

海天瑞声(688787)

计算机

发布时间：2025-05-12

证券研究报告/公司深度报告

买入

国内领军数据服务商，人工智能产业卖铲人

上次评级：买入

---海天瑞声公司深度报告

报告摘要：

数据服务商直接受益于人工智能产业发展，人工标注短期仍不可替代。自动化标注技术在现阶段还存在一定的局限性，如对某些类型的数据的识别准确率不高、对复杂场景的处理能力有限等。对于一些需要专业知识判断的标注任务，如医学影像标注、法律文本标注等，人工标注能做出更准确、合理的标注。大模型领域的发展对数据服务提供商主要的推动力包括：（1）预训练阶段，需要数据服务商提供大量的原始数据；（2）强化学习阶段，大模型的对齐需要数据服务商提供调优、反馈等服务；（3）应用阶段，企业需要针对自身业务特点和需求训练专用的小模型，需要数据服务商提供专业化、全流程数据服务。

随着 AI 大模型训练的深化，数据服务商行业开始壁垒增长、附加值提高、集中度提高，行业头部企业将更受益于行业的发展。数据服务商在根据不同行业和企业要求进行改造，并提供需求数据的同时，还应该基于行业理解，与企业共同制定相关的标注规范和规则，从而让算法发挥出更高的识别效率。此外，还要深度参与到客户人工智能算法从预研立项、定制开发到部署投放的全过程中，凭借自身的数据服务能力，在不同阶段为客户提供自制数据集，定制化的数据采集、数据清洗和数据标注服务以及数据中台集成能力。我们认为，随着行业飞轮效应的逐步体现，数据服务商行业集中度将进一步提高，行业头部企业将更受益于行业的发展。

海天瑞声是 AI 数据服务的国内头部企业。公司深耕数据服务行业 20 年，从做智能语音领域的数据服务起家，公司主要业务是 AI 训练数据的研发设计、生产及销售业务。技术能力方面和场景资源方面，公司在基础研究、平台工具、训练数据生产方面均有技术积累，并公司持续在智能语音、计算机视觉、自然语言、以及智能驾驶为代表的垂直领域进行技术拓展；软实力方面，公司具有国资股东，拥有较强的项目管理能力和全流程的数据服务能力。这些优势有望使得公司在行业的竞争中进一步脱颖而出，进一步享受行业发展红利。

盈利预测：预计公司 2025-2027 年营业收入分别为 3.33/4.66/5.43 亿元，归母净利润分别为 0.27/0.46/0.61 亿元。公司所处赛道目前处于赛道正在大爆发早期阶段，应更关注公司收入及订单的增长而非利润，给予“买入”评级。

风险提示：地缘政治风险、海外业务拓展不及预期、市场竞争加剧、估值模型与盈利预测不及预期

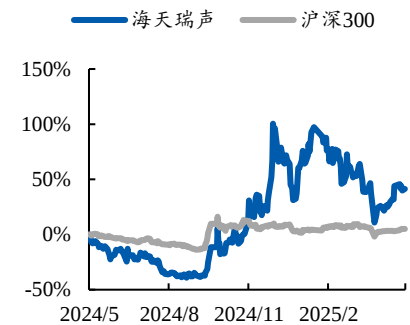
财务摘要（百万元）	2023A	2024A	2025E	2026E	2027E
营业收入	170	237	333	466	543
(+/-)%	-35.33%	39.45%	40.33%	40.00%	16.68%
归属母公司净利润	-30	11	27	46	61
(+/-)%	-203.16%	—	140.90%	66.97%	34.17%
每股收益（元）	-0.50	0.19	0.45	0.76	1.01
市盈率	(144.20)	518.53	211.73	126.81	94.52
市净率	5.56	8.00	8.02	8.48	9.17
净资产收益率(%)	-3.77%	1.49%	3.79%	6.68%	9.71%
股息收益率(%)	0.00%	0.37%	0.87%	1.46%	1.96%
总股本(百万股)	60	60	60	60	60

股票数据

2025/05/12

6 个月目标价（元）	
收盘价（元）	95.85
12 个月股价区间（元）	41.30~136.08
总市值（百万元）	5,782.17
总股本（百万股）	60
A 股（百万股）	60
B 股/H 股（百万股）	0/0
日均成交量（百万股）	1

历史收益率曲线



涨跌幅(%)	1M	3M	12M
绝对收益	14%	-20%	50%
相对收益	11%	-19%	45%

相关报告

《海天瑞声（688787）：数据标注基地建设拉动国内市场新需求》

--20240603

《医疗是 AI 的重点落地场景，“AI+医疗”将重塑行业生态》

--20250512

《中国 AI 崛起：技术突破与应用落地》

--20250326

《华为中国合作伙伴大会+英伟达 GTC 大会，AI 行业再迎国内外催化》

--20250317

目录

1.	海天瑞声：国内数据服务头部企业，充分受益与行业发展.....	4
1.1.	深耕行业 20 年，从智能语音领域拓展至各 AI 核心领域.....	4
1.2.	外部环境的影响减弱，业务规模化回暖.....	5
1.3.	数据要素提供发展新动能.....	9
2.	AI 大模型带动数据服务需求爆发，个性复杂化业务加深行业壁垒.....	11
2.1.	数据服务商：AI 产业链里的“大炼化”赛道.....	11
2.1.1.	训练阶段，提供大量的原始数据.....	12
2.1.2.	强化学习阶段，提供模型的反馈及打分.....	13
2.1.3.	应用阶段，提供专业化、全流程的高质量数据训练.....	15
2.2.	壁垒加深，个性化需求带动数据服务商向复杂化、专业化方向发展.....	17
2.2.1.	在 AI 的辅助下，标注效率程度变高，难度变大.....	17
2.2.2.	数据标注类型复杂化，行业化、垂直化属性逐渐增强.....	20
2.3.	政策持续催化，行业发展进入快速发展期.....	22
3.	他山之石：美国的发展情况.....	25
3.1.	ScaleAI：从数据标注到数据分析.....	25
3.2.	InnoData：帮助企业构筑更好的大模型.....	28
4.	盈利预测与投资评级.....	30
5.	风险提示.....	31

图表目录

图 1：公司发展历程.....	4
图 2：公司股权结构.....	4
图 3：公司主要客户.....	5
图 4：2019-2025Q1 公司营业收入及同比增速.....	6
图 5：2019-2025Q1 公司归母净利润及同比增速.....	6
图 6：2021-2024 公司分地区毛利率.....	6
图 7：2019-2024 公司毛利率及净利率及合计费用率.....	6
图 8：2019-2024 各项费用率.....	7
图 9：2020-2024 研发费用及研发人员数量.....	7
图 10：研发成果.....	7
图 11：公司业务全场景布局.....	9
图 12：刘烈宏主持召开数据标注领域民营企业座谈会.....	10
图 13：人工智能基础数据产业链.....	11
图 14：AI 领域对数据服务商需求.....	11
图 15：人工智能算法模型开发过程中需要结构化数据做支撑.....	12
图 16：2005-2024 数据标注企业数量.....	13
图 17：Deepseek 训练数据需求.....	13
图 18：大模型强化学习过程.....	14
图 19：RHLF 过程（以 GPT 为例）.....	14
图 20：大模型强化学习过程.....	15
图 21：AI 应用的下游行业.....	15
图 22：大模型驱使下的全生命周期管理.....	16
图 23：百度智能云数据闭环方案.....	17
图 24：大模型将提高自动标注的效率.....	18
图 25：目前 AI 基本可以给所有类型的数据打标.....	19

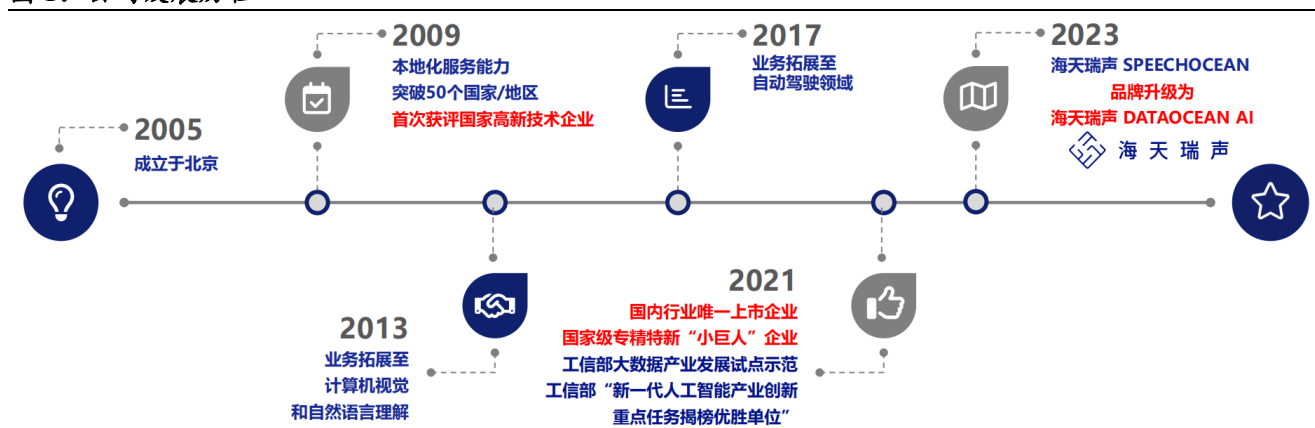
图 26: 自动标注对于效率的提升	19
图 27: 自动标注对于成本的节约	19
图 28: 传统业务类型: 边界标注	20
图 29: 专业化业务类型: 3D 标注	20
图 30: 数据标注行业需求变化	20
图 31: 医学影像标注示例	21
图 32: 文学评价标准示例	21
图 33: 数据服务飞轮效应	22
图 34: 国家数据局发布 7 个标注基地	24
图 35: Scale AI 业务情况	25
图 36: Scale AI 的 Nucleus 产品	26
图 37: Scale Donovan 产品	27
图 38: ScaleAI 年度经常性收入	27
图 39: Innodata 发展历程	28
图 40: Innodata AI 相关业务	29
图 41: Innodata 近期业绩情况	29
表 1: 公司核心技术	8
表 2: 全国各地提及数据标注的政策文件一览	23
表 3: 公司 2024-2026 年收入预测	30

1. 海天瑞声：国内数据服务头部企业，充分受益与行业发展

1.1. 深耕行业 20 年，从智能语音领域拓展至各 AI 核心领域

公司深耕数据服务行业 20 年，从做智能语音领域的的数据服务起家，目前已发展成为 AI 产业链数据服务领域国内头部企业。公司成立于 2005 年，成立之初从事语音识别、语音合成等智能语音领域的的数据服务，并不断拓展服务地区与业务领域。公司业务于 2013 年开始拓展至计算机视觉、自然语言理解领域，于 2017 年拓展至自动驾驶领域。2021 年，公司于科创板上市，成为国内数据服务行业唯一上市企业。目前，公司已经实现了标准化产品、定制化服务、相关应用服务全覆盖，所提供的训练数据涵盖智能语音（语音识别、语音合成等）、计算机视觉、自然语言等多个核心领域，全面服务于人机交互、智能家居、智能驾驶、智慧金融、智能安防等多种创新应用场景。

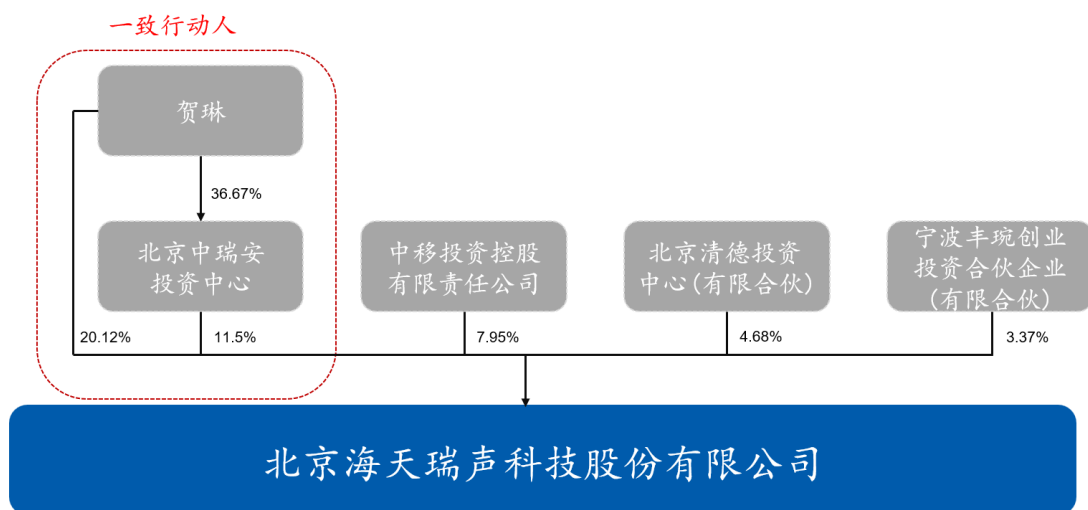
图 1：公司发展历程



数据来源：公司官网，东北证券

公司董事长、实控人为创始人贺琳女士。公司创始人、实际控制人贺琳直接持有公司 1,213.76 万股股份（占总股本的 20.12%），并通过控制中瑞安（员工持股平台）间接控制公司 693.58 万股股份（占总股本的 11.50%），合计控制公司 31.62% 的股份，为公司实际控制人。三股东中移投资为中国移动全资子公司，2019 年通过定增进入，带来国资背景和合作机会。

图 2：公司股权结构



数据来源：Wind，东北证券

公司传统业务为 AI 训练数据的研发设计、生产及销售业务，主要产品覆盖了智能语音、计算机视觉及自然语言处理三大 AI 核心领域。公司通过设计数据集结构、组织数据采集、对取得的原料数据进行加工，最终形成可供 AI 算法模型训练使用的专业数据集，通过软件形式向客户交付。公司所提供的训练数据涵盖智能语音（语音识别、语音合成等）、计算机视觉、自然语言等多个核心领域，全面服务于人机交互、智能家居、智能驾驶、智慧金融、智能安防等多种创新应用场景。

目前公司客户包括国内外互联网厂商、科技公司、教育科研机构等。公司的产品和服务已获得阿里巴巴、腾讯、百度、科大讯飞、海康威视、字节跳动、微软、亚马逊、三星、中国科学院、清华大学等国内外客户的认可，应用于其研发的个人助手、智能音箱、语音导航、内容生成、搜索服务、短视频、虚拟人、智能驾驶、机器翻译等多种产品相关的算法模型训练过程中。目前公司客户累计数量超过 930 家，覆盖了科技互联网、社交、IoT、智能驾驶、智慧金融等领域的主流企业，教育科研机构以及部分政企机构。

