

人工智能技术（数据分析师）

职业能力等级评价标准

（试行稿）

1 项目概况

1.1 项目名称

人工智能技术（数据分析师）

1.2 项目定义

从事运用人工智能算法、机器学习模型及数据分析工具，进行数据采集、清洗、建模、分析与可视化等工作的专业人员。

1.3 能力等级

本项目共设三个等级，分别为：初级、中级、高级。

1.4 能力特征

具备运用人工智能算法与机器学习模型，针对不同业务场景设计数据处理与分析架构，完成数据源筛选、数据清洗策略制定及分析模型选型的能力。具备使用 Python、SQL 等工具进行数据采集、存储及预处理，以及对数据可视化工具进行配置联调，确保分析流程正常运行的能力。具备运用模型评估指标与数据分析工具诊断模型偏差、数据异常等问题，定位算法缺陷、数据质量等根源并实施修复的能力。具备制定数据分析项目计划，协调算法、开发、业务等多方资源，把控分析进度、结果质量与资源成本的能力。

1.5 职业能力等级评价要求

1.5.1 申报条件

具备以下条件之一者，可申报初级：

- （1）累计从事相关职业工作 1 年（含）以上。
- （2）相关专业在校学生。

具备以下条件之一者，可申报中级：

- （1）取得本项目或相关职业初级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作 2 年（含）以上。
- （2）累计从事相关职业工作 4 年（含）以上。
- （3）取得相关专业毕业证书。

具备以下条件之一者，可申报高级：

(1) 取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作 3 年（含）以上。

(2) 累计从事相关职业工作 6 年（含）以上。

(3) 具有高等职业学校、高级技工学校、技师学院相关专业毕业证书，并取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）。

(4) 具有大专及以上学历相关专业毕业证书，并取得本项目或相关职业中级评价证书（含职业资格证书、职业技能等级证书等）后，累计从事相关职业工作 1 年（含）以上。

1.5.2 申报条件注释

(1) 满足本项目高级别申报条件可申报本项目低级别。

(2) 相关职业:智能交互产品设计相关职业。

(3) 参考《普通高等学校高等职业教育专业目录（2021 年）》涉及相关专业的归类¹。

1.5.3 评价方式

职业能力等级评价考试包括理论知识、技能操作两个科目，较高等级必要时可增加综合评审。

理论知识考试以笔试为主，可以机考，条件成熟时试点开展网络考试，主要考核从业人员从事本职业应掌握的基本要求和相关知识要求。技能操作考核主要采用现场操作、模拟操作、面试答辩等方式进行，主要考核从业人员从事本职业应具备的技能水平。综合评审通常采取审阅申报材料、技术答辩等方式进行全面评议和审查。理论知识考试和技能操作考核均采用百分制，成绩达到 60 分以上者为合格。

理论知识考试和技能操作考核均实行百分制，成绩皆达 60 分（含）以上者算合格。

1.5.4 监考人员、考评人员与考生配比

理论知识考试和技能操作考核中的监考人员与考生配比不低于 1: 15，且每个考场不少于 2 名监考人员。技能操作考核中考评人员为 3 人以上单数。

1.5.5 评价时间

理论知识考试时间不少于 90 分钟；技能操作考核时间:初级不少于 90 分钟，中级/高级不少于 120 分钟。

1.5.6 评价场所设备

理论知识考试：在标准教室或标准联网多媒体计算机教室进行。

¹ 根据《普通高等学校高等职业教育专业目录（2021年）》，涉及的相关专业有人工智能技术应用（510209）、大数据技术（510205）、云计算技术应用（510206）、信息安全技术应用（510207）、物联网应用技术（510102）、软件技术（510203）、计算机网络技术（510202）、数字媒体技术（510204）、工业互联网技术（510211）、区块链技术应用（510212）、电子信息工程（080701）、电子科学与技术（080702）、通信工程（080703）、人工智能（080717T）、计算机科学与技术（080901）、软件工程（080902）、信息安全（080904K）、物联网工程（080905）等。

