

# 中国软件行业协会

## 关于举办大模型时代的 AI 数据工程与模型训练暨人工智能训练师中级/高级专题培训班的通知

各有关单位：

当前，人工智能正加速向千行百业渗透，大模型、多模态技术对高质量训练数据的需求呈爆发式增长。数据标注、模型训练、质量管控及行业部署已成为决定 AI 项目成败的关键环节。为系统培养高水平的人工智能训练人才，提升各单位从数据标注到模型部署的全链路工程能力，我协会举办“《大模型时代的 AI 数据工程与模型训练》暨《人工智能训练师（中级/高级）》”专题培训班。

现将有关事项通知如下：

本次研修分为中级和高级两个独立课程体系。中级课程侧重机器学习标注基础、质量管控与行业场景应用；高级课程侧重复杂场景标注统筹、AI 项目部署配合、前沿技术（大模型/多模态）应用及项目管理。课程采用“理论讲解+工具实战+行业案例”相结合的方式，覆盖金融风控、医疗影像、制造检测等典型场景，帮助学员掌握从标注规范到模型部署反馈的全流程技能。

研修班特邀人工智能训练、数据工程与模型部署领域资深专家团队主讲。专家团队在数据标注规则设计、质量管控体系搭建、大模型微调数据筛选及行业落地方向拥有丰富的  
一线实战经验，将结合金融、医疗、制造等行业典型案例进行系统化讲授与实战指导。

通过本次研修，学员将系统掌握 AI 数据工程与模型训练的核心技术体系，获得从标注规则设计到模型部署反馈的闭环实战能力，为各单位提升 AI 项目落地质量与效率、抢占大模型与多模态数据标注先机提供关键技术支撑。

敬请各相关单位积极参加！

联系电话：010-85913702 郭老师

联系电话：010-62118502 张老师

联系邮箱：csia\_org@yeah.net

附件：《大模型时代的 AI 数据工程与模型训练》  
暨《人工智能训练师（中级/高级）》专题培训班简章



# 《大模型时代的 AI 数据工程与模型训练》暨《人工智能训练师（中级/高级）》专题培训的通知简章

## 一、时间与方式

项目	内容
培训时间	2026 年 7 月、9 月、11 月。
培训形式	线上课程+ 答疑。
中级课程	人工智能训练师中级共计 70 个课时（学习累计时长约 24 小时）
高级课程	人工智能训练师高级共计 80 个课时（学习累计时长约 30 小时）
<b>批次</b>	<b>统一培训时间与考试安排</b>
7 月班	培训时间 7 月 17 日-23 日
9 月班	培训时间 9 月 11 日-17 日
11 月班	培训时间 11 月 20 日-26 日
备注	每期满 30 人开班，如未滿 30 人将自动顺延至下一期。

## 二、培训对象

面向从事大模型时代的 AI 数据工程、模型训练、人工智能训练等领域的相关单位及人员，具体包括：

1、人工智能企业、数据服务商、标注外包公司、金融科技企业、医疗 AI 企业、智能制造厂商、电信运营商、广电运营商、云厂商、互联网公司、IT 公司、机器人公司、智能交互公司、智慧城市、自动驾驶、教育科技等单位的 AI 训练师、标注工程师、质检专员、规则制定者、机器学习工程师、数据科学家、产品经理、项目经理、技术管理者及科研人员；；