图 3：公司主要客户

国内科技企业	 Alibaba Group 阿里巴巴集团	 HUAWEI	 Tencent 腾讯	 Bai 百度	 京东	
	 DJI	 MI	 ByteDance 字节跳动	 快手	 科大讯飞 IFLYTEK	 中国移动 China Mobile
国外科技企业	 Google	 Meta	 amazon	 NVIDIA	 intel	
	 Microsoft	 TOSHIBA	 IBM	 SAMSUNG	 SONY	
知名车企	 SAIC 上汽集团 SAIC MOTOR	 BYD	 滴滴	 BMW	 HYUNDAI	
高校	 清华大学 Tsinghua University	 北京大学 PEKING UNIVERSITY	 中国科学院 CHINESE ACADEMY OF SCIENCES	 NUS National University of Singapore	 JOHNS HOPKINS UNIVERSITY	

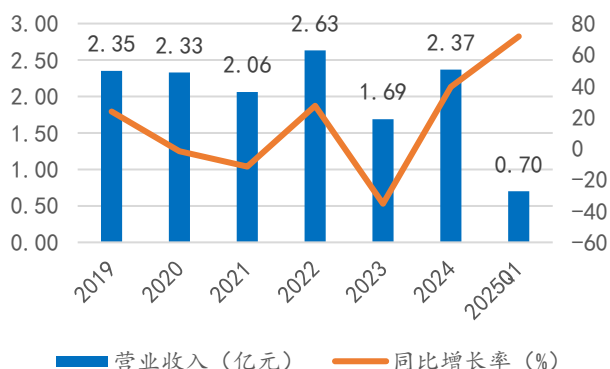
数据来源：公司年报，东北证券

1.2. 外部环境影响减弱，业务规模化回暖

外部环境影响减弱，公司业绩逐步回暖。公司近年境内营收基本保持稳定，国外营收波动较大，主要系疫情以及客户预算影响。2021 年，受疫情影响，公司海外客户业务预算减少，海外收入下滑；2022 年，由于公司进一步加强境外销售团队建设，并在营销推广方面持续发力，通过展会、社交媒体、搜索引擎等方式拓宽客户触达渠道，取得较好成效，2022 年境外收入同比增长 102.04%；2023 年，境外两家大客户经历阶段性裁员、业务方向及研发节奏周期性调整等大变动，暂缓业务预算，同时叠加 2023 年上半年数据出境相关法规落地实施的阶段性影响，公司境外收入同比大幅下滑，境内业务方面，国内客户对 AI 数据的研发投入仍持谨慎态度，同时行业内竞争加剧致价格有所下降，综合导致境内收入同比下滑。

2024 年开始，公司基本面逐步恢复。公司 2024 年营业收入达 2.37 亿元，同比增长 39.45%；归母净利润 1133.61 万元，同比增长 137.31%，成功实现扭亏为盈。2024 年一季度营收 6980.95 万元，同比增长 71.75%，延续了 2024 年的增长势头。归母净利润扭亏为盈至 37.16 万元，核心业务盈利能力逐步改善。

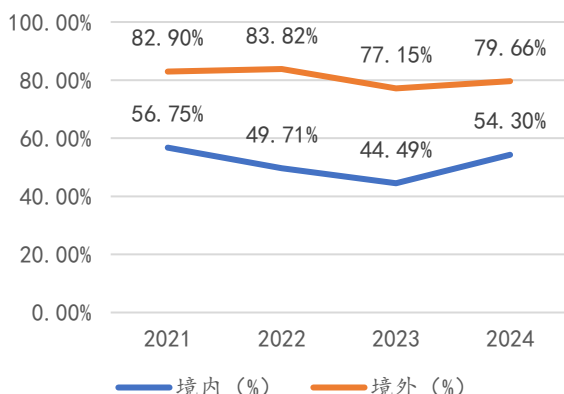
图 4：2019-2025Q1 公司营业收入及同比增速



数据来源：iFind，东北证券

公司毛利率受收入结构影响，2019-2023 毛利率呈下滑趋势主要系客户预算缩紧、数据产品业务占比波动影响，2024 开始有所反弹。分地区看，境外毛利率高于国内毛利率，主要系境外业务收入中毛利率较高的训练数据产品业务的占比相对较高，国内仍以项目制为主，毛利率随项目定价有波动，近年来国内业务竞争逐渐激烈，毛利率有所降低。后续毛利率有望随着公司数据产品业务的成熟而有所提升。

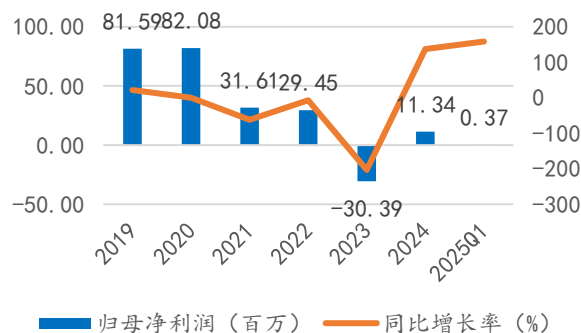
图 6：2021-2024 公司分地区毛利率



数据来源：iFind，东北证券

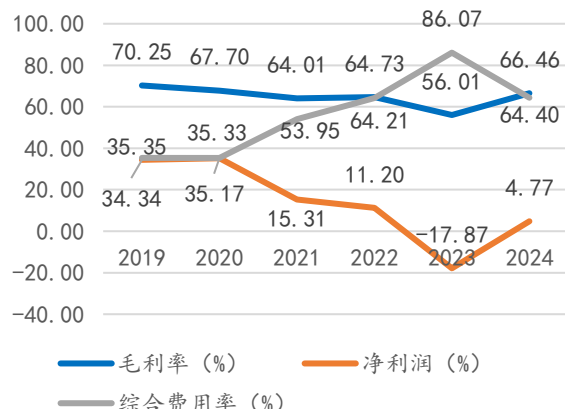
费用率方面，公司费用率 2024 年有所回落，主要系公司配合业务拓展目标，布局销售、研发能力，提质增效所致。销售费用方面，2022、2023 年增长主要系公司进行营销体系建设，在销售团队建设及激励机制优化方面加大投入所致；管理费用方面，2022 年有大幅增长，主要系为配合整体战略发展及业务拓展目标，公司加大人员储备、升级管理架构及激励体系，增加临时性场地租赁面积带来一定的房租增长，并完成新办公楼购置所致，新办公楼的折旧摊销将持续影响后续管理费用；财务费用变化主要系公司持有外币资产汇率影响。2024 年开始，公司主动收缩经费，提质增效，在业绩增长的同时，费用得到了有效控制。

图 5：2019-2025Q1 公司归母净利润及同比增速



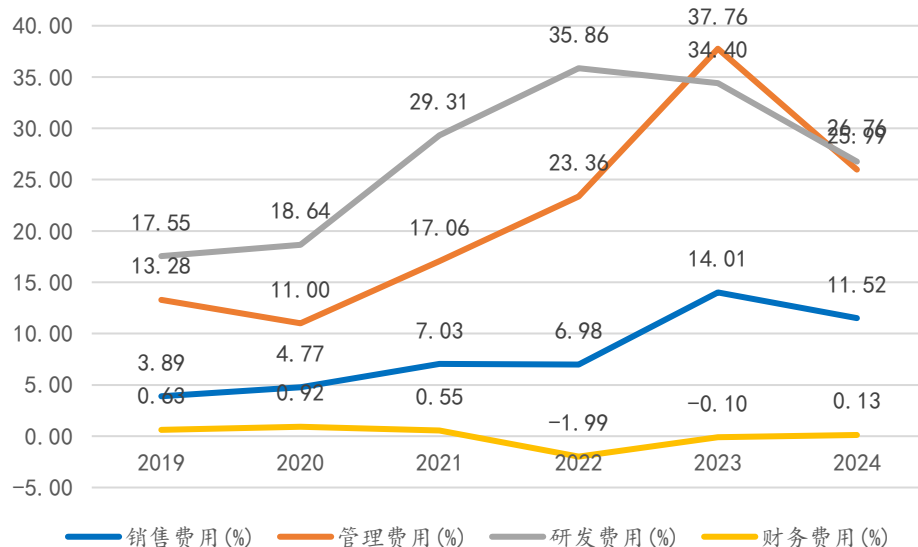
数据来源：iFind，东北证券

图 7：2019-2024 公司毛利率及净利率及合计费用率



数据来源：iFind，东北证券

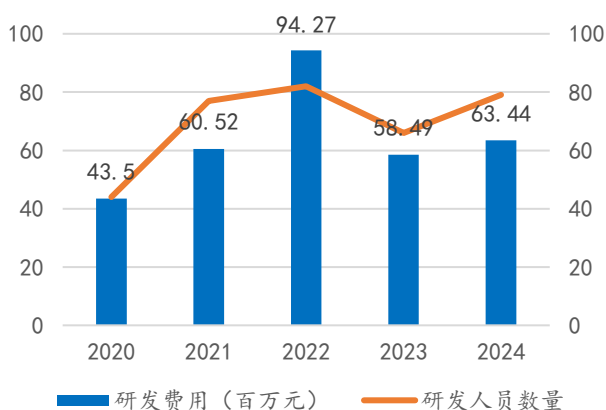
图 8：2019-2024 各项费用率



数据来源：iFind，东北证券

公司重视研发投入，研发费用、研发人员数量 2019-2022 年持续增长，2023 年公司基础能力建设完成，研发费用有所回落。2019-2022 年，公司研发费用持续上升，主要是为了快速响应行业发展和变化，提升市场占有率和研发领先实力，并加大前期投入大、后期毛利率高的数据集产品储备；2023 年，研发费用开始下滑，主要是由于公司在数据生产智能化、以及标准化数据集产品建设等已达到阶段性成熟，标准化数据集集中建设于 2023 年 5 月结项，投入强度逐渐趋于平稳。2024 年恢复上涨，公司研发人员达到 79 人，其中本科及本科以上人数占比达到 90%。

图 9：2020-2024 研发费用及研发人员数量



数据来源：iFind，东北证券

图 10：研发成果



数据来源：公司官网，东北证券

我们认为公司主要竞争优势在于：

- (1) 技术能力方面和场景资源方面，公司在基础研究、平台工具、训练数据生产方面均有技术积累，并公司持续在智能语音、计算机视觉、自然语言、以及智能驾驶为代表的垂直领域进行技术拓展；
- (2) 软实力方面，公司拥有较强的项目管理能力和全流程的数据服务能力。

较强的技术能力和丰富的场景资源。技术能力方面，公司在基础研究、平台工具、训练数据生产方面均有技术积累；场景资源方面，公司持续在智能语音、计算机视觉、自然语言、以及智能驾驶为代表的垂直领域进行技术拓展，并针对近年来对数据服务商需求的复杂化、专业化趋势，提供多样化的合作方式以满足不同客

户的需求，包括定制数据服务、售卖成品数据库、提供软件方案、提供标注基地等，目前已在以上四个方向拥有较全面的服务能力，有望受益于下游需求爆发。

表 1：公司核心技术

	公司核心技术	核心技术项下的细分技术示例
基础研究	语音识别算法	语音数据库质量预估技术、语音识别系统评测技术
	语音合成算法	语音合成数据库评估技术、说话人自适应语音合成技术、语音合成系统评测技术
	计算机视觉算法	人脸检测和识别技术、物体识别技术、光学字符识别技术、场景分割技术、运动轨迹跟踪技术、3D 点云目标检测技术、3D 点云语义分割技术
	语音语言学基础研究	基于词典与模型的发音预测技术
	训练数据集设计技术	多语种文本正则化技术、基于语言模型的文本易读性评测技术、多语种情感及领域分类技术
平台工具	一体化数据处理平台	基于 C/S 架构和 B/S 架构的大规模语音处理平台，基于 C/S 架构的音素标注技术，多语言分布式文本处理技术，基于 C/S 架构和 B/S 架构的文本标注技术，基于 C/S 架构和 B/S 架构的图像标注技术，基于 B/S 架构的 3D 点云标注技术，大规模数据采集及标注平台
训练数据生产	多语种多模态训练数据设计技术	多语种语料清洗技术、多语种语音库设计技术、多语种音素均衡语料设计技术、混合语言语料设计技术
	多模态多通道数据采集技术	多通道录音技术、移动设备上的语音采集技术、分布式图片内容分析技术、移动设备上的图片采集技术、多语言手写体数据采集技术、多模态数据采集技术
	数据同步技术	音频对齐技术
	实时采集质控技术	语音信号质量检测技术
	大数据驱动的高效数据处理技术	多语种拼写检查技术、视频流中特定帧定位技术、音素边界自动预测技术、基于语音特征的韵律预测技术、基于 CRF 的韵律预测技术、基于 HMM/CRF 的词性预测技术、基于 CRF 的命名实体识别技术、2D 障碍物自动检测和跟踪技术、2D 车道线自动检测技术、3D 点云自动贴合技术、3D 点云障碍物自动检测和跟踪技术
	分布式高性能自动校验技术	音素标注正确率校验技术、相似说话人自动筛查技术、音字一致校对技术

数据来源：公司公告，东北证券

拥有全流程数据服务能力，包括数据生产、平台工具、算法、数据合规能力等，可满足下游客户多种需求。数据生产环节，公司可针对公司需求进行数据定制，并完成数据集设计、数据采集、数据加工、生成质检等流程，交付定制训练数据；平台工具方面，公司拥有 ADS、VDS、数据中台管理模块；算法方面，公司形成多样化的算法矩阵，覆盖语音识别、自动驾驶、OCR 等领域。数据合规方面，针对部分客户对数据安全的要求，公司数据标注平台支持私有化部署，确保全流程数据安全合规。公司平台提供专属供专属 OSS 桶，只读权限，不可下载；标注平台只可获取 url，且 url 不可下载；终端作业人员浏览器仅可读，无法下载；项目结束后，定时删除所有数据。

图 11：公司业务全场景布局



数据来源：公司年报，东北证券

1.3. 数据要素提供发展新动能

数据要素被视为数字经济的核心，对传统生产方式变革具有重大影响。数据要素是指那些以电子形式存在的、通过计算的方式参与到生产经营活动并发挥重要价值的的数据资源，2020 年，国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，首次明确提出要培育数据要素市场。数据要素是数字经济的核心，对传统生产方式变革具有重大影响。它通过提高生产效率与资源配置效率，促进跨领域活动过程中资源的高效流动，从而全面提高生产效率与资源配置效率。数据要素的合理配置和利用，对于推动经济高质量发展、实现新旧动能转换具有不可替代的作用。

2025 年 3 月 19 日上午，国家数据局党组书记、局长刘烈宏主持召开数据标注领域民营企业座谈会，与海天瑞声、数据堂、百度、安徽飞数、小咖邦、乐普通盈、砺英数智、谱蓝科技、希尔贝壳、整数智能等 11 家民营企业代表座谈交流。会上，各民营企业代表结合自身企业实际经营情况，分析研判数据标注产业的发展现状、面临的困难与挑战，并对推动数据标注产业高质量发展提出具有针对性的意见建议。国家数据局成立以来高度重视民营企业意见，推动建立企业联系制度，多次召开企业家座谈会，着力为民营企业发展创造良好政策环境。下一步，国家数据局将深入学习贯彻习近平总书记关于民营经济发展的系列重要讲话和指示批示精神，认真研究吸纳本次座谈会上民营企业代表提出的相关意见建议，在政策制定、技术攻关、标准制定等方面重点发力，为数据要素市场化配置改革提供好的政策环境和技术环境，推动数据“供得出、流得动、用得好、保安全”。同时，要加快培育壮大数据标注产业多元化经营主体，形成良好的数据标注产业生态，充分发挥数据要素叠加倍增作用。