技能操作考核：在配备必要办公软件及技能操作所需的工具的联网多媒体计算机教室进行，考试结束后能还原考试环境。



工业和信息化部教育与考试中心
EDUCATION & EXAMINATION CENTER OF MINISTRY OF INDUSTRY AND INFORMATION TECHNOLOGY

2 基本要求

2.1 职业道德

2.1.1 职业道德的基本知识

- (1) 遵纪守法，诚信从业。
- (2) 安全第一，预防为主。
- (3) 质量至上，精益求精。
- (4) 保密义务，尊重隐私。
- (5) 公平竞争，合作共赢。
- (6) 遵循伦理，合法合规。

2.1.2 职业守则

- (1) 专注专业，持续学习。
- (2) 勇于创新，严谨负责。
- (3) 客户至上，服务优质。
- (4) 遵规守纪，安全合规。

2.2 基础知识

2.2.1 人工智能与大模型理论知识

- (1) Transformer 模型架构及注意力机制的核心原理与运行逻辑。
- (2) 扩散生成模型的核心原理与运行逻辑。
- (3) 大语言模型、多模态模型的结构组成与工作机制。
- (4) 大模型数据处理流程、训练算法及模型评估指标体系。
- (5) 模型量化、剪枝和蒸馏等压缩技术的原理与适用场景。
- (6) 分布式训练与推理的实现方式及数据同步方法。
- (7) 模型参数计算、计算复杂度分析与性能评估方法。

2.2.2 统计学与机器学习理论知识

- (1) 正态分布、二项分布等概率分布知识
- (2) 均值、方差、分位数等描述性统计指标
- (3) T 检验、卡方检验等假设检验知识
- (4) 置信区间的核心原理与业务场景
- (5) 线性回归、逻辑回归、K-Means 等监督与无监督学习算法
- (6) 解决过拟合与欠拟合现象的方法

2.2.3 数据分析工具与工程技术知识

- (1) 窗口函数、CTE（公共表表达式）等 SQL 高级查询与数据转换技巧
- (2) Pandas、NumPy 等库的核心数据结构与操作方法
- (3) BI 智能工具的主要功能及操作方法
- (4) Hadoop、Spark 等大数据技术的基本架构与工作原理
- (5) 维度建模、事实表、维度表等数据仓库的基本概念

2.2.4 数据治理与安全知识

- (1) 数据治理的核心框架，
- (2) 数据治理工作的实施方法。
- (3) 数据质量的评估、监控及提升的基本方法。
- (4) 数据加密、数据脱敏、访问控制等数据安全技术的应用原则。

2.2.5 项目管理知识

- (1) 敏捷项目管理在 数据分析中的应用场景
- (2) 创新设计思维在数据分析项目中的应用
- (3) 数据分析项目风险管理与应急预案
- (4) 跨领域、跨团队的数据分析项目中资源协调与管理策略相关知识
- (5) 知识产权保护与商业价值评估相关知识

2.2.6 法律法规知识

- (1) 《中华人民共和国个人信息保护法》中与人工智能技术相关的知识。
- (2) 《中华人民共和国网络安全法》中与人工智能技术相关的知识。
- (3) 《中华人民共和国数据安全法》中与人工智能技术相关的知识。
- (4) 《中华人民共和国知识产权法》中与人工智能技术相关的知识。

