

从数据“搬运工”到专家“培养者” AI训练师逐渐走红

写诗作画样样精通、诊病解题事事都行……如今，“十项全能”的AI(人工智能)令人惊叹不已。作为幕后推手，AI训练师逐渐走入大众视野。

从数据“搬运工”，到专家“培养者”，AI训练师见证了人工智能的飞速发展，自身也正经历着日新月异的变化，甚至悄然催生出新的培训需求。



半途出家 从数据标注做起

“这两年国内 AI 越来越火,相关岗位也变多了。”今年初,物联网专业本科毕业的小雅正式进入互联网大厂外包公司,成为一名 AI 训练师。“主要做数据标注,我分到的是数学方向,用 LaTeX 这种工具写解题过程。”

小雅坦言,这项工作并不需要太多专业背景,“有规则文档,对照着把步骤整理出来。”

去年底,美术专业出身的小琳也从UI(用户界面)设计师转行成为AI训练师。“之前做了10年设计工作,后来觉得这个行业快要被AI取代,于是选择改行。”小琳发现,如今的工作与设计基本没什么关系,“我主要做文本类的数据标注,按照公司规则对模型生成的内容打分、改写,从而训练大模型。”对于这份新工作,小琳有着自己的理解,“可以把大模型想象成一个小孩,你要教这个小孩怎么输出更合理的内容。如果小孩犯错,还要给纠错。”

相比起这两名刚入行的“新手”来说,小磊俨然已经是行业里的“老兵”。作为金属材料专业的肄业生,小磊手中的最高学历只有高中。2018年,从事客服工作的他偶然接触到智能客服产品,“当时在一个外包项目里负责辅助互联网大厂的训练师做数据标注,一年后被训练师介绍过去专门做这个。”

在当时,这项工作十分冷门,“对方看我有工作经验,也就没有卡学历。”就这样,小磊抱着学习的态度踏入AI行业。

“最初主要为智能产品做个性化需求落地,以问答为主,可以算小模型,通过搭建知识库、调整问答策略来提升回答质量。”2022年以来,小磊转向从事与大模型有关的数据标注,主要负责提供各种语料,并带领资源团队进行数据生产。“需要与算法技术团队密切沟通,将他们对数据的需求转化为具体的标注规则。”

以常见的问答文本标注为例,小磊不仅要制定答案质量的打分标准,如考量回答的正确性、全面性等,还要对低分答案进行修改,将其转化为高分答案。“规则都是五花八门的,永远不知道下一批任务的规则是什么。”

现阶段,小磊面对的数据标注规模动辄上万甚至几十万条,需要团队协作完成。“团队规模根据项目需求动态调整,小的时候可能只有五六个人,大的时候则可达上百人。”

岗位细分
薪资待遇差别大

在行业内打拼多年，小磊目睹了大模型出现对AI训练师带来的影响。“在大模型出现之前，AI训练师的工作相对聚焦。”小磊回忆说，“以前输出内容主要依赖知识库检索，数据标注只是重新设计表现不好的语料，标注工作相对轻量，问题触发的内容也都是行业内部知识，可控性比较强。”

自从大模型诞生,这项工作变得复杂许多。“现在任务类型多种多样,除了文本以外,还可能有图像、音频等,大模型根据语料推理出来的答案也更加不可控。”

小磊提到,DeepSeek 的出现也给行业带来很大变化。“过去大家都在堆语料,觉得越多越好,但现在要打个问号,思考是不是应该调整方向。”小磊说,以往做小模型的知识库构建时也曾遇到过类似问题,“一开始确实堆的知识越多效果越好,但超过临界值以后,反而会出现意图缠绕,知识变得混淆不清。”

尽管大模型的功能日益强大,但小磊发现问题也逐渐凸显。“大模型确实很擅长公文写作,只是精准度依然有限。像央企国企的公文,通常要求较高,大模型暂时还很难达到这个标准。”