图 12：刘烈宏主持召开数据标注领域民营企业座谈会



数据来源：数据局，东北证券

我们认为海天瑞声作为国内行业首家上市的数据供应商，且有国资背景，有望在数据要素市场发力。作为国内唯一一家人工智能训练数据上市企业，公司有望肩负“构筑人工智能数据基石”的使命，积极参与标准制定和实施，不断提升数据的质量和可靠性，规范数据服务的过程和结果，更好地满足企业及机构的需求，推动数据要素市场化配置标准化工作的发展。

2. AI 大模型带动数据服务需求爆发，个性复杂化业务加深行业壁垒

2.1. 数据服务商：AI 产业链里的“大炼化”赛道

传统意义上的数据服务商位于人工智能基础数据产业链中游。从数据生产到算法应用构成了人工智能基础数据产业链，其中上游负责数据的采集和标注，中游提供数据处理服务，下游则进行算法研发和应用。

图 13：人工智能基础数据产业链



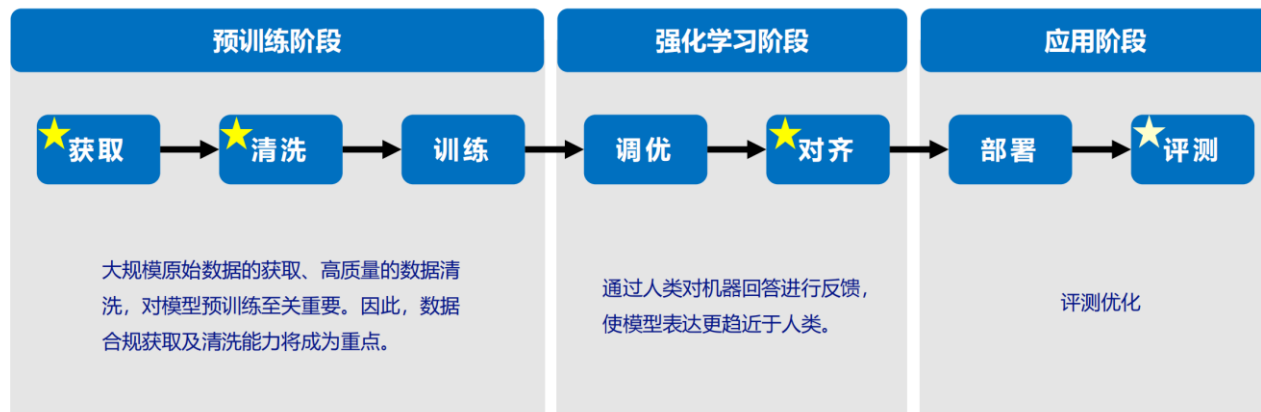
数据来源：德勤，东北证券

没有标注数据，AI 模型就像是没有教材的学生，无法学习。数据标注就是给原始数据加上标签，告诉计算机这些数据是什么，数据标注通常包括文本、图像、音频或视频等数据的分类、注释或标记，这些工作可以是人工完成的，也可以使用自动化工具来实现。数据标注为机器学习算法提供了训练数据，帮助算法理解数据的特征和关系，从而提高模型的准确性和可靠性。

在 AI 的发展趋势中，数据质量是决定 AI 能力的重要一环，这也是为何数据服务提供商商会直接受益于高质量的人工智能产业发展。数据服务提供商通常需要完成数据集结构/流程设计、数据处理、数据质检等工作，为下游客户提供通用数据集、定制化服务、数据闭环工具链等，其在产业链中各环节的任务分别为：

- （1）预训练阶段，需要数据服务提供商提供大量的原始数据；
- （2）强化学习阶段，大模型的对齐需要数据服务提供商提供调优、反馈等服务；
- （3）AI 应用阶段，企业需要针对自身业务特点和需求训练专用的小模型，需要数据服务提供商提供专业化、全流程数据服务。

图 14：AI 领域对数据服务商需求



数据来源：公司公告，东北证券

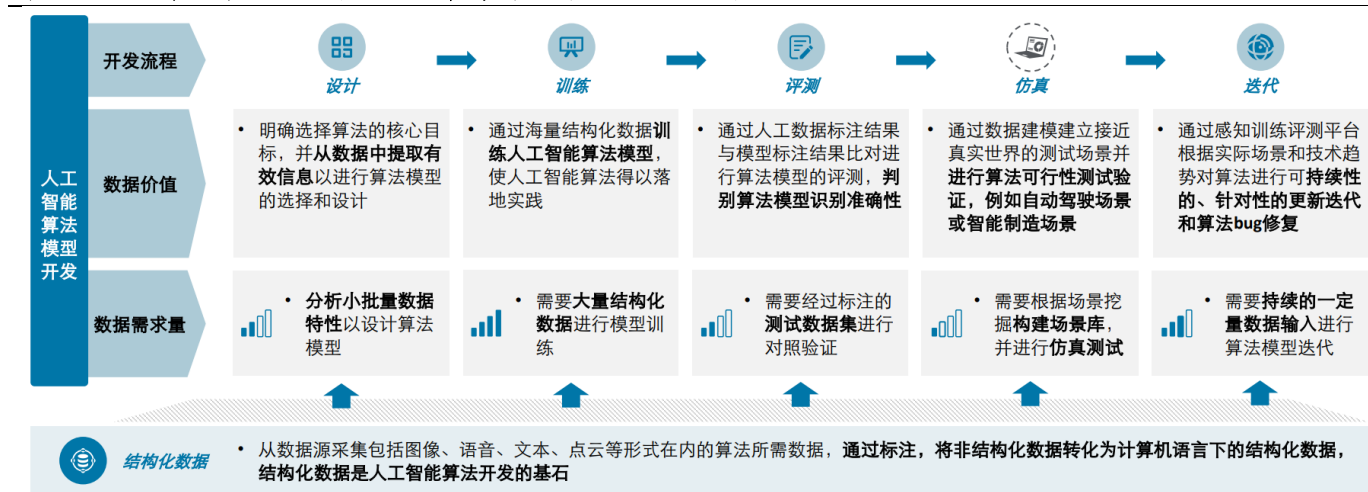
从不同阶段的数据标注服务需求来看，在研发阶段，企业一般使用开源数据集或数据集产品进行训练，数据标注需求量大，标注内容倾向于标准化；在训练阶段，企业通常需要部署私有化标注平台以及较强的数据安全流程，对数据标注内容需求更多、更丰富，准确性要求更高；在产品落地阶段，企业不仅需要部署私有化标注平台，对采标的数据内容有特殊指向，对服务方稳定性和效率也有较高要求。

未来，“基础模型+微调”将成为 AI Agent 开发的新范式，定制化服务在数据标注市场将占据更大比例。定制化的数据标注服务将成为市场需求的主流。这一阶段，在人工智能技术广泛应用于图像识别、语音处理、自然语言处理等领域的背景下，对高质量标注数据的需求持续增长；数据标注行业正逐步规范化，不合规的中小供应商面临被淘汰的风险，市场将更加青睐具有稳定数据来源、强大技术支持和规范管理体系的服务商；随着人工智能技术的深入应用，企业对数据服务的需求更加多样化和定制化，能够提供专业、精准、高效数据标注服务的企业将更具市场竞争力。

2.1.1. 训练阶段，提供大量的原始数据

高质量的数据是人工智能算法训练的重要底座，原始数据的获取、清洗、标注对大模型训练效果至关重要。比如数据清洗则是检查和修正数据中的错误，确保数据准确无误。这包括去除重复数据、填补缺失值、处理异常值和转换数据格式等操作。数据清洗的目的是提高数据的准确性和可靠性，为后续的数据分析和建模提供坚实的基础。数据清洗确保了训练数据的准确性，避免了因为数据错误导致的模型偏差。干净、准确的数据是训练出高性能 AI 模型的关键。数据标注的工作是从数据源采集包括图像、语音、文本、点云等形式在内的算法所需数据，通过标注，将非结构化数据转化为计算机语言下的结构化数据。结构化数据是人工智能算法开发的基石它帮助机器学习算法理解数据的特征、属性、结构和关系等方面的信息，从而更好地进行数据分析和建模。

图 15：人工智能算法模型开发过程中需要结构化数据做支撑



数据来源：Ofweek，东北证券

因此，随着大模型对数据需求呈指数增长，也将会带动常规数据训练服务的需求增长。2024 年底，我国数据标注企业数量逾 4000 家，年均复合增长率达 11.3%。其中“专精特新”企业占比 12%，上市公司（包括海外上市）占比 3%。从增长趋势上看，近 20 年可以分为 3 个阶段，第一阶段 2005-2013，企业数量年增长率处于 10%~12%区间，第二阶段 2014~2018，数据标注新增企业数量快速增长，2018 年增长率高达 17%；第三阶段 2019-至今，增长逐步缓和，受制于宏观经济和外部冲击，2022 年增长率降至 5%。这三个阶段分别对应了数据标注产业的成长期、快速增长期、平稳高质量发展期。值得注意的是，新增企业数量在经历了 2021-2023 年

的快速回落后，于 2024 年出现了明显的反弹，背后可能的原因是 2024 年我国人工智能产业的爆发式增长对数据标注行业有效拉动的结果。

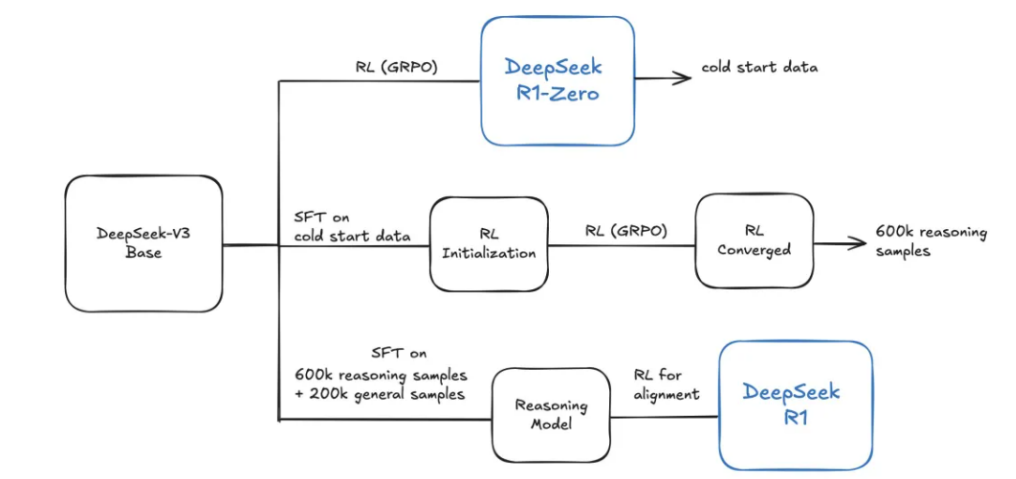
图 16：2005-2024 数据标注企业数量



数据来源：上海数据交易所，东北证券

即使是以监督学习为主的 DeepSeek，在训练初期也用了冷启动数据。及时人工标注几千条高质量解题示范（带标准格式和清晰步骤），让 AI 先临摹，再自主发挥。在 DeepSeek-R1 的研发过程中，为了避免强化学习训练初期的这种不稳定冷启动阶段，通过构建长思维链数据和采用特定的数据收集与处理方法，有效地缓解了这一挑战。这些数据通过带有详细反思和验证步骤的提示生成，并用于微调模型。同时，DeepSeek 还收集了 DeepSeek-R1-Zero 的输出，并由人类标注者进行精炼和优化。人类标注者在这里扮演了重要的角色。他们不仅需要对收集到的数据进行清理和整理，还需要对数据进行进一步的优化和精炼，以确保模型在训练过程中能够充分利用这些数据。

图 17：Deepseek 训练数据需求



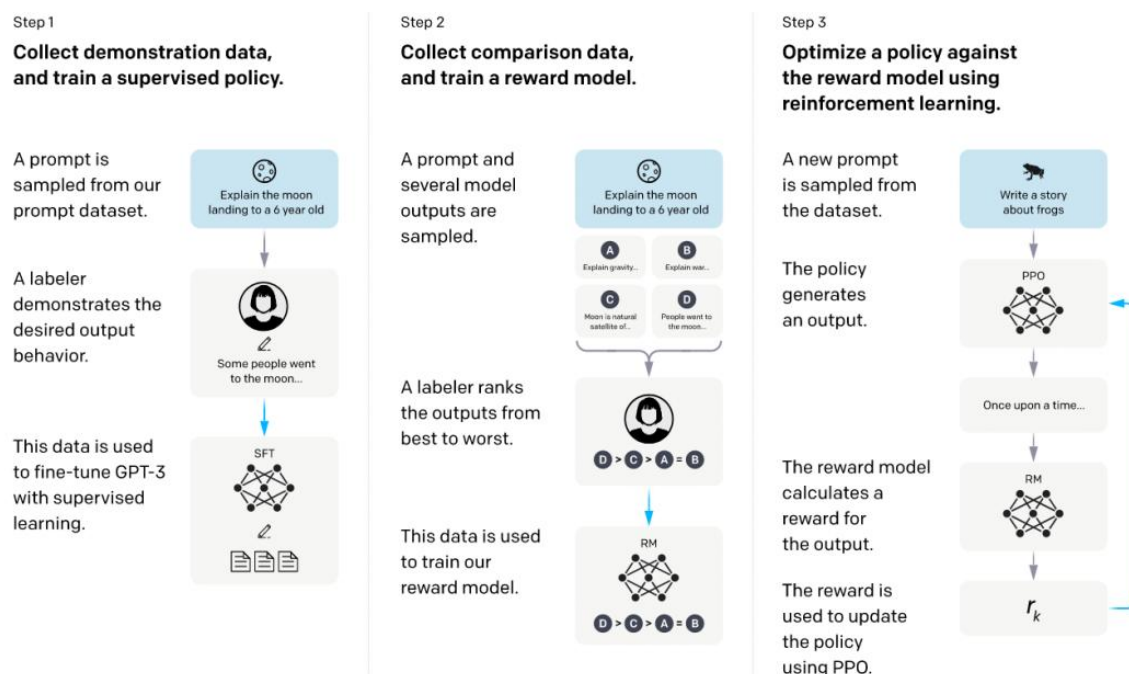
数据来源：DeepSeek，东北证券

2.1.2. 强化学习阶段，提供模型的反馈及打分

目前主流大模型普遍采用无监督自动学习机制，然而对其进行对齐、调优仍需要收集人类反馈。BERT、GPT、Transformer 等模型已经可以直接处理原始的非结构化数据，通过自注意力机制、位置编码等技术，能够捕捉到非结构化数据中的复