2、三大运营商、云厂商、算力服务商、数据中心企业、算力调度企业中从事 AI 训练数据准备、标注、质量管控的技术、产品、运营及管理人员；

3、金融机构（银行、保险、证券、消费金融、互联网金融）风控、反欺诈、信用评分等场景的 AI 数据标注与质量管控人员；

4、医疗健康领域（医院影像科、病理科、AI 辅助诊断中心、医疗 AI 企业）的医学影像标注、电子病历结构化、伦理合规审核人员；

5、制造与工业检测企业（汽车制造、消费电子、半导体、工业视觉检测）的产品缺陷标注、工业视觉训练数据准备及标注与产线对接的工程师；

6、高校及科研院所从事机器学习、计算机视觉、自然语言处理、数据科学、智能医学工程等方向教学与研究的教师、研究员及研究生；

7、标注工具、训练辅助工具开发公司的产品经理、技术支持工程师，以及从事多工具协同集成、专业数据格式适配的技术人员；

8、各级政府大数据局、科技园区、行业协会负责人工智能产业促进、标准制定及AI伦理合规监管的工作人员。

9、全国各类高等院校、职业院校、技工院校的在校学生（含专科、本科），希望系统学习AI数据工程与模型训练基础知识，提升就业竞争力的学生群体。学生可报名中级课程，符合条件的可享受官方政策补贴。

### 三、培训大纲

---

※ 说明：本课程分为中级和高级两个独立大纲。

中级课程：实操技能，机器学习标注基础、质量管控、行业落地应用。

高级课程：综合管理能力，复杂场景标注、项目部署、前沿技术、项目及合规管理。

注：两级课程内容相互独立，学员结合自身岗位与从业经验自主选报。

#### （中级）课程大纲

---

- 模块一：机器学习应用深化  
（监督学习进阶标注、无监督学习预处理、数据反馈机制、训练数据准备）
- 模块二：模型训练辅助基础  
（训练过程配合、训练结果分析、指标体系）
- 模块三：标注质量管控  
（管控流程设计、问题诊断与解决）
- 模块四：行业场景应用  
（金融行业、医疗行业、制造行业）
- 模块五：综合项目实战  
（流程整合、质量管控实战、合规风险规避）
- 模块六：复杂场景标注应用

（多目标标注、遮挡场景标注、专业领域适配）

- **模块七：训练辅助工具应用**  
（数据统计工具、质量反馈工具、进度配合工具）
- **模块八：标注质量管控应用**  
（校验工具逻辑、批量检测方法、问题追溯修正）
- **模块九：行业场景项目应用**  
（金融行业适配、医疗行业适配、制造行业适配）
- **模块十：综合项目全流程应用**  
（工具整合逻辑、协同应用原则、反馈工具应用、行业典型案例、问题答疑）

## （高级）课程大纲

---

- **模块一：复杂场景标注统筹**  
（金融行业、医疗行业、制造行业）
- **模块二：AI 项目应用部署配合**  
（金融行业、医疗行业、制造行业）
- **模块三：前沿技术应用认知**  
（大模型标注、多模态数据标注、行业适配趋势）
- **模块四：项目统筹与管理**  
（目标设定与拆解、资源配置与协调、进度与风险管控）
- **模块五：AI 伦理与合规**  
（金融行业、医疗行业、制造行业）
- **模块六：复杂场景标注统筹应用**  
（金融行业、医疗行业、制造行业）
- **模块七：部署配合应用**  
（金融行业、医疗行业、制造行业）
- **模块八：跨领域项目模拟应用**  
（标注规范适配、数据质量管控、合规要求整合）
- **模块九：项目管理应用**  
（计划制定与执行、团队协作管理、成果评估与优化）
- **模块十：合规审核应用**  
（金融行业、医疗行业、制造行业）

