

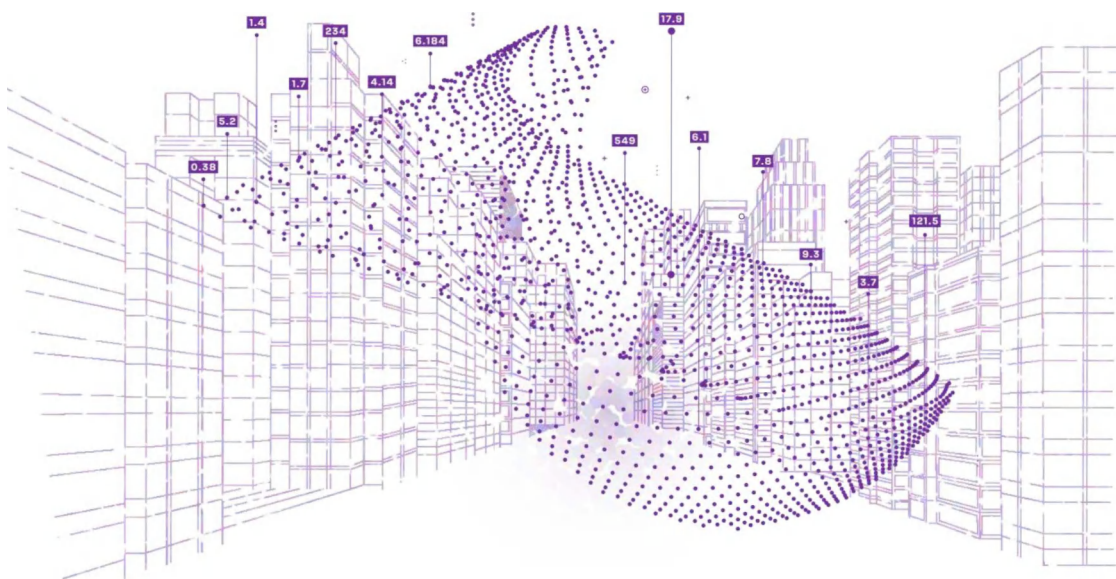


清华大学数字政府与治理研究院
Institute for Digital Government and Governance, Tsinghua University



清华大学社科学院数据治理研究中心
Center on Data and Governance, School of Social Sciences, Tsinghua University

智能数据标注产业发展观察报告



发布单位：清华大学数字政府与治理研究院
江苏省数字化协会
支持单位：上海数据交易所
华东江苏大数据交易中心

2025 年 4 月

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

智能数据标注产业发展观察报告

执行摘要

- 随着人工智能技术进入大规模应用阶段，数据标注作为人工智能模型训练的基础环节，其重要性日益凸显。特别是在生成式人工智能和大模型快速发展的背景下，数据标注的规模和质量要求都达到了前所未有的高度。与此同时，传统人工标注模式面临效率低下、成本高昂等挑战，智能数据标注技术成为突破行业发展瓶颈的关键路径。课题组基于对产业实践的观察，系统分析智能数据标注产业发展前景，撰写本报告。
- 本报告基于全国企业招聘数据分析和深度产业观察，旨在系统把握数据标注产业的发展脉络。研究重点聚焦三个方面：首先，通过用工需求的时空变化分析产业发展动态；其次，考察不同区域和行业的标注需求特征；最后，深入剖析智能标注企业的技术创新和应用实践。这三个维度相互支撑，共同构成了理解产业现状与未来的完整框架。
- 基于用工需求与企业发展数据的时空行业特征分析发现，数据标注产业用工需求呈现明显的阶段性特征，这种波动与人工智能技术发展周期高度吻合。东部地区在产业规模上保持绝对优势，但近年来中西部地区保持较高增长态势。行业需求方面，信息传输与软件服务业占据主导地位。
- 观察智能标注技术应用与企业实践创新发现，头部企业已普遍采用人工智能辅助标注技术，通过预标注、主动学习等方法提升人工标注效率。在商业模式上，智能数据标注企业已在实践中形成四种差异化发展路径：头部势力具备全栈技术能力，以稳定的服务质量和跨领域服务能力为核心优势；创新势力专注技术创新提升标注效率；众包团队通过“核心+弹性”模式优化成本；场景团队依托行业数据标注积累深耕专业领域。这四种模式既独立又互补，推动智能标注向专业化、规模化发展，满足多元化市场需求。

版权声明

本报告著作权完整归属于《智能数据标注产业发展观察报告》课题组，并受法律保护。转载、摘编、编译或以其他方式使用本报告的全部或部分内容的，应标注“课题组：《智能数据标注产业发展观察报告》”。对于侵权行为，课题组及其成员单位保留追究法律责任的一切权利。

特别说明

- 本报告在展示评估结果时，优先呈现两位小数；如末位小数为零，则做去零处理。
- 本报告所用图表均为课题组自制。

目 录

1. 导言	1
1.1 产业实践背景	1
1.2 政策支撑	2
1.3 研究问题与目的	5
2. 数据标注产业发展实践	6
2.1 数据标注产业演进	6
2.2 数据标注产业流程及主体	7
3. 数据标注产业用工需求观察：时空行业演进分析	12
3.1 数据标注员市场需求分析	12
3.2 用工数据标注企业趋势分析	22
3.3 数据标注产业发展挑战	28
4. 智能标注前沿技术与企业实践观察	30
4.1 智能数据标注技术发展前沿	30
4.2 智能标注企业实践及分类分析	32
5. 智能数据标注典型企业实践	37
5.1 东部地区的全面发展	37
5.2 中部地区的场景挖掘	45
5.3 西部地区的技术探索	48
5.4 东北地区的协同实践	49
6. 研究展望与总结	51
6.1 未来展望	51
6.2 研究总结	53

1. 引言

在人工智能蓬勃发展的浪潮中，数据标注作为其关键环节，正逐渐成为推动技术进步的核心力量。随着人工智能应用的广泛普及，从自动驾驶到智能语音助手，从医疗影像分析到自然语言处理，数据标注产业迎来了前所未有的增长机遇。数据标注不仅是为机器学习提供“燃料”的基础性工作，更是人工智能模型能够理解世界、学习规律并做出智能决策的前提条件。

1.1 产业实践背景

高质量的数据标注能够显著提升模型的性能，使其更接近人类的认知水平；而低质量的标注则可能导致模型学习到错误的模式，影响其应用效果。因此，数据标注产业发展不仅关乎人工智能技术的进步，更对数字经济、产业升级乃至社会生活的智能化转型具有深远意义。以生成式人工智能的典型应用为例，这类模型通过生成自然语言文本，能够完成从写作、翻译到编程等多种复杂任务，展现出强大的语言理解和生成能力。然而，其性能高度依赖于语料基础，尤其是标注语料对技术发展和性能提升起到了至关重要的作用。

生成式人工智能语料体系包括预训练语料和微调语料两部分。预训练语料主要由开发团队从书籍、杂志、百科、论坛等渠道收集的海量无标注文本数据组成，为模型提供了广泛的知识背景和语言模式，使其能够学习到语言表达的逻辑和结构。微调语料则包括从开源代码库爬取、专家标注、用户提交等方式收集的高质量有标注文本数据，用于针对性的微调，进一步增强模型在特定任务上的表现。高质量的数据标注能够使人工智能算法模型学习到正确的规律，从而提高模型的性能。数据标注需要把语音、图像、文本等数据人为地打上正确的标签，然后把这些数据灌输给人工智能模型。模型算法在周而复始地“学习”这些内容后，最终使得人工智能实现自主识别成为可能。

数据标注产业链涵盖了从数据采集、标注到应用的全过程。上游包括劳动力和计算机软硬件设备供应商，为数据标注提供基础资源；中游是数据标注的核心环节，涉及专业服务商和科技巨头自建平台，负责将原始数据转化为有价值的信息；下游则涵盖多个应用领域，如计算机视觉、智能安防、智能语音等，为人工智能技术的落地提供了广阔空间。随着人工智能产业的发展，数据要素的高质量供给成为技术发展与产业落地的关键瓶颈。数据标注作为人工智能发展的基

石，其重要性愈发凸显。

大语言模型的快速发展已经暴露了诸多问题，其中数据标注环节尤为关键。数据标注是人工智能模型训练的基础，标注过程中的主观性、标注标准的不统一以及标注数据的偏差，都可能导致模型学习到错误或有偏见的信息。智能数据标注是一种融合了人工智能技术与传统数据标注方法的创新性数据处理方式，旨在通过智能化手段提升数据标注的效率、质量和准确性。它利用机器学习模型、自动化工具以及人工智能辅助技术，对数据进行高效处理和精准标注，从而为人工智能模型的训练提供高质量的“燃料”。这种标注方式不仅能够显著减少人工标注的工作量，提高工作效率，还能通过自适应学习和实时反馈机制，确保标注数据的一致性和准确性，满足复杂场景下对数据标注的高要求。

随着人工智能技术的不断发展，智能数据标注将成为未来数据标注产业的重要发展方向。它不仅能够满足大规模数据标注的需求，还能为复杂的人工智能模型提供高质量的数据支持，推动人工智能技术在各领域的广泛应用和深度发展。

1.2 政策支撑

在全球范围内，各国政府纷纷将人工智能视为未来科技竞争的战略高地，积极出台多项政策以推动其发展。近年来，随着人工智能技术的快速发展，我国政府高度重视其在推动科技创新、产业升级和社会进步中的重要作用，出台了一系列政策文件，为人工智能产业的发展提供了坚实的政策支持。这些政策不仅涵盖了人工智能技术发展的宏观指导，还深入到数据要素的管理和隐私保护等关键领域，为数据标注产业的高质量发展营造了良好的政策环境。

2017年7月国务院《新一代人工智能发展规划》多次强调数据在人工智能发展中的基础性作用。《规划》提到大数据是人工智能发展的关键驱动力之一，而数据标注是将原始数据转化为可供机器学习模型使用的结构化数据的重要环节。《规划》指出人工智能在多个领域的应用，如智能医疗、智能教育、智能交通等。这些领域的智能化应用都需要大量的标注数据来训练模型，以实现精准识别、预测和决策。同时，《规划》强调要构建开放协同的人工智能科技创新体系，推动人工智能产业生态的建设。可见，数据标注作为数据处理的重要环节，是人工智能产业生态中不可或缺的一部分，为技术研发和应用提供了基础支持。

2022年1月，国务院发布《“十四五”数字经济发展规划》，强调数据要

素的作用，提出提升数据管理水平、促进数据要素市场化流通、探索数据资产定价机制等措施，为数据要素的合理流动和高效利用奠定了基础。2022年4月，中共中央、国务院联合发布《关于加快建设全国统一大市场的意见》，提出加快培育数据要素市场，建立健全数据安全、权利保护、跨境传输管理等基础制度和标准规范，推动数据资源的开发利用。2022年12月，《关于构建数据基础制度更好发挥数据要素作用的意见》进一步明确了数据产权制度，推进公共数据、企业数据、个人数据分类分级确权授权使用，建立数据资源持有权、数据加工使用权、数据产品经营权等分置的产权运行机制，为数据要素的管理和利用提供了更加完善的制度保障。2023年7月，《生成式人工智能服务管理暂行办法》规定在生成式人工智能技术研发中，数据标注提供者应制定清晰、具体、可操作且符合相关管理办法要求的标注规则；开展数据标注质量评估，通过抽样核验标注内容的准确性；同时对标注人员进行必要培训，提升其尊法守法意识，并监督指导其规范开展标注工作。

2024年12月，国家发展改革委等部门联合发布《关于促进数据标注产业高质量发展的实施意见》，为数据标注产业的未来发展提供了全面的政策指导。该意见明确了数据标注产业的定义和发展目标，强调数据标注产业作为新兴产业，对提升数据供给质量、推动人工智能创新发展具有重要支撑作用。意见提出，到2027年，数据标注产业的专业化、智能化及科技创新能力显著提升，产业规模大幅跃升，年均复合增长率超过20%。为实现这一目标，意见从深化需求牵引、增强创新驱动、培育繁荣生态、优化支撑体系、促进标注产业安全发展等方面提出了具体措施。例如，通过释放公共数据标注需求、挖掘企业数据标注需求，推动数据标注产业与实体经济的深度融合；开展关键技术攻关，健全数据标注标准体系，打造高水平创新载体；培育数据标注龙头企业，完善产业生态，推动产业链上下游协同发展；加大财税金融支持力度，提升标注公共服务能力，加强标注人才队伍建设等。这些政策措施的实施，将有力推动数据标注产业的高质量发展，提升我国人工智能产业的整体竞争力。

2025年3月发布的《合肥数据标注产业发展规划（2025-2027年）》提出了一系列具有创新性和可操作性的政策措施，为产业发展提供了全方位的保障。合肥市立足本地资源和产业基础，链接跨部门、跨区域、全球化的创新要素，构建

高效、智能、安全的数据产业生态，这将为数据标注产业的发展提供丰富的资源和广阔的空间。《规划》明确了六大重点任务，包括攻关数据标注关键共性技术、建设支撑项目、培育市场主体、赋能行业发展、营造发展生态以及创新运营机制。这些任务涵盖了技术创新、平台建设、市场主体培育、行业应用、生态营造等多个方面，为数据标注产业的全面发展提供了有力支撑。

2025 年 3 月，数据标注基地先行先试现场会在四川省成都市召开。会议强调，人工智能作为新一轮科技革命和产业变革的核心驱动力，具有强大的“头雁”效应，而高质量数据集是人工智能发展的基石，数据标注工作则是构建高质量数据集的关键环节。因此，加快培育和发展数据标注产业，推进数据要素市场化配置改革，对于促进数据开发利用、赋能经济社会发展、培育数据标注新业态、布局数字科技新赛道以及构建产业国际竞争新优势，均具有极为重要的战略意义。同时，会议提出了一系列明确指引。下一步，各地需以习近平新时代中国特色社会主义思想为指导，因地制宜、多措并举，大力推进数据标注产业发展。一是加快推动数据标注产业政策落地实施，各地应全面贯彻落实《关于促进数据标注产业高质量发展的实施意见》，在科技创新引领、财税金融支持、人才培养体系、产业监测服务等方面，进一步夯实产业发展基础，为产业的持续发展提供全方位保障。二是加快推进高质量数据集建设，各基地要聚焦技术创新、行业赋能、生态培育、标准应用、人才就业、数据安全等六大建设任务，推动工业、金融、医疗、交通、教育等多领域的高质量数据集建设，为人工智能的高水平发展筑牢数据底座。三是因地制宜培育数据标注产业，各地应立足自身优势和资源禀赋，找准发展定位，探索特色发展路径，通过差异化定位和特色化发展，形成优势互补、协同发展的产业格局，推动数据标注产业在全国范围内实现均衡、高效发展。

由此可见，我国政府通过一系列政策文件，从人工智能技术发展、数据要素管理到数据标注产业的高质量发展等方面，构建了全方位的产业发展政策支持体系，为人工智能产业的健康、可持续发展创造了良好的政策环境。在政策的引导和支持下，数据标注产业有望在未来几年内实现跨越式、智能化发展，为我国数字经济的腾飞提供强大动力。

1.3 研究问题与目的

随着人工智能技术的飞速发展，数据标注作为其发展的关键支撑环节，正面临着前所未有的机遇与挑战。数据标注不仅是人工智能模型训练的基础，更是实现人工智能应用落地的重要前提。然而，当前数据标注产业在发展过程中暴露出诸多问题，如标注效率低下、标注质量参差不齐等，制约了行业的健康发展。与此同时，随着人工智能技术的不断演进，尤其是大语言模型和生成式人工智能的兴起，数据标注产业也在经历着技术的快速迭代和产业格局的深刻变革。在此背景下，本报告旨在深入探讨数据标注产业的发展现状、智能标注技术演进趋势以及未来机遇，以期为智能标注产业发展提供有价值的参考。

具体而言，本报告将围绕以下几个核心问题展开：一是数据标注产业当前的发展现状如何，包括市场规模、地域行业差异等方面；二是智能数据标注企业实践与技术演进趋势如何，特别是人工智能辅助标注、合成数据等前沿技术的发展应用及其对行业的影响；三是数据标注产业未来的发展方向和机遇在哪里，包括技术突破、市场竞争等方面。通过对这些问题的深入研究，本报告旨在明确智能数据标注产业的发展方向。

综上所述，本报告将深入剖析数据标注产业的现状、智能标注技术实践与发展趋势，为推动产业高质量发展提供全面的理论支持与实践指导。

2.数据标注产业发展实践

数据标注作为人工智能发展的基石，是将原始数据进行加工处理，如分类、拉框、注释、标记等操作，转换成机器可识别信息的过程。这一过程不仅为机器学习提供了“燃料”，更是人工智能模型能够理解世界、学习规律并做出智能决策的前提条件。数据标注项目的整体流程涵盖了从数据准备、标注规范制定、标注员招募与培训，到数据标注、质检、结果导出以及数据管理和存储等多个环节，每一个环节都至关重要，共同确保了数据标注项目的顺利进行和高质量完成。

随着人工智能技术的飞速发展，数据标注的需求也在不断增长。从图像到文本，从语音到视频，数据标注的类型和任务划分日益多样化。例如，文本标注包括词性标注、分类标注、情绪标注等；图像标注涵盖图像分类、语义分割等任务。这些标注任务不仅需要专业的技术工具，还需要标注员具备相应的专业知识和技能。此外，数据标注产业还遵循着“二八定律”，即在一个人工智能项目中，数据准备工作通常占据 80%的时间，而模型训练和部署仅占 20%。这一规律凸显了数据标注在整个人工智能项目中的重要性。

2.1 数据标注产业演进

在 2016 年之前，数据标注产业处于人力标注阶段，数据标注主要依靠人工完成，以训练任务和算法模型为导向，进行简单的图像标注工作。随着数据需求的增长，人工标注的局限性促使自动化标注技术开始发展，众多人工智能数据初创公司随之诞生。

从 2016 年到 2022 年，产业进入平台工具标注阶段，自动驾驶等场景的爆发推动了行业发展，数据标注从纯人工向自动化和半自动化转变，平台和工具逐渐成熟，国内人工智能初创公司大量涌现，市场需求激增。

2023 年起进入产业重组阶段，数据质量成为关键驱动力。更多企业参与数据标注，创业门槛提高。垂直大模型加速落地，数据处理范式和标准基本确定。2024 年及以后，行业进入知识密集阶段，人机协同关系进一步发展，数据标注将更加注重质量、专业化，以适应更复杂的应用场景和更高的技术要求。

市场规模方面，数据标注产业呈现出快速增长的趋势。《合肥数据标注产业发展规划（2025—2027 年）》指出到 2027 年底，合肥将实现多语种标注和语音标注能力国际领先，构建超 11 个行业高质量数据集，标注数据规模达 3000TB，

