

海天瑞声(688787)

报告日期: 2023年01月04日

国内 AI 训练数据龙头提供商, 自动驾驶业务打开成长空间

——海天瑞声深度报告

投资要点

- 公司是国内领先的人工智能数据标注提供商, 提供 AI 算法模型开发所需的专业数据集, 业务覆盖智能语音、计算机视觉、自然语言等场景, 并积极布局自动驾驶业务, 未来有望打开公司成长空间。公司产品服务在人机交互、智能家居、智慧城市等领域均有深度应用, 对标海外龙头 Scale AI 和 Appen, 我们认为公司有望充分受益于数据要素生态构建和人工智能产业加速发展, 并且在自动驾驶新业务加持下, 实现业务的快速增长。**
- AI 进入数据智能时代, 质量+规模拉动数据标注需求快速成长**
 人工智能行业持续成长, 模型复杂度和算力提升拉动了训练数据需求量的加速提升, 根据 Congnilytica 预测, 到 2027 年全球 AI 训练数据市场需求将达到 220 亿美元, 而观研天下数据中心预测到 2029 年, 我国数据标注行业市场规模将达到 204.3 亿元。数据规模以及质量需求升高将拉动数据标注行业的快速成长。
- 自动驾驶渗透率提升, 公司智能驾驶数据业务有望迎来爆发**
 公司积极布局自动驾驶相关业务, 于 2022 年 6 月上线第三代智能驾驶标注平台并发力研发第四代产品, 截至 2022 年第三季度, 公司已验收订单以及在手订单合计约 5000 万元, 同比增长超 200%。我们认为随着自动驾驶渗透率提升, 叠加技术向 L3 及以上级别发展, 数据标注需求将快速成长, 公司已设立独立的自动驾驶事业部并配置专项研发、销售、产品和项目交付团队, 未来相关业务有望实现高速增长。
- 持续布局语音语义赛道, 多语种为公司核心技术壁垒**
 公司基于多年语音识别及合成领域的技术积累, 在多语种领域构筑深厚技术壁垒。截至 2022H1, 公司拥有 190 个语种/方言的覆盖能力, 累积词条数超过 1000 万。公司持续布局海外市场, 2022 年 H1 境外收入占总收入达 57.14%, 境外市场签单额同比增长近 2.5 倍, 前三季度境外基础训练数据业务收入同比增长近 140%。我们认为随着公司海外业务的加速布局, 能够更好地发挥公司在语音和自然语言领域的多语种优势, 打开公司广阔成长空间。
- 盈利预测与估值**
 我们预计公司 2022-2024 年实现营业收入分别为 2.67/4.00/5.48 亿元, 同比增长分别为 29.36%/49.67%/37.00%; 对应归母净利润分别为 0.14/0.45/1.02 亿元, 同比增速分别为 -56.75%/230.52%/126.32%, 对应 EPS 为 0.32/1.06/2.39 元, 对应 PE 分别为 205.04/62.04/27.41 倍。首次覆盖予以“买入”评级。
- 风险提示**
 1、技术迭代风险; 2、汇率波动风险; 3、新客户拓展不及预期带来的风险; 4、市场竞争加剧风险;

投资评级: 买入(首次)

分析师: 程兵
 执业证书号: S1230522020002
 chengbing01@stocke.com.cn

基本数据

收盘价 ¥ 65.48
 总市值(百万元) 2,802.54
 总股本(百万股) 42.80

股票走势图



相关报告

财务摘要

(百万元)	2021A	2022E	2023E	2024E
营业收入	206.48	267.09	399.75	547.65
(+/-) (%)	-11.53%	29.36%	49.67%	37.00%
归母净利润	31.61	13.67	45.18	102.24
(+/-) (%)	-62.85%	-56.75%	230.52%	126.32%
每股收益(元)	0.74	0.32	1.06	2.39
P/E	88.67	205.04	62.04	27.41

资料来源: 浙商证券研究所

投资案件

● 我们与市场的观点的差异

(1) 市场普遍认为 AI 训练数据标注行业属于重劳动力行业，对标注人员需求量大，认为赛道内厂商的产品服务受制于标注人员的标注效率和准确度，业务天花板不高。

我们认为公司基于智能化标注平台的不断迭代和研发，人工标注比例持续下降，标注效率和准确度有望持续提升。同时，公司持续储备标准化的数据产品，坚持标准化产品销售和定制化服务相结合的商业模式，保证公司的毛利率、净利率维持在良好水平，未来有望随着新业务的持续开展，不断提升行业内的竞争力。

