

人工智能应用工程师（初级）考试大纲

（试行版）

人工智能应用工程师是能够利用人工智能相关技术进行应用研发，并开展各类工作的从业人员统称。人工智能应用工程师未来发展大致分为三个方向，并对应了职业技能考评的三个级别：

一是高级技术人员，对应人工智能应用工程师（高级），例如算法研究人员、高级架构人员；

二是业务管理人员，对应人工智能应用工程师（中级），例如产品经理、技术经理、项目经理等；

三是应用技术人员，对应人工智能应用工程师（初级），例如数据标注人员、知识库开发人员、应用开发人员等。

人工智能应用工程师（初级）的主要职能包括：1. 使用标定工具对数据内容（文章、图片、日志等）进行分类和标签标注；2. 调整改进标注结果质量；3. 管理标注进度，保证按期完成；4. 针对负责的优化标注标准和标注平台提出改进建议。

按照以上岗位职能要求，参考人工智能应用领域的行业特点提出了如下考核内容。

一、考核对象

1. 计算机、软件、数学、统计等相关专业学生
2. 人工智能企业的技术开发人员、数据标注人员
3. 人工智能领域相关企业的销售、产品、项目管理等相关岗位从业人员

二、考核目标

1. 掌握人工智能的概念、历史沿革、产业发展情况及各国人工智能发展战略
2. 掌握人工智能的逻辑的概念及特点
3. 掌握机器学习的概念,以及归纳学习和神经网络两种主流学习方法的概念和特点
4. 掌握数据标注的概念、起源及发展
5. 掌握数据采集的概念、分类及相关规范
6. 了解数据分类、数据获取组件的概念
7. 掌握视频采集、文本采集、音频采集的概念、特性及采集方法
8. 掌握数据采集工具的概念、种类
9. 掌握主流数据采集工具的使用方法及采集、调优规范
10. 掌握数据标签的概念、分类
11. 掌握数据标注的规范
12. 掌握数据标签优化规范
13. 掌握数据收集及数据归纳的规范以及数据归纳优化方法
14. 掌握数据标注质量标准及相关的检验方法
15. 掌握单列、多列数据汇总和分析的方法
16. 掌握数据质量的概念及数据分析原则
17. 掌握数据分析工具的使用方法并形成报告
18. 掌握知识库的概念和搭建原则
19. 掌握知识库的搭建规范以及搭建工具
20. 掌握内容管理的概念、原则及相关规范

21. 掌握知识库日常运维的概念及工具的使用方法

22. 掌握数据标注的应用领域，并能够熟练使用数据标注工具进行自动驾驶、智能安防、智能医疗及智能机器人等领域的数据进行标注

三、考核内容

(一) 人工智能概述

1. 人工智能的概念

- 人工智能的概念
- 思维、智能的概念
- 图灵测试的概念，图灵测试的由来以及其与人工智能的关系
- 强人工智能与弱人工智能的区别

2. 人工智能的历史沿革

- 人工智能的早期历史
- 人工智能发展的里程碑：博弈、专家系统、神经计算、进化计算、自然语言处理、生物信息学的概念
- 新千年人工智能的发展特点
- 美国、英国、日本、欧盟等国家的人工智能发展战略
- 我国人工智能发展战略