产业规模突破 30 亿元，带动相关产业形成千亿级体量。2025 年 3 月，数据标注基地先行先试现场会指出，各地方主动作为，在数据标注产业发展中取得了阶段性成果，7 个基地的数据标注规模达到 17282TB，形成了涵盖医疗、工业、教育等多个行业的 335 个高质量数据集，为 121 个大模型研发提供了有力赋能，引进和培育了 223 家标注企业，从业人员数量达到 5.8 万人，带动相关产值 83 亿元。

如上产业发展成绩表明，数据标注产业不仅已经具备了坚实的发展基础，还展现出巨大的发展潜力。同时，各地在生态培育、技术创新等方面积极探索实践，构建了标准、应用、人才一体化保障体系，推动数据标注向智能化、自动化方向发展，并形成了一批可复制推广的发展经验，为产业拓展奠定了坚实基础。

2.2 数据标注产业流程及主体

数据标注产业通过标准化的数据处理流程，将原始数据转化为可供机器学习的高质量训练数据，其核心参与者包括数据标注企业、专业技术团队及经过系统培训的标注人员。随着人工智能技术的快速发展，数据标注产业正逐步从劳动密集型向技术密集型转型，对流程标准化、质量控制和人员专业化提出了更高要求。

2.2.1 数据标注工作流程

数据标注工作流程是人工智能模型训练中不可或缺的一环，它涉及到将原始数据转化为机器可识别信息的全过程。这一流程开始于对所需标注数据的明确定义和预估数据量，包括分析数据类型、量级、用途及应用场景等。随后，需要整理数据，明确数据与标签文件的存放目录结构，并制定数据与标签文件的命名规则，以便于数据追踪和标注追踪。在准备工作完成后，制定详细的标注说明规则，这些规则包括项目背景、意义、数据应用场景、标注工具、任务描述、标注方法、正确示例和常见错误等内容。这些规则应具备可变更性，以适应项目需求的变化。同时，需要对参与标注的工作人员进行培训，确保他们理解标注规则并满足技能要求。根据任务的数据量级、保密性、资质要求、对业务规程的理解程度、成本预算以及交付时间等因素，确定标注人力的供给方式，如内部自营标注、第三方标注或众包标注等。同时，标注工作需选择合适的标注工具和平台，这些工具和平台应满足易操作性、规范性和高效性等条件。

创建标注任务前，需将待标注数据上传至平台，并明确任务的基本信息，包括任务目的、需求、描述等。然后，根据任务需求匹配不同的标注工具，添加相

关标签，上传数据路径至平台，并进行版本控制。标注任务的分发涉及到将任务分配给标注人员和审核人员，明确参与标注人数、子任务数量、每人每天工作量、回收子任务时间点和任务结束时间点等参数。

在标注任务开展阶段，数据标注方法主要分为全人工标注和半自动标注。全人工标注依靠人力进行，而半自动标注结合人工与自动化工具。全人工标注和半自动标注是数据标注流程中的两种主要方法，它们各自适用于不同的场景和需求。全人工标注依赖人力对数据进行详尽的分析和标记，这种方法通常用于对标注精确度要求较高的任务。在全人工标注中，标注员会直接与数据交互，运用专业的知识对文本、图像、语音或视频等数据进行归类、整理、编辑、纠错、标记和批注等操作。这种方式能够确保数据的高质量标注，但当处理的数据量较大时，会消耗较多的人力资源，增加时间和成本。

相对于全人工标注，半自动标注则结合了人工操作和自动化工具的优势，以提高标注效率。在半自动标注过程中，通常会先使用训练有素的模型对目标数据进行初步的检测和标记，然后由标注员使用标注工具对模型的输出结果进行校正和完善。这种方法适用于数据量较大且标注任务相对简单的场景。半自动标注建立在较成熟的模型基础上，如果模型检测结果的准确度不够，可能需要人工进行较多的修正工作，从而增加工作量。

在全人工标注中，若对标注结果的准确率要求较高，通常会在标注前对标注人员进行相关任务的培训，包括标注工具或平台的使用方法、标注的任务目的、标注内容和标准等。此外，建立标注者与标注数据使用者之间的反馈机制，确保信息同步，也极为重要。在半自动标注中，由于模型已经完成了部分工作，因此对标注员的培训可能更侧重于如何使用标注工具以及如何处理模型输出的结果。无论采用哪种方式，都需要在标注前对标注人员进行相关任务培训，并建立反馈机制，确保信息同步。

标注任务完成后，项目负责人需要回收标注作业，并保证任务的完整交付。这包括自营标注团队直接收取、第三方标注服务公司按时交付或众包平台的合理回收。标注结果的质量检查和控制是确保数据标注结果有价值的关键环节。在进行质量检查时，项目负责人需要根据项目的特性和要求，选择适当的检查方法。这些方法可能包括对标注数据的全面检查，即逐条核实每一条标注信息，这适用

于对数据准确性要求极高的项目。另一种方法是按比例抽查，也就是从全部标注数据中抽取一部分样本进行核查，这种方法适用于数据量大、时间有限的情况。此外，还可以采用分层抽样或系统抽样等统计学方法，以确保样本的代表性。

在实际操作中，质量检查与控制还需要建立一套明确的合格标准，这些标准应可量化，便于在检查和控制过程中进行评估。例如，在医学影像标注中，需要由具有专业资质的专家对标注结果进行验证，确保其符合临床诊断的标准。在自然语言处理领域，需要检查文本标注是否准确地反映了语言的语义和语境。

最后，标注结果的输出交付要求包括图像、文本、语音、视频等不同类型的数据，每种类型的数据都有其特定的输出格式和组成部分。数据交付时，标注团队需对最终提交的数据量进行说明，交付的内容应包括标注结果、交付和说明文档、关于标注数据的元数据以及原始数据等。数据需求方在数据验收期内完成对数据标注结果的验收工作，若验收数据质量未达到预期值，可要求数据服务提供商对数据进行修正。整个流程需要严格管控，以确保数据标注的质量和效率。

2.2.2 数据标注企业

数据标注产业的快速发展吸引了众多参与者，形成了从上游的资源汇聚源头到中游的核心加工枢纽，再到下游的广泛应用场景的完整产业链结构。上游部分主要涉及劳动力资源和计算机软硬件设备供应商，为数据标注产业提供基础资源。中游则是数据标注的核心环节，包括专业服务商和科技巨头自建平台。下游部分则涵盖了数据标注成果的应用领域，如计算机视觉、智能安防、智能语音等，这些领域对标注数据的需求持续增长，推动了整个行业的快速发展。

数据标注企业作为人工智能产业链中的关键环节，承担着将海量原始数据转化为机器可识别信息的重要任务。这些企业通过集中化的生产模式和标准化的操作流程，高效地完成从数据接收、预处理、标注、质检到最终交付的全过程。它们不仅是数据标注的执行者，更是人工智能模型训练的重要支撑。数据标注企业的组织架构通常围绕高效的生产流程展开。管理层负责整体运营和战略规划，确保项目的顺利推进和资源的合理分配。标注员团队是企业的核心力量，他们经过专业培训，具备处理多种数据类型的能力，从文本、图像到语音和视频，他们都能精准地完成标注任务。质检团队则专注于标注结果的质量把控，通过严格的审核流程，确保数据的准确性和一致性。此外，技术支持团队负责开发和维护标注

工具，引入先进的技术手段，如人工智能辅助标注，以提升标注效率和质量。

在运作模式上，数据标注企业首先接收来自客户的原始数据，这些数据可能包含文本、图像、语音等多种类型。数据进入企业后，会经过预处理，包括数据清洗、格式转换等操作，以确保数据的质量和可用性。随后，预处理后的数据被分配给标注员进行标注。标注任务的分配会根据数据类型、标注难度和标注员的专业能力进行合理安排，以保证标注工作的高效进行。在标注过程中，管理层和技术支持团队会实时监控标注进度，及时解决标注员遇到的问题。质检团队则会定期对标注结果进行抽检，确保标注质量符合标准。标注完成后，标注数据经过最终审核，确认无误后交付给客户。客户在使用标注数据的过程中可能会提出反馈意见，企业需要根据这些反馈及时调整标注流程和方法，以满足客户需求。

为有效应对行业竞争与数据质量等挑战，数据标注企业也在不断探索未来的发展方向。随着人工智能技术的不断进步，企业将越来越多地采用智能化、自动化标注技术。这些工具和技术可以自动完成部分标注任务，提高标注效率和质量，降低人工标注的工作量。与此同时，企业还将更加注重专业化和精细化发展，根据不同的数据类型和应用场景，提供更加专业化的标注服务。

2.2.3 数据标注员

数据标注员在数据标注产业中扮演着至关重要的角色，他们承担着将原始数据转化为机器可识别和学习信息的重任。他们的工作内容包括但不限于：分类标注，即将数据分配到预定义的类别中，例如在图像中识别和分类不同的对象；拉框标注，在图像或视频数据中绘制边界框以标识特定对象的位置和大小；注释和标记，为数据添加描述性标签或标记，如文本中的情绪分析或语音中的关键事件；序列标注，对序列数据（如文本或时间序列）进行标注，以识别特定的模式或事件。数据标注员不仅需要具备一定的专业知识和技能，还可能需要对特定领域有所理解，例如医疗影像分析中的医学知识，以及对标注工具的熟练操作。2020年“人工智能训练师”正式被纳入国家职业分类目录，其核心工作内容包括数据标注。这不仅提升了数据标注员的社会地位，也为行业的发展提供了规范和指导。

随着人工智能技术的不断进步，对数据标注员的技能要求也在不断提高。他们可能需要掌握更复杂的标注任务，如自然语言处理中的语义角色标注或语音识别中的音素标注，以适应行业发展的需求。与此同时，数据标注员的学历与技能

要求随着人工智能技术的发展和行业的转型而不断演变。在传统数据标注领域，工作技术门槛相对较低，主要依赖专科生或经过短期培训的低技能劳动力，尤其是在发展中国家，这些地区的劳动力成本优势使得其成为数据标注产业的主力军，在国内，部分数据标注项目工作与扶贫工作紧密相连。然而，随着大模型时代的到来，数据标注产业正从劳动密集型向知识密集型转变，对标注从业人员的学历和技能要求也显著提高。如今，本科以上学历逐渐成为部分地区标注员招聘的硬性要求，同时，他们还需要具备多领域的专业知识，如医学、金融等，以应对日益复杂的标注任务。在技能方面，传统数据标注员主要掌握基本操作技能，如拉框、描点、转写等，并熟练使用数据标注工具。但在大模型时代，标注工作变得更加复杂，标注员需要具备排序、改写、多轮对话、评估等高级技能，同时还要运用一定的认知能力来理解数据的语义和逻辑结构。此外，随着人工智能辅助标注工具的广泛应用，标注员必须学会、理解、使用这些工具来提高标注效率和质量，并具备质量控制意识，以确保标注的准确性和一致性。

为了满足这些新的要求，数据标注员需要通过专业培训来提升技能，包括入职培训和持续培训，以适应技术发展和任务复杂化的需求。实践经验的积累同样重要，通过参与实际项目和跨领域任务，标注员可以拓宽知识面并提升综合技能。同时，利用人工智能辅助标注工具和专业数据标注平台，标注员能够更好地管理任务并提升标注质量。

可见，随着技术的不断进步，数据标注员的角色将更加关键，智能标注工具的广泛应用将进一步提升标注效率和质量，但同时也对标注员的技能提出了更高要求。标注员需要不断学习，掌握人工智能辅助标注技术，提升跨领域知识和高级认知能力，以适应产业发展需求。

3.数据标注产业用工需求观察：时空行业演进分析

招聘数据作为反映产业发展的关键指标，能够从人才供需角度揭示数据标注产业的真实发展态势。报告依托 Resset 中国企业大数据平台招聘数据库，通过对数据标注相关岗位的深度挖掘，为观察产业发展提供了独特的分析视角。系统分析招聘数据，不仅能够勾勒出数据标注产业的发展现状，更为预判产业未来走向提供了重要参考。

3.1 数据标注员市场需求分析

数据标注员的市场需求分析是理解数据标注产业发展态势的重要窗口。通过时间、地域和行业三个维度的分析，可以全面把握数据标注人才市场结构性特征。

3.1.1 数据标注员年份需求分析

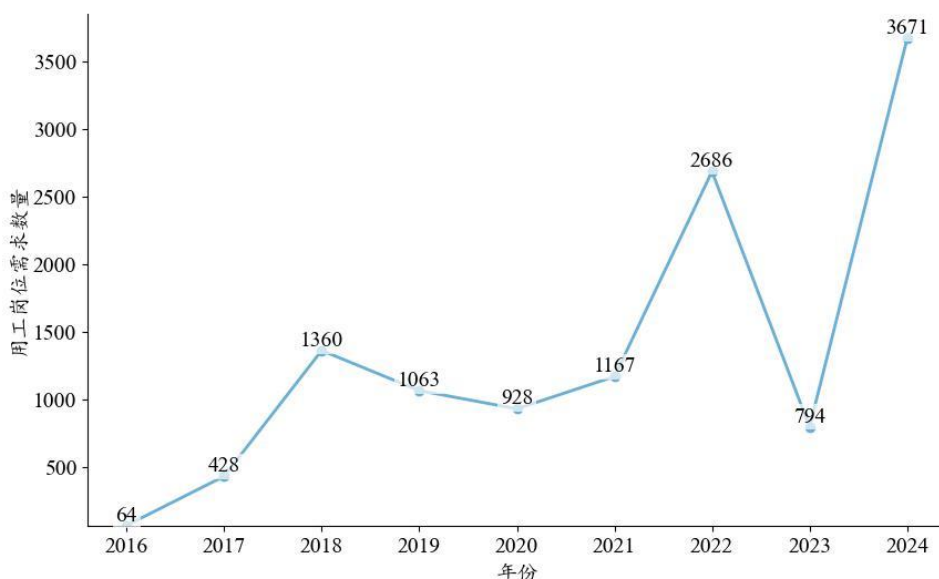


图 3-1 用工岗位年度趋势

2016 年首次出现数据标注相关招聘需求，此时需求量较低，2016 年的用工需求仅为 64 人，随后在 2017 年显著增长至 428 人，显示出该产业的初步发展势头。2018 年，用工需求进一步激增至 1360 人，2019 年至 2021 年，用工需求保持在千人水平。到 2022 年，ChatGPT 等生成式人工智能兴起，文本标注、指令微调等新需求的涌现使得用工需求大幅上升至 2686 人。2023 年用工人数下降至 794 人，可能与市场在政策推动后的自我调整有关，产业在快速发展过程中，需

要时间来消化和整合资源，同时也在探索更加高效的发展模式。

2024 年，数据标注产业用工人数量急剧上升，这一现象可以从技术发展和政策支持两个方面进行分析。从技术发展来看，2024 年 OpenAI 推出了多模态大模型 GPT-4o，该模型具备处理文本、图像、音频等多种数据类型的能力，标志着人工智能技术在多模态交互方面的重大突破。与此同时，国内的深度求索（DeepSeek）也取得了显著进展，DeepSeek-V3 是全球首个全开源的 MoE（混合专家）模型，其高性能和低成本的特点进一步推动了行业的发展。多模态大模型技术突破使得数据标注的需求大幅增加。从政策支持来看，《关于促进数据标注产业高质量发展的实施意见》的发布为数据标注产业提供了发展方向和政策支持，吸引了更多的企业和资本进入这一领域，从而带动了用工需求的快速增长。

整体来看，数据标注产业的用工需求呈现出波动性增长的趋势，反映了该产业随着技术进步和市场需求变化而不断调整和发展的动态过程。

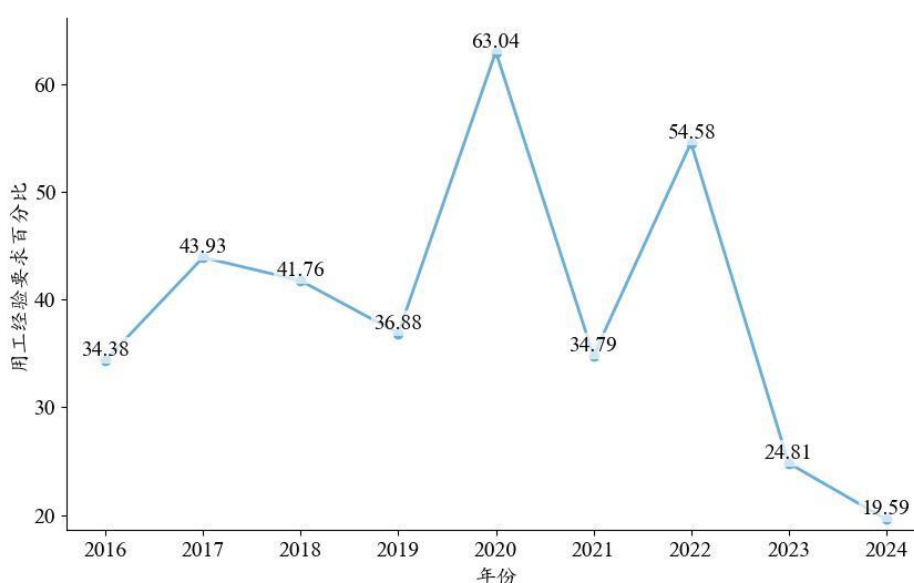


图 3-2 平均用工经验要求年度趋势

用工经验要求体现着企业对标注质量、复杂场景处理等能力的重视。在计算各年份数据标注产业招聘用工经验要求百分比的过程中，首先需对招聘岗位的“经验要求”进行分类。将岗位分为“无经验”与“有经验”两类，其中“无经验”岗位表示不要求应聘人员具有工作经验，以数字“0”标识；“有经验”岗位则表示要求应聘人员具有一定时间的工作经验，以数字“1”标识。在此基础上，以年为单位进行统计。在统计每年数据标注产业的招聘岗位总数，同时分别

统计“无经验”岗位与“有经验”岗位的数量。进而计算各年份“有经验”岗位的占比，即“有经验”岗位数量与该年份所有岗位数量之比。该比值即为该年份数据标注产业招聘用工经验要求的百分比。

就数据标注产业用工经验要求需求看，在 2016 年，平均用工经验要求为 34.38%，这一较低的百分比可能反映了初期数据标注工作的技术门槛较低。2017 年，这一比例上升至 43.93%，显示出随着产业的发展，对工人经验的要求逐渐提高。2018 年，经验要求略有下降至 41.76%，而 2019 年进一步下降至 36.88%，这可能与技术的成熟和标准化有关。然而，2020 年经验要求急剧上升至 63.04%，这主要源于专业领域对高质量标注数据的迫切需求。2021 年，经验要求下降至 34.79%，2022 年又上升至 54.58%，显示出经验要求的波动性。2023 年，经验要求大幅下降至 24.81%，而 2024 年进一步下降至 19.59%，这可说明技术的进一步发展使得数据标注工作变得更加自动化和简化，从而降低对工人经验的要求。

