

工业和信息化人才培养工程培训课程标准

人工智能数据处理工程师

(试行版)

 EIAEC 工业和信息化部教育与考试中心

工业和信息化部教育与考试中心

二〇二二年十二月

说明

为贯彻落实《关于加强和改进工业和信息化人才队伍建设的实施意见》（工信部人〔2022〕138号），立足新发展阶段、贯彻新发展理念、构建新发展格局，工业和信息化部教育与考试中心依据数字技术、智能制造等行业发展人才实际需要，积极整合行业教育资源优势，组织行业专家、教育专家持续研发《工业和信息化人才培养工程培训课程标准》（以下简称“标准”），用于指导工业和信息化人才培养工程相关培训课程建设，高质量推动工业和信息化人才培养工程发展。

《标准》以客观反映现阶段行业技术发展水平和从业人员能力要求为目标，在充分考虑经济发展、科技进步和产业结构变化的基础上，对课程的等级、模块划分进行定义，对培训内容要求、专业能力要求、知识要求和考核权重进行了详细说明。

《标准》组编遵循了有关技术规程的要求，既保证体例规范，又体现以专业活动为导向、以专业技术技能为核心的特点，模块化的结构使其具有根据技术发展进行调整的灵活性和实用性，符合培训工作的需要。

本《标准》编制工作由工业和信息化部教育与考试中心具体组织实施。参与标准编制单位有广州粤嵌通信科技股份有限公司、中国科学院计算机网络信息中心、北京软件和信息服务交易所有限公司、中国软件评测中心、数记科技（北京）有限公司、珠海科技学院、深圳职业技术学院、重庆电子工程职业学院、广东交通职业技术学院、广州番禺职业技术学院、辽宁职业学院、包头职业技术学院。参与编制人有王欣欣、邓人铭、胡良霖、于铁强、李天池、张志强、王兵、张毅恒、刘仁、刘杰、陆晓梅、梁艳春、林世霖、万光海、余明艳、蔡运富、郝丽萍、管明雷。邓人铭和林世霖完成汇编与校稿工作。

本《标准》经工业和信息化部教育与考试中心批准，自颁布之日起施行。

人工智能数据处理工程师

培训课程标准

1 职业概况

1.1 职业名称

人工智能数据处理工程师

1.2 课程定义

本课程面向信息技术行业人工智能相关方向的从业人员，培养其掌握人工智能应用方法和数据分析工具，负责数据的分析、整合、迁移、转换以及实施等数据相关的工作。

1.3 课程等级

本课程共设三个等级，分别为：初级、中级、高级。

1.4 能力要求

具有较强的学习能力、研究分析能力；具有一定的理解、判断和表达能力；具有较强的分析解决问题的能力 and 沟通能力。

1.5 普通受教育程度

高中及以上文化程度（或同等学历）。

1.6 课程培训要求

1.6.1 培训期限

初级课程不少于线上或线下 80 标准学时；中级课程不少于线上或线下 120 标准学时；高级课程不少于线上或线下 160 标准学时。

1.6.2 培训教师

承担初级、中级理论知识或专业能力培训任务人员，应具有相关课程培训经验 1-3 年。

承担高级理论知识或专业能力培训任务人员，应具有相关课程培训经验 5 年以上，或具有相关职业高级专业技术等级、相关专业高级职称二者之一。

1.6.3 培训场所设备

理论知识培训应有可容纳 30 人以上学员的教室，并配有满足教学需要的网络环境和学习软件、设施等。培训所需软件：Python、PyCharm、Jupyter Notebook、MySQL 等。

2 基本要求

2.1 专业守则

- (1) 遵纪守法，爱岗敬业
- (2) 精益求精，勇于创新
- (3) 诚实守信，恪守职责
- (4) 遵守规程，安全操作
- (5) 认真严谨，忠于职守

2.2 基础知识

2.2.1 基础理论知识

- (1) 操作系统基本知识
- (2) 计算机网络基本知识
- (3) 编程基础知识
- (4) 数据结构与算法基本知识
- (5) 数据库基本知识
- (6) 软件工程基本知识
- (7) 人工智能基本知识

2.2.2 技术基础知识

- (1) Python、PyCharm、Jupyter Notebook 环境安装
- (2) 人工智能架构知识
- (3) Python 编程语言开发知识
- (4) PythonWeb 全栈开发知识
- (5) Python 网络爬虫知识
- (6) 数据计算与数据指标知识
- (7) 常用数据分析与挖掘方法
- (8) 人工智能与深度学习的特点