3 工作要求

本标准初级、中级、高级的技能要求和相关知识要求依次递进，高级别涵盖低级别的要求。

3.1 初级

职业功能	工作内容	技能要求	相关知识要求
1. 需求分析	1.1 业务需求调研	1.1.1 能使用WPS、Excel等工具模板，记录与业务部门沟通的业务需求，并生成结构化的需求文档和会议纪要 1.1.2 能使用Excel、WPS等工具，生成需求访谈纪要，基于“用户增长率”、“转化率”等业务指标生成可支持绩效度量的“指标清单”	1.1.1 基础业务指标的定义与业务含义 1.1.2 数据分析的基本流程
	1.2 数据现状勘察	1.2.1 能根据“指标清单”，查询相关表和字段，生成初步的“数据源映射表” 1.2.2 能使用豆包、Deepseek等LLM（大语言模型）查询与业务指标相关的表和字段，验证并补充“数据源映射表”	1.2.1 关系型数据库的基本概念 1.2.2 数据字典与元数据的作用
	1.3 技术可行性评估	1.3.1 能在高级分析师指导下，使用WBS方法拆解分析任务，生成“任务步骤分解表” 1.3.2 能使用Excel、WPS等表格，通过任务分解表估算工时，生成“工时估算报告”	1.3.1 数据分析项目中各环节的主要工作内容 1.3.2 工作量估算的方法
2. 数据准备	2.1 数据采集规划	2.1.1 能使用人工智能辅助生成SQL语句，执行数据提取操作，获得原始数据集 2.1.2 能使用SQL客户端，将查询到的原始数据集导出为CSV、Excel等格式文件	2.1.1 SQL的基础查询语法 2.1.2 数据库客户端的数据导出功能
	2.2 数据清洗与预处理	2.2.1 能使用Excel和WPS表格的IF、VLOOKUP等函数完成去重和空值填充，生成干净数据集 2.2.2 能使用Kimi、Deepseek等LLM（大语言模型）分析复杂数据问题，依据建议修正数据集	2.2.1 数据清洗的基本方法 2.2.2 常见数据质量问题的类型
	2.3 数据架构设计	2.3.1 能Kimi、Deepseek等LLM（大语言模型）询问在电子表格中合并数据的方法	2.3.1 数据关联与合并的基本概念

		法，并根据其提供的步骤或公式，完成多源数据的关联与整合 2.3.2 能使用Excel的透视表功能，生成多维度数据摘要表	2.3.2 据透视表的原理与构成要素
3. 数据分析	3.1 模型选定	3.1.1 能根据分析目标，从对比分析、分组分析等基础分析方法中，选择出最适合的分析方法 3.1.2 能使用Kimi、Deepseek等LLM（大语言模型）建议的多种分析路径，生成“分析路径方案”	3.1.1 描述性统计分析的核心概念（如均值、中位数、众数等） 3.1.2 常用基础分析方法的适用场景（如对比法、分组法、结构分析法等）
	3.2 功能开发	3.2.1 能根据选定的分析路径，使用Kimi、Deepseek等LLM（大语言模型）进行统计计算，生成初步分析结果 3.2.2 能使用Excel、WPS等电子表格软件，对人工智能给出的计算结果进行交叉验证，并输出经过核查的“分析结果数据表”	3.2.1 人工智能生成内容的常见谬误类型（如事实错误、逻辑不一致等） 3.2.2 数据交叉验证的基本逻辑与方法
	3.3 成果交付	3.3.1 能向Kimi、Deepseek等LLM（大语言模型）清晰地描述数据和期望的图表类型，以生成基础的可视化图表 3.3.2 能将人工智能生成的图表和针对图表的解读整合进报告模板中，完成“数据分析报告”	3.3.1 常用可视化图表的含义和适用场景 3.3.2 结合文本与图像进行多模态提问的基本方法