3.1.2 数据标注员省份需求分析

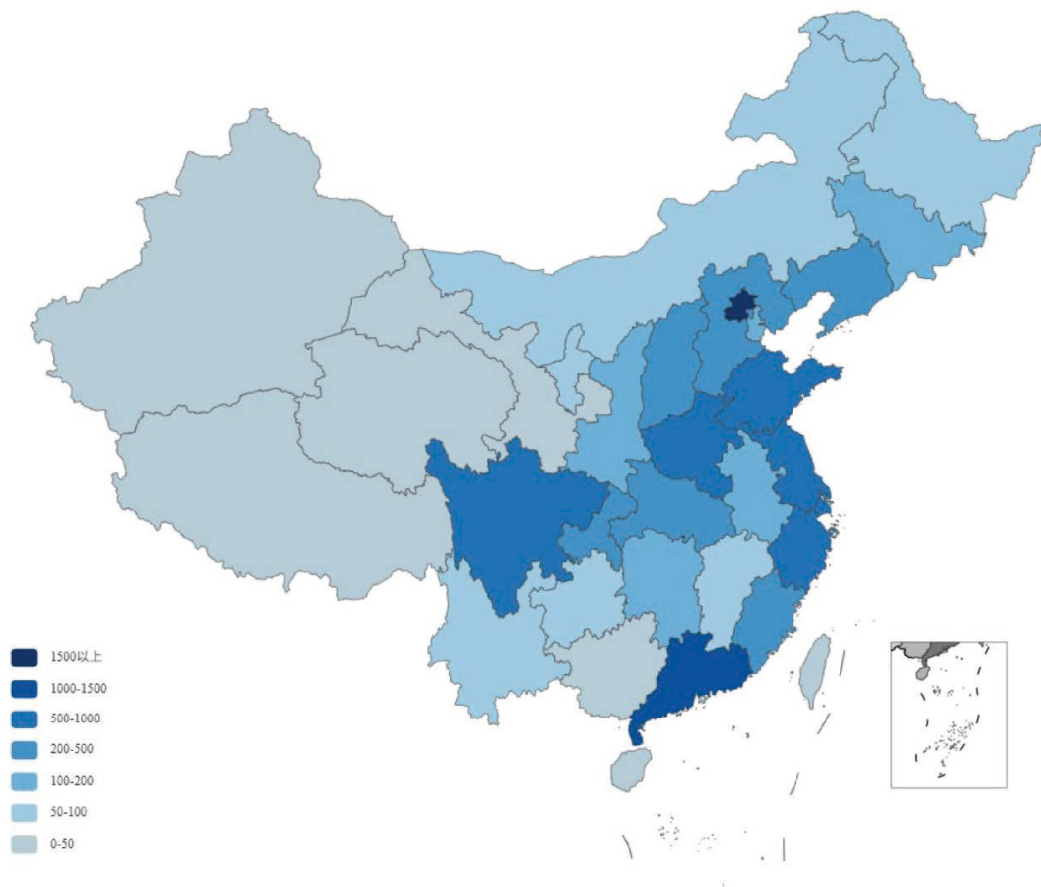


图 3-3 各省份用工岗位分布

数据标注产业中，各省份的招聘用工人数呈现出显著的地域差异^①。北京以 2555 人的招聘需求位居首位，显示出其在数据标注领域的重要地位和对人才的强烈需求。广东省和河南省分别以 1496 人和 992 人的招聘需求紧随其后，这与这些地区的经济发展水平和对数据标注服务的高需求有关。此外，上海市和江苏省的招聘需求也较为显著，分别为 911 人和 762 人。

相比之下，一些西部和边远地区的招聘需求则相对较低，如西藏自治区、青海省和新疆维吾尔自治区，招聘人数不超过 50 人。这些地区的招聘需求较低可能与经济发展水平、产业结构以及人才资源的分布有关。

总体而言，东部沿海地区及部分中部地区在数据标注产业的招聘需求方面表现出显著优势，市场地位更为突出。

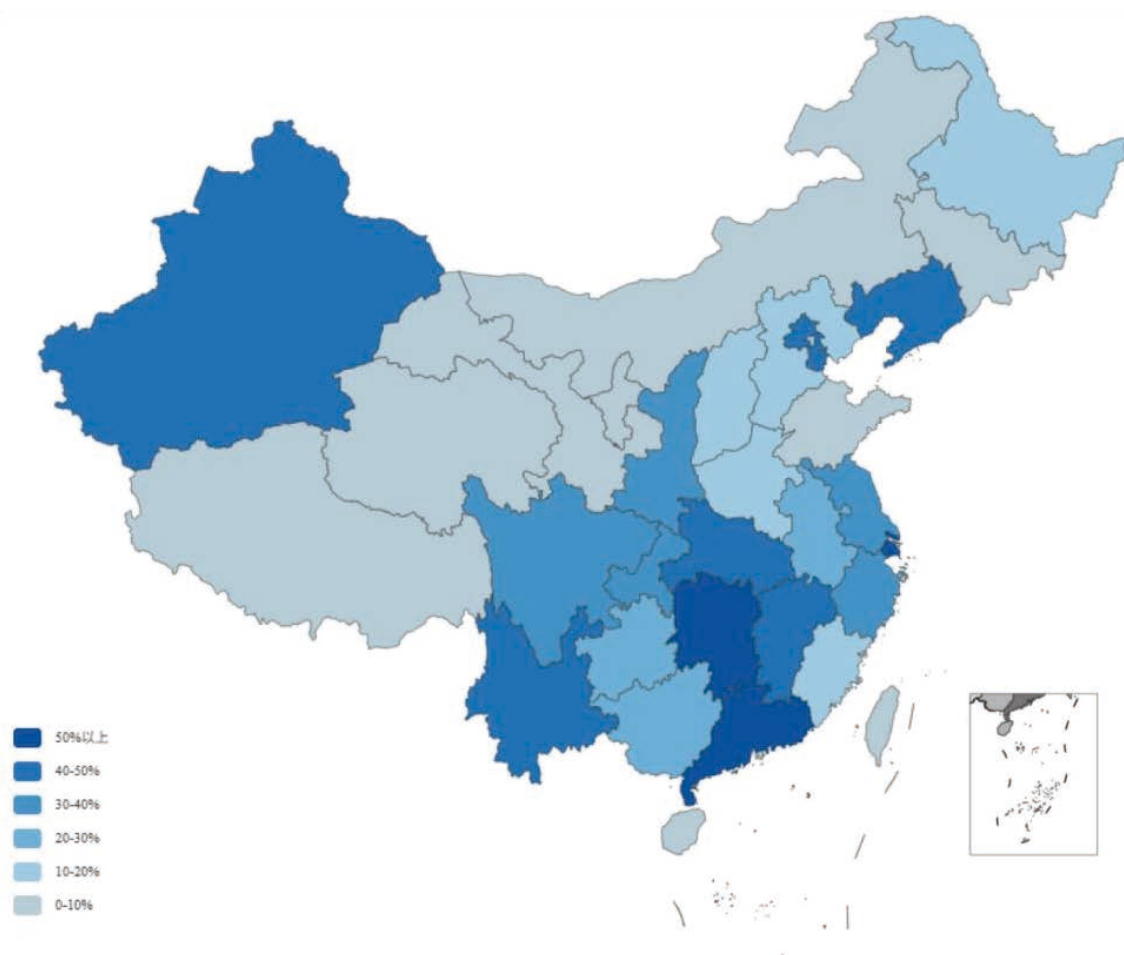


图 3-4 各省份平均用工经验要求

^① 鉴于数据的可得性，本报告对我国 31 个省级数字政府（包括 22 个省、5 个自治区与 4 个直辖市）的招聘用工数据展开分析，尚未覆盖我国的港澳台地区，特此说明。

在数据标注产业的招聘用工经验要求中，各省份呈现出不同的趋势。广东省和湖南省对经验的要求最高，分别达到了 55.01%和 50.75%，这反映了这些地区对数据标注工作的专业性和质量有着更高的期望。紧随其后的是辽宁和新疆，分别为 48.84%和 48.15%，显示出这些地区对经验丰富的数据标注人员的偏好。相比之下，宁夏、青海、内蒙古和甘肃等地区在数据标注产业的招聘中对经验的要求相对较低。这或许反映出这些地区在人才储备和产业成熟度上与东部沿海地区存在一定差距，因此更倾向于招聘新手或对工作经验要求不高的求职者，以满足其当前产业发展的需求。此外，北京、上海和天津等一线城市对经验的要求也相对较高，这可能与这些城市的经济发展水平和对高技能人才储备充分有关。

总体来看，东部沿海地区和部分西部地区对数据标注人员的经验要求较高，而一些中部和东北地区则相对较低。这种分布规律可能与各地区的经济发展水平、产业结构以及对数据标注服务模式有关。

3.1.3 数据标注员行业需求分析

在数据标注就业市场中，不同行业的用工需求呈现出显著的差异性。通过分析各行业的用工人数、用工经验要求，可以更深入地了解这些差异，并从中总结出产业发展规律。

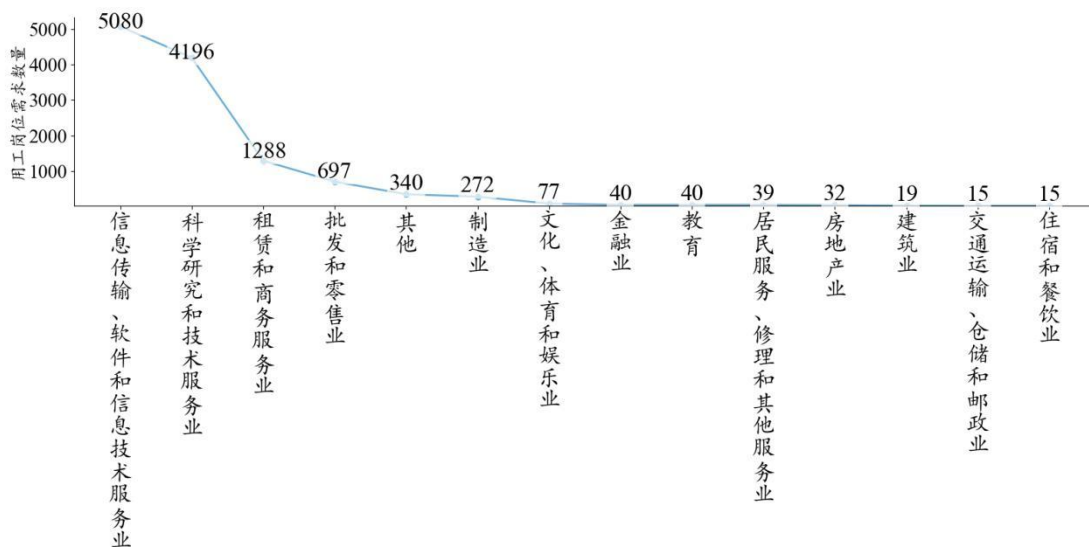


图 3-5 用工岗位行业差异

首先，从用工人数来看，信息传输、软件和信息技术服务业以及科学研究和技术服务业的用工需求最为旺盛，分别为 5080 人和 4196 人。这两个行业的高需求反映了当前社会对高科技和研发能力的重视。租赁和商务服务业以及批发和零

售业的用工需求也相对较高，分别为 1288 人和 697 人，这可能与这些行业的服务性质和市场需求有关。相比之下，制造业行业、文化、体育和娱乐业行业的用工需求则相对较低，分别为 272 人、77 人。而金融业、教育等行业的用工需求则更为有限，均在 50 人以下。

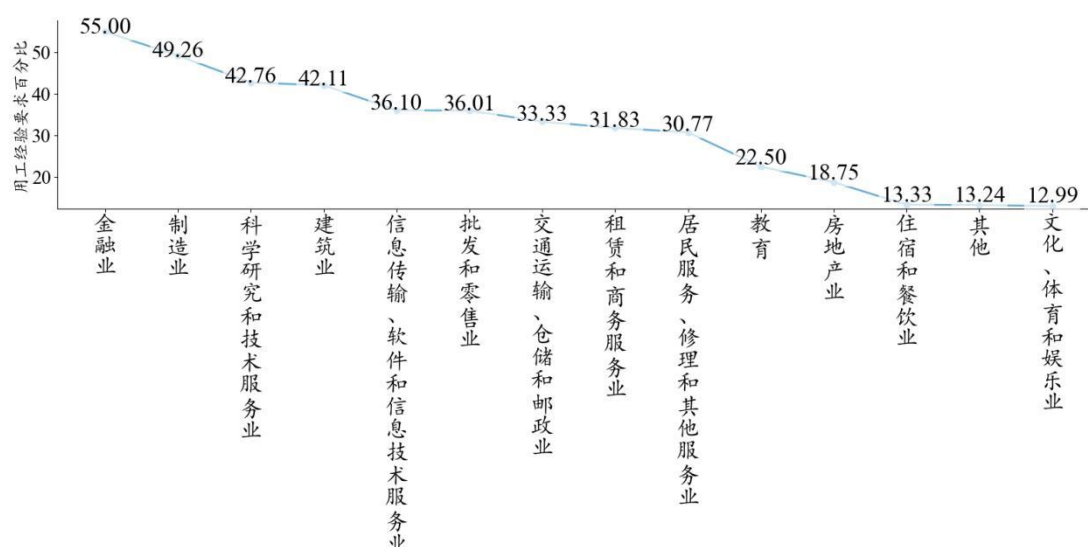


图 3-6 用工经验要求行业差异

从用工经验要求来看，金融业对员工的经验要求最高，达到 55%，这可能与金融行业的专业性和风险性有关，需要经验丰富的员工来确保业务的稳定性和安全性。制造业、科学研究和技术服务业的经验要求也相对较高，分别为 49.26% 和 42.76%。这些行业的高经验要求可能与技术复杂性和创新需求有关。相比之下，住宿和餐饮业以及文化、体育和娱乐业的用工经验要求则低于 15%，这反映出相关行业的标注任务技能要求相对较低，更倾向于招聘新手。

3.1.4 “年份-省份-行业”交叉分析

从 2016 年至 2024 年的数据标注产业招聘用工的“省份-年份”数量分布看，北京、广东、上海、江苏和山东等东部沿海地区在数据标注产业中占据显著位置，显示出这些地区在数据标注产业中的活跃程度和市场需求。北京在 2022 年达到 728 人的高峰，尽管之后有所波动，但 2024 年仍然保持在 394 人，显示出其在数据标注产业中的重要地位。广东在 2022 年达到 445 人，2024 年略有下降至 421 人，但整体需求依然强劲。中部地区如河南等省份的用工人数近年保持稳定增长趋势，西部地区的用工人数虽然相对较少，但也显示出一定的增长趋势。

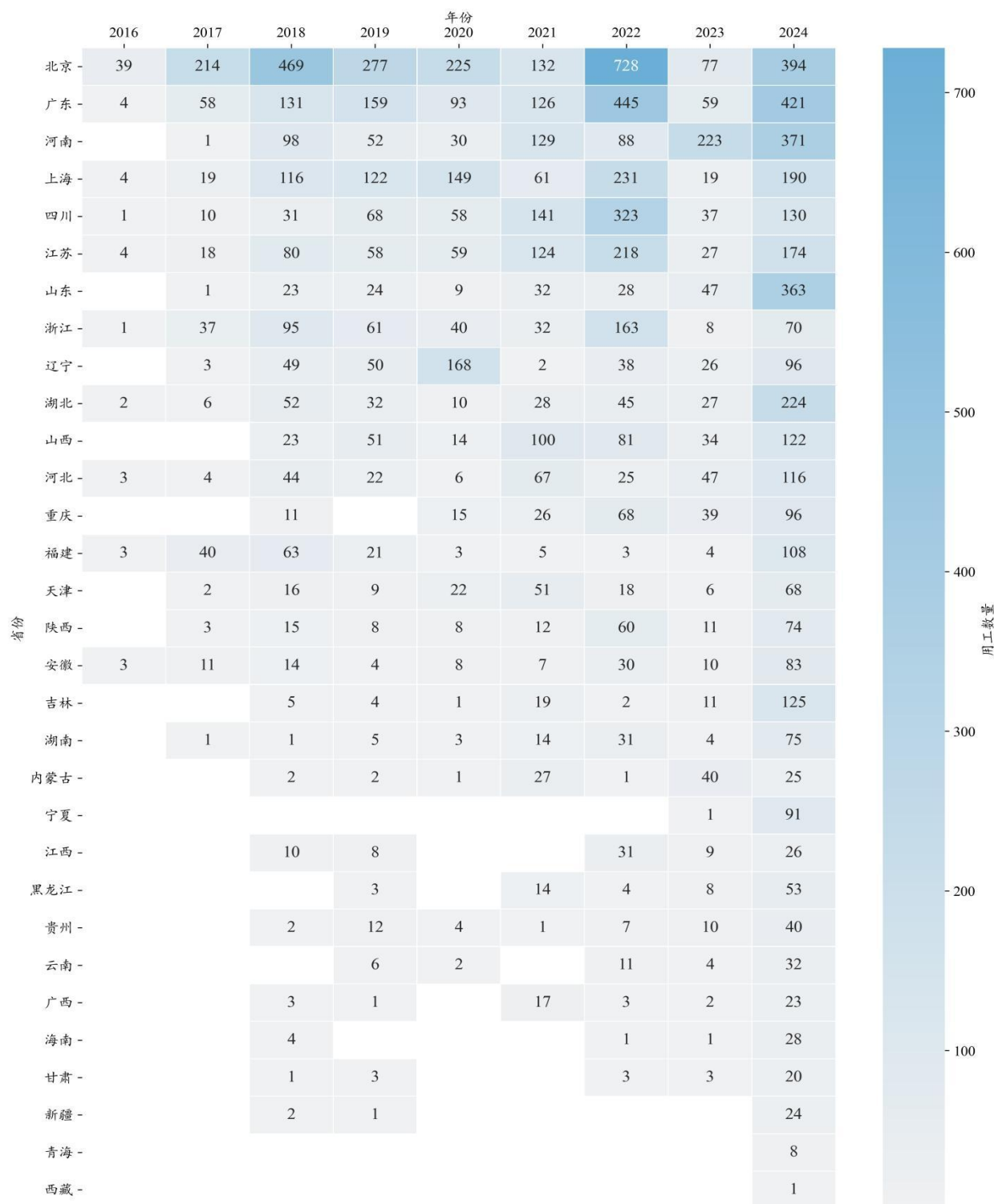


图 3-7 各省份各年份用工数量分布示意图

总体来看，数据标注产业的用工需求在各地区呈现出一定的差异性，东部地区由于经济发达、科技产业集中，在数据标注需求上占据优势。中部地区虽然起步较晚，但随着科技产业的发展，需求也在逐渐增长。西部地区虽然需求相对较低，但随着国家政策的支持和科技产业的布局，未来可能会有更大的发展空间。