(2) 市场普遍认为 AI 标注行业厂商众多，考虑近年来大量初创公司涌入该赛道，公司的行业竞争力以及竞争壁垒可能受到挑战。

我们认为对于 AI 数据标注行业，业务智能化和数据安全合规将决定公司的竞争壁垒。在业务智能化方面公司通过多年研发已构建高效数据标注平台，标注能力较行业竞争对手具备显著优势。同时 2022 年自然资源部先后发布文件，针对数据安全合规提出更高要求，未取得测绘资质的企业将不得从事自动驾驶舱外数据的处理工作，而公司于 2022 年 6 月获批乙级测绘资质，为自动驾驶领域业务的开展奠定良好基础。

● 股价上涨的催化因素

高级别自动驾驶渗透率提升带动训练数据需求加速成长。随着自动驾驶技术向 L3 及更高级别发展，AI 训练模型复杂度提升叠加单车传感器数量增加，将带动训练数据需求量的爆发成长，公司发力布局智能驾驶赛道，并与国内外车企、Tier1 厂商建立良好合作关系，未来有望受益行业加速成长。

数据要素生态持续完善。2023 年 1 月 3 日发改委在《求是》撰文指出，应加快构建中国特色数据基础制度体系，促进全体人民共享数字经济发展红利。随着国家政策引导以及数据产权制度、数据流通和交易、受益分配制度等环节的持续完善，数据要素生态有望迎来加速成长期。公司作为 AI 训练数据提供商，业务设计数据采集、标注和交易等多个环节，并且是上海数交所和北数所首批数据厂商，未来有望受益行业加速成长。

● 风险提示

技术迭代风险：由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，公司业务相关的各项技术面临快速迭代更新的压力。如果公司无法及时判断行业需求的发展趋势，合理规划研发人才及研发投入，可能导致在技术迭代进程中失去技术优势以及对新需求、新业务场景的先发优势，进而对公司经营造成不利影响。

汇率波动风险：公司积极拓展境外客户，2022H1 公司境外收入占总营收 57.14%，该收入采用外币结算，并受人民币汇率水平变化的影响，因此汇率波动可能对公司境外收入造成不利影响。

新客户拓展不及预期带来的风险：2022 年上半年，公司收入主要来自老客户、占比达到 95.82%，新客户收入占比较低。若老客户降低对公司产品、服务的采购量或公司未能维持与老客户的合作关系，则公司的经营情况或将受到不利影响。

市场竞争加剧风险：近年来 AI 基础数据服务领域公司不断增多由于整体行业较新，行业标准和法律监管仍处于规范阶段。公司如果不能持续有效地制定并实施业务发展规划，将受到竞争者的挑战，影响公司的盈利能力和发展潜力。

正文目录

1 AI 进入数据智能时代，质量+规模拉动数据标注需求快速成长	6
1.1 人工智能发展，算法+算力+数据三轮驱动	6
1.2 AI 进入商业化时代，与行业和场景的紧密结合催化高质量训练数据需求	8
1.3 数据标注行业发展迅速，智能化+数据安全将决定公司竞争力	9
2 自动驾驶渗透率提升，公司智能驾驶数据业务有望迎来爆发	12
2.1 智能驾驶从 L2 向更高级别过渡，AI 数据训练需求将加速成长	12
2.2 公司智能驾驶业务快速发展，智能标注平台技术优势显著	14
2.3 对标海外厂商，公司发展前景广阔	15
2.3.1 对标 Sacle.AI，基于 AI 辅助工具大幅提升数据标注效率	15
2.3.2 对标海外龙头厂商 Appen，公司盈利能力占优前景广阔	16
3 持续布局语音语义赛道，多语种为公司核心技术壁垒	18
3.1 公司在智能语音领域积累深厚，技术优势显著	18
3.2 持续布局海外市场，标准化产品赋能盈利能力提升	19
4 海天瑞声，国内 AI 训练数据提供商龙头	20
4.1 深耕行业近二十年，成为 AI 产业核心价值提供者	20
4.2 公司结构稳定，核心团队背景奠定行业竞争优势	21
4.3 疫情影响下公司营收承压，智能化平台有望驱动公司快速成长	22
5 盈利预测与估值	23
5.1 盈利预测	23
5.2 估值与投资建议	24
6 风险提示	25