## 四、研修收益

---

## （中级）课程收益

### 1. 夯实机器学习标注基础：

系统学习监督学习与无监督学习场景下的标注方法，掌握标注误差控制、复杂分类标注策略及不平衡样本优化逻辑，为后续进阶打下扎实根基。

### 2. 构建训练数据全流程能力：

从训练数据准备（分层抽样、数据增强）到训练过程配合（实时识别数据问题），再到训练结果分析（模型评估指标与标注质量关联）建立闭环的数据反馈意识。

### 3. 掌握标注质量管控核心技能：

学会设定准确率、Kappa 系数等量化指标，设计抽检、全量校验与问题整改流程，并能诊断标注错误类型并采取针对性缓解措施。

### 4. 实战三大行业标注需求：

深入金融风控、医疗影像、制造缺陷检测等场景，理解行业标注规则、合规红线及标注与业务流的关联，具备直接上岗的基本能力。

### 5. 熟练运用标注全流程工具链：

涵盖数据统计、质量反馈、进度配合、自动校验等工具，掌握批量检测、问题溯源、多工具协同等实战技能，提升标注项目效率与质量。

## （高级）课程收益

### 1. 攻克复杂场景标注难题：

掌握金融（跨产品交易、高频交易）、医疗（多病灶、病灶遮挡）、制造（多缺陷并存、微小缺陷）等复杂场景的标注规则分层设计与多目标优先级逻辑。

### 2. 具备 AI 项目部署配合能力：

学习部署前数据校验、格式适配、部署后模型反馈的标注优化，能够无缝衔接算法团队与业务系统，确保标注数据在真实环境中的有效性。

### 3. 引领前沿技术标注实践：

深入理解大模型标注数据需求、微调数据筛选逻辑及多模态数据（文本+图像+语音）协同规则，站在技术前沿，为企业大模型应用做好准备。

### 4. 掌握项目统筹与管理技能：

从目标设定、资源调配、跨部门协作到里程碑管控、风险识别与进度跟踪，全面提升项目领导力，胜任标注团队负责人或项目总监角色。

### 5. 精通 AI 伦理与合规审核：

系统学习金融、医疗、制造三大行业的 AI 伦理要求（公平性、脱敏、偏见防控、知识产权保护等），并能独立执行多源合规审核流程。

### 6. 跨领域项目模拟应用：

通过真实跨行业项目演练，提炼通用标注规范，解决规范冲突，整合多行业合规要求，提升复杂环境的综合决策能力。

## 五、研修详细内容

---

※ 说明：本课程分为中级和高级两个独立大纲。

中级课程：实操技能，机器学习标注基础、质量管控、行业落地应用。

高级课程：综合管理能力，复杂场景标注、项目部署、前沿技术、项目及合规管理。

注：两级课程内容相互独立，学员结合自身岗位与从业经验自主选报。

## 人工智能训练师（中级）研修内容

机器学习应用 深化	监督学习进阶 标注	监督学习中标注误差的控制方法
		复杂分类任务的标注策略（多分类标注的标签层级设计）
		不平衡样本的标注优化逻辑（少数类样本强化标注）
	无监督学习预 处理	无监督学习的数据清洗进阶要求（异常值精准识别）
		数据特征工程与标注的关联逻辑
	数据反馈机制	模型训练结果对标注数据的反馈逻辑（错误标注识别、标注规则优化）
基于模型预测结果的标注迭代方法		
模型训练辅助 基础	训练数据准备	标注数据的增强逻辑（不改变核心标签的衍生数据）
		训练数据的分层抽样规则（训练集、验证集、测试集划分）
		数据量与模型复杂度的匹配逻辑
	训练过程配合	模型训练中的数据加载与标注格式适配
		训练过程中数据问题的实时识别（如数据格式错误导致的训练中断）
	训练结果分析	模型评估指标（准确率、召回率等）与标注质量的关联分析
模型错误预测结果对应的标注问题排查逻辑		
标注质量管控	指标体系	标注准确率的量化计算方法（正确标注样本/总样本）
		标注一致性的评估指标（Kappa 系数等）
	管控流程设计	标注前的规则培训与考核流程
		标注中的抽样检查频率与比例设定
		标注后的全量校验与问题整改流程
	问题诊断与解 决	标注错误的类型划分（规则理解错误、操作失误、数据本身问题）
不同类型错误的根源分析方法		
行业场景应用	金融行业	金融风控场景的核心标注需求（风险等级划分、欺诈行为识别）
		金融交易数据的标注规则（交易行为特征提取、风险标签定义）
		金融标注的合规红线（客户隐私保护、交易数据保密）
	医疗行业	医疗影像标注的核心需求（病变区域识别、病灶特征提取）
		医学术语与标注标签的精准对应规则
	制造行业	产品缺陷检测的标注需求（缺陷类型、严重程度、位置）