杂模式和关系。然而，机器学习在预训练、无监督学习过程中容易产生模型幻觉、价值取向等等问题，需要人类反馈来进行对齐。RHLF（Reinforcement Learning from Human Feedback，人类反馈强化学习）指模型需要加入人工打分环节，以保证模型能够给出符合逻辑的高质量答案，能让大模型更好地与人类指令保持一致。

图 18：大模型强化学习过程



数据来源：CSDN，东北证券

图 19：RHLF 过程（以 GPT 为例）



数据来源：安全智汇，东北证券

针对此类需求，数据服务商也会提供数据、反馈、打分等服务来强化训练模型。给大模型输出打分是为了评估其生成内容的质量和准确性。具体来说，打分可以帮助我们理解模型的优缺点，并指导模型的改进。例如图片描述可以从识别准确

度、内容涵盖度、表述质量等几个维度进行打分，并输出最终整体打分结果，将人类的偏好转化为数字奖励信号，进而引导模型的行为更加符合人类的期望。

图 20：大模型强化学习过程

➤ 示例_图文评测



【question】
请详细描述这张图片

【answer1】

这幅图描绘了一片宁静的自然风光，中心是一条蜿蜒的河流，河水清澈见底，河岸两侧是茂密的绿色植被。河岸左侧是一片开阔的草地，而右侧则是一排整齐排列的树木，它们的树干粗壮，树冠茂密，覆盖着翠绿的树叶。天空呈现出晴朗的蓝色，点缀着几朵洁白的云彩，为这幅画面增添了一抹生动的色彩。整体而言，这幅图展现了一种平和与和谐的氛围，让人感受到大自然的宁静与美丽。

【answer2】

这张图片展示了一个风景优美的自然景观。图中天空湛蓝，几乎没有云朵，呈现出一种宁静和晴朗的氛围。在画面的中央和右侧，可以看到一排排整齐的树木，树叶翠绿，似乎是在春夏季节。树木形成了一道长长的绿色屏障，与蔚蓝的天空形成鲜明的对比。树木的前方有一条静谧的水面，可能是一条小河或湖泊，水面旁边长有高高的干草，呈现出金黄色，与周围绿色植被形成对比，增添了画面的色彩层次。左侧的绿色草地向下倾斜，表明这里可能是一个小坡地形。整体上，这张图片给人一种平和宁静的感觉，是一个适合散步、放松的好地方。

answer1识别准确度>>4
answer1内容涵盖度>>4
answer1表述质量>>5
answer1整体打分>>4

answer2识别准确度>>4
answer2内容涵盖度>>5
answer2表述质量>>5
answer2整体打分>>4

对比结果>>answe2好

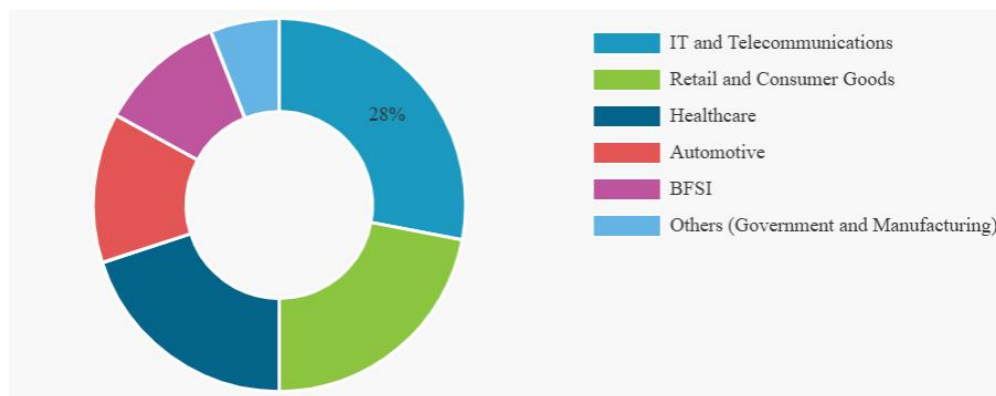
数据来源：海天瑞声、东北证券整理

RLHF 标注需要非常专业的人士，而大模型厂商通常会将此类工作外包给人员管理能力更强的第三方数据服务商。例如，OpenAI 招了几十名 PhDs 来做 RLHF 的标注，数据服务公司 ScaleAI 作为 OpenAI 的上游供应商，也招聘了几十名 PhDs 为 OpenAI 服务，具体的分工是 ScaleAI 负责标注，而 OpenAI 负责质量检测。

2.1.3. 应用阶段，提供专业化、全流程的高质量数据训练

随着人工智能进入多元行业和场景落地阶段，自动驾驶、医疗、法律、金融等有一定专业性要求的垂直化场景将成为主要需求，从而带动行业进一步朝着垂直化、定制化方向发展。这将使得专业性要求越来越高，市场准入门槛显著提高。数据服务商在根据不同行业和企业要求进行改造，并提供需求数据的同时，还应该基于行业理解，与企业共同制定相关的标注规范和规则，从而让算法发挥出更高的识别效率。此外，还要深度参与到客户人工智能算法从预研立项、定制开发到部署投放的全过程中，凭借自身的数据服务能力，在不同阶段为客户提供自制数据集，定制化的数据采集、数据清洗和数据标注服务以及数据中台集成能力。

图 21：AI 应用的下游行业



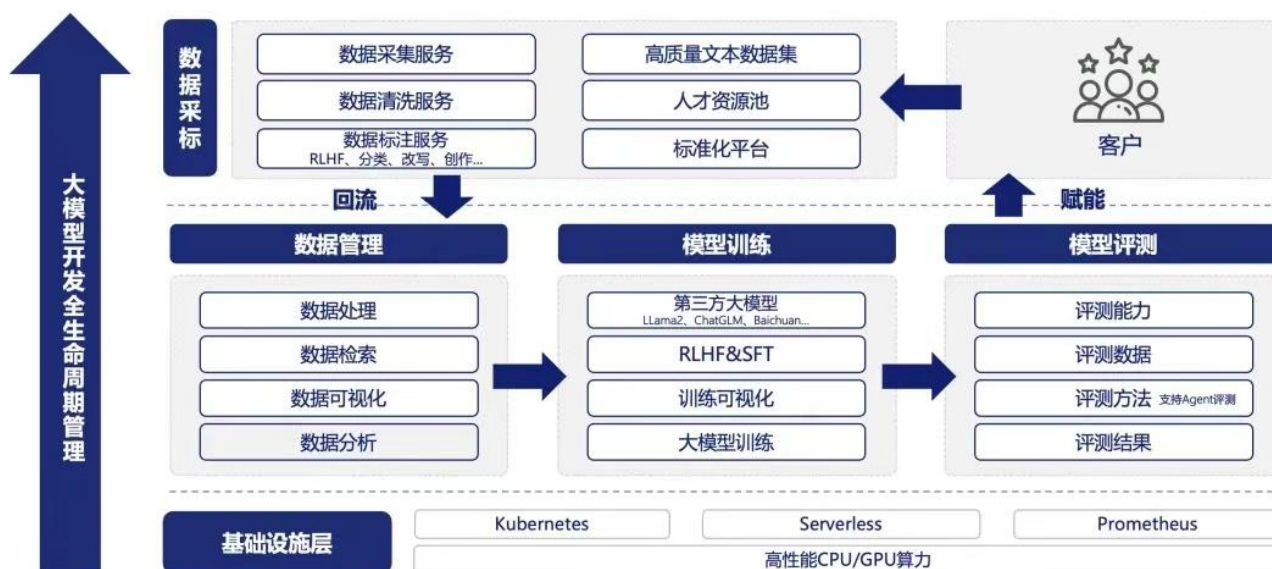
数据来源：Fortune Business Insights，东北证券

数据是 AI Agent 发展的关键要素：AI Agent 的构建主要依赖于大型语言模型和交互能力两大核心要素，而高质量的数据是训练高效大语言模型的关键。AI Agent 行业对数据的需求不仅在于数量，更在于多样性和质量，以确保模型能够适应不同的语言环境和应用场景。

- **确保数据质量**：数据标注厂商具有专业的标注团队和严格的标注流程，能够对数据进行清洗、分类、标注等处理，确保数据的准确性、一致性和规范性。海天瑞声在数据清洗和标注方面持续发力，通过制定严格的数据清洗规则和标注流程，能够为客户提供高质量的数据，满足大模型训练的需求。
- **实现数据多样性**：数据标注厂商可以采集和标注各种类型的数据，包括文本、图像、音频、视频等，为 AI Agent 厂商提供丰富的数据来源，帮助模型更好地学习和理解不同模态的信息，提升 AI Agent 在多模态任务中的表现。
- **节省成本和时间**：数据标注是一项耗时费力的工作，需要大量的人力和物力投入。AI Agent 厂商如果自行建立数据标注团队，需要花费大量的时间和资源来招聘、培训人员，购买设备和软件等。而与数据标注厂商合作，可以将这些工作外包出去，节省自身的成本和时间，专注于 AI Agent 的核心技术研发和产品创新。

同时部分有 AI 需求的下游非科技型企业并不具有算法研发能力，这也需要第三方厂商提供帮助提供全生命周期服务，为数据服务商开拓新的市场空间。在研发阶段，企业一般使用开源数据集或数据集产品进行训练，数据标注需求量大，标注内容倾向于标准化；在训练阶段，企业通常需要部署私有化标注平台以及较强的数据安全流程，对数据标注内容需求更多、更丰富，准确性要求更高；在产品落地阶段，企业不仅需要部署私有化标注平台，对采标的数据内容有特殊指向，对服务方稳定性和效率也有较高要求。

图 22：大模型驱使下的全生命周期管理

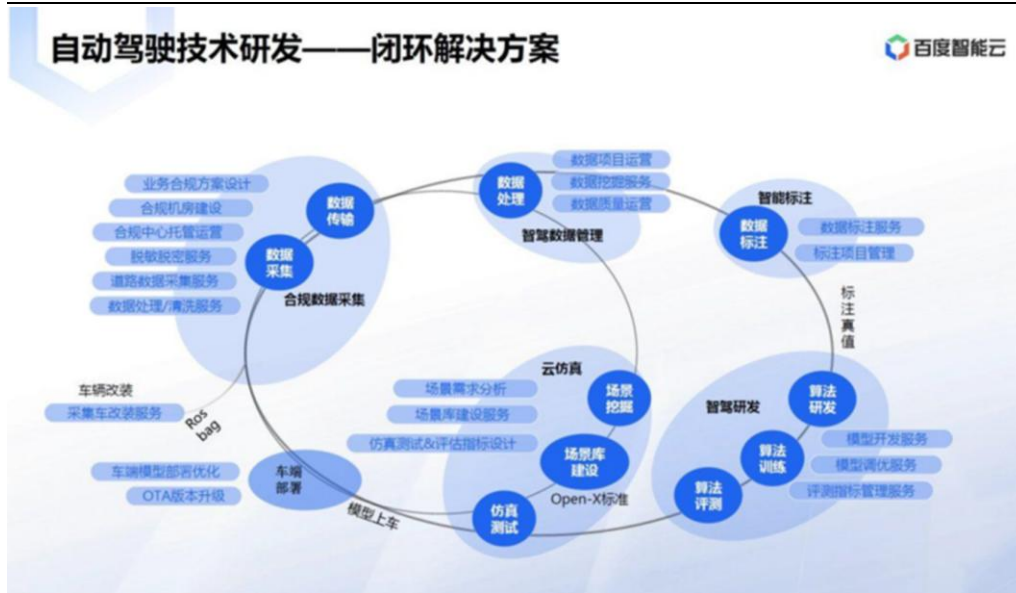


数据来源：海天瑞声，东北证券

同时在目前的产业模式中，产业链上中下游界限逐渐模糊，数据服务商能力开始向上下游渗透。从商业模式来看，当前基础数据服务的商业模式包括两大类。一类是提供基础的数据服务模型，如数据出售、数据 API 服务、数据标注服务、数据工具和软件出售等，这类商业模式适用于通用类的数据服务，其定价依据数据大小量、标注难度等，价格在几千至几十万之间。这一类商业模式适合数据采集和标注的门槛较低的场景，以及通用性较强的场景。第二类是全栈式商业模式，如

百度智能云提供的数据闭环解决方案及工具链平台，研发闭环包括两个环，一个是中间层小环，数据产生之后，经过数据管理平台，包括一些仿真测试、仿真云、再到车端部署，实现仿真的场景。另一个是外面的大环，主要做一些数据管理的工作，经过标准、加工、模型训练以及最后的车端部署，来实现这样一个大环。将上面的数据闭环链路拆解到架构层来看，可以把自动驾驶的数据研发分为如下五层：从最底下的基础设施，到中间的 PaaS 工具链，再到上面的一些 AI 运营服务。

图 23：百度智能云数据闭环方案



数据来源：百度智能云，东北证券

随大模型向行业应用发展的趋势，数据服务商将从提供基础数据服务模式为主转变为提供全栈式服务模式。随着下游应用方的下场参与和 AI 进入到 AI Agent 的时代，下游应用方算法能力积累较少，而同时面临算法能力快速部署与迭代的需求，进而催生出了对包括云资源、算法工具链等全栈式工具服务的需求。数据服务商在根据不同行业和企业要求进行改造，并提供需求数据的同时，还应该基于行业理解，与企业共同制定相关的标注规范和规则，从而让算法发挥出更高的识别效率。此外，还要深度参与到客户人工智能算法从预研立项、定制开发到部署投放的全过程中，凭借自身的数据服务能力，在不同阶段为客户提供自制数据集，定制化的数据采集、数据清洗和数据标注服务以及数据中台集成能力。

2.2. 壁垒加深，个性化需求带动数据服务商向复杂化、专业化方向发展

在 AI 大模型发展的催化下，数据服务商的也在逐步变的更加专业、更加复杂，其行业发展的趋势体现在：

- (1) 标注自动化：从提供人工标注的服务模式为主转变为提供 AI 服务模式；
- (2) 需求专业化：数据标注复杂化、行业化、垂直化属性逐渐增强；
- (3) 集中度提高：行业集中度将进一步提高，头部的品牌数据服务商将受益。