3 课程内容要求

本标准对初级、中级、高级人工智能工程师的专业能力要求依次递进，高级别涵盖低级别的要求。

3.1 初级

课程模块	培训内容	专业能力要求	相关知识要求
1. Python 编程基础	1.1 Python 环境搭建	1.1.1 能掌握下载安装 Python 开发环境的方法 1.1.2 能掌握下载安装 PyCharm 工具的方法 1.1.3 能掌握下载安装 Jupyter Notebook 工具的方法	1.1.1 Python 开发工具的下载与安装 1.1.2 PyCharm 工具的下载与安装 1.1.3 Jupyter Notebook 工具的下载与安装
	1.2 Python 基础语法	1.2.1 能掌握 Python 基本数据类型 1.2.2 能掌握 Python 数据结构 1.2.3 能编写 Python 流程控制 1.2.4 能编写自定义函数 1.2.5 能掌握 Python 异常处理 1.2.6 能读写文件和编写常用操作 1.2.7 能掌握 Python 面向对象	1.2.1 Python 基础知识 1.2.2 Python 数据结构 1.2.3 程序流程控制语句 1.2.4 函数 1.2.5 Python 异常处理 1.2.6 文件基础 1.2.7 面向对象编程
	1.3 Python 高级语法	1.3.1 能掌握进程的开发方法 1.3.2 能掌握线程的开发方法 1.3.3 能掌握线程池的开发方法 1.3.4 能掌握 Python 的 TCP 编程方法 1.3.5 能掌握 Python 的 UDP 编程方法	1.3.1 Python 的进程控制 1.3.2 Python 的线程处理 1.3.3 Python 的线程池操作 1.3.4 Python 的 TCP 编程 1.3.5 Python 的 UDP 编程
2. 人工智能基础	2.1 人工智能概要	2.1.1 能掌握人工智能基本概念 2.1.2 能掌握人工智能分支 2.1.3 能掌握人工智能应用场景	2.1.1 人工智能基本概念 2.1.2 人工智能分支 2.1.3 人工智能应用场景
	2.2 科学计算库	2.2.1 能掌握 NumPy 计算数据方法 2.2.2 能掌握 Pandas 分析数据方法 2.2.3 能掌握 Matplotlib 绘制图像方法	2.2.1 NumPy 常用函数 2.2.2 NumPy 基本运算 2.2.3 NumPy 拆分合并矩阵 2.2.4 Pandas 读取、展示数据 2.2.5 Matplotlib 绘制图像
	2.3 MySQL 数据库基础	2.3.1 能安装配置 MySQL 2.3.2 能操作 MySQL 数据库、表、关系等 2.3.3 能使用 SQL 语句完成数据的增、删、改、查等操作	2.3.1 关系型数据库与非关系型数据库 2.3.2 MySQL 安装与配置 2.3.3 表结构及 DDL 2.3.4 数据类型及 DML 2.3.5 复杂查询及实际案例应用

3.2 中级

课程模块	培训内容	专业能力要求	相关知识要求
1. Python Web 全栈	1.1 HTML 标签与 CSS 样式	1.1.1 能掌握 HTML 主体结构 1.1.2 能掌握 HTML 文字处理标签应用 1.1.3 能掌握 HTML 图片标签应用 1.1.4 能掌握 HTML 链接标签应用 1.1.5 能掌握 HTML 表格标签应用 1.1.6 能掌握 HTML 表单标签应用 1.1.7 能掌握 CSS 样式 1.1.8 能掌握 CSS 盒子模型	1.1.1 HTML 结构 1.1.2 HTML 文字标签 1.1.3 HTML 图片标签 1.1.4 HTML 链接标签 1.1.5 HTML 表格标签 1.1.6 HTML 表单标签 1.1.7 CSS 样式使用 1.1.8 CSS 盒子模型 1.1.9 CSS 选择器

