



## HCIE-AI Solution Architect V1.0 版本说明

版本	日期	说明
V1.0	2024年6月28日	启用V1.0培训材料、实验手册及题库。

### 一、培训教材变更说明

培训教材：

知识点	V1.0	说明
1.大模型业务场景流程综述	10%	AI 应用发展现状 大模型分类和特点 主流大模型介绍 大模型应用 大模型训练及推理流程介绍 大模型业务流程
2.大模型应用的数据	3%	大模型数据集与小模型数据集的区别 数据集标注工具 数据类型和深度学习中的数据结构
3.数据工程	6%	LLM 训练数据及预处理流程介绍 数据预处理工具 微调数据构造 Tokenizer
4.大模型算法架构	18%	深度学习算法发展及瓶颈 Transformer 详解 Transformer 中的问题及优化方式 大语言模型架构介绍 多模态大模型架构介绍 MoE 结构介绍
5.AI Agent 发展趋势	2%	AI 发展之路 大模型的局限性 AI Agent 的优势与发展 模型协同技术与案例
6.昇腾大模型解决方案	7%	大模型并行计算 CANN 介绍 MindX 介绍 MindStudio 介绍 算力运营
7.MindSpeed 领域加速库介绍	2%	大模型分布式训练的趋势及挑战 昇腾分布式加速库 MindSpeed 详解



8.华为智算中心解决方案	10%	数据中心基础设施系统和常用的指标介绍 智算中心配电及制冷方案介绍 数据中心基础设施选址及可用性评估 智算中心存储解决方案 智算中心网络解决方案
9.行业模型选型	6%	开源模型选择 PyTorch 模型迁移分析 迁移适配 精度调试 性能调优及算子开发
10.DeepSpeed 进阶	7%	DeepSpeed 简介 DeepSpeed 训练特性 DeepSpeed 推理特性
11.MindFormers 实战进阶	9%	MindFormers 介绍 设计与特性介绍 模型调试调优 套件组合 大模型并行特性 最佳实践
12.大模型微调和安全治理	12%	LLM 微调概念 LLM 微调流程与 SFT 微调方法介绍 微调框架介绍 LLM 安全治理 最佳实践
13.提示工程原理	2%	提示工程的由来 提示工程发展历史 提示工程相关研究
14.大模型压缩与部署	6%	模型压缩与部署概念 模型压缩技术 模型压缩工具介绍 大模型部署框架介绍 大模型应用开发

## 二、实验手册变更说明

实验手册：

知识点	V1.0	说明
数据工程实验	3%	数据集格式转换
ChatGLM2-6B 迁移实验	10%	大模型迁移实验



基于 MindFormers 的大模型推理实验	3%	通过 MindFormers 实现大模型的推理
大模型微调实验	6%	通过 MindFormers 实现大模型的 Lora 微调
大模型推理实验	3%	使用 MindFormers 实现大模型基于 Pipeline 和基于 Generate 的推理
大语言模型综合实验	31%	基于 InternLM-7B 进行微调和 InternLM-Chat-7B 进行推理
视觉大模型综合实验	13%	使用 DeepSpeed 加速库进行 BEiT V2 视觉大模型预训练实验； Vision Transformer 预训练实验
多模态大模型综合实验	31%	LLaVA 模型的训练和推理； Open-Sora-Plan 文生视频实验

### 三、考点变更说明

认证考点：

章节/版本	笔试知识点占比	实验知识点占比
	V1.0	V1.0
1.大模型业务场景流程综述	10%	/
2.大模型应用的数据	3%	/
3.数据工程	6%	10%
4.大模型算法架构	17%	/
5.AI Agent 发展趋势	2%	/
6.昇腾大模型解决方案	6%	5%
7.MindSpeed 领域加速库介绍	2%	5%
8.华为智算中心解决方案	10%	5%
9.行业模型选型	8%	5%
10.DeepSpeed 进阶	6%	5%
11.MindFormers 实战进阶	10%	10%
12.大模型微调和安全治理	12%	15%
13.提示工程原理	2%	/
14.大模型压缩与部署	6%	10%
大模型综合实验（语言/视觉/多模态）	/	30%