EDUCATION & EXAMINATION CENTER OF MINISTRY OF INDUSTRY AND INFORMATION TECHNOLOGY

3.2 中级

职业功能	工作内容	技能要求	相关知识要求
1. 需求分析	1.1 业务需求调研	1.1.1 能独立主持与业务部门的需求研讨会，将业务目标转化为可量化的分析指标，并形成“指标规范文档” 1.1.2 能围绕“指标规范文档”，设计并撰写“数据分析方案”	1.1.1. 常用业务分析模型的构成与应用场景 1.1.2 所在行业的核心业务逻辑和商业模式
	1.2 数据现状勘察	1.2.1 能针对“数据分析方案”中的核心指标，对候选数据表进行内容、范围探查，并形成“数据现状评估纪要” 1.2.2 能根据探查结果，识别出潜在的数据质量问题，并形成“数据现状评估纪要”	1.2.1 数据剖析的核心技术与指标 1.2.2 数据时效性、覆盖率、一致性的定义与评估标准
	1.3 技术可行性评估	1.3.1 能使用WBS（工作分解结构）工具分解数据分析项目，生成“任务分解列表” 1.3.2 能根据“任务分解列表”，对数据分析实施总周期进行准确预估，并提出包含应对预案的“风险清单”	1.3.1 项目管理基础知识 1.3.2 风险管理的基本流程与矩阵
2. 数据准备	2.1 数据采集规划	2.1.1 能编写JOIN连接和子查询的SQL语句，整合来自多个数据表的信息，生成分析所需的宽表数据集 2.1.2 能根据分析需求，向数据工程团队提出清晰的数据提取需求，并输出包含表、字段和数据采集频率的“数据需求单”	2.1.1 SQL的多表连接（JOIN）与子查询语法及原理 2.1.2 ETL/ELT流程的基本原理。数据集成中增量与全量抽取的方法
	2.2 数据清洗与预处理	2.2.1 能使用Python编写脚本，完成复杂的数据清洗和特征工程，生成可供建模的干净数据集 2.2.2 能将包含复杂问题的代码片段粘贴给LLM（大语言模型），并根据其优化建议，完成对数据处理脚本的重构	2.2.1 Python科学库的用法 2.2.2 数据处理脚本的重构方法
	2.3 数据架构设计	2.3.1 能根据维度建模理论，独立编写高效的多维分析查询，以获取所需的聚合数据 2.3.2 能根据业务分析需求，向数据工程师提出构建分析专用数据集的具体需求，并输出“专用数据集需求文档”	2.3.1 数据仓库的维度建模理论。（钻取、切片、旋转） 2.3.2 OLAP（联机分析处理）的核心分析操作

3. 数据分析	3.1 模型选定	3.1.1 能根据分析目的（如归因、分类），从多维分析、漏斗分析等多种分析模型中，选择并组合出最合适的分析框架 3.1.2 能通过“角色扮演”等提示技巧，引导LLM（大语言模型）生成多个维度的业务假设，并为验证这些假设选择相应的分析方法，形成“假设验证方案”	3.1.1 多维分析的核心方法（钻取、切片、旋转）与漏斗分析模型 3.1.2 Role-playing（角色扮演）与Few-shot Prompting（少样本提示）的提示词设计模式
	3.2 功能开发	3.2.1 能根据“假设验证方案”，独立编写窗口函数等复杂SQL查询语句，以完成用户行为路径、漏斗转化等分析 3.2.2 能编写Python脚本，进行探索性数据分析，并输出“探索性数据分析报告”	3.2.1 用户行为路径分析与漏斗分析模型的原理与SQL实现方法 3.2.2 Python核心数据分析与可视化库（如Pandas, Matplotlib, Seaborn）的使用方法
	3.3 成果交付	3.3.1 能独立设计数据故事，并使用Python编写脚本，生成清晰表达分析过程与结论的静态图表 3.3.2 能独立撰写完整的“专项分析报告”，并清晰地呈现从问题定义到最终结论的全过程	3.3.1 高级可视化图表（如箱线图、热力图）的制作方法 3.3.2 数据故事讲述的技巧与原则