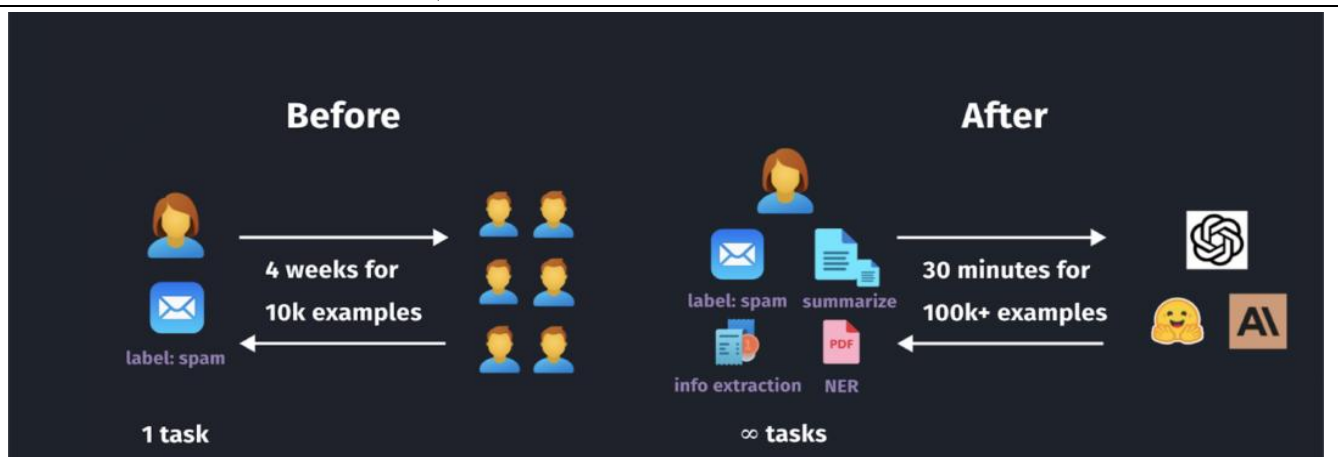
2.2.1. 在 AI 的辅助下，标注效率程度变高，难度变大

当批量处理的规模效应使每千次错误的修正成本低于人工执行的机会成本时，AI 幻觉便从缺陷转化为特定场景下的“效率催化剂”。这或许正是大模型落地的务实路径：在容错边界内，让技术红利跑赢误差成本。基于统计预测的大模型架构使

得 AI 幻觉成为技术内生的“创造性副产品”，比如 DeepSeek-R1，其在复杂推理中出现了 14.3% 的幻觉率。当模型尝试填补训练数据的空白时，难以消除生成虚构细节的逻辑。然而，在即将到来的批量处理场景中，例如电商客服话术生成、短视频脚本初筛、基础数据标注等容错率大于或等于 15% 的业务领域，这种系统性误差反而具有经济合理性。以某零售平台的实践为例，部署大模型后，单条客服响应成本从 1.2 元降至 0.15 元。尽管内容偏差率达到 8%，但通过预设“敏感词过滤+人机协同池”机制，仅需 5% 的人工复核即可交付。按日均处理 10 万条计算，单日可节约人力成本 9750 元，而错误修正的边际成本仅为 1200 元，规模效应形成的成本差达到 8 倍。

这种“用 1% 的错误率换取 90% 效率提升”的模式，在内容生产、数据标注等劳动密集型场景中尤为显著。我们认为这本质上是人机协同的成本再分配。正如医疗影像初筛允许 10% 的误判率（需人工复核），却将医生日均阅片量从 30 例提升至 200 例，行业已经验证了这种容错的经济性。对于下周即将落地的千万级文本标注项目，我们的测算显示，若采用“大模型初稿+抽检复核”模式，整体成本将降至纯人工的 1/7，而误差率（约 12%）完全处于业务可接受区间。在非关键决策场景中，与其追求零幻觉的技术乌托邦，不如拥抱“可控误差下的效率革命”。

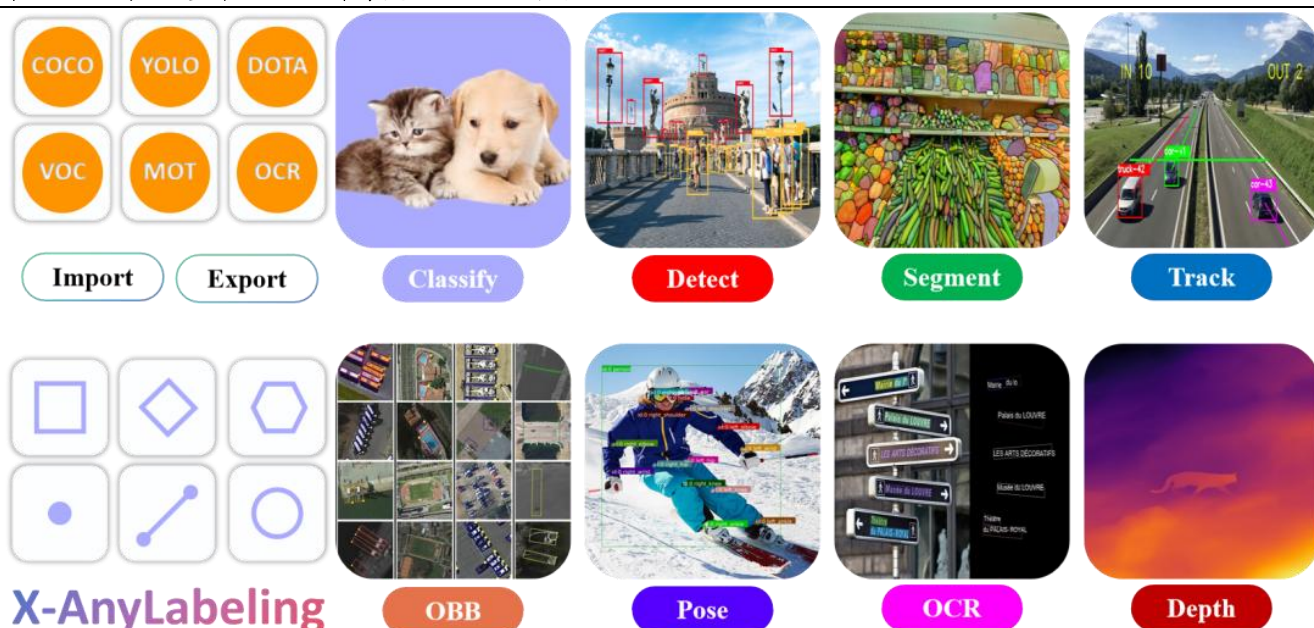
图 24：大模型将提高自动标注的效率



数据来源：Refuel，东北证券

数据标注以及文字识别是目前用的最多的场景。在大模型的催化下，人类能够标注的东西，AI 基本都能标注，甚至在某些领域，AI 的标注速度和质量已经超越了人类的平均水平。从简单的文本分类到复杂的语义分析，从图像识别到多模态数据处理，大模型能够高效地处理海量数据，并精准地标注出其中的关键信息。它不仅能够快速学习人类标注的规则和模式，还能通过不断优化自身的算法，逐步提高标注的准确率和效率。这不仅极大地提高了数据标注的效率，降低了成本，也为人工智能的进一步发展提供了更坚实的数据基础，推动着整个行业迈向智能化、自动化的全新阶段。

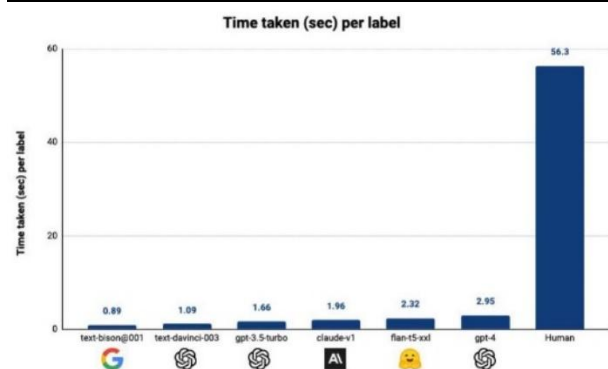
图 25：目前 AI 基本可以给所有类型的数据打标



数据来源：X-AnyLabeling，东北证券

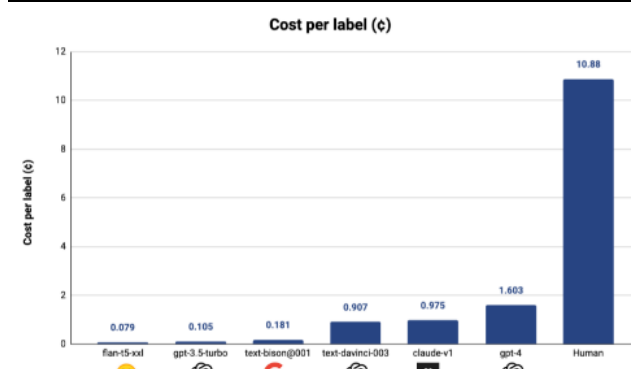
AI 方案能明显提高数据的标注效率。用自动化的方式标注数据，相比于人工标注，效率最高可以提高 100 倍，就算按照使用成本最高的 GPT-4 来算，采用 Autolabel 标注的成本只有使用人工标注的 1/7，而如果使用其他更便宜的模型，成本还能进一步降低

图 26：自动标注对于效率的提升



数据来源：Refuel，东北证券

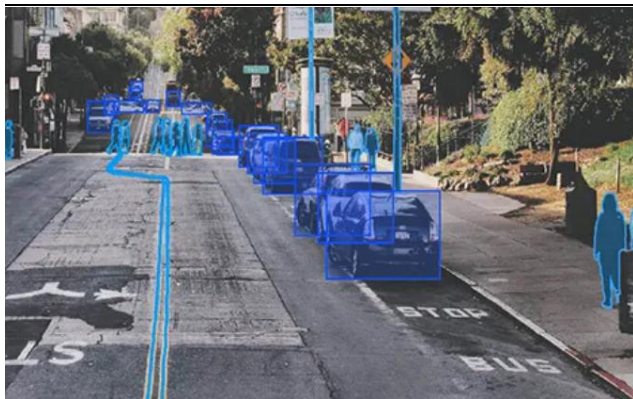
图 27：自动标注对于成本的节约



数据来源：Refuel，东北证券

AI 能完成人类也无法完成复杂的标注，比如在自动驾驶的辅助场景中的 3D 以及 4D 标注。自动驾驶场景的数据标注也对 AI 的要求逐渐提高，也从传统的车道线、路面标识等 2D 标注扩展至为点云语义分割、点云 BEV 等 3D 标注场景以及行车场景、泊车场景等 4D 标注场景。例如车辆信息标注场景，传统标注方式为拉框标注，近年来点云标注成为主流，对标注员的专业性要求提高。拉框标注主要处理二维图像数据，通过矩形框标注出图像中的指定目标对象，通常使用简单的图像编辑软件，主要要求标注员具备图像识别和目标定位的能力；而点云标注处理的是三维数据，通过激光雷达等传感器采集到的密集点云进行分类和标注，需要使用专业的点云处理软件，能够处理大量的三维数据，同时需要标注员具备三维空间感知和数据处理的能力，这点人就远远不如 AI。

图 28：传统业务类型：边界标注



数据来源：博远数栈，东北证券

图 29：专业化业务类型：3D 标注



数据来源：博远数栈，东北证券

2.2.2. 数据标注类型复杂化，行业化、垂直化属性逐渐增强

AI 的发展使得自动标注工具开始代替人工完成大部分低技术含量的标注工作，但 AI 无法对于高知识价值场景进行标注，行业从开始劳动密集型朝知识密集型方向发展。随着人工智能进入多元行业和场景落地阶段，自动驾驶、医疗、法律、金融等有一定专业性要求的垂直化场景将成为主要需求，从而带动行业进一步朝着垂直化、定制化方向发展。这将使得专业性要求越来越高，市场准入门槛显著提高。

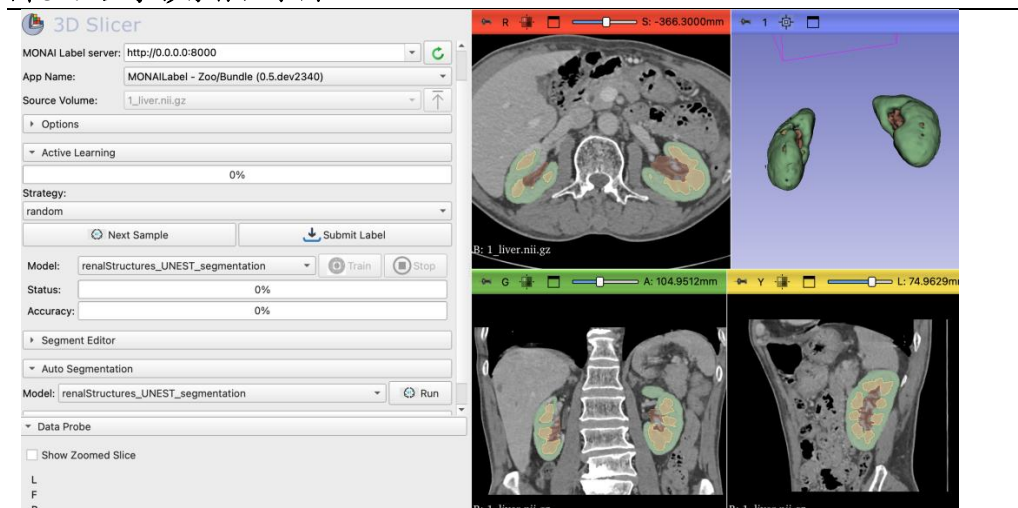
图 30：数据标注行业需求变化

	传统数据标注	大模型数据标注
领域划分	按不同领域或任务划分	按不同阶段划分
具体实操	拉框、描点、转写等操作	排序、改写、生成等操作
标注要求	偏客观	偏主观
评价指标	准确率+效率	难以对齐标准
解决方案	工具/平台标注+人类质检	专业培训、定期开会对齐等举措
人才要求	专科为主	本科以上，多领域专业人才
标注角色	按职能划分 标注员、质检员、管理员	按阶段划分 AI训练师、模型精调师、指令工程师、红队测试军团等。
覆盖区域	主要集中在三四线城市	重新打散

数据来源：量子位，东北证券

如对更复杂路段的自动驾驶、语言文学类大模型等的标注，会需要更专业、更高学历的数据标注师来完成，这也是专业机构的价值所在。例如在自动驾驶中，标注人员需要准确地标示出道路上的障碍物、行人、车辆等，任何错误都可能导致严重的交通事故，因此在 AI 标注后，需要高级人员进行质控来保障业务的稳定，并且需要定期进行错误分析，记录错误点；某些领域需要标注人员具备相关的专业知识，例如在医疗影像标注中，标注人员需要具备医学知识；在自然语言处理中，标注人员需要具备语言学以及地方文化知识，这些也是一般 AI 所无法直接具备的。