图表目录

图 1: 中国人工智能市场规模及增速.....	6
图 2: 2016-2022H1 中国算力总规模 (单位: EFlops)	7
图 3: 2021-2027 全球 AI 训练数据市场需求 (单位: 十亿美元)	8
图 4: 人工智能技术广泛渗透进经济生产活动主要环节	8
图 5: 机器学习项目中不同环节耗费时间的结构比重	9
图 6: 中国数据标注行业市场规模发展趋势	9
图 7: 2021 年我国数据标注行业下游需求占比	10
图 8: 2017-2029 年中国数据标注行业平均价格趋势及预测	10
图 9: 2021 年我国数据标注与审核细分产品市场结构	10
图 10: 2021Q1-2022Q1 中国 L2 级自动驾驶乘用车渗透率	12
图 11: 2021-2030 年全球自动驾驶 渗透率预测	12
图 12: 2D 点云标注示例	13
图 13: 3D 点云标注示例	13
图 14: 自动驾驶技术训练量需求	13
图 15: 毫末智行 4D 点云标注示例	13
图 16: 公司 3D 点云标注平台	14
图 17: 公司一体化数据处理平台示意图	14
图 18: Scale.AI 历轮融资情况	15
图 19: Scale AI 3D 点云标注平台	15
图 20: Scale AI 稠密建图工具	15
图 21: Appen 自研 MatrixGo 数据标注平台	16
图 22: Appen2018-2021 年营收总收入及净利润趋势	17
图 23: Appen、海天瑞声净利率、毛利率比较	17
图 24: 2017-2022 年中国智能语音市场规模趋势及预测	18
图 25: 语音识别加工	18
图 26: 语音合成加工	18
图 27: 2018-2022H1 公司境内及境外收入占比情况	19
图 28: 公司境内外业务毛利率对比	19
图 29: 公司发展历程	20
图 30: 公司产品服务矩阵	20
图 31: 公司研发人员数量大幅提升	21
图 32: 公司研发人员学历构成	21
图 33: 公司 2018-2022Q3 营业总收入	22
图 34: 公司 2018-2022Q3 归母净利润	22
图 35: 公司 2018-2022Q3 毛利率	22
图 36: 公司 2018-2022Q3 研发、销售、管理费用率	22
表 1: AI 神经网络模型参数量及复杂程度	7
表 2: 自然资源部测绘工作准入资质相关文件	11
表 3: 各级别自动驾驶对传感器数量的需求	12
表 4: 公司智能驾驶数据解决方案	14
表 5: MatrixGo 数据标注平台功能点	16

表 6: 公司及国内外竞品公司对比.....	19
表 7: 公司核心管理团队.....	21
表 8: 可比公司估值分析.....	24
表附录: 三大报表预测值.....	26

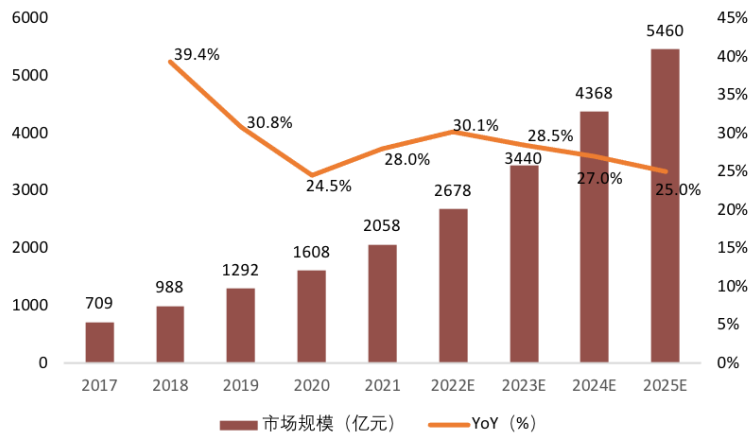
1 AI 进入数据智能时代，质量+规模拉动数据标注需求快速成长

1.1 人工智能发展，算法+算力+数据三轮驱动

移动互联网时代的到来为 AI 发展提供数据资源，结合算力和模型持续迭代，AI 呈现爆发式成长。数据爆炸、算力提升和新算法出现引导了 AI 近十年来的发展浪潮。AI 研究领域逐步扩大到专家系统、机器学习、计算机视觉、自然语言处理等，并与金融、政府、教育等行业结合愈发紧密。