		1.1.9 能掌握 CSS 选择器应用	
	1.2 JavaScript 与 jQuery	1.2.1 能掌握 JavaScript 基本类型 1.2.2 能掌握 JavaScript 分支结构 1.2.3 能掌握 JavaScript 循环结构 1.2.4 能掌握 JavaScript 的 DOM 对象 1.2.5 能掌握 JavaScript 事件操作 1.2.6 能掌握 jQuery 语法 1.2.7 能掌握 jQuery 选择器应用 1.2.8 能掌握 jQuery 事件应用 1.2.9 能掌握 jQuery AJAX 应用	1.2.1 JavaScript 基本类型 1.2.2 JavaScript 分支结构 1.2.3 JavaScript 循环结构 1.2.4 JavaScript 的 DOM 对象 1.2.5 JavaScript 事件操作 1.2.6 jQuery 语法 1.2.7 jQuery 选择器 1.2.8 jQuery 事件 1.2.9 JQuery AJAX
	1.3 Bootstrap	1.3.1 能掌握 Bootstrap 排版样式 1.3.2 能掌握 Bootstrap 表格和按钮 1.3.3 能掌握 Bootstrap 表单和图片 1.3.4 能掌握 Bootstrap 栅格系统 1.3.5 能掌握 Bootstrap 按钮组件 1.3.6 能掌握 Bootstrap 导航与标签组件	1.3.1 Bootstrap 排版样式 1.3.2 Bootstrap 表格和按钮 1.3.3 Bootstrap 表单和图片 1.3.4 Bootstrap 栅格系统 1.3.5 Bootstrap 按钮组件 1.3.6 Bootstrap 导航与标签组件
	1.4 Django 框架	1.4.1 能掌握 Django 环境搭建方法 1.4.2 能掌握 Django MVT 框架 1.4.3 能掌握 Django 模型 1.4.4 能掌握 Django 路由映射 1.4.5 能掌握 Django 视图应用 1.4.6 能掌握 Django 表单应用 1.4.7 能掌握 Django 模板应用 1.4.8 能掌握 Django 插件应用 1.4.9 能掌握 Django 的 Cookie 和 Session 应用	1.4.1 Django 环境搭建 1.4.2 Django MVT 框架 1.4.3 Django 模型 1.4.4 Django 路由映射 1.4.5 Django 的 FBV 视图 1.4.6 Django 的 CBV 视图 1.4.7 Django 表单 1.4.8 Django 模板 1.4.9 Django 插件 1.4.10 Django 的 Cookie 和 Session
	1.5 Linux 基础	1.5.1 能掌握 Ubuntu 系统安装 1.5.2 能掌握 Linux 基本命令 1.5.3 能掌握 Linux 进程管理 1.5.4 能掌握常用服务器 1.5.5 能掌握 Nginx 部署	1.5.1 VMware 安装 1.5.2 ubuntu 系统安装 1.5.3 ubuntu 系统网络配置 1.5.4 Linux 目录结构 1.5.5 Linux 常用指令 1.5.6 Linux 进程管理 1.5.7 ftp、Samba 服务器 1.5.8 DNS、DHCP 服务器 1.5.9 Nginx+uwsgi 部署 Django
2. Python 网络爬虫	2.1 网络爬虫概述	2.1.1 能掌握爬虫概念与原理 2.1.2 能掌握请求头和响应头 2.1.3 能掌握 Cookie 和 Session 应用 2.1.4 能掌握抓包工具 Fiddler 应用	2.1.1 爬虫概念与原理 2.1.2 请求头和响应头 2.1.3 Cookie 和 Session 2.1.4 抓包工具 Fiddler
	2.2 基础爬虫	2.2.1 能掌握 urllib3 模块应用 2.2.2 能掌握 requests 模块应用 2.2.3 能掌握伪装爬虫技术应用 2.2.4 能掌握正则表达式解析方法 2.2.5 能掌握 Xpath 解析方法	2.2.1 urllib3 模块 2.2.2 POST 和 GET 请求 2.2.3 IP 代理伪装网 2.2.4 正则表达式解析 2.2.5 Xpath 解析

		2.2.6 能掌握 Beautiful Soup 解析方法 2.2.7 能掌握保存文件方式	2.2.6 Beautiful Soup 解析 2.2.7 保存到本地或数据库中
	2.3 Scrapy 框架	2.3.1 能掌握 Scrapy 框架安装配置 2.3.2 能掌握 Scrapy 组件的 Download 2.3.3 能掌握 Scrapy 组件的 Spider 2.3.4 能掌握 Scrapy 组件的 pipeitem	2.3.1 Scrapy 框架安装配置 2.3.2 Scrapy 组件的 Download 2.3.3 Scrapy 组件的 Spider 2.3.4 Scrapy 组件的 pipeitem

