168		卷积神经网络中使用池化层可以缩减模型的大小，从而提高计算速度。
169		卷积神经网络的全连接层中，每个神经元与其前一层的所有神经元进行全连接，用来把前边提取到的特征综合起来。
170		全连接层如果输出值采用 softmax 逻辑回归进行分类，该层也称为 softmax 层。
171		激活函数给神经元引入了非线性因素，使得神经网络可以任意逼近任何非线性函数。
172		模型训练中导致过拟合的原因往往是模型过分地考虑了训练集中已知数据的自身特性（如噪声数据），而导致对未知的测试集中新样本的预测能力降低。
173		神经网络中损失函数的作用一般是衡量神经网络的输出与预期值之间的距离，以便控制、调节参数。
174		可从训练数据集中切分出验证集，验证集中的数据将不参与训练模型，而是用于验证训练后的模型，从而增强其泛化能力。
175		使用 EasyDL 平台内置的 OCR 识别模块来识别图片中的文字信息，不需要对模型进行训练就可直接使用。
176		Transformer 最初应用于自然语言处理并取得较好的效果
177		Transformer 内部由编码器和 Decoder 构成
178		深度学习的神经网络，通常是指包含多个隐藏层的神经网络。
179		在深度学习中，循环神经网络 RNN 由于其具有记忆性的特点，常用于自然语言处理等领域。
180		长短期记忆网络 LSTM 是一种特殊的循环神经网络，可以较好地解决长时依赖问题，适合于处理和预测时间序列中间隔和延迟非常长的重要事件。
181		票据 OCR 识别中的文本检测任务是要定位图片中的文本框坐标，对检测图片进行包含文字部分的和不包含文字部分的二分类。
182		EAST 文本检测算法预测的文本框形状可以是常规的水平矩形框或是旋转的文本框。
183		目前中文票据 OCR 识别中主流的文本识别算法均采用基于深度学习的方法。
184		假设 paddlepaddle 中文票据 OCR 识别预训练模型 ocr 的 recognize_text() 方法返回对象 result，则 results[0]['data'] 对应第一幅输入票据识别出来所有文本信息。

185		Attention-OCR 算法是在卷积、循环神经网络提取特征的基础上引入加入注意力机制的编解码模型。
186		Transformer 模型是一种采用自注意力机制的深度学习模型。
187		能够实现脑和外部设备间的双向信息交换的脑机接口，一方面大脑要发出指令实现对外部设备的控制，另一方面外部设备也要不断地给大脑发送各种各样的反馈信息，让大脑及时调整控制策略。
188		人与智能的视觉交互是包括人脸识别、手势识别、表情识别、姿态识别等以图像识别为特征的交互方式。
189		人体姿态识别是指通过计算机视觉技术，对人体在图像或视频中的姿态进行自动识别和分析。姿态识别算法可以对人体的关键点、姿态、动作等进行识别和分析，从而实现对人体运动和行为的自动识别和监测。
190		任务或目标型的对话系统一般需要通过多轮交互实现一个特定的任务或目标。
191		人机交互是人工智能最具挑战性、最综合性的技术之一，涵盖了语义理解、知识表示、语言生成、逻辑与推理等各个方面。
192		ChatGPT 能够通过理解和学习人类的语言来进行对话，甚至能完成撰写邮件、论文、程序代码等通常需要人类才能完成的工作。
193		一般来说，知识库中存储的问答信息越多，涉及的知识越广泛，智能客服机器人能回答的问题也越多。
194		收集问答和建立知识库是实现智能客服机器人的基础。
195		在为智能客服机器人问答知识库添加问答条目时，一般都需要扩展标准问题的相似性问法。
196		如果智能客服机器人与用户的交互方式是语音，则最后一步是将生成的回答合成语音后反馈给用户。
197		智能客服机器人通过自主深度学习来不断积累新知识和完善知识库。其中一种较新的深度学习方法是使用图神经网络来构建推荐算法。
198		智能客服机器人在回复用户时，如果用户的问题在各层级知识库中均未找到匹配，则一般转人工处理。
199		编写讲义时，应该确定培训对象的学习风格。

200		演示法的演示装置移动不方便，不利于培训场所的变更
201		数据采集和处理技术可以使用开放源代码技术。