3.3 高级

职业功能	工作内容	技能要求	相关知识要求
1. 需求分析	1.1 业务需求调研	1.1.1 能结合跨领域业务知识，设计全面的指标体系，输出“业务指标体系白皮书” 1.1.2 能基于行业趋势和业务洞察，主动挖掘分析课题，撰写包含商业价值和实施计划的“商业分析计划书”	1.1.1 所在行业及相关领域的商业分析方法论 1.1.2 构建公司级指标体系的方法论
	1.2 数据现状勘察	1.2.1 能针对业务域，通过系统性地使用数据地图或血缘追踪工具，梳理出核心的数据资产分布和关键链路，并形成“业务数据资产地图” 1.2.2 能主动评估关键数据的可获得性与可信度，并对数据质量问题可能给分析结论带来的影响进行量化预警，输出“数据风险预警报告”	1.2.1 数据血缘与数据地图在分析工作中的应用 1.2.2 分析结论的敏感性与稳健性检验方法
	1.3 技术可行性评估	1.3.1 能主导大型、跨部门数据分析项目的可行性评估，并输出包含商业价值、技术难度、资源投入等维度的“综合决策报告” 1.3.2 能使用成本效益分析方法，量化分析项目的预期ROI（投资回报率），并形成“项目ROI分析报告”	1.3.1 敏捷项目管理的核心概念与流程 1.3.2 成本效益分析方法
2. 数据分析	2.1 数据采集规划	2.1.1 能从业务分析的视角，评估数据采集方案的完整性和合理性，并向数据工程团队提供书面的“数据采集建议” 2.1.2 能定义数据产品的核心业务指标（如数据时效性、覆盖率），并与数据工程师协作推动监控仪表盘的建立	2.1.1 业务需求到数据埋点的映射与设计原理 2.1.2 数据质量的核心评估维度
	2.2 数据清洗与预处理	2.2.1 能针对用户反馈文本、网页长文等高度非结构化数据，设计并实现数据清洗与特征提取流水线 2.2.2 能通过评估不同的异常值处理方法，选择最适合当前业务场景的技术，并阐明其对后续分析可能产生的影响	2.2.1 数据清洗技术与数据插补方法 2.2.2 异常值检测的统计方法
	2.3 数据架构设计	2.3.1 能针对用户行为、商品生命周期等业务主题，设计逻辑清晰、易于理解的数据分析模型，并输出“逻辑数据模	2.3.1 业务过程到数据实体与关系的抽象建模原理 2.3.2 ERD（实体-关系图）

		型设计文档” 2.3.2 能与数据工程师协作，基于设计好的逻辑数据模型，在数据仓库中建立可供多人复用的分析专用数据集	的绘制规范与核心要素
3. 数据分析	3.1 模型选定	3.1.1 能针对预测、归因等复杂的业务问题，通过评估不同统计学或机器学习模型的优劣，选择最能揭示业务规律的模型，并撰写“分析模型选型报告” 3.1.2 能主导完整的A/B测试实验方案的设计，并产出详细的“A/B测试实验设计方案”	3.1.1 常用机器学习预测模型的原理 3.1.2 实验设计的原理与统计功效分析
	3.2 功能开发	3.2.1 能依据“分析模型选型报告”，使用Java、Python等开发语言完成数据建模、训练和评估的全过程，并输出“模型评估报告” 3.2.2 能为分析任务，设计CoT(思维链)提示词模板，并将其作为可复用资产纳入团队知识库	3.2.1 CoT（思考链）提示词框架的原理与应用 3.2.2 构建和维护团队级Prompt（提示词）知识库的方法
	3.3 成果交付	3.3.1 能面向公司高层，制作并呈现富有洞察力和影响力的“战略数据分析报告” 3.3.2 能将数据洞察转化为可落地的商业策略，并设计用于追踪策略效果的指标体系，形成“策略效果追踪方案”	3.3.1 商业演示的结构与叙事技巧 3.3.2 战略咨询与商业案例分析的基本框架

4 权重表

4.1 理论知识权重表

项目		技能等级		
		初级	中级	高级
		(%)	(%)	(%)
基本要求	职业道德	5	5	5
	基础知识	20	20	20
相关知识	需求分析	10	10	10
	数据准备	30	25	20
	数据分析	35	40	45
合计		100	100	100

4.2 技能要求权重表

项目		技能等级		
		初级	中级	高级
		(%)	(%)	(%)
技能要求	需求分析	20	20	20
	数据准备	35	40	40
	数据分析	45	40	40
合计		100	100	100