		制造数据的标注与生产流程的关联逻辑
综合项目实战	流程整合	跨场景项目的标注规则统一逻辑(通用规则+场景专项规则)
		多类型数据混合标注的流程衔接
	质量管控实战	项目级标注质量的目标设定(准确率、一致性阈值)
		大规模标注项目的质量抽检方案
合规风险规避	综合项目中的多源数据合规审核流程	
	跨团队协作中的数据安全管控	
复杂场景标注应用	多目标标注	多目标标注的标签区分规则
		目标重叠场景的标注优先级逻辑
		多目标标注的效率提升与质量保障平衡策略
	遮挡场景标注	遮挡场景的可见区域判定标准
		部分遮挡与完全遮挡的标注差异
	专业领域适配	专业领域标注工具的功能拓展(如医疗影像专用标注模块)
工具与专业数据格式(如DICOM医疗影像格式)的适配逻辑		
训练辅助工具应用	数据统计工具	训练数据的数量与质量统计指标(样本分布、标签分布)
		统计结果的可视化呈现逻辑
	质量反馈工具	工具的错误标注识别功能逻辑
		标注质量问题的分类统计与反馈流程
	进度配合工具	标注进度的实时跟踪指标(完成率、效率)
		进度与质量的联动监控逻辑
		进度滞后的预警与资源调配辅助功能
标注质量管控应用	校验工具逻辑	自动校验工具的核心算法逻辑(标签冲突检测、格式错误识别)
		校验工具的参数配置(如冲突判定阈值)逻辑
		人工校验与自动校验的协同流程
	批量检测方法	批量数据的质量检测批次划分规则
		批量检测中的异常数据标记与隔离流程
	问题追溯修正	标注问题的溯源路径(标注人员、时间、规则版本)
批量错误标注的统一修正规则		
行业场景项目应用	金融行业适配	金融标注工具的脱敏功能适配要求
		交易数据标注的批量处理功能逻辑
	医疗行业适配	医疗标注工具的医学术语库关联功能
		影像标注工具的图层管理与测量功能逻辑
	制造行业适配	医疗工具的患者隐私保护功能(如影像脱敏)
		制造标注工具的工业图像适配逻辑(高分辨率、复杂背景处理)
		缺陷标注工具的尺寸与位置精准标注功能
综合项目全流程应用	工具整合逻辑	项目中多工具的功能互补选择
		工具间数据格式的兼容与转换规则
	协同应用原则	标注工具与质量管控工具的实时联动

		训练辅助工具与标注工具的数据同步逻辑
	反馈工具应用	项目数据质量的反馈指标设定
		反馈结果的分级处理机制（紧急问题、一般问题）
行业典型案例	金融行业案例	金融风控跨区域项目标注统筹案例
问题答疑	核心问题解答	跨行业标注规则冲突的解决方法