图 31：医学影像标注示例



数据来源：医学图像人工智能实战，东北证券

图 32：文学评价标准示例

整体得分	说明
5分	完美，生成内容不包含与图片不符的事实错误，且完全涵盖图片中的重要信息与明显信息，语言通顺流畅。
4分	caption基本能够准确识别和理解图片内容，可以允许少量不重要信息的精度错误。
	caption能够涵盖画面一半以上的明显信息、重要信息。
	caption语言通顺、容易理解，可以允许小幅度的模版感。
	整体达到了较好水平，可用性较高
3分	caption表述信息存在较多错误，但不影响对图片主要信息的阐述、整体内容仍与图片信息不相悖。
	caption除能够涵盖画面内容的核心内容信息以外，还能兼顾到图片上的其他明显信息。
	caption表达内容可以让人理解，允许出现少量表达重复、错别字等瑕疵性问题。
	整体水平达到了小作修改后可用的程度，存在可容忍程度的信息不准确、不全面、不流畅。
2分	caption中存在较大量的错误，但关于图片核心信息的内容生成正确。
	caption的内容能够涵盖图片核心信息。
	caption出现较大的歧义、混乱，需要通过推测来判断其表达含义。
	整体水平存在问题，如包含严重错误、严重不全面、严重不流畅等。
1分	caption信息完全错误。
	caption内容未涵盖图片核心信息。
	caption表达完全崩坏，大量重复信息、语句混乱，无法通过推测得出其含义。
	图片内容识别错误、理解错误、表达错误

数据来源：海天瑞声，东北证券

行业集中度将进一步提高，头部的品牌数据服务商将受益，数据服务商主要竞争点在于技术能力、场景资源。技术能力包括：数据闭环工具链的智能化水平、对大模型/AI 算法的理解、数据工程化能力、基础设施建设等；场景资源能力主要包括高质量的场景数据以及场景人才（领域专家、深度用户等），这与行业持续多年的深耕不无关系——企业懂得行业 know-how，能够根据客户需求，快速找到并利用与场景最为贴合的数据和人才资源。

数据服务商的技术和场景，具有飞轮效应，数据服务商行业集中度将进一步提高。以往靠销售渠道、人力成本等驱动因素，服务商实现业务量增长，进而赢得市场获得口碑，获得新客户也会越来越容易。如今在技术和场景资源能力双重驱动下，数据处理能力越强，大模型标注经验越丰富，落地案例越来越多，数据处理的可扩展性和灵活性也越高，头部的品牌数据服务商由于具备较强算法能力和稳定训练数据资源，有实力打造 AI 辅助标注、自动化标注平台工具，标注效率、质量和成本优势将愈发凸显。新创业公司入局门槛将进一步提高，专业数据服务提供商更多机会将在垂直场景，帮助企业完成私有化部署。数据服务商行业集中度较低，集中度仍有较大的提升空间，例如，2019 年中小数据供应商的市场份额高达 47%，百度市场份额为 11%，而 2022 年百度的市场份额跃居到 18%。

图 33：数据服务飞轮效应



数据来源：量子位，东北证券

2.3. 政策持续催化，行业发展进入快速发展期

数据标注产业作为人工智能发展的基石，已上升至国家战略层面。2024 年 12 月，国家发展改革委联合国家数据局、财政部、人力资源社会保障部发布《关于促进数据标注产业高质量发展的实施意见》（以下简称《意见》），明确提出到 2027 年产业规模年均复合增长率超过 20% 的发展目标，并系统性部署了需求牵引、创新驱动、生态培育、支撑保障四大任务。《意见》的出台不仅标志着我国数据标注产业进入规范化发展的新阶段，更为行业迭代指明了方向——从传统劳动力密集型标注向智能化、标准化、生态化跃迁。本报告基于工商企业数据、专利数据、全国各地政策文件等，对我国数据标注产业进行了初步分析，涵盖产业规模、空间分布、产业关联、技术创新、制度建设等方面，旨在揭示我国数据标注产业的典型事实，并提出相关政策建议。2025 年 1 月 13 日，国家发展改革委等部门发布了《关于促进数据标注产业高质量发展的实施意见》。

从政策层面来看，我国数据标注产业政策呈现出以下导向：

- **国家战略引领，构建顶层框架：**国家层面以《促进数据标注产业高质量发展的实施意见》为核心，形成了“标准建设 + 技术攻关 + 生态培育”三管齐下的政策体系。
- **区域差异化布局，打造特色产业集群：**各地基于产业基础与资源禀赋，形成了差异化的发展路径。作为劳动力大省，河南利用其人力资源优势，提出了“数据标注乡（村）”的概念，推进数据标注产业集聚发展；山东则依托海洋优势，建设海洋数据标注基地。
- **技术应用深度融合：**政策普遍强调数据标注与人工智能、大模型的协同发展，支持自动化标注、AI 辅助标注技术研发，强化数据标注在智能制造、车联网等场景的应用支撑，推动数据标注从基础服务向智能化升级，构建“数据标注 - 模型训练 - 场景应用”的闭环。
- **标准体系建设加速推进：**国家层面加快研制数据采集、标注、治理等全流程标准，各地推进数据标注相关地方标准与行业标准建设，逐步形成了“国家标准 + 行业标准 + 地方标准”的立体化标准建设体系。
- **要素保障持续强化：**在人才培养方面，河南开展大数据标注基础人才培养，贵州强调培养数据标注工程师人才；在数据供给方面，安徽通过开放公共数据资源吸引数据标注企业，广东建立人工智能产业数据资源清单，提升人工智能数据标注库的规模和质量；在资金支持方面，安徽对在本省从事数据标注的企业按数据交易额与企业绩效进行补贴与奖励。

表 2：全国各地提及数据标注的政策文件一览

政策文件	地区	相关阐述
《国家发展改革委等部门关于促进数据标注产业高质量发展的实施意见》	国家	全篇
《国家数据基础设施建设指引》	国家	在数据标注产业的生态构建、能力提升和场景应用等方面先行先试。确保数据标注的准确性和数据模型的专业性。制定高质量数据标注与交付规则，提高训练数据质量。
《新材料大数据中心总体建设方案》	国家	强化工业数据字典、数据标注等标准研制
《国家人工智能产业综合标准化体系建设指南（2024 版）》	国家	规范人工智能研发、测试、应用等过程中涉及数据服务的要求，包括数据采集、数据标注、数据治理、数据质量等标准。
《关于促进数据产业高质量发展的指导意见》	国家	强化数据标注、数据合成等核心技术攻关。推动数据清洗、质量检测、数据加工、数据标注、数据集成等技术和业态发展，创新数据开发治理一体化模式，支持人工智能技术在自动化数据处理、数据标注、模型构建、预测分析等领域的应用。
《关于促进企业数据资源开发利用的意见》	国家	培养数据治理、数据分析、数据合规、数据标注、人工智能训练等方面的技术技能人才
《合肥数据标注产业发展规划（2025 - 2027 年）》	合肥市	全篇
《郑州市数据要素市场化配置改革行动计划（2023—2025 年）》	郑州市	推动数据挖掘和统计分析核心算法、关系网络、数据标注、舆情分析、搜索等数据应用引擎开发及应用
《重庆市元宇宙产业发展行动计划（2023—2025 年）》	重庆市	推进数据交易、数据标注等领域快速发展
《北京市数字经济全产业链开放发展行动方案》	北京市	探索搭建人工智能数据标注库或知识生产平台，面向社会提供数据服务
《长春市支持数字产业发展若干政策》	长春市	数字产业企业在开展数据标注、数据分析处理、数据采集、模型训练等数据开发加工产业基地建设、运营的，享受数字经济园区同等政策
《青岛市数据要素市场化配置改革三年行动方案》	青岛市	开展数据标注、清洗、脱敏、分析挖掘、流通交易、安全与隐私保护、可信数据空间、大模型等核心技术攻关
《济南市推动数据要素市场化配置改革加快数字经济发展行动方案（2024—2025 年）》	济南市	以数据要素开发利用与市场化配置为牵引，带动人工智能、数据中心、数据标注、信创、车联网等产业发展。构建“2+N”数据要素产业发展布局，形成数据要素相关产业的集聚效应，2024 年，建设人工智能大模型创新工厂，推动大模型集成开发与训练产业创新发展，争创国家级数据标注基地
《打造通用人工智能产业创新和应用高地若干政策》	安徽省	对提供数据采集、数据清洗、数据标注等服务的在皖数据生产处理供应商，鼓励各市按企业当年数据交易额的一定比例给予补贴，省级根据绩效给予各市综合奖补
《安徽省通用人工智能创新发展三年行动计划（2023—2025 年）》	安徽省	推动数据的采、存、洗、标、训等业务全面发展，促进数据清洗、标注等方面商业合作。搭建数据集精细化标注众包服务平台，鼓励不同领域专业人员参与数据标注等工作，招引数据标注企业，通过开放公共数据资源、完善数据交易机制等方式招引专业领域数据标注企业和创新创业团队来皖落地。
《甘肃省“数据要素 X”三年行动实施方案（2024—2026 年）》	甘肃省	提升服务能力，围绕数据清洗、数据标注、数据分析、数据可视化等需求，加快培育多层次数据服务，发展智能服务、价值网络协作、开发运营一体化等数据服务新模式新业态，支持企业“上云用数赋智”，推动价值链延伸。
《广东省人民政府关于加快建设通用人工智能产业创新引领地的实施意见》	广东省	基于隐私计算支撑样本数据流通安全，搭建可信数据标注和模型训练环境跨行业数据融合。建设高质量数据集和精细化标注平台，推动成立数据标注联盟，形成数据标注行业标准，建立人工智能产业数据资源清单，汇聚行业数据资源，提升人工智能数据标注规模和质量。
《2024 年广东省数字经济工作要点》	广东省	建设省级标注数据管理平台，高效开展公共数据标注攻坚行动，持续提升公共数据标注质量。
《贵州省推动人工智能高质量发展行动方案（2025—2027 年）》	贵州省	做大面向人工智能应用的数据采集、数据标注、数据集建设、数据交易等数据服务产业，建设数据标注产业基地。支持各市（州）围绕数据标注、行业大模型应用、数据安全等人工智能产业链错位布局产业园区。
《扎实推进 2024 年下半年经济稳向好若干措施》	河南省	大力培养数据标注、模型算法、模型微调工程师等产业人才
《河南省大数据产业发展行动计划（2022—2025 年）》	河南省	开展省级数据标注基地建设试点
《河南省“十四五”数字经济和信息化发展规划》	河南省	鼓励发展面向数据清洗、数据标注、数据分析、数据可视化等需求的平台化服务。到 2025 年，全省数据服务能力全面提升，数据标注、数据安全等产业规模全国领先。推广“众包”“众包+工厂”“机器+人工”等数据标注发展新模式，发展数据标注产业。充分发挥人力资源优势，推进数据标注产业集聚发展，建设一批数据标注乡（村）。
《2024 年河南省大数据产业发展工作方案》	河南省	鼓励发展面向数据清洗、数据标注、数据分析、数据可视化等需求的平台化服务。支持商丘北航天空创新园数据加工基地、安阳大河数据标注产业园等园区打造省级数据标注基地，数据标注产业规模突破 5 亿元。开展大数据标注、分析、分类、管理等基础人才培养
《湖北省数字经济发展 2024 年工作要点》	湖北省	争创国家数据标注基地，打造高质量数据集。

《湖北省大数据产业“十四五”发展规划》	湖北省	支持建设数据清洗和分析应用中心，面向数据供应方和需求方，提供数据清洗、数据标注、数据分类、数据评估、数据分析等数据服务。
《江苏省“十四五”数字经济发展规划》	江苏省	创新数据服务模式，强化数据清洗、数据标注、数据加工等数据服务供给
《关于加快释放数据要素价值培育壮大大数据产业的意见》	江苏省	支持有条件地区发展数据标注产业
《宁夏回族自治区数字经济发展“十四五”规划》	宁夏回族自治区	引进和培育一批数据清洗、数据标注等数据加工企业
《山东省大数据产业发展 2023 年工作要点》	山东省	鼓励发展数据标注、数据清洗、数据交换等新业态
《山东省“十四五”数字强省建设规划》	山东省	大力发展数据标注、数据清洗、数据交换等新商业模式
《关于加快推进数据要素市场化配置改革的实施意见》	山东省	依托优势产业建设海洋等数据标注基地
《山西省数字企业发展 2022 年行动计划》	山西省	巩固提升数据标注发展水平，打造专业数据集。发展先进计算、数据标注、软件服务、网络安全等主导产品
《山西省关于促进数据要素市场化配置改革的实施意见》	山西省	打造行业数据采集、存储、清洗、标准化、标注等治理能力，服务行业大模型等人工智能技术研发创新。支持我省数据标注基地提质升级，积极争创国家级数据标注基地。
《关于加快数字经济高质量发展的实施意见》	四川省	加快建设成都国家数据标注基地试点，推进“蓉数公园”、四川数据要素产业园等特色数据产业园区建设
《云南省数字经济发展三年行动方案（2022—2024 年）》	云南省	建设数据标注基地、标注培训服务、数据采集服务、算法库等。提供 AI 模型训练、数据标注等基础业务支撑
《浙江省人民政府办公厅关于加快人工智能产业发展的指导意见》	浙江省	支持建设数据标注众包服务平台，提供数据采集、清洗、标注、脱敏、存储、溯源、质量评估等工具

数据来源：公司公告，东北证券

数据标注基地持续建设，各地政策持续落地中。探索建设国家级数据标注基地，充分发挥地方配套支撑作用，在数据标注产业的生态构建、能力提升和场景应用等方面先行先试，集聚头部企业，促进区域人工智能产业生态发展。目前我国七个数据标注基地分别位于四川成都、辽宁沈阳、安徽合肥、湖南长沙、海南海口、河北保定和山西大同，数据标注总规模达到 17282TB，相当于中国国家图书馆数字资源总量的 6 倍左右。目前已形成医疗、工业、教育等行业的高质量数据集 335 个；赋能 121 个国产人工智能大模型研发；引进和培育标注企业 223 家；标注从业人员达 5.8 万人；带动数据标注行业相关产值超过 83 亿元。国家数据局表示，未来将进一步畅通数据采集、标注、人工智能应用产业链，重点推动工业、金融、医疗、交通、教育等几大领域的高质量数据集建设，促进数据标注产业高质量发展。