3. 人工智能产业发展概况

- 自动驾驶技术的发展及应用
- 智能机器人的发展及应用
- 智能医疗的发展及应用
- 虚拟现实和增强现实的发展及应用

- 智能家居的发展及应用
- 无人飞行器的发展及应用

(二) 人工智能中的逻辑

1. 逻辑和表示

- 逻辑和表示的概念
- 逻辑和表示的关系

2. 命题逻辑

- 命题逻辑的概念及特点
- 命题逻辑的基础
- 命题逻辑中的论证方法的概念及主要内容

3. 谓词逻辑

- 谓词逻辑的概念及特点
- 谓词逻辑中的合一的概念
- 谓词逻辑中的反演的概念
- 将谓词表达式转换为子句形式的方法

4. 其他逻辑

- 二阶逻辑的概念和特点
- 非单调逻辑的概念和特点
- 模糊逻辑的概念和特点
- 模态逻辑的概念和特点

(三) 机器学习

1. 机器学习概述

- 机器学习的概念以及机器学习的历史沿革

- 机器学习系统中反馈的作用

2. 归纳学习

- 归纳学习的概念
- 利用决策树学习的方法
- 适用于决策树的不同问题
- 熵的概念及特点

3. 神经网络

- 神经网络的概念及历史沿革
- 人类大脑与神经元模型的关系
- 网络结构的概念及特点
- 感知器学习规则的概念及特点
- 增量规则的概念及特点
- 反向传播的概念及特点
- 实现关注点的概念及其方法，包括模式分析和训练方法
- 神经网络的主要应用领域

(四) 数据标注的起源与发展

- 数据标注的概念
- 数据标注的分类
- 数据标注的流程
- 数据标注的主要应用场景
- 数据标注与人工智能的关系

(五) 数据采集

1. 标注对象

- 主要的数据来源

- 常见的标注数据

2. 数据标注分类

- 图像标注的概念及应用领域

- 语音标注的概念及应用领域

- 文本标注的概念及应用领域

3. 数据采集规范

- 数据的概念、数据的分类以及属性

- 数据的获取组件以及标称的概念及属性

- 数据文本的采集概念及采集规范

- 音频文本的采集概念及采集规范

- 视频文本的采集概念及采集规范

- 数据资源的统筹及调研方法

4. 数据清洗

- 数据清洗的方法

- 数据清洗的流程

- MapReduce 数据去重

5. 数据采集工具；

- 数据探针、网页采集、日志收集、流处理等工具的概念

- 数据采集工具的使用规范

- 数据采集工具日常维护、日志收集、数据库仓库的概念

- 数据采集工具安全规范、流的处理以及信息的检索方法

- 数据采集工具的调试优化方法

(六) 数据归纳

1. 数据标签；

- 数据标签的概念、元数据的标注、语料库的概念
- 数据标注的类型、范围
- 标注数据、连接范围标注、结构化模型标注的概念及使用方法

2. 数据归纳；

- 数据度量、合概率分布的概念
- 数据收集规范、计算出现次数的方法
- 数据矩阵与相异性矩阵的概念
- N次元、标准属性的概念及规范
- 语言模型的使用方法
- 二元属性的邻近性度量方法

(七) 数据分析

1. 数据分析；

- 数据的质量、数据搜索、数据预处理的概念
- 列中包含的值、数据变换与数据离散化的概念
- 数据分析的原理、EXCEL 以及集成数据的清理方法
- 数据分析、基础图标绘制、数据集成的方法
- 问题模板编写、问题报告总结方法

2. 数据标注质量标准

- 图像标注质量标准
- 语音标注质量标准
- 文本标注质量标准

3. 数据标注质量检验方法

- 实时检验
- 全样检验
- 抽样检验

(八) 知识库建设

1. 知识库构建；

- 知识库、数据检索、标量、向量矩阵的概念
- 知识库搭建规范
- 知识库搭建工具的使用方法
- 数据格式化转换方法、向量空间的基本了解

2. 知识点关联；

- 内容管理的概念
- 数据文本合并、随机变量的的概念
- 知识点关联、过程文件撰写、正态分布的方法
- 知识点输入和输出的方法

3. 知识点模板；

- 导数和积分、二元分类的概念
- 线性回归、复合函数与链式法则、窗口效应的概念
- 掌握多元函数、逻辑分布的概念
- 知识点模板的概念
- 知识点模板的设计方法及知识点模板基本管理方法
- 知识录入工具及知识点模板的使用方法

4. 知识库运维；

- 知识库日常维护的概念及核心要素
- 知识库日常维护工具的使用方法
- 知识库日常运维的关键因素
- 知识库日常监控以及统计方法代码实现方法
- 撰写知识库日志报告的常规方法

(九) 数据标注的应用

1. 自动驾驶的数据标注

- 行人图像标注方法
- 人脸数据标注方法
- 车牌图像标注方法

2. 智能安防的数据标注

- 遥感影像标注方法

3. 智能医疗的数据标注

- 医疗影像标注方法

4. 智能机器人的数据标注

- 日常习惯用语的标注方法

*具体考核内容以最终官方发布为准