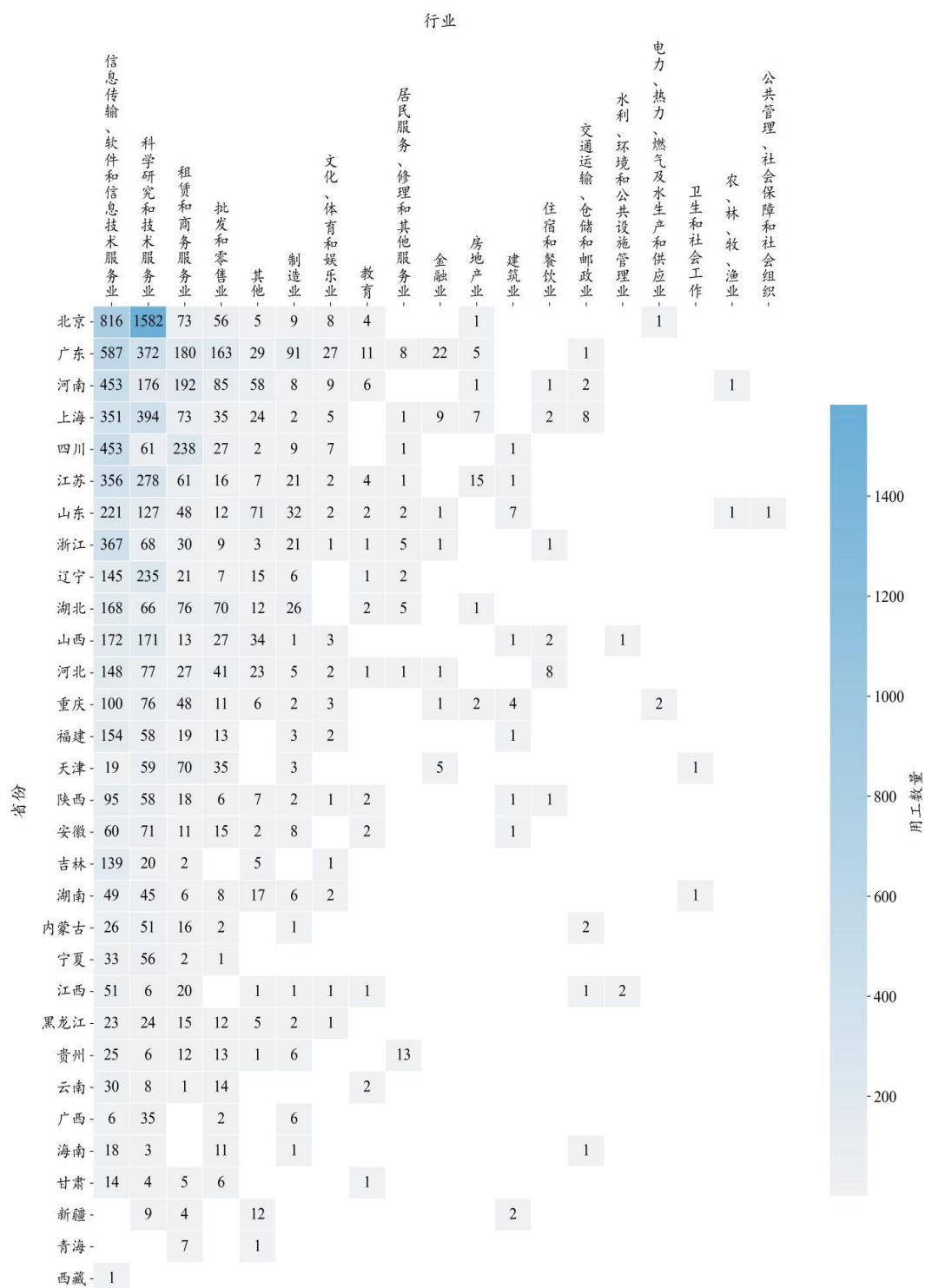


图 3-8 各省份各行业用工数量分布示意图

从“省份-行业”用工数量分布可以看出，数据标注产业在中国各省份的用工需求呈现出显著的地域和行业分布特点。北京和广东作为经济和科技中心，在数据标注产业中表现尤为突出。北京的信息传输、软件和信息技术服务业用工数量达到 816 人，科学研究和技术服务业更是高达 1582 人，显示出北京在高新技

术产业方面的强劲需求。广东的信息传输、软件和信息技术服务业用工数量为 587 人，科学研究和技术服务业为 372 人，租赁和商务服务业为 180 人，批发和零售业为 163 人，这些数据表明广东在多个行业中对数据标注的需求都较为旺盛。

河南是中部地区代表性省份，河南的信息传输、软件和信息技术服务业用工数量为 453 人，科学研究和技术服务业为 176 人，这些数据表明河南在数据标注产业中的需求正在逐步增长。湖北的信息传输、软件和信息技术服务业用工需求分别为 168 人，显示出中部地区在数据标注产业相关行业中的发展潜力。从整体上看，西部地区各行业用工需求相对东中部地区都无明显优势，四川是西部地区的代表，四川的信息传输、软件和信息技术服务业用工数量为 453 人，科学研究和技术服务业为 61 人，租赁和商务服务业为 238 人。

从行业分布来看，信息传输、软件和信息技术服务业在各省份中普遍占据主导地位，因为数据标注是信息技术服务的基础部分，因此该行业在各省份中需求较大。此外，科学研究和技术服务业、批发和零售业、租赁和商务服务业等行业也有较多的企业需求数据标注服务，北京的突出特点是科学研究和技术服务业所需数据标注人员占比极高，其数据标注需求主要服务知识创新与技术服务。

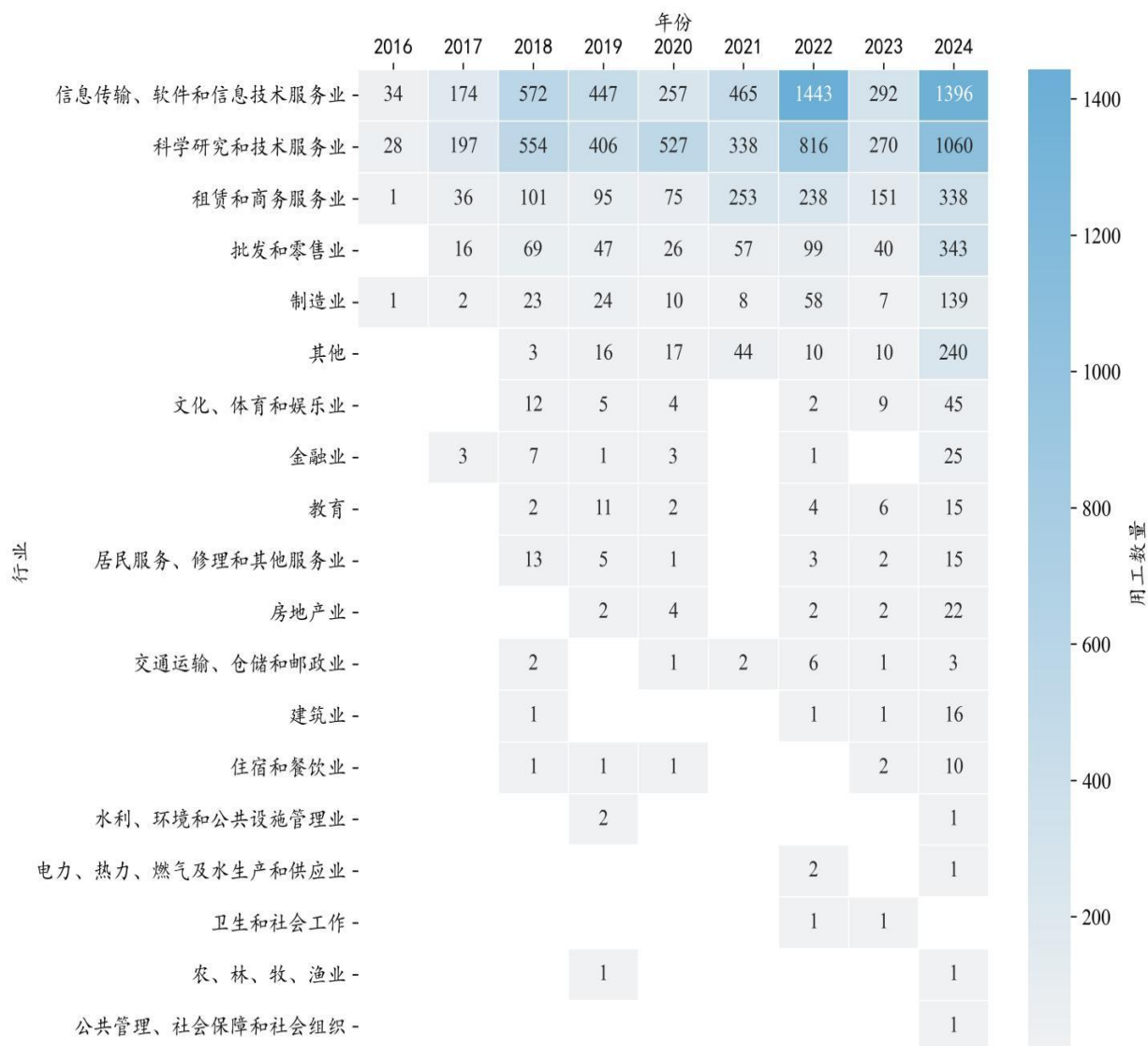


图 3-9 各行业各年份用工数量分布示意图

由“年份-行业”用工数量分布可知，信息技术以及科学研究在整个观测时间段内始终是数据标注用工需求的主要来源，保持着增长趋势。而 2022 年至 2024 年用工需求波动也主要受这两类行业用工需求波动影响。科学研究和技术服务业对数据标注员专业技能和经验的要求更高，2020 年科学研究和技术服务业用工需求的增长也侧面反映为何当年用工经验需求显著提高。租赁和商务服务业、批发和零售业的用工需求自 2016 年起基本保持增长趋势，2024 年超 300 人，这可能与电子商务的快速发展有关，因为电商平台需要大量的数据标注来优化推荐系统和用户体验。

总体来看，数据标注产业的需求在各行业中呈现出一定的差异性，但随着社会数字化转型的推进，预计未来各行业对数据标注的需求将继续增长。

3.2 用工数据标注企业趋势分析

分析用工企业相关数据趋势，考察企业数量变化、地域分布和行业差异，能够揭示产业规模动态变化和产业集聚特征。

3.2.1 用工数据标注企业单维分析

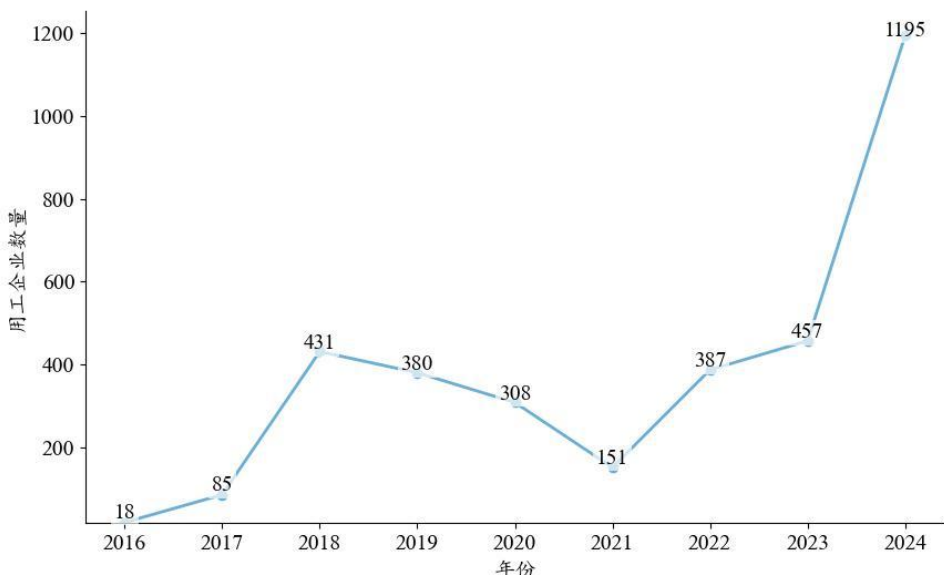


图 3-10 用工企业数量年度趋势

通过分析 2016 年至 2024 年的数据标注产业中具有招聘用工需求的企业数量，可以观察到一个较为显著的增长趋势。2016 年，仅有 18 家企业有数据标注的招聘需求。随后，在 2017 年，这一数字显著增加至 85 家，显示出数据标注需求开始受到市场的关注。2018 年，企业数量激增至 431 家。人工智能技术在自然语言处理和计算机视觉等领域取得显著进展，人工智能应用场景的多样化，如自动驾驶、医疗影像分析等，进一步扩大了数据标注的市场需求。

2019 年至 2021 年，企业数量从 380 家下降至 151 家，在前技术发展突破引致数据标注需求被逐步消化，具有招聘用工需求的企业数量下降。受生成式人工智能等关键技术突破的影响，从 2022 年开始，企业数量再次显著上升，2022 年为 387 家，2023 年增至 457 家，到 2024 年更是达到了 1195 家，这一增长趋势反映了数据标注在多个行业中的重要性日益增加。总体来看，数据标注产业的企业数量经历了初期的快速增长、中期的调整和近期的再次快速增长，显示出该行业的市场需求和发展潜力的不断变化。

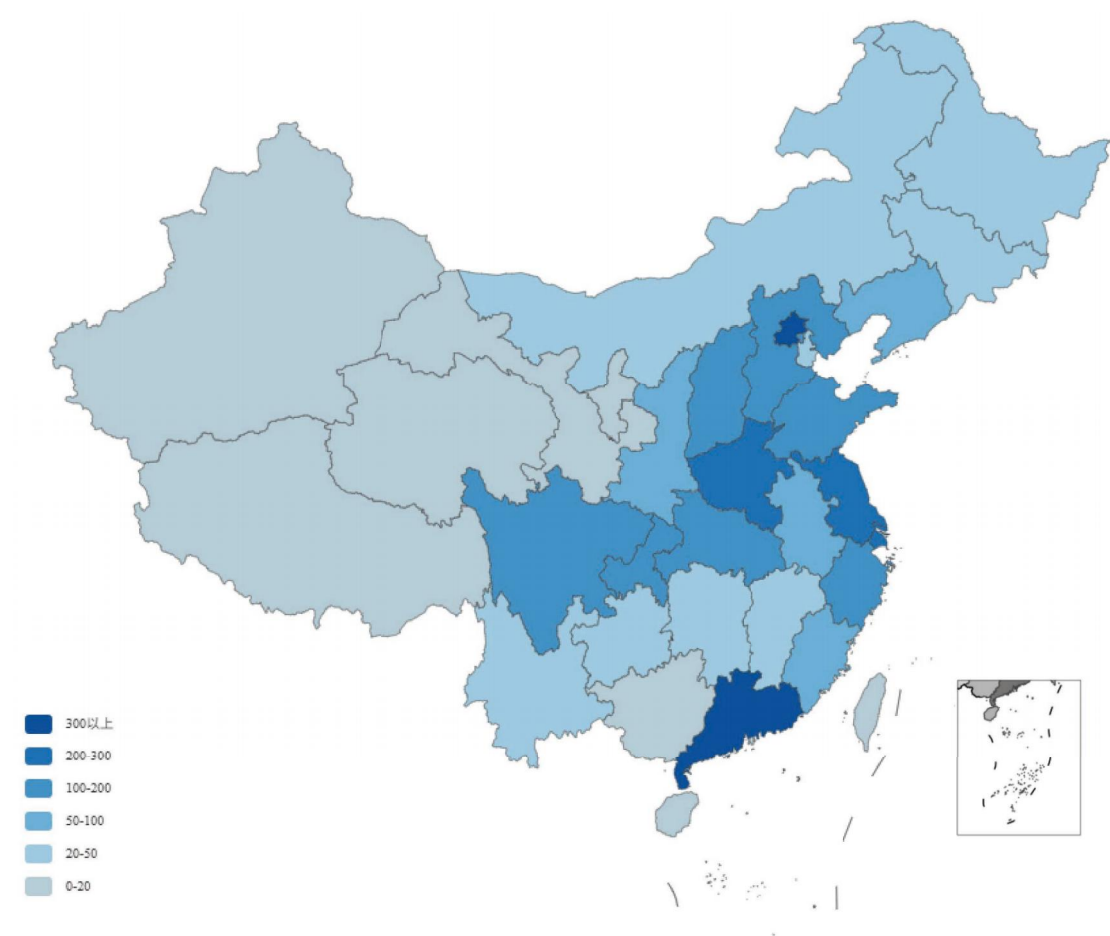


图 3-11 各省份用工企业分布

从具有招聘用工需求的企业省份分布看，北京市拥有最多的数据标注企业，数量达到 530 家，这可能与其作为中国的政治、文化和科技中心的地位有关，吸引了大量的数据标注企业。广东省以 425 家企业紧随其后，这两个省份的经济发达，科技产业集中，对数据标注服务的需求较大。河南和上海的数据标注企业数量也较多，分别为 276 家和 264 家，显示出这两个地区在数据服务领域的活跃度。

相比之下，西藏自治区、青海省和宁夏回族自治区的数据标注企业数量较少，均不足 20 家，这可能与这些地区的经济发展水平和科技产业的集中度有关。此外，甘肃省、新疆维吾尔自治区和海南省的企业数量也相对较少。

总体来看，具有招聘用工需求的企业省份分布与数据标注员省份需求分布极为相似，东部沿海地区由于经济发达，科技产业集中，对数据标注服务的需求较大，因此吸引了更多的数据标注企业。而西部和边远地区由于经济发展水平相对较低，科技产业的集中度不高，对数据标注服务的需求较小，因此具有招聘用工需求的企业也相对较少。

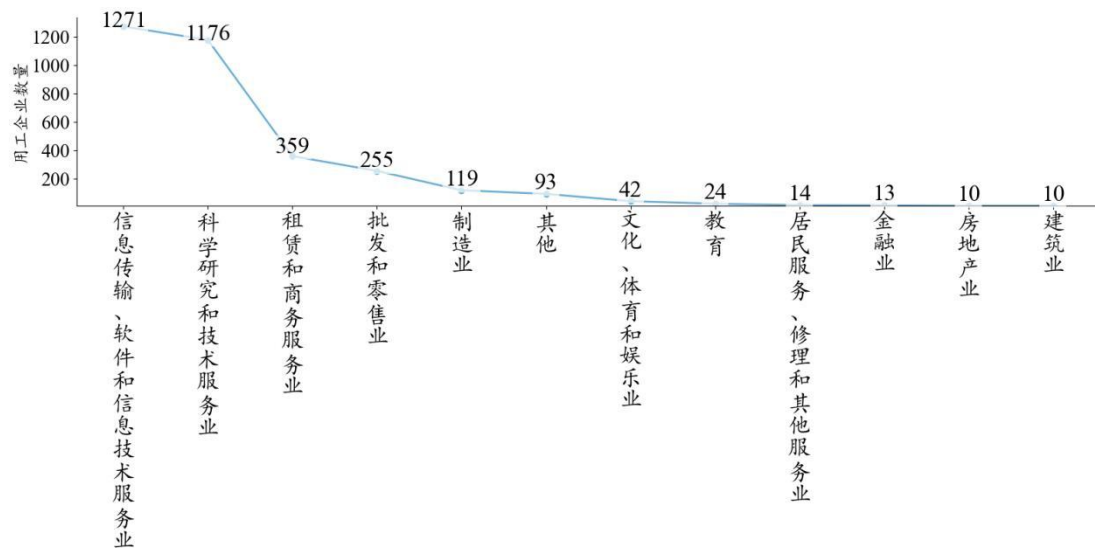


图 3-12 用工企业数量行业差异

从行业层面看，具有招聘用工需求的企业行业分布与数据标注员行业需求分布具有相似性，信息传输、软件和信息技术服务业以及科学研究和技术服务业在数据标注招聘用工需求方面表现突出，分别有 1271 家和 1176 家企业，这反映了这两个行业在数据处理、分析和人工智能应用方面的广泛需求，对数据标注服务有高依赖性。租赁和商务服务业以及批发和零售业的数据标注需求也相对较高，分别有 359 家和 255 家企业，这可能与这些行业在客户服务、市场分析和供应链管理中数据标注的需求有关。制造业的数据标注需求为 119 家，这可能与制造业在产品开发、质量控制和智能制造中对数据标注的需求有关。相比之下，文化、体育和娱乐业、教育等行业的数据标注需求相对较低。