图 34：国家数据局发布 7 个标注基地



数据来源：国家数据局，东北证券

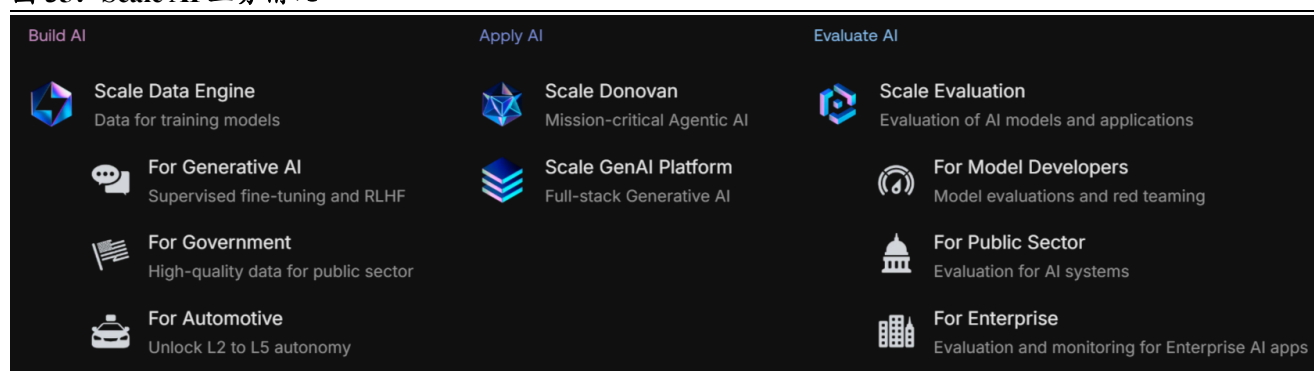
3. 他山之石：美国的发展情况

3.1. ScaleAI：从数据标注到数据分析

ScaleAI 与海天瑞声业务相似度高，从 ScaleAI 发展历程可以推测海天瑞声后续业务发展趋势。ScaleAI 成立于 2016 年，专注于提供数据标注服务，初期服务于自动驾驶领域，后扩展到政府、电商等多个行业。公司提供的数据标注服务几乎已覆盖所有领先的 AI 模型，致力于在各个行业更快地实现 AI 技术的应用和突破。ScaleAI 在 2024 年完成了 10 亿美元的 F 轮融资，估值达到 138 亿美元。本轮融资由 Accel 领投，投资者包括 Y Combinator、Nat Friedman、英伟达等现有投资者，同时思科投资、英特尔资本、AMD 风投、亚马逊、Meta 等新投资者也参与其中。

ScaleAI 的核心业务为数据标注，从自动驾驶场景起家，后切入政府、电商、机器人、大模型等场景。Scale AI 通过提供各种产品解决方案来管理机器学习生命周期的每个步骤，包括数据标注、数据管理、自动化数据提取、模型评估和合成数据生成。Scale AI 的主要产品是其数据引擎，公司可用其来构建和训练算法。数据引擎收集、筛选和标注数据以训练和评估模型。公司提供的产品和服务包括 Scale Rapid、Scale Studio、Scale Instant ML 等，涵盖了数据标注、内容理解、模型管理和大规模机器学习等领域。除了提供高质量的数据标注服务，Scale AI 拥有广泛的客户群，从 OpenAI、微软、Meta 等科技巨头，到谷歌的自动驾驶子公司 Waymo、通用汽车的 Cruise、丰田汽车 Toyota、Uber；机器人公司 Orchard Robotics 和 Ambi Robotics；美国政府（公共部门、美国各军种）等都是其合作伙伴。这种广泛的认可为 Scale AI 业务带来了强有力的支撑。

图 35：Scale AI 业务情况



数据来源：Scale AI，东北证券

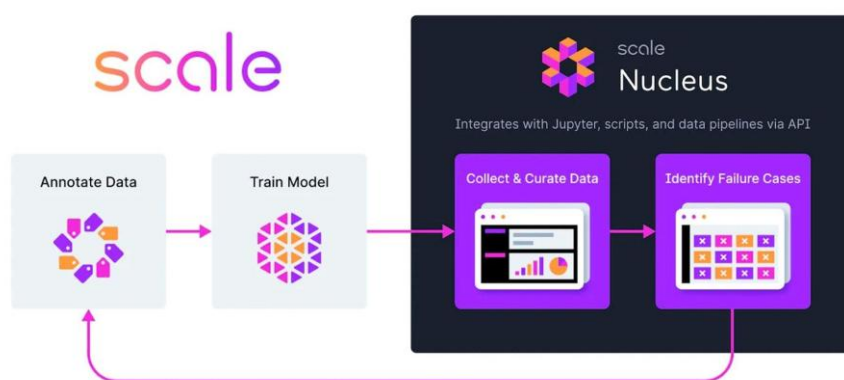
Scale AI 常规的 AI 的核心产品是其数据引擎。它依靠遍布肯尼亚、菲律宾和委内瑞拉的 24 万名员工，通过其子公司 RemoTasks 进行数据标注。企业使用数据引擎构建和训练机器学习算法。该数据引擎使客户能够收集、整理和注释数据，以训练和评估模型。包括 Lyft、丰田、Airbnb 和通用汽车在内的公司向 Scale AI 付费，以获得由人工承包商、机器学习算法或两者结合标注的高质量标注数据。Scale AI 作为数据引擎产品的一部分，提供了三种不同的数据标注解决方案，分别针对不同的需求：Scale Rapid、Scale Studio 和 Scale Pro。

- **Scale Rapid：**是一款专为机器学习团队设计的产品，专注于快速开发生产级质量的训练数据。用户可以在短短几小时内上传数据、设置标注指令，并获得初步标注的反馈与校准，从而实现数据标注流程的快速扩展，满足大规模需求。作为一款无最低限额的自助服务平台，用户能够灵活上传数据、选择或创建标注用例、向 Scale 团队分配任务，并在数小时内获取高质量的标注数据，使其成为快速项目周转的理想选择。Scale AI 提供专业的标注人员，确保数据标注的准确性和效率。

- **Scale Studio:** 的目标是最大化用户自有标注团队的工作效率。它允许用户上传数据、选择或创建标注用例，并使用自己的团队来执行标注任务，同时监控项目绩效。该产品主要面向希望内部管理标注流程并提升生产力的组织。Scale Studio 提供了用于跟踪和可视化标注人员指标的工具，以及机器学习辅助标注功能，以加速标注速度。它能够跟踪吞吐量、效率和准确性等关键指标。
- **Scale Pro:** 专为需要可扩展、高质量、支持复杂数据格式的 AI 企业量身定制。它具备 API 集成功能，配备专业的项目经理，支持定制化的项目设置，能够处理海量生产数据，并通过服务级别协议（SLA）确保最高质量，为用户提供优质、全面托管的标注体验。

Scale Studio/Pro 与 Scale Rapid 的主要区别** 在于数据标注的执行方式。Scale Rapid 的数据标注由 Scale AI 完成，而 Scale Studio 和 Scale Pro 则要求企业自带标注人员。不过，这三种产品均属于 Scale Data Engine 的产品体系。

图 36: Scale AI 的 Nucleus 产品

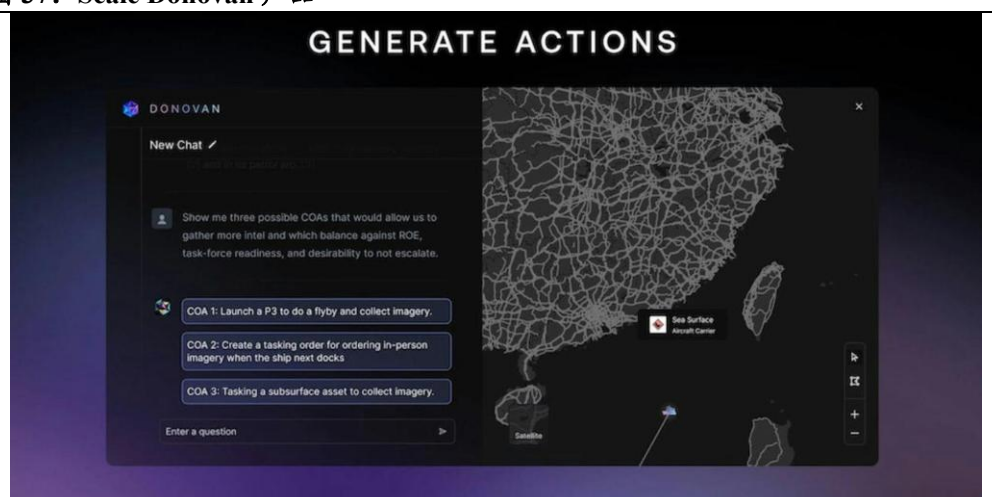


数据来源：Scale AI，东北证券

Scale Donovan 是联邦政府推出的人工智能工具套件。它能够从云端、混合云以及本地数据源中采集数据，并对数据进行组织，使其具备交互性。Donovan 支持操作员和分析师向传感器数据以及地图和模型数据提出问题，帮助他们快速获取所需信息。此外，Donovan 还能够生成行动方案、总结报告以及其他可操作的洞察，助力操作员实现任务目标。

Donovan 集成了检索增强生成（RAG）技术，允许用户通过大型语言模型（LLM）与任务相关的信息进行交互。它还配备了一个聊天界面，能够从文档中提取信息，并利用自然语言语义进行翻译。由于用户可以根据需求选择任何合适的模型，因此能够轻松找到最适合其特定用例的模型选项。

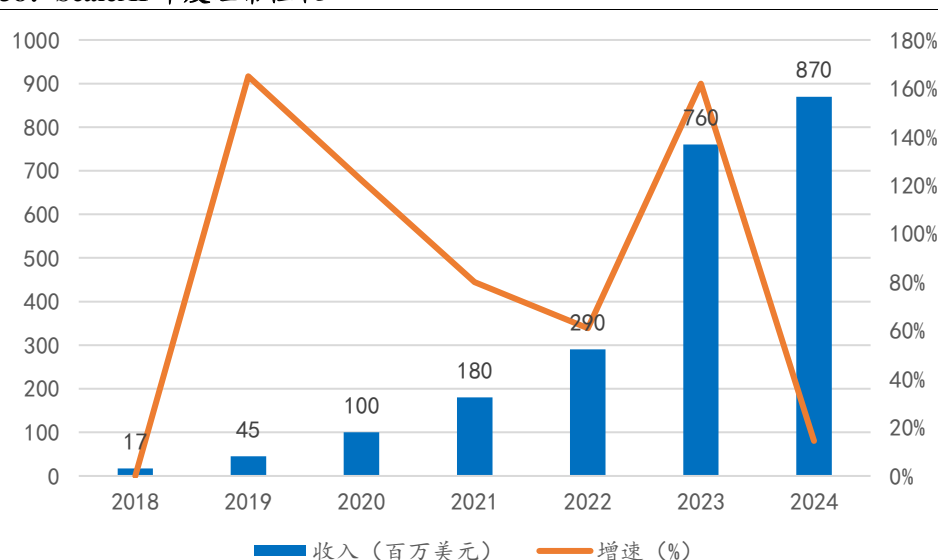
图 37: Scale Donovan 产品



数据来源: Scale AI, 东北证券

在 AI 浪潮的需求下, ScaleAI 收入快速增长, 估值持续提升。2023 年, Scale AI 的年收入达到 7.5 亿美元, 净利润达到 8000 万美元。在定价方面, Scale AI 官网为每款产品提供了标准化的按需定价模式。例如, Scale Image 的起价是每张图片 2 美分, 每条标注 6 美分; Scale Video 的起价是每帧视频 13 美分, 每条标注 3 美分; Scale Text 的起价是每项任务 5 美分, 每条标注 3 美分; Scale Document AI 的起价是每项任务 2 美分, 每条标注 7 美分。此外, 还有针对企业级项目的定制收费方式。2024 年, Scale AI 预计约 8.7 亿美元的收入。

图 38: ScaleAI 年度经常性收入



数据来源: ScaleAI, 东北证券

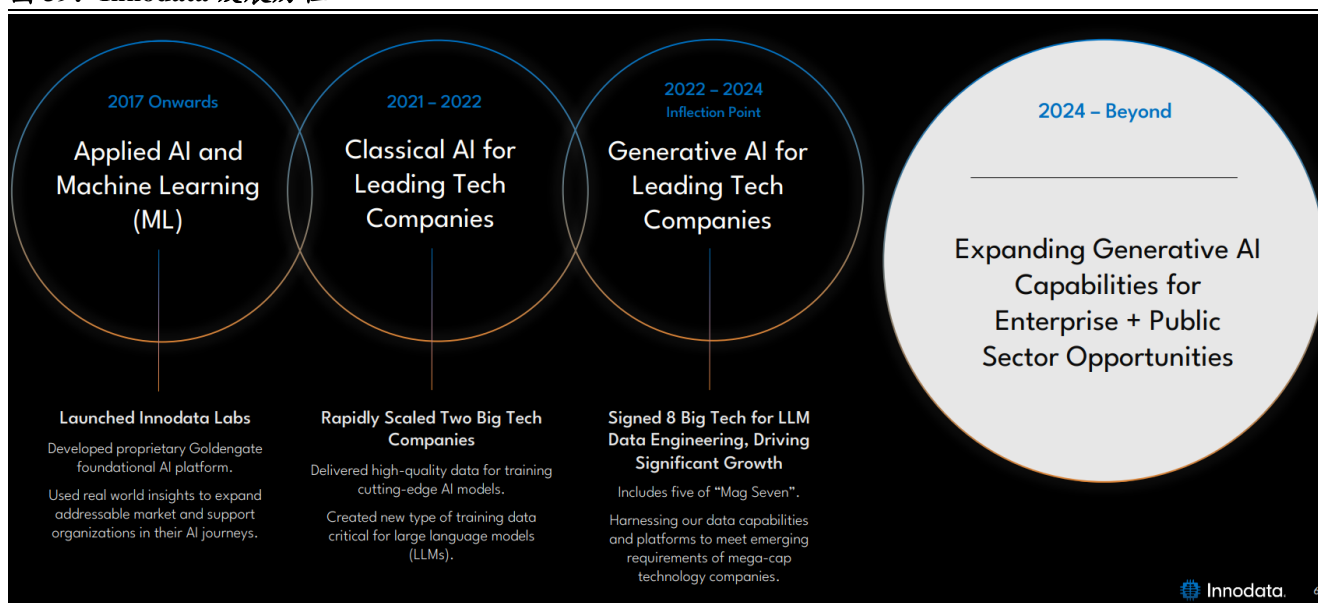
Scale AI 的核心竞争力可以归结为: 自动化、持续迭代。一方面, 通过标注大量数据来训练 AI 算法, 进而提高效率并降低成本。另一方面, 一旦行业有新趋势, Scale 都能很快捕捉到信号, 并快速推出相应的数据标注产品。Scale AI 致力于通过自动化系统完成大部分的标记和识别工作, 并推出了"Scale Rapid"快速数据标注服务, 其自动化质量标注工具可将效率提高 10 倍。公司的标注人员主要来自委内瑞拉、肯尼亚、菲律宾等劳动力成本较低的国家, 而客户主要是美国的高端企业, 这种商业模式实现了全球化套利, 具有较高的毛利率。我们认为这也是未来的发

展方向。2024 年 5 月，Scale AI 宣布获得 10 亿美元融资，估值为 138 亿美元（约为人民币 1000 亿元）。

3.2. InnoData：帮助企业构筑更好的大模型

Innodata 是一家成立于 1988 年的美国数据工程公司，该公司专注于通过人工智能与专业人力结合。公司为全球客户解决数据工程难题，业务涵盖数据标注、训练数据供应等领域，服务于机器学习、深度学习及计算机视觉等应用。谷歌、亚马逊、微软等多家全球知名科技企业都是其客户。2024 年，Innodata 营收表现亮眼，全年收入达 1.705 亿美元，同比大幅增长 96%，第四季度收入 5920 万美元，同比增长 127%。Innodata 在数据服务领域的市场拓展及与头部科技企业的合作，使其在 AI 数据产业链中占据一定位置。