## 人工智能训练师（高级）研修内容

复杂场景标注统筹	金融行业	金融复杂场景类型（跨产品交易、高频交易）
		金融复杂场景的标注规则分层设计（核心标签+辅助标签）
		金融多目标标注的优先级与关联性规则
	医疗行业	医疗复杂场景类型（多病灶、病灶遮挡、罕见病例）
		医学影像复杂场景的标注规范（病灶边界精准性、多病灶区分）
		医疗专业术语的标注标准化（避免歧义）
	制造行业	制造复杂场景类型（多缺陷并存、缺陷微小、背景复杂）
		动态生产场景中的标注时效性与质量平衡
		制造复杂场景的标注精度要求（尺寸、位置误差控制）
AI 项目应用部署配合	金融行业	金融 AI 项目部署的前置数据准备要求（标注数据的最终校验）
		部署环境中标注数据的格式适配（与部署模型的输入兼容）
		部署后数据反馈的标注调整流程（模型运行中的数据问题修正）
	医疗行业	医疗 AI 项目部署的标注数据标准化（与医院系统对接）
		部署后临床反馈与标注规则优化的联动
		部署前标注数据的临床验证关联逻辑
	制造行业	制造 AI 项目部署的标注数据与生产系统适配
		部署过程中标注数据的实时调用规则
	前沿技术应用认知	大模型标注
大模型微调中的标注数据筛选逻辑		
大模型标注的规则设计（适配模型的泛化能力）		
多模态数据标注		多模态数据（文本+图像+语音）的标注协同规则
		不同模态数据标注的优先级与关联逻辑
行业适配趋势		前沿技术在金融领域的应用（如大模型风控）标注需求变化
	前沿技术在医疗领域的应用（如 AI 辅助诊断）标注规范升级	
项目统筹与管理	目标设定与拆解	跨领域项目的总体标注目标（质量、进度、成本）设定
		拆解目标的量化指标与考核标准
		目标拆解的核心维度（按行业、按任务、按团队）
	资源配置与协调	跨领域项目的人员配置（通用标注员+行业专业标注员）
跨部门协作（技术部、业务部、合规部）的沟通机制		

		标注工具与数据集的资源整合(多行业数据兼容、工具适配)
	进度与风险管控	项目里程碑的设定(数据准备完成、标注完成、质量达标)
		常见风险类型(数据质量风险、合规风险、进度风险)识别 进度跟踪的核心指标(完成率、延期预警阈值)
AI 伦理与合规	金融行业	金融 AI 伦理的核心要求(公平性、透明性、保密性)
		金融敏感信息的深度脱敏规则
		金融标注中的偏见防控(如性别、地域偏见)
	医疗行业	医疗 AI 伦理的核心要求(生命至上、隐私保护、专业严谨)
		医疗标注中的伦理边界(如胎儿影像标注的特殊要求)
	制造行业	制造 AI 伦理的核心要求(安全第一、技术合规、环保适配)
制造标注中的知识产权保护(如专利技术相关数据)		
复杂场景标注 统筹应用	金融行业	金融复杂项目的标注团队分工(规则制定组、执行组、校验组)
		大规模金融数据标注的进度与质量平衡策略
	医疗行业	医疗复杂项目的标注专家参与机制(临床专家审核)
		多中心医疗标注项目的统一规范执行监督
	制造行业	制造复杂项目的标注与生产流程同步机制
		制造复杂缺陷标注的技术交底流程(工程师与标注员对接)
多工厂协同标注项目的统一标准执行		
部署配合应用	金融行业	金融部署前的标注数据最终校验(合规+质量)
		部署后模型反馈数据的标注优化(如新增风险类型标注)
		金融跨区域标注项目的协同管理(统一规则、实时沟通)
		部署过程中标注数据的调用与更新规则
	医疗行业	医疗部署前标注数据与临床系统的适配测试
		医疗复杂标注的质量抽检方案(高风险病例全覆盖)
		部署后临床反馈的标注规则迭代流程
	制造行业	制造部署前标注数据与产线 AI 模型的兼容性验证
部署后产线反馈的标注精度优化		
跨领域项目模 拟应用	标注规范适配	跨行业标注通用规范的提炼方法
		规范冲突的协调原则(优先级设定、折中方案)
		行业专项规范的差异化适配逻辑(如金融合规与医疗隐私的不同要求)
	数据质量管控	跨领域数据的统一质量标准(基础指标+行业补充指标)
		跨领域数据质量问题的分类处理流程
		不同行业数据的质量抽检比例设定
	合规要求整合	跨领域项目的合规要求梳理(提取共性、明确差异)
		合规审核流程的整合设计(一次审核覆盖多行业要求)
		多行业合规的优先级排序(如医疗隐私优先于通用合规)
项目管理应用	计划制定与执行	跨领域项目的详细计划制定(时间节点、任务分工、资源配置)