由此可见，数据标注需求在不同行业间存在显著差异，这可能与各行业的业务性质、技术应用水平和数据处理需求有关。信息技术行业和科研行业的需求最为旺盛，而传统行业的需求相对较低。

3.2.2 用工数据标注企业多维分析

基于年份、省份、行业可以交叉分析具有招聘用工需求企业的分布结构细节。

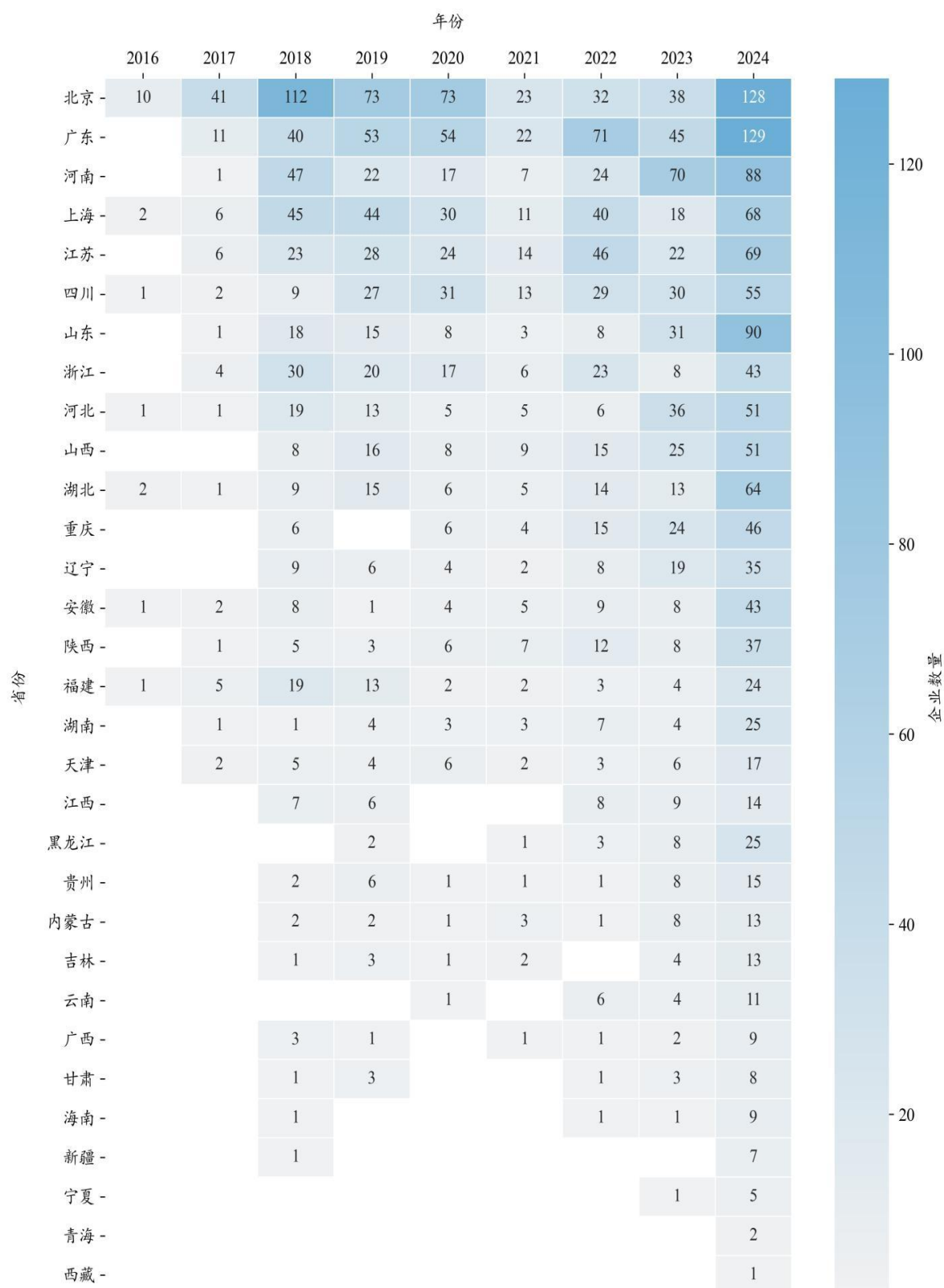


图 3-13 各省份各年份企业数量分布示意图

从 2016 年至 2024 年，各省份的企业数量呈现出明显的变化趋势。北京和广东作为经济发达地区，在整个时间段内一直保持着较高的企业数量。东部沿海地区如上海、江苏、山东等地的企业数量也相对较多，显示出这些地区在数据标注产业中具有较强的竞争力和市场需求。中部地区如河南等省份的企业数量相对东部地区较少，但也呈现出逐年增长的趋势，尤其是河南具有招聘用工需求的企业在 2024 年达到 88 家。西部地区如四川、重庆等省份的企业数量相对较少，但也在逐步增长。总体来看，东部地区在数据标注产业中占据主导地位，中部和西部地区虽然起步较晚，但也在逐步追赶。

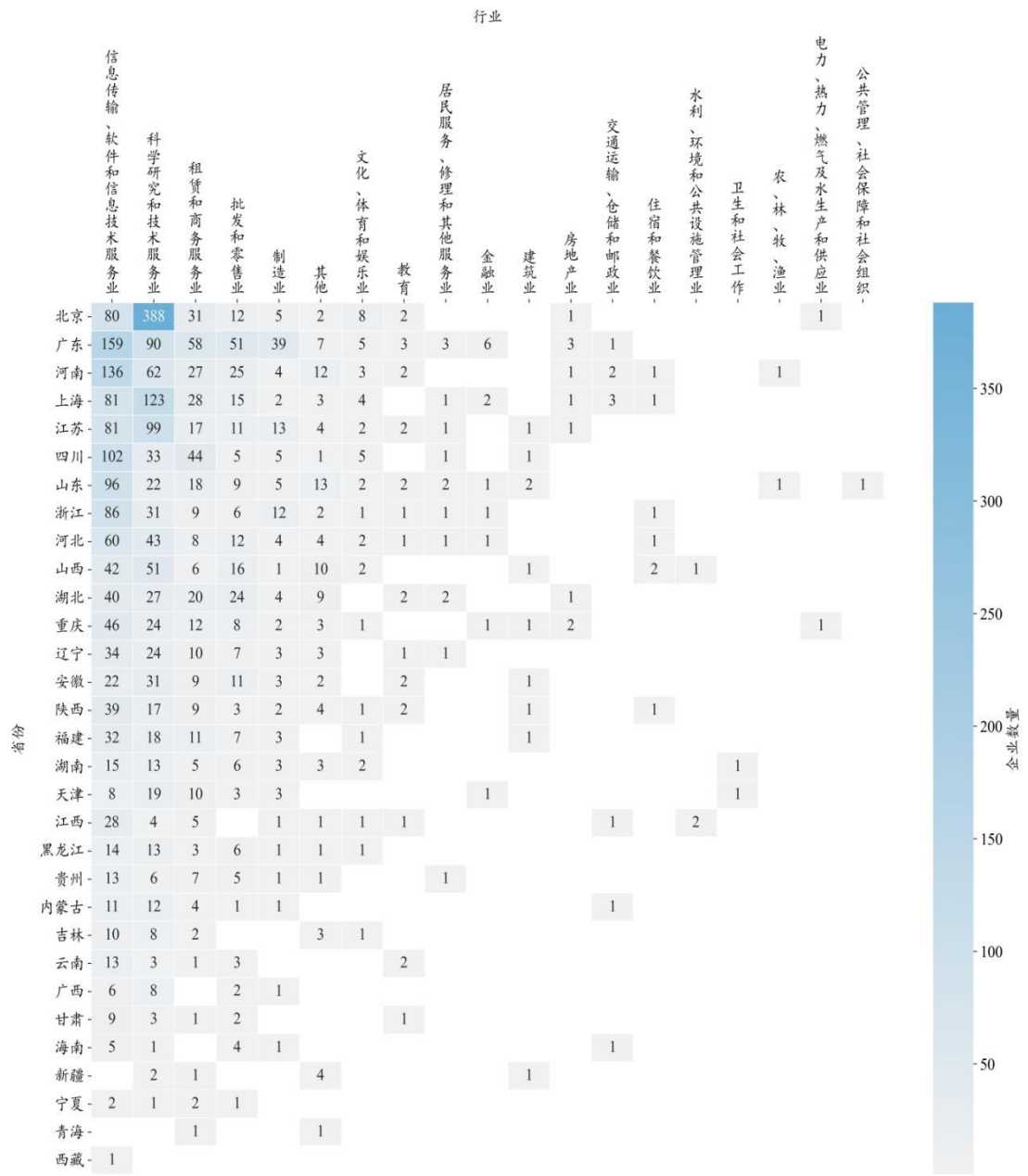


图 3-14 各省份各行业企业数量分布示意图

根据提供的各省份各行业企业数量分布可以观察到数据标注产业在不同地区和行业的分布情况呈现出一定的规律性。北京在科学研究和技术服务业中表现尤为突出，有 388 家，显示出北京在技术创新、高新技术产业方面的强大投入和产业集聚效应。广东紧随其后，在信息传输、软件和信息技术服务业有招聘用工需求的企业为 159 家，科学研究和技术服务业有 90 家。

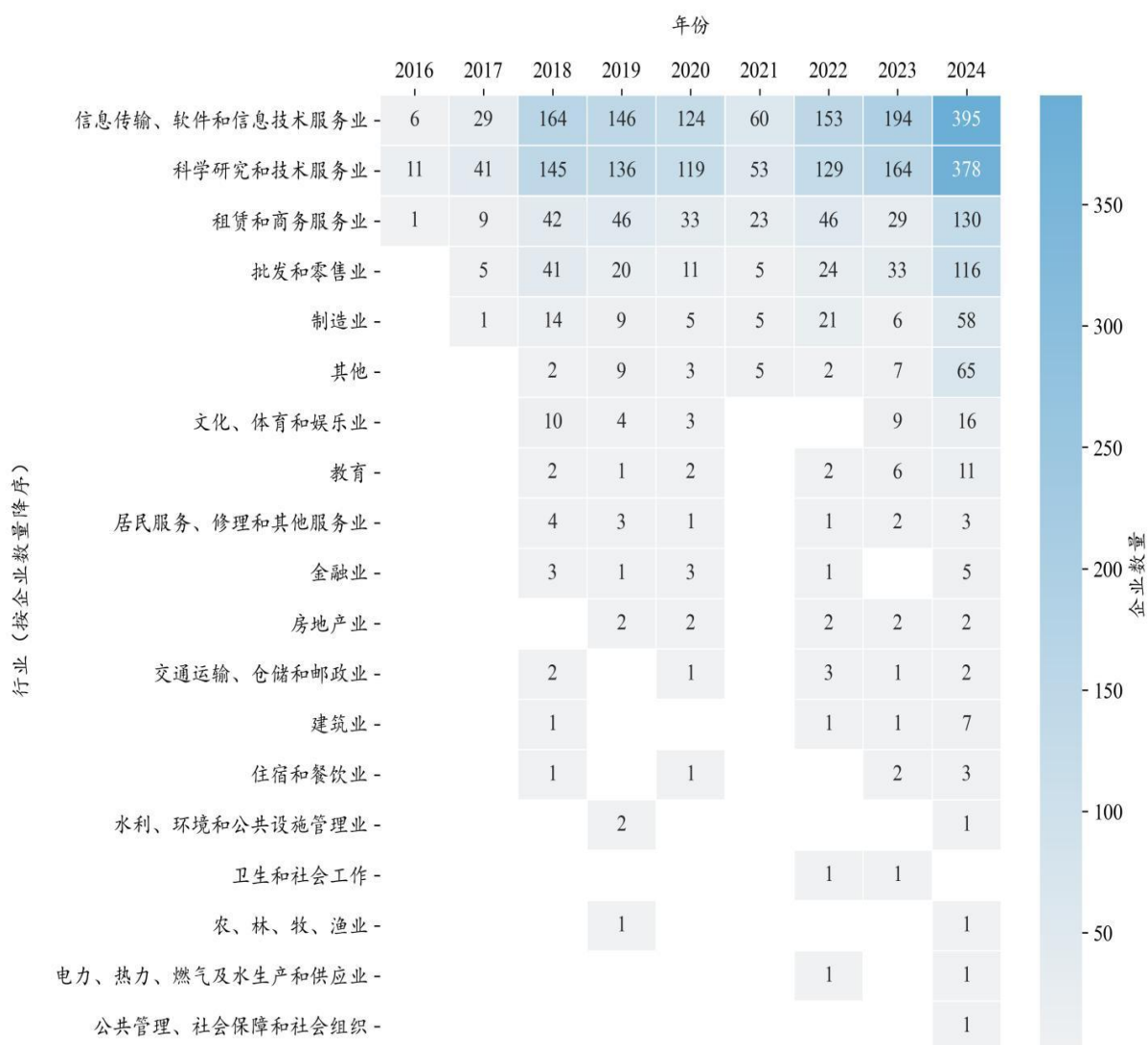


图 3-15 各行业各年份企业数量分布示意图

从 2016 年至 2024 年数据标注产业的企业数量变化来看，信息传输、软件和信息技术服务业一直是数据标注用工需求的主要来源。该行业在 2024 年具有招聘用工需求的企业达到了 395 家，显示出显著的增长趋势，反映出信息技术行业的快速发展和对数据标注服务的持续高需求。科学研究和技术服务业也是重要的用工需求方，2024 年有 378 家企业需要数据标注服务，这一增长趋势与科技创

新和研发活动的增加密切相关。此外，信息传输、软件和信息技术服务业与科学研究和技术服务业的企业分布保持类似的时间变动趋势。租赁和商务服务业、批发和零售业、制造业的企业数量也基本保持增长态势。

3.3 数据标注产业发展挑战

近年来，数据标注产业呈现出技术驱动型增长的特点，其发展轨迹与人工智能技术的突破高度相关。从招聘数据来看，2016 年行业萌芽期仅有 64 人的用工需求，2022 年 ChatGPT 等生成式人工智能的兴起催生了文本标注、指令微调等新需求，用工需求激增至 2686 人，但 2023 年又回落至 794 人，到 2024 年受 GPT-4o 多模态大模型和国内 DeepSeek-V3 等技术突破的推动，用工规模呈现指数级扩张。这说明产业受技术迭代影响显著，呈现周期性波动。目前，信息传输、软件和信息技术服务业占比最高，科学研究和技术服务业紧随其后，表明数据标注已成为人工智能产业链的关键支撑环节。尽管数据标注产业规模持续扩大，但产业发展仍面临多重挑战：

首先，数据标注产业正面临激烈的市场竞争格局。随着人工智能应用场景的持续拓展，数据标注市场需求呈现快速增长态势，吸引了大量企业争相进入这一领域。然而，缺乏核心竞争力的企业往往将价格战作为主要竞争策略，通过大幅压低服务报价来争夺市场份额。从招聘数据来看，2022 年生成式人工智能技术突破引发行业热潮，企业数量呈现爆发式增长，但部分技术储备薄弱、资金实力有限的中小企业将要面临严峻的生存竞争挑战。这种市场格局的变化，反映出数据标注产业应由粗放式增长向高质量增长转型。

其次，用工需求波动大，就业稳定性差。数据标注产业的用工需求呈现出显著的波动性特征，这种不稳定性直接影响了从业者的就业质量。随着人工智能技术迭代加速，行业用工需求呈现周期性变化，2022 年生成式人工智能爆发带动用工需求激增，但 2023 年随即大幅下降。在用工模式方面，众包模式虽然提高了就业灵活性，但也带来了一系列就业质量问题，如工作时间碎片化严重、从业者缺乏必要社会保障、难以获得系统的技能培训和职业晋升。这不仅影响个人职业成长，也制约了行业人才队伍的稳定性建设。随着人工智能技术向专业化、精细化方向发展，这种低质量的就业模式已难以满足产业升级需求，亟需通过技术创新和模式优化来提升就业质量。

最后，应以技术为核心竞争力，传统人工标注模式面临根本性变革。随着人工智能模型复杂度的提升，传统人工标注模式在效率、成本和质量上均遭遇瓶颈。例如，自动驾驶需要像素级标注精度，大模型训练需要百万级指令微调样本，这些需求已远超传统人工标注的能力边界。招聘数据显示，2024 年企业对标注员的经验要求较 2020 年明显降低。这一趋势可能源于多重因素：一方面，随着人工智能标注工具的成熟，部分标准化工作正逐步由算法接管；另一方面，行业分工的细化使得新手也能快速胜任基础标注任务。当前行业更显著的特征是人机协作模式的深化：智能工具处理可规范化的基础工作，而人工则专注于需要领域知识和质量把控的关键环节。这种分工优化既提升了整体效率，也为标注员创造了向“人工智能训练师”等更高价值岗位转型的空间。

面对上述挑战，智能数据标注成为行业升级的关键路径。它通过人工智能辅助标注、自动化质检、合成数据生成等技术，大幅提升效率与质量。例如，人工智能辅助标注利用预训练模型完成基础标注工作，人工仅需进行质量审核，显著降低人力成本；自动化质检系统通过算法检测标注错误，将错误率控制在低水平，确保数据质量；合成数据技术生成多样化训练样本，解决罕见场景数据稀缺问题，同时避免隐私合规风险。可见，智能标注技术的成熟，将推动数据标注从“人力密集型”向“技术驱动型”转变，成为人工智能时代数据基础设施的核心环节。

4.智能标注前沿技术与企业实践观察

分析智能标注技术发展与企业实践，旨在揭示人工智能产业链中数据标注服务环节的技术演进路径与商业应用模式。通过考察智能标注技术的前沿发展与典型企业实践案例，能够为产业把握技术发展趋势、优化标注流程、提升数据质量提供重要参考，同时为相关企业的战略布局和技术选择提供决策依据。