此外,“AI幻觉”也成为备受关注的热门话题。“大模型在回答时出现错误,甚至‘一本正经地胡说八道’,背后原因可能是多种多样的,其中就包括语料的质量问题。”小磊说,需要针对案例进行排查,确认到底是哪个环节出了问题,然后再调整大模型本身的策略。

为了提升语料的质量,小磊会严格筛选可靠信源。“比如,文本类主要选择权威新闻、期刊和实体书等,而不是一些营销号的内容。”在标注过程中,小磊也会对标注人员进行规范培训,要求他们按照标准流程进行信息采集和处理。同时,团队还会参考其他大模型的回复,“相互借鉴,不断优化自己的模型。”

在大模型的推动下，AI 训练师的岗位不断细分。“之前给 AI 当老师，主要是教基本知识，有点像教小孩子学常识。现在 AI 具备了一定的认知能力，就需要有更高水平或者更专业的人来教它高阶思维。”据小磊介绍，现在大模型涉及越来越多专业领域知识，也就需要更多具备医疗、教育、法律等专业知识背景和从业经验的人才加入进来，“这些专业性强的语料生成，外行确实做不了。”

在小磊看来,岗位细分的同时,AI训练师的薪资待遇差距也在拉大。“像基础的数据标注岗位,月薪可能只有几千元,而大公司里要求较高的岗位,月薪可以达到三五万元。”

学费较高
没必要盲目跟风

记者在调查中了解到,随着AI训练师岗位需求不断增加,相关培训迅速兴起。

“现在很多人转行做这个，培训班上什么专业的都有。有刚毕业的，也有工作好多年的。”上岗前，小雅就报班参加了AI训练师课程。“一共也就10次课，报班就是为了心里有个底，其实也可以自己投简历，面试多了自然就知道会问什么，面试通过以后可以在干中学。”

转行前，小琳同样选择先报班“补课”。“算是回流了，2013年刚毕业那会儿就在这个机构学的UI设计师。这两年好多设计师的工作干不下去，机构就又开了AI训练师的课，让大家学完尽快转行。”说到这里，小琳不由得苦笑，“我要是有讲课的天赋，我也去开班了，还是开班赚钱。”

记者在社交平台上搜索,发现不少分享转行经验的AI训练师都在小琳所说的这家机构报过课。

“我们的AI训练师课程是就业导向的，需要线下学习，在北京、深圳、杭州都设有校区，这些城市的工作岗位也比较集中。”据机构工作人员介绍，培训持续5周，上课时间安排在周末，共10次课。从大纲来看，课程相当紧凑，学员需要在一天之内学习大模型行业通用规则、单轮对话项目实训、多轮对话项目实训和AI训练师必学代码课。“学完专业课，会分配一对一的就业辅导老师。”工作人员称，学费每期涨1000元。“现在报名的话是第13期，3月29日开课。原价22999元，最近几天有一个拼团折扣价，优惠完20999元。第14期的话是5月10日开课，差不多就是这么一个招生节奏。每期50名学员，分成两个班。”

在去年 10 月的培训中,小琳从课上了解到 Reward 模型、RAG 项目等内容。“但这些只是基础知识,到了公司还是要靠自己在实际工作中摸索。”如何在简历关和面试关解决零经验的问题?小雅从培训中学到了“应对策略”,“可以结合以前的经验,把 AI 融入进去。”

作为过来人,小磊发现,很多数据标注方面的外包公司招聘需求很大。“这项工作相对比较枯燥,成长性也有限,很多人干一段时间就走了。”小磊说,如果本身学历有限,对这一行感兴趣,暂时也没有太多机会和选择,那么不妨尝试从数据标注做起,“起码也算接触了这个行业,后期通过工作经验积累,也能慢慢往更高阶的训练师方向转。”

至于要不要报班,小磊表示没必要盲目跟风。“如果只是做数据标注,其实要求没那么高。就算是为了应对面试时的问题,也未必需要花高价报班,完全可以找一些相关知识自学一下。” 来源:新华网

现状

变化

提醒