		项目例会与沟通机制（同步进度、解决问题）
		计划执行中的进度跟踪与偏差纠正
	团队协作管理	跨领域团队的角色定义与职责划分
		团队激励与考核机制（质量、进度、合规维度）
		团队协作中的冲突解决机制（如行业认知差异）
	成果评估与优化	跨领域项目的成果评估指标（质量达标率、进度达成率、合规通过率）
项目成果与预期目标的差距分析		
合规审核应用	金融行业	金融标注数据的来源合规性审核（授权文件核查）
		金融标注结果的合规使用范围审核
		金融敏感信息脱敏效果审核标准
	医疗行业	医疗标注数据的患者授权审核
		医疗标注的专业准确性与伦理合规双重审核
	制造行业	制造标注数据的生产授权与技术保密审核
制造标注规则的合规性（如环保、安全标准）审核		

## 六、培训费用

项目	说明
评测	完成全部课程学习，将进行专业测评考试，通过考试的学员，将由工信部教考中心颁发职业资格证书；实行统一编号登记管理证书可在官方网站查询，纳入工业和信息化技术技能人才库。可作为专业能力考核、投标加分、培训资格证明、岗位聘用、定级和晋升职务的重要依据。
证书类型	由工业和信息化部教育与考试中心颁发《人工智能训练师（中级/高级）》职业资格证书。
费用标准	高级：3980 元/人（含报名费、培训费、专家费、考核建档及证书费）
	中级：3480 元/人（含报名费、培训费、专家费、考核建档及证书费）
	会员单位或三人团报，可享受 9 折优惠
官方补贴	补贴条件：22 周岁及以下，持备案有效在校证明的在校生，仅限报考中级； 补贴规则：可享受 1000 元/人的政策补贴 备注：无在校证明、年龄超 22 周岁，非全日制不在补贴范围内。
缴费方式	开户名：中国软件行业协会 开户行：中国工商银行北京海淀西区支行营业室 账号：0200004509014490109

## 七、咨询方式及报名流程

---

联系人	王老师/郭老师
电 话	010-85913702
手机/微信	17610440521 / 17610465556
报名邮箱	csia_org@yeah.net
填写回执	下载并填写《报名回执表》（见附件），确保信息真实完整。
提交材料	将《报名回执表》、2寸电子版证件照（蓝底或白底）及身份证扫描件发送至报名邮箱。
缴纳费用	通过银行汇款缴纳相应培训费用，汇款时请备注“姓名”。
培训安排	报名截止时间为每次开班前1天，教务老师将统一对接学员并建立学习群。

## 人工智能训练师（中级/高级）--报名回执表

单位名称							
收件地址				姓名		电话	
参会费用	<input type="checkbox"/> 高级：3980 元/人。费用包含参会费用（考核建档及两本证书费等）						
	<input type="checkbox"/> 中级：3480 元/人。费用包含参会费用（考核建档及两本证书费等）						
	<input type="checkbox"/> 在校学生：（享补贴）费用包含参会费用（考核建档及两本证书费等）						
参会情况	人数：（ ）人，费用：（ ）元人民币						
学员姓名	类别	性别	学历	部门/职位	手机号码	电子邮箱	身份证号
开票信息	发票类型： <input type="checkbox"/> 培训费 <input type="checkbox"/> 其他 （如需其他类别请注明）						
	<input type="checkbox"/> 增值税普通发票			公司名称：			
				纳税人识别号：			
				单位地址及电话：			
开户行及账号：							
<input type="checkbox"/> 增值税专用发票				公司名称：			
				纳税人识别号：			
				单位地址及电话：			
				开户行及账号：			
汇款须知	开户名：中国软件行业协会 开户行：中国工商银行北京海淀西区支行营业室 账号：0200004509014490109						
注意事项	1. 参训的学员需将报名回执表送至会务组或协会指定邮箱 2. 培训前 1 天建立学习群并告知详细课程安排等事宜 3. 汇款后需提交汇款凭证（传真或电子邮件均可）						
联系人：王老师/郭老师				电 话：010-85913702			
手 机：17610440521 / 17610465556				邮 箱：csia_org@yeah.net			