4.1 智能数据标注技术发展前沿

数据标注产业作为人工智能发展的关键支撑，正面临着转型升级的重要机遇，而智能标注技术作为转型升级的重要支撑，近年来取得了显著进展。

4.1.1 数据智能自动标注技术

数据智能自动标注技术是智能标注的基础支撑技术，通过算法自动生成、优化标注数据，显著提升模型训练效率与质量。其核心涵盖三大方向：合成数据技术模拟真实场景，解决特定场景数据稀缺等问题；领域自适应算法通过迁移学习实现无标注数据的精准分类，降低标注成本；大模型驱动查询式摘要借助提示工程与自我校验机制，生成高质量标注数据集。这些技术共同推动人工智能模型在医疗、金融、工业等场景的快速落地，成为数据要素市场化的重要引擎。

合成数据是通过算法程序人工生成的数据，旨在模拟真实世界情况。与真实数据相比，合成数据具有诸多优势，例如可以避免隐私问题、降低数据采集成本、提供更丰富的数据场景等。合成数据的主要应用场景涵盖自动驾驶、机器人、生物医药、金融、工业等领域，具体应用包括数据增强、模型验证、可解释人工智能、具身智能、AR/VR 等。

随着人工智能技术的发展，合成数据的潜力逐渐被挖掘。例如，合成数据可以用于生成罕见场景或极端条件下的数据，帮助模型更好地学习和适应各种情况。此外，合成数据还可以通过数据增强技术，提高模型的泛化能力。尽管合成数据在某些方面具有优势，但其生成过程中的偏见和质量问题也需要引起关注。例如，合成数据可能携带生成算法中的偏见，或者由于生成过程中的随机性导致数据质量不稳定。因此，合成数据的生成和使用需要严格地监控和评估机制，以确保其对人工智能模型训练的有效性和安全性。

随着大数据时代的到来，数据标注成本高昂的问题日益突出。领域自适应算法作为一种迁移学习方法，通过将已标注数据集（源领域）的知识迁移到无标注

数据集（目标领域），显著降低了标注成本。这种算法不仅能够提高模型在无标注数据集上的分类性能，还能扩展深度学习算法的应用范围，例如在医学影像分析、情感分析等领域的应用。

领域自适应算法的核心在于学习源领域和目标领域之间的通用特征，从而减少领域偏移的影响。通过引入对抗学习机制，算法能够同时学习源领域和目标领域的特征，生成适用于目标领域的数据表示。这种方法不仅提高了模型的泛化能力，还为无标注数据的分类问题提供了有效的解决方案。

基于大模型和提示工程的查询式摘要数据集是一种新兴的数据标注方法。这种方法通过设计生成能够自我校验的提示链，利用大模型作为“数据标注员”，实现自动化数据标注。提示链包括角色定义、任务描述、任务要求、生成示例以及给定文档等部分，引导模型生成符合要求的查询式摘要。通过自我校验机制对生成的摘要进行一致性校验，校验通过的数据被纳入最终的数据集，未通过的则被丢弃，从而确保数据集的质量。这种方法不仅能够显著提高标注效率，还能借助生成式模型的强大语言理解能力，生成高质量的标注数据。此外，基于提示工程的数据标注方法还可以根据不同的应用场景和需求，灵活调整提示模板，生成多样化的标注数据，满足不同人工智能模型的训练需求。

4.1.2 人机协同自动化标注技术

人机协同自动化标注技术通过人工智能预标注与人工校验的闭环机制，实现了标注效率与质量的协同提升。当前技术已形成三大创新方向：众包任务推荐引擎、智能质检系统、模块化标注系统集成。这些技术创新不仅大幅提升了标注效率与质量，更为人工智能的规模化工业应用构建了关键基础设施支撑。

首先是众包平台个性化题目推荐系统。众包平台的个性化题目推荐系统是提升标注效率和质量的重要手段。通过分析标注员的历史行为和偏好，推荐系统可以为标注员提供符合其能力和兴趣的标注任务。这种个性化推荐不仅能够提高标注员的工作满意度和效率，还能减少标注员在任务选择上的时间和精力消耗。个性化推荐系统的设计基于用户画像和任务特征的匹配算法。通过收集标注员的基本信息、标注历史、任务完成时间和质量等数据，系统可以构建标注员的用户画像。同时，系统还会对标注任务进行特征提取，包括任务类型、难度、领域等信息。通过匹配算法，系统能够为标注员推荐最适合的任务，从而实现标注任务的

高效分配和标注员的高效工作。

其次是人机协同改进数据标注项目质量管理。在数据标注项目中，人员质量控制是确保标注数据准确性和一致性的关键环节。传统的人工标注模式依赖于大量标注员和质检员，不仅效率低下，还容易因主观差异导致标注质量参差不齐。随着人工智能技术的发展，人机协同模式逐渐成为主流。通过引入智能标注工具和算法，标注员可以借助预标注功能快速完成标注任务，而质检员则可以利用算法辅助质检，快速发现并纠正错误。这种人机协同模式不仅提高了标注效率，还通过算法的客观性和一致性，显著提升了标注质量。此外，智能标注系统还能够实时监控标注过程，通过数据分析和反馈机制，为标注员和质检员提供实时指导和改进建议。例如，系统可以自动识别标注员的常见错误，并提供针对性的培训内容，帮助标注员提升技能水平。这种基于数据驱动的质量管理方式，不仅能够有效减少标注错误，还能持续优化标注流程，提升整体项目质量。

此外，数据采集标注系统搭建也至关重要。数据采集标注系统是数据标注产业的核心工具，其设计和实现直接影响标注效率和数据质量。现代数据采集标注系统通常采用 B/S 架构（Browser/Server Architecture，浏览器/服务器架构），包括基础设施层、数据资源层、服务资源层、系统应用层和用户表现层。基础设施层负责系统的稳定运行和维护更新；数据资源层用于存储海量数据信息；服务资源层利用数据资源为用户提供各项服务；系统应用层实现系统各项业务功能；用户表现层则涵盖界面设计与展示、知识图谱构建以及推荐列表可视化等。

数据采集标注系统的关键技术包括数据预处理、标注工具的开发、算法辅助标注、数据质量控制和数据安全保护。数据预处理通过清洗、归一化等操作，确保数据的质量和可用性。标注工具的开发需要结合人工智能技术，提供高效、易用的标注界面和功能。算法辅助标注通过预标注和智能推荐，减少人工标注的工作量。数据质量控制通过多轮审核和算法质检，确保标注数据的准确性和一致性。数据安全保护则通过加密、去标识化等技术手段，保护数据的隐私和安全。

4.2 智能标注企业实践及分类分析

分析智能标注企业的实践模式与分类特征，旨在总结领先企业的智能化实践并分类研究不同类型企业的发展策略，为企业定位和资源投入提供决策依据。

4.2.1 智能标注企业实践模式总结

随着人工智能技术的快速发展，智能数据标注产业正经历着从传统劳动密集型向技术驱动型的深刻转型。在这一转型过程中，智能技术的应用已全面渗透到数据标注工作流程的各个环节，包括数据准备、标注说明培训、任务创建与派发、标注执行以及质量检查等，显著提升了标注效率和质量，同时降低了人力成本。通过对行业实践的观察可以发现，领先的智能标注企业通过技术创新和模式优化，已经形成了一套成熟且高效的智能标注实践模式。

在数据准备环节，智能技术的应用主要体现在数据预处理和合成数据生成两个方面。传统的数据准备需要大量人工进行数据清洗和整理，而现代智能标注系统通过自动化工具实现了数据的去重、去噪、归一化等预处理操作，大幅提高了数据质量。更为前沿的是合成数据技术的应用，通过算法程序生成模拟真实场景的数据，不仅解决了数据隐私问题，还能生成罕见场景或极端条件下的数据样本，为模型训练提供更全面的数据支持。例如，在自动驾驶领域，合成数据可以模拟各种天气条件和交通场景，极大丰富了训练数据的多样性。

标注说明培训环节的智能化主要体现在知识传递方式的创新。传统的标注培训依赖纸质文档或简单视频，而现代智能标注平台通过构建基于知识图谱的交互式学习系统，结合虚拟人讲解和智能问答功能，能够为标注人员提供个性化的学习路径。这种人工智能助学模式不仅缩短了培训周期，还能根据标注人员的认知特征动态调整教学内容，确保培训效果。例如，平台通过分析标注员的历史行为数据，为其推荐最适合的学习内容和练习任务，显著缩短新标注员的上手时间。

在完成标注任务阶段，预标注技术的广泛应用是这一环节的重要特征，通过训练有素的模型对数据进行初步标注，大幅减少了人工标注的工作量。例如，在图像标注任务中，目标检测算法可以预先标出物体的大致位置，标注员只需进行微调即可，效率提升可达 40%。多模态大模型的应用使得平台能够快速完成文本、音频、视频、3D 点云等不同模态数据的预标注工作，显著提升数据处理效率。

质量检查环节的智能化转型最为显著，传统的人工全检或抽样检查正被多层次的智能质检体系所取代。初级质检由算法自动完成，通过一致性检查、逻辑验证等规则对标注结果进行快速过滤。中级质检采用模型辅助方式，例如通过训练专门的质检模型预测标注错误概率，指导人工复检重点样本。高级质检则针对争

议样本引入专家仲裁机制，确保最终质量。个别情形下，还可采用“埋雷”机制，督促标注员保证标注质量。

从整体来看，智能标注企业的实践呈现出几个显著特征：首先是技术应用的全面性，从数据准备到质量检查的全流程都实现了智能化升级；其次是模式创新的多样性，不同企业根据自身技术积累和市场定位，发展出各具特色的智能标注解决方案；最后是生态建设的协同性，领先企业普遍注重与上下游合作伙伴共建数据标准和技术生态。这些实践表明，智能数据标注已经发展成为一个技术密集、创新活跃的领域，其发展水平直接影响着人工智能产业的整体进步。未来随着大模型技术的普及和具身智能等新领域的兴起，对高质量标注数据的需求将持续增长，智能标注企业需要进一步强化技术研发，深化行业应用，推动人机协同标注模式向更高水平发展，为人工智能产业提供更加强大的数据支撑。

4.2.2 智能标注企业分类分析

从产业实践来看，智能数据标注领域的企业发展路径呈现出多元化特征。基于代表性企业的技术路线与商业模式分析，可以观察到不同企业根据其技术储备和应用场景的差异化布局，形成了各具特色的发展模式。这种分类方式有助于理解企业在产业链中的定位与价值创造逻辑。

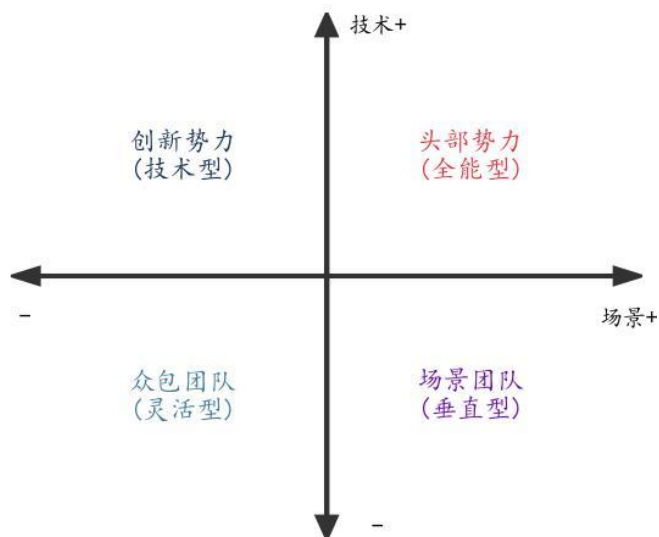


图 4-1 场景与技术分类分析

第一类公司是具备技术优势与丰富场景积累的头部势力。这些企业是数据标注产业的领军者，凭借强大的技术实力和丰富的场景落地经验，在市场中占据主导地位。它们不仅拥有先进的数据标注技术，能够高效地处理各种复杂的数据标

注任务，还具备丰富的行业应用场景，能够为不同领域的客户提供定制化的数据标注解决方案。例如，科技巨头通过自建数据标注平台，为自身及本行业其他客户项目提供高质量的数据支持。这些头部公司凭借技术与场景的双重优势，树立了行业标杆，吸引了大量企业用户，推动了整个数据标注产业的发展。

第二类公司是以独到技术为特征的创新势力。这类企业通常由技术专家创立，专注于某一特定领域的技术创新，如自动驾驶、AIGC 等。它们在技术上具有独特的优势，能够开发出高效、精准的数据标注工具和算法，为行业带来新的技术突破。然而，由于缺乏足够的场景落地经验，这些创业公司在市场推广和客户拓展方面可能面临一定的挑战。例如，专注于自动驾驶领域的创业公司，通过研发先进的视觉识别算法和数据标注技术，能够为自动驾驶企业提供高质量的标注数据，但在将技术应用于其他行业场景时，可能需要进一步的探索和积累。尽管如此，这些创新势力凭借其技术实力和创新能力，依然在市场中占据了一席之地，为行业的技术进步注入了新的活力。

第三类公司以中小众包团队为主，这类企业是数据标注产业的基础力量，通常以灵活的运营模式和低成本的优势在市场中寻找机会。中小众包团队通过组织大量的标注人员，以众包的方式完成数据标注任务，能够快速响应市场需求，提供大量的标注数据。这些企业在技术和场景方面可能相对薄弱，但在特定细分领域或项目中，能够凭借其灵活性和成本优势，满足客户的基本需求。例如，一些小型的众包平台，通过吸引大量的兼职标注人员，能够快速完成一些简单的标注任务，为初创企业提供低成本的数据标注服务。众包平台通过内部优化，提高数据标注效率，为自身的产品迭代提供支持。

第四类企业是场景团队，其专注于垂直行业场景的数据服务能力建设。这类企业通常在特定领域持续深耕，积累了丰富的行业数据集和场景化经验，对业务需求和应用痛点具有深刻理解，能够为客户提供高度适配的数据解决方案。不同于通用型数据服务平台，场景团队的优势在于对垂直行业的深度聚焦。以医疗影像标注为例，这类企业通过与医院、科研机构长期合作，建立了符合临床标准的医学影像数据库，为 AI 医疗企业提供精准标注数据，有效提升诊断模型的准确性和可靠性。此外，依托上海数据交易所和华东江苏大数据交易中心等国家级交易平台的基础设施服务，东部地区企业能够更高效地实现行业数据的合规积累。

通过将行业经验转化为标准化服务能力，场景团队在医疗、金融、工业等专业领域形成差异化优势，成为企业智能化转型中值得信赖的数据服务伙伴。

总体来看，数据标注产业的竞争格局正在发生变化，数据需求方越来越重视数据质量、场景多样性和可扩展性。随着大模型时代的到来，数据标注企业需要不断提升技术水平，优化标注流程，培养高素质的人才队伍，以适应市场的变化和 demand。同时，企业还需要加强与上下游企业的合作，形成良好的产业生态，共同推动数据标注产业的高质量发展。

5.智能数据标注典型企业实践

我国智能数据标注产业呈现鲜明的区域发展特色^①：东部地区依托技术、资本和人才优势，形成了覆盖全产业链的智能标注生态体系，在自动驾驶、大模型等高价值领域实现技术突破与商业化落地；中部地区立足产业需求，通过建设分布式标注基地和产教融合模式，深耕垂直场景的数据服务创新；西部地区发挥算力资源优势，在金融、医疗等专业领域开展智能化标注技术探索；东北地区则通过产学研协同和特色基地建设，推动行业标准制定与残健融合等社会价值实践。各区域差异化发展共同构建了多层次、专业化的智能数据标注产业体系。

5.1 东部地区的全面发展

东部地区作为我国数字经济发展的前沿阵地，在智能数据标注领域展现出全方位领先优势。依托雄厚的产业基础和技术创新能力，该区域已构建起覆盖数据采集、标注加工、质量评估、模型训练的全生命周期服务体系。在技术层面，通过大模型、深度学习等前沿技术的深度融合，实现了从传统人工标注向智能化标注的转型升级，支持 2D/3D/4D 多模态数据处理。服务能力方面，形成了“技术+资源+质量”的三重保障体系，通过规模化标注团队和严格的质量管控，确保高效可靠的服务交付。商业化应用成效显著，在自动驾驶、智能客服、内容审核等领域树立了多个标杆案例。

百度智能云

百度智能云数据标注服务作为百度人工智能战略的核心业务板块，经过十余年的深耕发展，已构建起行业领先的全栈式数据标注解决方案。依托百度的技术积累和生态优势，该服务已形成覆盖数据采集、标注加工、质量评估、模型训练的全生命周期服务体系，为各行业人工智能落地提供坚实数据支撑。

在技术能力建设方面，百度智能云打造了业界领先的智能标注平台，该平台深度融合了百度自研的大模型技术，通过飞桨深度学习框架的底层支持，实现了从传统人工标注向智能化标注的转型升级。平台支持 2D 图像标注、3D 点云标注、语音标注、文本标注、视频标注等全模态数据处理能力，特别是在自动驾驶

^① 根据国家统计局的划分方案，空间区域被划分为东部、中部、西部和东北四大区域，东部 10 省（市）包括北京、天津、河北、上海、江苏、浙江、福建、山东、广东和海南；中部 6 省包括山西、安徽、江西、河南、湖北和湖南；西部 12 省（区、市）包括内蒙古、广西、重庆、四川、贵州、云南、西藏、陕西、甘肃、青海、宁夏和新疆；东北 3 省包括辽宁、吉林和黑龙江。

场景实现了 100% 的覆盖。其创新的 4D 智能标注系统基于 VIS 大模型底座，采用 Transformer 网络结构，通过千万级多源数据训练，使算法整体准召率达到 85% 以上，标注效率提升最高达 60%。

百度智能云在数据标注领域的技术突破主要体现在三个维度：首先是预标注算法的创新，针对检测、分割、车道线等场景研发了基于大模型的预标注技术；其次是专项优化能力，通过伪标签、数据增强等手段实现不同场景和设备的专项训练；最后是后处理优化，引入地面贴合和长时序全局优化策略，使高质量标注结果占比从 5% 提升至 30% 以上。这些技术创新使百度在自动驾驶数据标注领域保持绝对领先优势。

在服务能力方面，百度智能云建立了完善的标注资源体系。目前拥有业内规模最大的标注团队，包括 2000 万注册众包人员和 500 家签约代理商，形成了覆盖全国的服务网络。同时，百度构建了严格的质量管控体系，通过多轮质检和专家复核确保数据质量。这种“技术+资源+质量”的三重保障，使百度能够为客户提供高效可靠的标注服务。

商业化落地成果显著是百度智能云数据标注服务的突出特点。该服务已广泛应用于自动驾驶、智能客服、内容审核等多个领域，服务客户包括高通、比亚迪、滴滴等头部企业。在自动驾驶方向，百度为多家车企提供高精度的环境感知数据标注；在大模型领域，通过高质量的文本标注助力客户模型训练。这些实践不仅验证了技术的实用性，也为行业树立了标杆案例。