3.3 高级

课程模块	培训内容	专业能力要求	相关知识要求
1. Python 网络爬虫	1.1 爬取动态渲染信息	1.1.1 能掌握 AJAX 数据的爬虫应用 1.1.2 能掌握 Selenium 爬虫应用 1.1.3 能掌握 Splash 爬虫应用	1.1.1 AJAX 请求 1.1.2 Selenium 自动化爬虫 1.1.3 Splash 爬虫
	1.2 多线程与多进程爬虫	1.2.1 能掌握进程与线程应用 1.2.2 能掌握线程间通信应用 1.2.3 能掌握进程间通信应用 1.2.4 能掌握多进程爬虫应用 1.2.5 能掌握多线程爬虫应用	1.2.1 进程与线程 1.2.2 线程间同步 1.2.3 进程间通信 1.2.4 多进程爬虫 1.2.5 多线程爬虫
	1.3 Scrapy-Redis 分布式爬虫	1.3.1 能掌握 Redis 安装配置方法 1.3.2 能掌握 Scrapy-Redis 模块应用 1.3.3 能掌握 MySQL 数据库应用 1.3.4 能掌握自定义分布式爬虫应用	1.3.1 Redis 安装配置 1.3.2 Scrapy-Redis 模块 1.3.3 MySQL 数据库 1.3.4 自定义分布式爬虫
2. 数据挖掘	2.1 数据处理	2.1.1 能使用 Python 读取数据、查看数据、验证数据 2.1.2 能使用 Python 处理数据异常值、缺失值和重复值等操作 2.1.3 能使用 Python 对数据进行建立索引、数据选取、排序、分组聚合、数据转换、统计分析等操作 2.1.4 能使用 Python 完成数据维规约、数值规约、特征构造等操作	2.1.1 读写不同类型的数据 2.1.2 Pandas 基本操作 2.1.3 转换与处理时间序列数据 2.1.4 使用分组聚合进行组内计算 2.1.5 创建透视表与交叉表 2.1.6 合并数据 2.1.7 数据分组集合 2.1.8 清洗数据 2.1.9 标准化数据 2.1.10 转换数据
	2.2 数据分析	2.2.1 能使用 Python 完成数据质量分析、关联分析、特征分析 2.2.2 能使用 Python 进行线性回归、岭回归 2.2.3 能使用 Python 进行决策树、逻辑回归、支持向量机、贝叶斯等分类预测 2.2.4 能使用 Python 进行 K 均值聚类、密度聚类、期望最大化聚类等聚类分析 2.2.5 能使用 Python 进行回归模型、分类模型、聚类模型的模型评估及参数调优	2.2.1 数据质量分析、关联分析、特征分析 2.2.2 线性回归、岭回归等回归预测 2.2.3 决策树、逻辑回归、支持向量机、贝叶斯等分类预测 2.2.4 K 均值聚类、密度聚类、期望最大化聚类等聚类分析 2.2.5 回归模型、分类模型、聚类模型的模型评估及参数调优