图 39：Innodata 发展历程



数据来源：Innodata，东北证券

通过“数据工程+垂直行业”的战略，公司在 AI 基础设施层建立竞争优势，未来增长取决于跨行业复制能力与大型客户深度绑定。

为基础模型训练客户提供基础数据：帮助基础模型商处理他们需要的大规模数据，解决非法、幻觉等不良信息，让客户能够快速、可靠且自信地在多个主题领域和任务中部署基础模型。公司的基础模型能够成功检测和去除有害内容，涉及语言包括英语、西班牙语、法语、德语、意大利语、印地语、日语、中文和阿拉伯语，拒绝不适当的用户输入，并过滤模型输出。

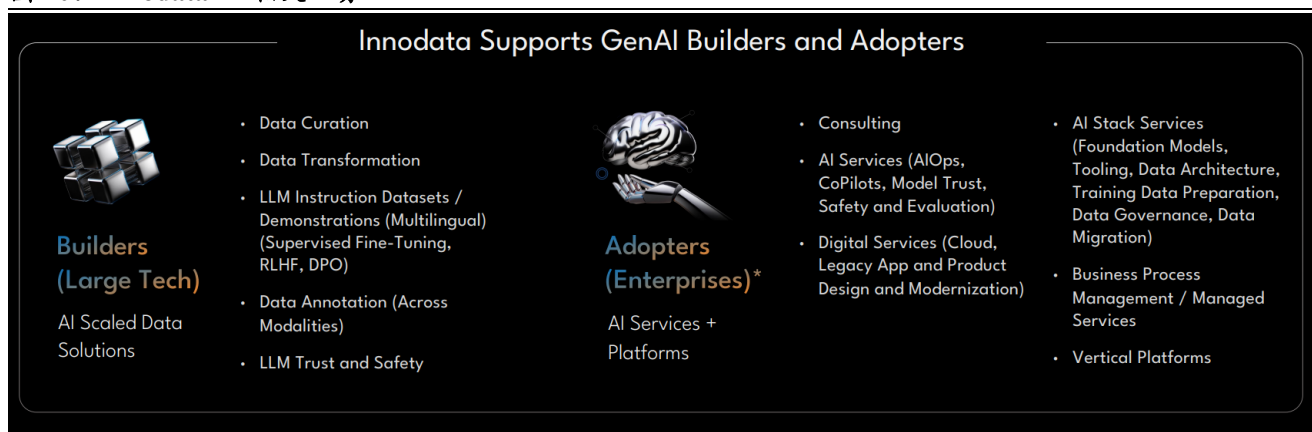
为世界模型客户提供监督微调：大语言模型用原始互联网内容进行预训练后，需用示范数据进行有监督的微调，公司能提供大量高质量、多样化的数据集，涵盖广泛的通用模型，以及为特定用例和领域（包括写作、数学、编程、法律、金融、银行和医疗等）构建的模型，涉及超 10 种语言。这些数据集用于在文本生成、文本摘要、问答、图像生成、代码生成、实体提取和检索增强生成（RAG）编排等特定任务上训练大语言模型。

帮助行业用户训练需要的模型：Innodata 提供咨询和实施服务。首先开展愿景研讨会，向客户进行知识普及并确定业务机会点。与关键利益相关者紧密合作，定义和评估为期三个月的概念验证（POC）。其中一个 POC 是创建德国法院案件摘要并与内部分类法匹配；另一个是从荷兰劳动法律书籍中提取关键词。最终实施采用

微调大语言模型（LLMs），结合密度链、提示工程、微调以及带相似性匹配的向量数据库等多种技术。

帮助行业用户搭建需要的平台：金融服务机构（以及其他受到高度监管的实体）最为关键的分析职能之一，是追踪监管变化，并对内部控制和流程做出必要调整。这一过程被称为“监管变革管理”和“视野扫描”，通常需要各国大量法律 / 监管专家团队参与。Innodata 实现了对全球超 2000 个发布机构的法律、规则、法规以及公告和其他变更通知的自动化监测和汇总。公司设计了一个基于人工智能的系统，并对生成式人工智能变换器模型进行微调，以识别文档中的引用，并将这些引用与该银行包含超 1500 万条引用的内部合规义务数据库进行匹配。

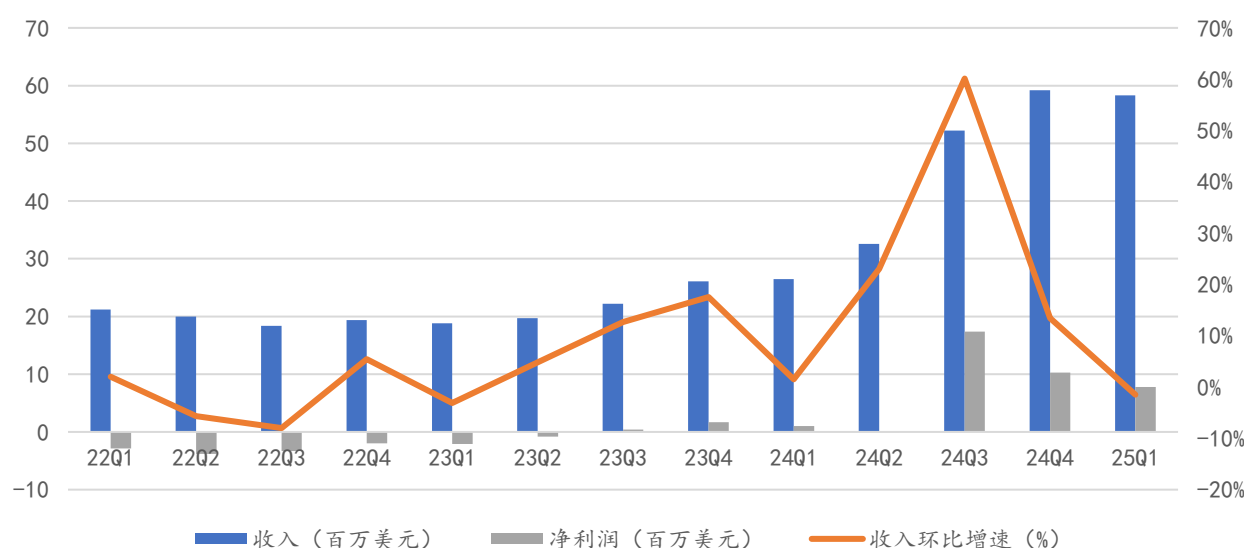
图 40：Innodata AI 相关业务



数据来源：Innodata，东北证券

受益于美国七巨头强劲的人工智能驱动资本支出，公司业绩自 24 年下半年启逐步爆发，屡创新高。公司 2024 年四季度营收为 5920 万美元，较去年同期的 2610 万美元增长 127%，2025 年第一季度收入为 5830 万美元，同比增长 120%。目前公司的客户聚焦美国大型科技企业，已覆盖“美股七巨头”中的五家。这些科技巨头围绕 AI 基建的资本支出扩张，为 Innodata 带来了大量业务。例如，其最大客户授予该公司价值约 2400 万美元的额外合同，使来自该客户的总年化运营收入达到约 1.35 亿美元。除最大客户外，另外七家大型科技公司客户的收入在 2024 年第四季度环比增长了 159%。

图 41：Innodata 近期业绩情况



数据来源：Innodata，东北证券

4. 盈利预测与投资评级

我们认为，新一代数据标注成为是大模型高质量发展的重要一环，行业壁垒正在逐步提升。在大模型时代，AI 开发以数据为核心，其发展速度迅猛。数据的多寡与质量直接关系到大模型的智能水平。对于大模型的数据标注，包括预训练数据、指令微调、提示工程和人类反馈强化学习等任务，这些任务不仅技术含量高，知识密集，而且具有很高的价值。数据标注工作需要更多的专业知识和技能，如领域知识、数据理解和分析能力等，数据标注从业者将出现高学历背景和多学科融合的特点。不再是简单的劳动密集型工作。

随着全球 AI 需求持续提升，整体数据供应需求上行。目前全球领先的大模型数据标注企业主要分布在北美，突出的特点是技术驱动导向，数据标注服务供给能力和质量较高，如 Scale AI 估值达 130 亿美金,2023 年营收超过 7.6 亿美元（ARR）。Scale AI 的竞争优势来自标注算法迭代后的规模经济效益，既有效率又有质量。海天瑞声也在不断完善算法技术，提高人机交互的数据处理效率及实现规模化效应，并将建立一个海外技术研发体系，紧跟全球 AI 的发展动态，积极开发与海外新兴技术相适应的 AI 数据解决方案，以不断增强公司在国际市场的竞争力

国家级数据标注基地的建设将显著提升数据标注行业的地位和影响力，也还将带动行业的快速发展。当前，大模型训练中高质量的中文语料资源相对不足。为了确保我国人工智能技术的可控性和可持续发展，必须构建一个有利于中文语境表达和产业应用的大模型预训练语料库。作为国家级的项目，政策上的大力支持和资源上的优先配置，将为数据标注行业的发展提供有力保障。同时，这也将引发社会对数据标注行业的更多关注和认可，提高行业的整体形象和地位，同时也会释放出大量的相关订单。

综上所述，预计公司 2025-2027 年营业收入分别为 3.33/4.66/5.43 亿元，归母净利润分别为 0.27/0.46/0.61 亿元。公司所处赛道目前处于赛道正在大爆发早期阶段，应更关注公司收入及订单的增长而非利润，维持“买入”评级。

表 3：公司 2024-2026 年收入、成本预测

总计	2024	2025E	2026E	2027E
收入（亿元）	2.37	3.33	4.66	5.43
成本（亿元）	0.8	1.22	1.71	1.96
毛利率(%)	66.24%	63.35%	63.35%	64.02%
增速(%)	39.41%	40.38%	40.00%	16.68%
国内				
收入（亿元）	1.23	1.85	2.58	2.84
成本（亿元）	0.92	0.85	1.19	1.31
毛利率(%)	54.30%	54.00%	54.00%	54.00%
增速(%)	11.82%	50.00%	40.00%	10.00%
海外				
收入（亿元）	1.14	1.48	2.07	2.59
成本（亿元）	0.23	0.37	0.52	0.65
毛利率(%)	79.66%	75.00%	75.00%	75.00%
增速(%)	90.00%	30.00%	40.00%	25.00%

数据来源：东北证券

5. 风险提示

地缘政治风险：目前，AI 行业本身的逻辑更多基于技术发展逻辑，而仍未大规模地受到政治逻辑的影响。但是 2024 年的政治风险也将给公司的海外业务带来不可小视的影响，尤其是全球大选年所导致的政局的变动可能会导致公司业务拓展不及预期。

海外业务拓展不及预期：由于 AI 大模型市场类别繁多，各个国家以及区域之间存在一定文化和购买习惯上的差异，因此可能存在部分区域业务拓展失败的风险。

市场竞争加剧：目前来看，在市场快速增长的吸引力下，市场的竞争愈发激烈，再加上不断升级的人力成本或将导致公司成本居高不下，盈利压力与日俱增等风险。

估值模型与盈利预测不及预期：估值模型中由于存在多个假设，因此存在一定的不确定性，比如可能会因为公司在获取和管理内部资源方面采取的独特管理举措以及更加激进的销售策略等因素所产生的成本变动导致估值模型与盈利预测不及预期。

附表：财务报表预测摘要及指标

资产负债表（百万元）	2024A	2025E	2026E	2027E
货币资金	56	59	41	30
交易性金融资产	272	282	287	292
应收款项	86	90	106	108
存货	23	21	33	40
其他流动资产	3	3	3	3
流动资产合计	447	474	495	499
可供出售金融资产				
长期投资净额	0	0	0	0
固定资产	256	257	257	256
无形资产	8	8	8	8
商誉	0	0	0	0
非流动资产合计	362	364	364	364
资产总计	808	838	860	864
短期借款	0	30	50	80
应付款项	23	37	57	71
预收款项	0	0	0	0
一年内到期的非流动负债	3	3	3	3
流动负债合计	63	115	175	230
长期借款	0	0	0	0
其他长期负债	2	2	2	2
长期负债合计	2	2	2	2
负债合计	65	117	177	232
归属于母公司股东权益合计	743	721	682	630
少数股东权益	0	0	1	1
负债和股东权益总计	808	838	860	864

利润表（百万元）	2024A	2025E	2026E	2027E
营业收入	237	333	466	543
营业成本	80	122	171	196
营业税金及附加	4	5	8	9
资产减值损失	0	0	0	0
销售费用	27	40	57	65
管理费用	62	87	114	130
财务费用	0	-1	0	1
公允价值变动净收益	1	0	0	0
投资净收益	7	23	33	38
营业利润	11	28	47	62
营业外收支净额	0	0	0	0
利润总额	11	28	47	62
所得税	-1	0	1	0
净利润	11	28	46	62
归属于母公司净利润	11	27	46	61
少数股东损益	0	0	0	1

现金流量表（百万元）	2024A	2025E	2026E	2027E
净利润	11	28	46	62
资产减值准备	2	2	8	10
折旧及摊销	20	0	0	0
公允价值变动损失	-1	0	0	0
财务费用	1	0	1	1
投资损失	-7	-23	-33	-38
运营资本变动	4	6	-3	5
其他	-2	-1	0	0
经营活动净现金流量	29	11	19	40
投资活动净现金流量	41	12	27	33
融资活动净现金流量	-53	-20	-65	-84
企业自由现金流	115	24	46	73

财务与估值指标	2024A	2025E	2026E	2027E
每股指标				
每股收益（元）	0.19	0.45	0.76	1.01
每股净资产（元）	12.32	11.95	11.31	10.45
每股经营性现金流量（元）	0.48	0.19	0.32	0.67
成长性指标				
营业收入增长率	39.5%	40.3%	40.0%	16.7%
净利润增长率	—	140.9%	67.0%	34.2%
盈利能力指标				
毛利率	66.5%	63.4%	63.4%	64.0%
净利率	4.8%	8.2%	9.8%	11.3%
运营效率指标				
应收账款周转天数	136.51	95.65	76.08	71.13
存货周转天数	62.35	65.12	57.03	66.79
偿债能力指标				
资产负债率	8.1%	14.0%	20.6%	26.9%
流动比率	7.08	4.12	2.84	2.17
速动比率	6.60	3.78	2.52	1.90
费用率指标				
销售费用率	11.5%	12.0%	12.2%	12.0%
管理费用率	26.0%	26.0%	24.5%	24.0%
财务费用率	0.1%	-0.2%	-0.1%	0.1%
分红指标				
股息收益率	0.4%	0.9%	1.5%	2.0%
估值指标				
P/E（倍）	518.53	211.73	126.81	94.52
P/B（倍）	8.00	8.02	8.48	9.17
P/S（倍）	25.07	17.38	12.41	10.64
净资产收益率	1.5%	3.8%	6.7%	9.7%

资料来源：东北证券