面向未来发展，百度智能云正在推进标注基地的全国布局，已在山西、山东、海南等地建设十余个基地，其中乌镇基地聚焦智能网联汽车赛道，累计孵化企业 11 家。同时，百度持续加强技术研发投入，重点突破自动化标注、合成数据生成等前沿技术。通过技术创新和生态建设的双轮驱动，百度智能云将持续巩固在人工智能数据服务领域的领导地位。

恒达地理

邯郸市恒达地理信息工程有限责任公司于 2005 年注册成立，是河北省高新技术企业，在地理信息、规划设计、土地服务及房地产开发等多领域展现卓越实力的综合性企业，凭借丰富的实践经验、强大的技术能力和全面的资质认证，在行业内树立了良好口碑。

恒达地理业务广泛且深入，核心业务覆盖地理信息系统工程、测绘、规划、自动驾驶数据以及土地相关服务等关键领域。恒达地理于 2023 年下半年开始重点布局数据要素业务，以高精地图，自动驾驶以及 AIGC 等标注为发力点，逐步聚焦政府、企业需求，打造数据处理、分析与治理一站式方案。依托数据工场运营、AI 训推一体建设、数据资产管理三大能力，以“技术+服务”双轮驱动，深挖数据价值，为各方发展注入澎湃动能，公司已与众多标注第三方平台以及汽车主机厂建立了持久而稳固的合作关系。

现阶段标注团队人数 300 余人，专注于自动驾驶数据以及大模型领域的标注。团队成员涵盖了数据采集、标注、分析、算法开发等多个专业领域，具备全方位解决自动驾驶数据难题的能力。凭借丰富的实践经验和精湛的技术水平，能够高效地处理海量的数据，并从中提取有价值的信息，为大模型和自动驾驶技术的发展提供坚实的数据支持。审核团队 50 人，涵盖平台广告及用户发布内容分类判断。此外恒达积极与众多院校开展校企合作。通过与高校和创业基地，产业园的紧密合作加强数据标注交付，比如，开悟智数众创空间、云创星河众创空间、邯郸市众鼎创业孵化基地、邯郸数字经济产业园、县域十余家数字经济产业园等。

恒达地理拥有多种数据标注业务优势。第一，交付体系完备。拥有完善的数据交付体系，借助自研管理平台，能提升交付效率与能力，确保大规模数据安全交付。第二，数据质量高：基于对多行业的深度理解，产出高质量数据，满足智能驾驶、智慧医疗等多行业复杂需求。第三，供应链灵活，可快速搭建供应链体系，持续扩充产能，具备较强的业务拓展与应变能力，能及时响应市场变化和客户需求。第四，专业领域资源储备丰富，与邯郸当地各大院校合作，拥有稳定的职业基地垂类人员大模型训练人员涉及 AIGC 等 5+ 领域。

数族科技

数族科技成立于 2016 年 5 月，专注于企业经营数据全生命周期管理，通过自主研发的核心技术体系赋能机构构建数据驱动决策能力。作为全链路数据要素服务商，公司聚焦数据资产构建与运营，推动数据资源市场化流通与价值转化。

在智能数据标注领域，数族科技基于“AI 驱动产业数据高质量标注”的研究，构建了一套创新性的 AI 驱动产业数据标注理论方法与技术体系。公司自主研发的产业数据标注智算引擎，通过多模态自动化标注、基于大模型的高效标注

以及低算力、高效率平台架构，显著提升了数据标注的智能化水平，将标注成本降低约 93%-95%，标注效率提升 1200%-1500%。

数族科技的智能标注技术具有多项创新特点：一是多模态语义解析技术，融合 NLP 与知识图谱，支持对文本、表格等多形式数据的语义解析与特征提取，构建 3000 余项细粒度标签；二是动态标签映射机制，构建跨领域标签体系，支持标签间语义关联，通过动态适配实现多场景数据兼容；三是自适应标注优化算法，采用自适应学习与迁移学习模型，动态优化标注过程中的语义理解能力；四是隐私保护与分布式架构，引入差分隐私与分布式计算，保障敏感数据的安全隔离与合规处理。同时，公司已成功构建了一个完整的数据经济生态体系，将分散的数据资源转化为可流通、可交易的市场化资产。通过创新的数据采集和治理技术，不仅推动了数据资产化的进程，更为产业与城市提供了一站式的数据流通与交易解决方案。数族科技的数据标注物加工中台支持本地大数据基座的私有化部署，以及通过 API 接口实现数据输出，推动数据资产流通，为企业招商、产业链优化、担保贷款评估、贷后风险跟踪等提供数据驱动的决策支持。

数族科技的智能标注技术在多个领域实现了创新应用：在金融领域，通过对企业经营数据的智能标注，构建了包含 1481 项“经营数据指标”和 1072 项“应用场景指标”的评价模型算法体系，为金融机构提供精准的企业信用评估服务；在产业分析领域，通过对产业链数据的智能标注，实现了产业链关键节点的自动识别与分析，为政府部门提供产业链“强链补链”决策支持；在数据资产化领域，通过标准化的数据标注流程，将非结构化数据转化为可计量、可定价的数据资产，推动数据要素市场建设。

公司的智能标注系统采用“人机协同”模式，通过五层架构设计（数据接入层、数据处理层、AI 标注层、人机协同层和数据服务层）实现高效标注。系统支持分布式训练框架，可根据不同领域数据特点自动选择最优标注策略，同时通过数据血缘与版本管理确保标注数据的可追溯性。在实际应用中，系统已成功处理超过 10PB 的多源异构数据，标注准确率达到 95%以上，为多个行业客户提供了高质量的数据标注服务。

作为江苏省高新技术企业和专精特新企业，数族科技已获得多项国家“两重”项目备案，包括重大资源储备项目“企业经营数据高质量数据集”以及数据基础

设施项目“产业数据人工智能标注协同引擎”等。

金盾检测

金盾检测技术股份有限公司成立于 2010 年，是一家专注于网络安全领域的国家级高新技术企业，同时也是国家级专精特新“小巨人”企业。公司主营业务涵盖合规检测（包括等保测评、商用密码应用安全性评估、软件测试等）、数据安全服务（如数据分类分级、数据安全风险评估、数据安全管理体系建设等），以及围绕网络和数据安全运营开展的安全运维、安全防护能力检测、渗透测试、攻防演练、安全竞赛和培训等，致力于为广大客户提供信息系统全生命周期的网络安全服务。

公司的商业模式主要依托于为数据标注等相关产业提供合规检测和安全测试服务，通过技术测试输出符合法律法规及行业标准的检测结果认定，并以此作为核心盈利来源。在竞争优势方面，金盾检测拥有超过 200 人的技术团队，其中包括 10 余名高级工程师和 50 余名中级工程师，团队成员不仅具备相关职业资格证书，还积累了丰富的工作经验，并在国内多项网络和数据安全竞赛中取得优异成绩，获得行业嘉奖。

在技术能力与创新方面，金盾检测围绕数据标注提供相关合规和安全配套支撑技术，拥有一支专业技能深厚、行业经验丰富的复合型网络安全人才队伍，团队成员持有网络安全等级保护测评师、密码测评师、软件测评工程师、注册信息安全专业人员（CISP）等专业资质证书，持证上岗经验丰富。截至目前，公司已获得近百项软件著作权、8 项发明专利及 10 项实用新型专利，展现了强大的技术研发实力。在行业应用与典型案例方面，金盾检测的服务覆盖医疗卫生、能源电力、金融、交通、教育等多个领域，协助各行业客户开展数据安全漏洞挖掘、威胁处置及应急支撑保障工作，积累了丰富的实践经验。公司战略规划聚焦于深度参与智能数据标注行业的安全漏洞挖掘和数据安全技术平台研发，为行业企事业单位提供安全标准制定和技术人才培养支持，致力于打造智能数据标注行业网络和数据安全服务的领先品牌。

德鲸未来

德州市德鲸未来数据科技有限公司，成立于 2024 年 11 月，公司位于山东省鲁北大数据中心核心承载区，这一省级数字经济基础设施枢纽为公司提供了海量

数据存储和百 P 级算力支撑，构建了从数据采集到应用的全链条赋能体系。得益于国有企业的战略参股，德鲸未来获得了雄厚的资金支持，用于拓展数据服务场景、购置高端算力设备及引进顶尖技术人才。同时，公司打通了政务数据合规流通渠道，形成了“政、产、学、研”四位一体的创新生态。

公司核心业务聚焦人工智能数据要素全生命周期管理，形成“数据加工处理-数据资产交易-数据资产入表-AI 场景落地”的闭环服务体系，涵盖计算机视觉、自然语言处理、语音工程等多模态数据处理领域。技术团队汇聚行业资深专家，自主研发数据智能处理平台，集成预识别算法引擎、智能质检系统及多维度数据标注工具，可支持 2D、3D 融合标注、3D 点云精细化处理、4D 时空序列分析等复杂场景作业。标注团队通过“AI 工具赋能+人工精细校验”双轮驱动模式，实现智驾、文本、音频等多类型数据的精准处理：在智能驾驶方向，具备单帧/连续帧标注、3D 目标检测框标定、4D 轨迹跟踪标注等核心能力，精细完成障碍物分类（含动态/静态目标）、交通信号灯状态识别、车道线曲率拟合等关键要素标注，累计为自动驾驶客户处理超 1.2 亿帧高复杂度交通场景数据；在汽车产业生态，公司深度融入智能驾驶技术创新链条，与头部车企及自动驾驶解决方案商建立战略合作伙伴关系，提供高精度标注的动态交通流数据、复杂路况场景库及传感器融合数据集，有力支撑智能驾驶算法训练与功能迭代，助力 L2+级以上智能驾驶技术加速落地。

公司积极响应国家数字经济人才培养战略，与德州学院、德州职业技术学院建立深度产教融合体系，共建“AI 数据训练师产业学院”。搭建校企双导师库，目前已累计提供超 200 个实习就业岗位，形成“人才培养-技能提升-就业服务”全链条闭环，成为区域数字经济人才供给的重要枢纽。

澳鹏中国

澳鹏（Appen）是全球人工智能数据服务领域的标杆企业，自 1996 年成立以来持续引领行业创新，2019 年澳鹏中国和田小鹏博士带领下快速崛起，2024 年实现 4.2 亿人民币营收，71% 的年增长率彰显强劲发展势头，其中大模型/AIGC 业务更以 526% 的增速成为核心增长引擎。公司业务覆盖自动驾驶、医疗金融、法律代码等九大垂直领域，通过数据标注、采集、成品数据集及增值服务的全链条解决方案，服务全球顶尖人工智能企业。在智能驾驶领域，澳鹏的 3D 动态障

碍物检测系统帮助头部车企提升 30%识别效率，其语义分割技术处理超 500 万点云数据时效率提升 40-50%。面向无人机企业开发的视觉影像识别系统，通过人工智能自动化标注实现 10-20%的效能突破。大模型服务方面，公司交付的 50 亿图文对数据达到 98%准确率，医疗垂类累计为 30 余家客户提供千万级专业医学数据，涵盖临床诊断、药学等细分场景。

技术创新构成了澳鹏的核心竞争力。公司自主研发的端到端预标注大模型开创行业先河，配合动态加载算法可高效处理 10 亿级点云数据，25%的效率提升树立行业标杆。其 2D/3D 融合检测技术显著增强目标追踪鲁棒性，而多模态工具链能在 5 分钟内适配 99%的标注需求。思维链标注工具和全流程人机协同系统，持续推动项目交付效率提升 20%以上。

依托全球 9 大交付基地和百万级多语言众包资源，澳鹏中国具备周产 20 万+数据的规模化能力。300 人医学团队、400 名金融持证专家构建的专业知识储备，配合 ISO 37301 认证的安全体系，形成“技术+场景+合规”三位一体优势。面向未来，公司计划重点开发医疗与具身智能专用平台，目标将自动化率提升至 90%，并通过生态合作推动中国大模型在通用与垂直领域的双重突破。

本原智数

上海本原智数科技有限公司是一家专注于智能数据标注和全链路多模态数据服务的高新技术企业，由来自上海交通大学、清华大学等知名院校的连续创业者联合创办。凭借雄厚的技术研发实力和丰富的产业实践积累，公司已迅速成长为业内领先的人工智能数据服务提供商，与多家政府机构及主流互联网企业建立了深度合作关系，为国内外市场提供端到端的智能数据解决方案。

公司聚焦自动驾驶、医疗健康、金融科技等垂直领域，提供涵盖图像、视频、语音、文本等多模态数据的精准标注服务，业务范围包括数据清洗、人工校验及人工智能辅助预标注等全流程环节。通过融合合成数据生成、自动化清洗和实时质检等增值技术，依托自主研发的数据管理平台与多任务协同系统，本原智数能够为客户量身定制数据闭环处理方案，高效构建大模型训练所需的高质量数据集。在医学影像标注领域，公司累计完成超过 500 万人次的标注任务，拥有 800 余名全职影像医生和约 5000 名兼职医生参与，标注准确率高达 99%以上；在自动驾驶数据标注方面，每月交付数据量达千万级别，由 500-600 名全职人员支持，

准确率超过 98%；此外，在教育垂类模型建设中，公司每月产出 500 余项标注任务，由博士级专家团队参与交付，整体质量水平稳定在 90%以上。

技术创新是本原智数的核心驱动力。公司通过建设自动化标注工具与人工智能辅助标注系统，搭建人工智能中台及自动化工具链，实现标注效率的翻倍提升。其开发的跨模态、多任务数据管理平台支持多人协作与实时反馈，形成数据标注的闭环管理能力。同时，公司搭建了合成数据生成与强化学习数据引擎，能够模拟复杂交互场景，显著降低数据生产成本并提升模型训练效果。在具身智能领域，本原智数构建了数据处理一体化平台，与中国信通院和人工智能行业协会合作推出本土化行业评测标准，其自研平台具备多模态模型性能评估功能，涵盖准确率、鲁棒性和偏差检测等关键指标。

规模化运营能力是本原智数的另一大优势。公司在全国拥有多个交付中心，线下标注团队规模达数千人，线上众包资源超过 400 万，专家网络覆盖 10 万人。凭借自研的自动化预标注工具、合成数据引擎和人工智能质检系统，公司积累了丰富的数据资产与模型能力，形成了坚实的技术壁垒。深耕自动驾驶、医疗、金融等垂直行业的经验，使其具备深刻的行业认知与成熟的解决方案能力，推动公司以每年超 100%的增速跻身国内人工智能数据标注领域一线阵营。

面向未来，本原智数计划通过持续完善人工智能中台和跨模态数据管理平台，构建规模与效率的增长飞轮。公司将进一步拓展具身智能和行业垂类模型应用，探索医疗、金融等新兴领域，并积极布局国际市场。同时，本原智数将深化与政府机构、科研院所及大型人工智能企业的生态合作，推动行业标准建立，加速具身智能与人工智能产业的整体发展。在商业模式上，公司正从人力密集型标注服务向自动化工具驱动、模型主导的创新模式转型，以提升市场竞争力。

宏景智驾

上海宏景智驾信息科技有限公司成立于 2018 年，致力于服务交通产业未来发展，是一家全栈式自动驾驶解决方案服务商，具备完全自主研发的车规级自动驾驶计算平台、全栈的软件算法和完整的系统集成能力，可针对不同客户需求提供定制化的高性能智能驾驶解决方案，全周期赋能 L1-L4 级别智能驾驶。

2023 年宏景战略布局智驾 AI 领域，着手构建了专业的 AI 相关数据团队，自研 4D BEV 感知真值系统 Hyper-GTMax，搭载 NVIDIA DRIVE Orin 系统级芯

片（SoC），结合云端加速计算与 AI 解决方案，构建了完整的智能驾驶感知闭环，解决了自动驾驶感知训练中 4D BEV 真值数据构建成本高、效率低的痛点，助力车企快速构建高阶感知能力，提升开发效率并降低量产成本；宏景自主研发的智能驾驶数据平台“景云”，专注于自动驾驶全流程数据闭环管理，景云平台构建了从“数据采集-存储-标注-训练-仿真-OTA”的完整闭环，支持量产路测数据与云端数据管理平台的协同，确保数据的高效利用与算法的持续优化。支撑 L2-L4 级自动驾驶技术的量产与迭代，加速了自动驾驶技术的商业化进程。

自动化标注是 AI 与智能驾驶行业从“数据驱动”到“价值创造”的桥梁。随着数据量的不断积累，宏景已实现了对目标检测、车道线等数据的自动化标注，大大降低了标注成本，提高了标注效率，缩短了算法迭代周期。在技术成果方面，宏景科技取得了令人瞩目的进展。2023 年公司成功实现了泊车感知模型的量产，这一成果标志着宏景在自动驾驶感知技术上迈出了重要一步，为智能泊车系统的商业化应用奠定了基础。随后，在 2024 年宏景再次宣布行车感知模型量产，这一持续的技术迭代与优化，不仅体现了公司在自动驾驶技术上的深厚积累，也彰显了其快速响应市场需求、持续推动技术创新的能力。

5.2 中部地区的场景挖掘

中部地区立足区域产业特色，深耕垂直场景应用，形成了差异化发展路径。该区域围绕智能驾驶、医疗健康、工业质检等具体场景需求，构建了专业化数据服务体系。通过产教融合模式培养场景适配人才，打造了覆盖多行业的标注服务矩阵。区域内企业注重场景化解决方案开发，与产业上下游深度协同，形成了“场景需求-数据服务-产业赋能”的良性循环。这种基于实际应用场景的发展模式，有效推动了区域特色产业的数字化转型。

飞数信息

安徽飞数信息科技有限公司是安徽言知科技有限公司旗下专注于人工智能数据服务的企业，自重组创立以来，凭借创新的技术能力和对行业的深刻理解，迅速成长为高质量数据生产领域的代表性企业。公司致力于构建智能数据生态体系，为人工智能产业发展提供专业化的数据支撑。

作为一家技术驱动型企业，安徽飞数拥有超过 2000 名专业标注人员，服务于全国 9 大标注基地，线上标注平台注册众包人员超 50 万人，分布在全国 300

多个城市及海外多个国家。其团队汇聚了多领域的行业专家，包括语言、科研、医学、金融等方向，能够满足不同场景下的复杂数据标注需求。公司已与科大讯飞、小米、蔚来、百度等知名互联网企业建立合作关系，服务中石油、中石化等大型央、国企。同时，还与多个政府保持合作与联系。此外，还与清华、中科大、哈工大等一流高校开展合作。

在业务布局上，安徽飞数构建了辐射全国的数据标注服务矩阵，具备多语言、多模态的数据生产能力。这一战略布局不仅提升了服务效率，也促进了区域数字经济的发展，为地方产业升级和人才就业提供了有力支持。公司由行业资深专家谭昶博士领衔，其在大数据、智慧城市及人工智能领域拥有丰富经验，曾主持多项国家级科研项目，并参与行业标准制定，为公司的技术研发和业务拓展提供了专业指导。