	2.3 可视化	<p>2.3.1 能使用 Matplotlib、Seaborn、pyecharts 等绘图模块或工具绘制柱状图、散点图、饼图</p> <p>2.3.2 能根据业务需求调整绘图参数</p> <p>2.3.3 能根据业务需求和数据可视化结果，撰写相应的数据分析报告</p>	<p>2.3.1 Matplotlib 数据可视化基础</p> <p>2.3.2 Matplotlib 绘图基础语法与常用参数</p> <p>2.3.3 常见图表类型实现（柱状图、散点图、饼图、箱线图）</p> <p>2.3.4 Seaborn 数据可视化进阶</p> <p>2.3.5 Seaborn 绘图基础</p> <p>2.3.6 pyecharts 交互式数据可视化</p> <p>2.3.7 pyecharts 绘图</p>
	2.4 模型构建	<p>2.4.1 能掌握线性模型、神经网络等分类与回归算法原理，并根据业务需求基于 Python 构建相应模型</p> <p>2.4.2 能掌握 K 均值聚类、密度聚类等聚类算法原理，并根据业务需求基于 Python 构建相应模型</p> <p>2.4.3 能掌握关联规则的算法原理，并根据业务需求基于 Python 构建相应模型</p> <p>2.4.4 能掌握智能推荐、时序模式等算法原理，并根据业务需求基于 Python 构建相应模型</p> <p>2.4.5 能根据业务需求基于 Python 完成文本分类、聚类等文本挖掘任务</p> <p>2.4.6 能根据业务需求基于 Python 的词典、主题模型等方法完成文本情感分析任务</p>	<p>2.2.1 线性模型、神经网络等分类与回归算法原理与 Python 实现</p> <p>2.2.2 K 均值聚类、密度聚类等聚类算法原理与 Python 实现</p> <p>2.2.3 关联规则的算法原理与 Python 实现</p> <p>2.2.4 智能推荐、时序模式等算法原理与 Python 实现</p> <p>2.2.5 Python 实现文本分类</p> <p>2.2.6 Python 实现文本聚类</p> <p>2.2.7 基于 Python 的词典、主题模型的文本情感分析实现</p>
3. 深度学习实战	3.1 数据处理	<p>3.1.1 能使用 Python 实现语音、图像、文本等数据的探索分析</p> <p>3.1.2 能使用 Python 实现语音、图像、文本等数据的清洗</p> <p>3.1.3 能使用 Python 实现语音、图像、文本等数据的标准化</p> <p>3.1.4 能使用 Python 实现语音、图像、文本等数据的特征选择与构造</p>	<p>3.1.1 语音、图像、文本等数据的探索分析</p> <p>3.1.2 语音、图像、文本等数据清洗</p> <p>3.1.3 语音、图像、文本等数据标准化</p> <p>3.1.4 语音、图像、文本等数据的特征选择与构造</p>
	3.2 模型构建	<p>3.2.1 能熟悉卷积神经网络、循环神经网络、长短时记忆网络等算法流程并能够根据业务需求基于 TensorFlow 实现相应模型构建</p> <p>3.2.2 能使用 TensorFlow 完成图像处理、图像识别、语音识别、自然语言处理等</p> <p>3.2.3 能使用 TensorFlow 搭建长短时记忆网络模型，完成文本分类、情感分析等</p> <p>3.2.4 能使用 TensorFlow 搭建序列</p>	<p>3.2.1 深度学习概念</p> <p>3.2.2 常见神经网络原理与实现</p> <p>3.2.3 卷积神经网络</p> <p>3.2.4 循环神经网络</p> <p>3.2.5 长短时记忆网络</p> <p>3.2.6 深度学习常见任务实现</p> <p>3.2.7 图像处理</p> <p>3.2.8 图像识别、语音识别</p> <p>3.2.9 文本分类</p> <p>3.2.10 情感分析</p>

		到序列模型完成语音翻译、中英文翻译等	3.2.11 语音翻译、中英文翻译
	3.3 模型优化与部署	3.3.1 能使用 TensorFlow 等深度学习框架实现模型搭建 3.3.2 能使用 TensorFlow 等深度学习框架实现模型评价与优化 3.3.3 能实现业务系统的模型部署与维护	3.3.1 模型部署 3.3.2 模型维护

4 考核权重表

4.1 理论知识权重表

课程模块 \ 级别		初级 (%)	中级 (%)	高级 (%)
基本要求	专业道德	5	5	5
	基础知识	15	10	10
理论知识要求	人工智能基础	30	-	-
	Python 编程基础	30	-	-
	PythonWeb 全栈	-	30	-
	Python 网络爬虫	-	30	20
	数据挖掘	-	-	40
	深度学习实战	-	-	15
	人工智能实战	20	25	10
合计		100	100	100

4.2 实操能力权重表

课程模块 \ 级别		初级 (%)	中级 (%)	高级 (%)
实操能力要求	人工智能基础	30	-	-
	Python 编程基础	30	-	-
	PythonWeb 全栈	-	35	-
	Python 网络爬虫	-	35	20
	数据挖掘	-	-	30
	深度学习实战	-	-	20
	人工智能实战	40	30	30
合计		100	100	100

附录

1 .术语和定义

国家、行业标准界定的以及下列术语和定义适用于本文件。

(1) 机器学习 Machine learning

功能单位通过获取新知识或技能，或通过整理已有的知识或技能来改进其性能的过程。

(2) 深度学习 Deep learning

深度学习是机器学习的分支，是一种以人工神经网络为架构，对数据进行表征学习的算法。

(3) 自然语言 Natural language

一种其规则是基于当前的用法且无需特别规定的语言。

2 .参考文献

- [1] GB/T 37721-2019 信息技术 大数据分析系统功能要求
- [2] GB/T35295-2017 信息技术 大数据 术语
- [3] GB/T35589-2017 信息技术 大数据 技术参考模型
- [4] T/CESA1040—2019 信息技术 人工智能 面向机器学习的数据标注规程
- [5] T/CESA 1039-2019 信息技术 人工智能 机器翻译能力等级评估