人工智能数据标注专项职业能力考核规范

**一、定义**

人工智能数据标注，应具备熟练掌握各种数据标注方法、用数据标注工具按照目标数据标注的要求，进行包括图像、语音、文本等在内的各种信息的标注的能力。

**二、适应对象**

运用或准备应用本项能力求职、就业的人员。

**三、能力标准和考核内容**

（一）理论知识考核内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **项目** | **考核范围** | **考核内容** | **考核比重** | **备注** |
| 基础知识 | 一、计算机操作基础与软件应用 | 1.计算机的发展与应用2.计算机基本组成3.计算机操作系统介绍4.计算机软件的概念及分类5.Office（Excel）应用 | 15% |  |
| 二、计算机网络基础 | 1.认识计算机网络2.网络数据通信基础3.互联网的使用 | 15% |  |
| 三、统计学数学基础 | 1.必然现象与随机现象2.随机试验与随机事件3.概率的统计定义4.概率的古典定义5.概率的性质6.概率分布 | 10% |  |
| 专业知识 | 数据标注、机器学习 | 1.数据标注概述2.数据采集与清洗3.数据标注分类4.数据标注质量检验5.数据标注管理6.机器学习相关知识 | 60% |  |

（二）操作技能考核内容

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **项目** | **考核范围** | **考核内容** | **考核比重** | **备注** |
| 数据标注实操 | 数据获取 | 1.图像数据获取2.语音数据获取3.文本数据获取 | 5% |  |
| 数据预览 | 1.图像数据预览2.语音数据预览3.文本数据预览 | 5% |
| 数据导入 | 1.图像数据导入2.语音数据导入3.文本数据导入 | 5% |
| 数据标注 | 图像标注实操 | 1.人脸分类标注2.物品分类标注3.细胞分类标注4.行人标框标注5.车牌标框标注6.人脸标框标注7.细胞标框标注8.人脸描点标注9.车辆多边形标注10.细胞多边形标注11.道路标志标线区域标注12.遥感影像区域标注13.医疗影像区域标注14.工业缺陷检测标注15.其他标注 | 50% |
| 语音标注实操 | 1.语音主角标注（性别）2.普通话标注3.背景音标注4.方言标注5.敏感词标注6.其他标注 | 15% |  |
| 文本标注实操 | 1.语句分词标注2.语义判定标注3.语句词性标注4.主题事件归纳5.情感色彩标注6.文本分类7.文本截取8.其他标注 | 15% |  |
|  | 数据导出 | 1.图像数据导出2.语音数据导出3.文本数据导出 | 5% |  |

**四、考核要求**

（一）申报条件

达到法定劳动年龄，具有相应技能的劳动者均可申报。

（二）考核方式、考核办法及考核时间

（1）考核办法

①考核内容包含理论知识考核和操作技能考核，理论知识考核和操作技能考核均实行百分制，两项成绩均达到60分及以上者为合格。

②理论知识考试采用闭卷考核方式；操作技能考核采用现场实际操作方式。

（2）考核时间

理论知识考试60分钟，操作技能考核90分钟。

（3）评分标准

在两项成绩均合格的基础上，最终得分 = 理论知识考核得分×20% + 操作技能考核得分×80%。

（三）考评人员与考生配比

（1）理论知识考试：每个考试室配备不少于2名监考员，监考员与考生配比为1:30。

（2）操作技能考核：每个试室配备不少于1名考务人员及1个考评组，每个考评组不少于3名考评人员，考务人员和考生配比为1:30，考评组人员和考生配比为1:20。

（四）考核场地及设备

1.场地

（1）每个考生有独立的工位，各工位之间操作间距不小于0.6m。

（2）工位数量满足同时至少50人考试需求。

（3）有良好的通风设施，场内必须干燥。

（4）有用电安全防护与设施。

（5）配备必要的灭火器材和消防设施，无易燃易爆物品。

 2.设备

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **名称** | **规格型号或技术要求** | **数量** | **备注** |
| 1 | 计算机 | 基本配置：操作系统WIN7或WIN10，内存4G，硬盘128G，显示器17英寸 | 每工位1台 |  |
| 2 | 常用数据标注工具、软件 | Office软件LabelImg工具Labelme工具Praat工具BRAT工具（虚拟机） | 每台电脑安装1套 |  |

**五、参考用书**

(一)《数据标注工程》，作者：刘鹏、张燕，出版社：清华大学出版社，书号：ISBN9787302493273，出版时间：2019年6月。