安徽飞数的主营业务涵盖数据标注全生命周期服务，包括数据集规划、采集治理、标注加工及流通应用等，同时提供合成数据、数据治理等增值服务。其核心产品“飞智标智能数据标注平台”通过人工智能助学、数据合成、人机协同等技术方案，有效解决了传统数据标注中的效率和质量问题。飞智标平台整合了智能化学习工具，帮助标注人员快速掌握专业知识，并与高校合作培养跨学科人才。在数据合成方面，公司采用先进的算法模型，生成高质量的训练数据，为人工智能模型的优化提供支持。此外，平台还集成了多种自动化标注工具，结合智能质检机制，确保数据生产的高效性和可靠性。

多鲸科技

河南多鲸信息技术有限公司成立于 2022 年 5 月，是一家专注于人工智能数据要素板块提供高质量数据服务、AI 应用和行业人才培养的产教融合型科技企业。创始团队由金融、科技与产业复合型人才组成，具备丰富的行业经验和技术创新能力。经过 3 年实践与探索，自营全国 9 大数据标注基地，700 余人全职交付团队覆盖长三角、京津冀及中部地区，交付链条覆盖数据采集、存储、清洗、标准化、标注、质检、训练、数据价值流通、人才供给等全链条，交付领域涵盖计算机视觉、大模型多模态以及行业垂类。

通过“工具赋能高质量数据服务+AI 场景应用+产教融合生态”的盈利模式，多鲸科技为自动驾驶、智慧医疗、智慧能源、智慧教育、智慧零售等领域提供全

链路算力+数据+场景一体化综合解决方案。累计交付超 1.5 亿帧国内自动驾驶头部车企 2D、3D 融合、3D 点云、4D 交通场景数据；超 150TB 耳鼻喉、心肺、消化、泌尿、普外、重症等医疗数据加工整理；服务国内 10 余家通用大模型公司、算法平台公司预训练、调优、拉齐；协助垂类模型公司完成了 AI 教学、智慧铺排货、智慧巡检等相关应用，并前瞻性积极布局具身智能数据训练。自研数据标注平台和多基地多项目大规模人力数据服务调度运营平台，搭建有企业项目管理和运营管理私有知识库，两平台两模型加持，效率超同业 20%。

同时，多鲸科技与 30 余所职业院校合作，进行校内数据基地建设及数字人才产业学院共建，推广“21 天集训+真实项目实训+实习就业保障”产教融合型数据标注人才培养模式，年培养 AI 训练师超 2000 人，3-5 人可快速复制拉动 50 人基地起盘，成为国家数据局数据标注人才培养优秀典型。

空港数据

河南空港数据科技有限公司成立于 2025 年，总部位于郑州航空港经济综合实验区，是一家以“数据标注+场景应用”为核心的新型人工智能服务企业。公司依托中部地区规模最大的智能算力中心和一级国资平台资源优势，整合区域算力中心资源，聚焦“高质量数据集生产”和“行业 AI 智能体开发”双核心业务，构建覆盖数据标注、模型训练、场景落地的全链条服务体系。为行业提供从数据生产到 AI 模型落地的轻量化解决方案。创始团队由数据服务行业资深从业者与算法工程师组成，已建成 100+人专业标注团队，初步形成覆盖 2D/3D/4D 数据标注、语音文本处理的基础能力。

公司以“算力协作、算法适配、场景深耕”为发展路径，与第三方算力平台建立战略合作，通过弹性调用 GPU 资源满足自动驾驶、工业质检等场景的标注需求。目前已构建三大核心能力：一是多类型数据处理体系，支持图像分割、语义标注、点云标注等主流标注类型；二是质检标准化流程，建立双层人工校验机制，标注准确率稳定在 98%以上；三是场景化数据服务，医疗影像领域完成 10 余个定制化数据集交付。

公司将分阶段推进能力升级：2025 年重点完善基础设施，扩建 500 人标注团队；2026 年深化场景落地，在医疗、金融、农业领域形成 3-5 个标准化数据集产品；累计赋能行业智能体 60 个，带动就业规模突破 2000 人；至 2027 年探索

数据要素流通新模式，联合行业伙伴共建区域数据服务生态，实现 AI 智能体应用场景超 100 个，累计解决就业 3000 人，形成“数据生产-智能应用-人才培养”的良性循环，全面支撑航空港区传统产业智能化转型。

公司将分阶段推进能力升级。初期，重点完善基础设施，扩建标注团队。随后，深化场景落地，在医疗、金融、农业等关键领域形成多个标准化数据集产品；累计赋能行业智能体若干，带动就业规模显著增长。最终，探索数据要素流通新模式，联合行业伙伴共建区域数据服务生态，实现 AI 智能体应用场景大幅拓展，全面支撑传统产业智能化转型。

5.3 西部地区的技术探索

西部地区依托本地算力资源与产业基础，在智能数据标注领域展现出独特的技术创新路径。以数字天蚂为代表的企业充分发挥后发优势，重点突破金融、医疗等专业领域的标注技术难点。该区域通过构建智能化标注工具与人机协同体系，显著提升了复杂场景下的数据标注效率与质量。同时，西部地区企业注重数据安全合规体系建设，积极参与行业标准制定。

数字天蚂

数字天蚂（重庆）信息技术有限公司作为蚂蚁科技集团旗下专注于人工智能数据服务的新锐企业，凭借集团强大的技术基因和资源优势，在短时间内快速构建起覆盖多行业、多场景的智能数据标注服务体系。公司立足于解决人工智能产业化过程中的核心数据需求，通过技术创新和流程优化，为各行业客户提供专业、高效、安全的数据标注解决方案。

在技术能力建设方面，数字天蚂打造了行业领先的大模型驱动标注操作系统，该系统深度融合了蚂蚁集团在人工智能领域的技术积累。平台支持图像标注、文本标注、语音标注、视频标注以及 3D/4D 动态场景分析等全模态数据处理能力，特别是在金融、医疗等专业领域展现出独特优势。其创新的人工智能辅助标注引擎通过知识库检索、关键信息识别等智能化功能，显著减少人工标注工作量，大幅提升了生产效率。

数字天蚂在数据标注领域的技术创新主要体现在三个维度：首先是智能化标注工具的研发，通过搭载大模型能力实现自动关联和辅助标注；其次是智能质检系统的构建，采用规则引擎与大模型双校验机制，确保数据质量与一致性；最后

是人机协同体系的完善，基于智能算法实现任务动态分配和进度监控。这些技术创新使数字天蚁在复杂场景数据标注中保持竞争优势。

在垂直行业应用方面，数字天蚁展现出深厚的专业积累。在自动驾驶领域，公司为某头部技术企业提供的标注服务，通过智能化培训 Agent 和质检插件，使产能爬坡时间缩短 40%，数据一致性显著改善；在金融大模型场景，创新的长思维链标注工具帮助某全球科技公司提升复杂金融问题的解决能力；在电商领域，图像预标注技术使某知名平台的商品识别准确率提升 13%。这些成功案例充分验证了数字天蚁的技术实用性和行业理解深度。

数据安全与合规是数字天蚁的核心竞争力之一。公司建立了完善的数据安全管理体系，严格按照《数据安全法》《个人信息保护法》等法律法规要求，实施数据加密、访问控制、操作审计等多重安全措施。同时，数字天蚁积极参与行业标准制定，推动数据标注服务的规范化发展，为客户提供安全可靠的数据保障。

5.4 东北地区的协同实践

东北地区在智能数据标注领域形成了特色鲜明的协同发展模式。以京数云为代表的企业通过建设残健融合的 AI 数据服务基地，创新性地实现了社会价值与产业发展的有机结合。该区域注重产学研协同创新，积极参与国家标准制定，推动行业规范化发展。同时，依托自主研发的智能标注平台，东北地区在内容审核、自动驾驶等专业领域形成了差异化服务能力，为区域数字经济发展注入新动能。

京数云

辽宁京数云大数据科技有限公司于 2021 年 11 月成立，公司专注于内容安全审核、AI 基础数据服务、数据采集、数据标注的科技企业，公司为万科、大唐移动、字节跳动、百度、科大讯飞、阿里、中软国际等提供服务。公司联合河南科技大学出版的《数据标注实用教程》已纳入 10 余所高校的数据标注专业教材。公司参与了《信息技术 众包服务平台 功能通用要求》国家标准制定、参与《信息技术 数据流程服务技术规范》团体标准制定。

京数云大数据建设了全国首家残健融合模式的 AI 数据服务基地，并于 2023 年 11 月被中国残疾人联合会评为首批国家级残疾人大学生实习基地。公司凭借前沿技术与专业团队，为众多客户提供精准高效的数据服务，交付能力备受认可，口碑卓越。在产学研合作方面，与清华大学、北京邮电大学建立了紧密的合作关

系，不断探索前沿技术，推动行业创新发展。在内容生态拓展上，与咪咕视频携手，为优质内容传播提供有力支持。主要业务如下：内容安全审核、大模型标注、2D/3D/4D 自动驾驶标注、医疗标注、图片标注等。

京数云大数据自研的数据标注实训平台实现系统培训、分布式管理、人机协同操作可以大大提升员工的工作效率以及准确率。自研的内容审核人机交互智能审核工具在项目上可以提升员工 30-40%的审核效率，20%-30%的审核准确率。京数云大数据致力于通过提供高质量 AI 基础数据服务实现数据智能。

6.研究展望与总结

本研究旨在系统梳理智能标注产业的发展现状与未来趋势，通过分析企业用工需求、企业实践等关键维度，为产业把握智能化转型路径提供理论支撑。总结研究成果不仅有助于厘清智能标注产业在人工智能产业链中的战略定位，更能为政策制定者、企业决策者提供前瞻性的发展建议，推动产业健康可持续发展。

6.1 未来展望

展望智能标注产业的未来发展，需从技术发展、企业竞争和人力资源三个维度展开分析。技术发展方向将决定产业升级的核心动能，企业竞争策略反映市场格局的演变逻辑，而人力资源管理则关乎产业转型的可持续性。

6.1.1 技术发展方向

为了更好地适应市场需求，提升数据标注的效率和质量，技术方向的探索与创新显得尤为重要。未来，应重点推动人工智能标注技术的研发，同时提升人工标注的质量，以实现智能数据标注产业的高质量发展。

首先，人工智能标注技术的研发是提升标注效率和质量的关键。当前，人工智能辅助标注技术已经在多个领域得到应用，但仍有较大的提升空间。未来，应进一步优化预训练模型和机器学习算法，使其能够更准确地进行初步标注。例如，引入更先进的深度学习模型，可以显著提高标注的准确性和效率。同时，应加强对标注数据的预处理和后处理技术的研究，确保标注结果的可靠性和一致性。此外，持续优化算法，使其能够更好地适应不同领域的标注需求，提升标注质量。

其次，提升人工标注质量是确保数据标注产业可持续发展的基础。尽管人工智能技术在不断进步，但在处理复杂任务时，人工标注仍然不可或缺。未来，应加强对标注员的培训，提升其对标注规范的理解和执行能力。例如，通过定期的培训课程和实践操作，帮助标注员掌握最新的标注技术和工具，提高标注的准确性和效率。同时，应建立完善的质量控制机制，通过多轮质检和反馈机制，确保标注结果的高质量。此外，企业应利用数据分析和反馈机制，为标注员提供实时指导和改进建议，帮助其不断提升技能水平。

最后，推动人工智能标注技术与人工标注的深度融合是未来发展的必然趋势。通过人机协同模式，标注员可以借助人工智能工具快速完成标注任务，而人工智能模型则可以通过标注员的反馈不断优化自身性能。例如，标注员可以在人

工智能预标注的基础上进行审核和修正，同时人工智能模型可以根据标注员的修改记录进行再训练，形成一个良性循环。这种人机协同模式不仅提高了标注效率，还通过算法的客观性和一致性，显著提升了标注质量。

6.1.2 企业竞争策略

为了在激烈的市场竞争中脱颖而出，企业需要制定具有前瞻性的策略，以适应行业的发展趋势。未来，企业应重点提供高质量的定制化服务，并加大对自动化标注技术的投入，以提升核心竞争力和市场地位。

首先，提供高质量的定制化服务是企业赢得市场的重要策略。随着人工智能应用的多样化和专业化，不同行业对数据标注的需求也日益多样化。企业应深入理解各行业的具体需求，开发针对性的标注工具和流程，以满足自动驾驶、医疗影像、自然语言处理等特定领域的高标准要求。

其次，加大对自动化标注技术的投入是提升企业效率和降低成本的关键。自动化标注技术不仅可以显著减少人工标注的工作量，还能提高标注的一致性和准确性。企业应积极研发和应用先进的预训练模型和机器学习算法，开发智能标注平台，实现从数据采集、预处理、标注到质检的全流程自动化。同时，企业应持续优化算法，使其能够更好地适应不同领域的标注需求，进一步提升标注质量。

此外，企业还应加强与高校、科研机构的合作，开展前沿技术研究，推动行业技术进步。通过合作，企业可以获取最新的研究成果，将其转化为实际应用，提升企业的技术创新能力。同时，企业应积极参与行业标准的制定，推动数据标注产业的规范化和标准化，提升整个行业的技术水平。

最后，企业应建立完善的质量控制体系，确保标注数据质量。通过多轮质检和反馈机制，企业可以及时发现并纠正标注中的错误，确保标注结果的准确性和一致性。同时，企业应利用数据分析和反馈机制，为标注员提供实时指导和改进建议，帮助其不断提升技能水平。

6.1.3 人力资源管理

在数据标注产业智能化转型进程中，人力资源管理创新正成为推动行业健康发展的核心驱动力。面对当前存在的劳动权益保障不足、标注员成长受限等挑战，需要构建一套兼顾效率与公平、技术与人文的现代化管理体系，实现从业者价值与产业发展的良性互动。

从职业发展维度来看，应当着力打造专业化的人才培养生态。通过建立“初级标注员-高级标注师-质量督导”的阶梯式成长通道，将标注质量、任务复杂度与职业发展深度绑定。配套开发模块化培训体系，重点培养人工智能工具操作、垂直领域知识和伦理规范等核心能力，同时推行“核心认证自由职业者”的混合用工模式，借助平台信用机制实现灵活就业与质量保障的动态平衡。

薪酬激励体系需要突破传统按件计酬的局限，转向更科学的价值评估。建议在保障基础生活需求的前提下，建立包含基础薪酬、质量奖金和长期激励的多层次回报机制。其中质量评估需引入准确率、任务难度系数等量化指标，对高质量产出实施阶梯式奖励，同时对稳定合作的核心人才提供社保补贴和技能提升津贴，构建可持续的福利保障网络。

在数据伦理治理方面，人力资源管理应当发挥更积极的作用。通过将人工智能伦理准则和偏见识别方法纳入必修课程体系，采用情景模拟和案例教学强化实践能力培养。配套实施“标注员自检-小组互审-人工智能复核”的全流程质量管控，在医疗、金融等关键领域引入专家评审机制。同时借助“最佳实践案例库”和“伦理标兵评选”等文化建设活动，持续培育客观公正的职业文化氛围。

智能化工具的应用为人力资源管理提供了新的可能性。搭建数字绩效看板可以实时追踪标注员的质量效率指标，并提供个性化改进建议；智能排班系统基于任务特征和员工能力实现最优人力配置；区块链技术的引入则为建立可信的职业发展档案提供了技术保障，有效提升行业人才的流动性。

6.2 研究总结

本研究全面探讨了与智能标注相关的产业发展现状、技术趋势、企业实践以及未来展望，旨在深入分析智能标注在人工智能发展中的关键作用及其面临的挑战与机遇。

从产业发展历程来看，数据标注经历了从早期人力标注到平台化工具标注，最终迈向知识密集阶段的演进过程。在这一过程中，技术驱动成为核心动力，推动标注需求从简单任务向复杂场景转变，标注质量逐渐成为行业竞争的关键指标。2024年，随着《关于促进数据标注产业高质量发展的实施意见》等政策文件的出台，以及多模态大模型技术的突破，产业规模快速扩张，用工需求显著增长，数据标注发展动态与人工智能技术周期高度关联。从区域分布来看，东部地

区如北京、广东、上海等地凭借技术集聚效应，在数据标注产业中占据主导地位。与此同时，中西部地区如河南、四川等地通过政策扶持和资源整合，实现了较快的增长。行业需求方面，软件和信息技术、科学研究是数据标注的主要需求方，数据标注成为人工智能技术发展的必要支撑。

智能标注作为人工智能发展的核心支撑环节，其重要性随着人工智能技术的广泛应用而日益凸显。通过对数据进行精准标注，智能标注为机器学习模型提供了高质量的训练数据，从而推动了人工智能在自动驾驶、医疗影像、自然语言处理等多个领域的快速发展。

在技术层面，智能标注技术的演进趋势表明，人工智能辅助标注技术的研发和应用是提升标注效率和质量的关键。通过引入先进的预训练模型和机器学习算法，智能标注工具能够显著减少人工标注的工作量，同时提高标注的一致性和准确性。人机协同模式的广泛应用，不仅提高了标注效率，还通过算法的客观性和一致性，显著提升了标注质量。此外，合成数据和领域自适应算法等前沿技术的发展，为智能标注行业带来了新的机遇，进一步推动数据标注技术的创新和应用。

在企业实践方面，智能标注企业通过提供高质量的定制化行业服务和加大对自动化标注技术的投入，提升了核心竞争力和市场地位。企业通过深入理解各行业的具体需求，开发针对性的标注工具和流程，满足了自动驾驶、医疗影像、自然语言处理等特定领域的高标准要求。同时，企业通过建立完善的质量控制体系，确保标注数据的高质量，为人工智能模型的训练提供了可靠的数据支持。

智能标注产业在人工智能的发展中扮演了不可或缺的角色。通过技术创新、企业实践、人力资源管理保障和政策支持，智能标注不仅能够提升数据标注的效率和质量，还能在实现数据标注员个人发展的基础上，推动人工智能技术的广泛应用和健康发展。随着技术的不断进步和政策的迭代完善，智能标注产业将迎来更加广阔的发展空间，为人工智能产业的持续发展提供坚实的数据及技术支持。

后 记

在《智能数据标注产业发展观察报告》的编撰过程中，相关单位与企业协助提供产业动态、典型案例及专业见解，为本报告的实证研究提供了关键支撑，使报告内容既保持了学术研究的严谨性，又具备了产业实践的指导价值。在此，课题组对上海数据交易所、华东江苏大数据交易中心两家支持单位，以及百度智能云、恒达地理信息工程有限责任公司、苏清（南京）数字产业有限公司、数族科技（南京）股份有限公司、金盾检测技术股份有限公司、德鲸未来数据科技有限公司、澳鹏（中国）、上海本原智数科技有限公司、上海宏景智驾信息科技有限公司、安徽飞数信息科技有限公司、河南多鲸信息技术有限公司、河南空港数据科技有限公司、数字天蚂（重庆）信息技术有限公司、辽宁京数云大数据科技有限公司的大力支持表示诚挚的感谢！

清华大学数字政府与治理研究院

清华大学社会科学学院数据治理研究中心

2025 年 4 月 20 日