AI 迭代带动商业化进程加速，应用场景日渐丰富，行业空间有望打开。德勤数据显示，2021 年中国人工智能市场规模达 2058 亿元，预计到 2025 年将达到 5460 亿元，2021-2025 年复合增长率约 27.63%。随着 AI 从感知智能逐步走向认知智能，未来有望有效赋能人的效率提升，创造更大的商业价值。

图1：中国人工智能市场规模及增速



资料来源：德勤，浙商证券研究所

AI 进入“大模型+小模型”时代，大规模参数设置与输入为 AI 解决复杂推理问题提供可行方案，为人工智能发展创造空间。大模型通过超大规模的参数设置和数据输入，具有基础、通用和综合的特点，而小模型基于实际任务场景的需求，通过高质量场景数据的输入和训练，使 AI 能够出色完成特定任务。在 NLP 领域，以 Google 发布的 BERT 模型和 Open AI 提出的 GPT 模型为典型代表。AI 通过大模型的预训练，能够在处理语言、推理、人机交互等领域任务上的表现实现极大提升，打开了 AI 渗透到各行业进行商业应用路径。

国内外大厂相继发布 AI 大模型，巨头发力布局人工智能有望主导行业发展与市场格局。2018 年以来，以 Google、Meta、微软、阿里巴巴为代表的国内外科技巨头相继发布自研 AI 大模型，为人工智能的商用奠定良好基础，未来各细分赛道公司有望从 AI 生态拥抱产业变革。

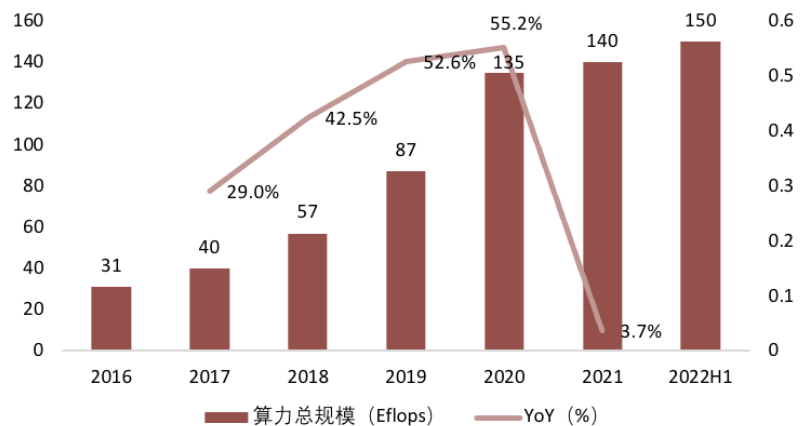
表1: AI神经网络模型参数量及复杂程度

类别	模型	公司	参数量
图像识别	AlexNet (2012)		6000 万
	VGG-16 (2014)		1.38 亿
	ResNet-50 (2015)		2500 万
自然语言模型	GPT (2018-2020)	Open AI	GPT-1 (2018): 1.5 亿 GPT-2 (2019): 15.42 亿 GPT-3 (2020): 约 1750 亿
	BERT (2018)	Google	3.4 亿
	MT-DNN (2019)	Microsoft	3.3 亿
	PERSEUS-BERT (2019)	阿里巴巴	1.1 亿
	Project Megatron (2019)	NVIDIA	83 亿
	T-NLG (2020)	Microsoft	172 亿
	源 1.0 (2021)	浪潮信息	2457 亿 (中文语言模型)

资料来源: CSDN, 浙商证券研究所

我国算力规模持续扩大, 保证 AI 大模型训练算力需求。“十三五”期间我国算力规模保持高速增长态势, 单台算力设备的性能也不断提升。前瞻产业研究院数据显示, 截至 2021 年底我国算力总规模达到 140EFlops (每秒 140 京次浮点运算次数), 占全球总算力约 31%。截至 2022H1, 我国算力总规模超过 150EFlops, 仅次于美国。

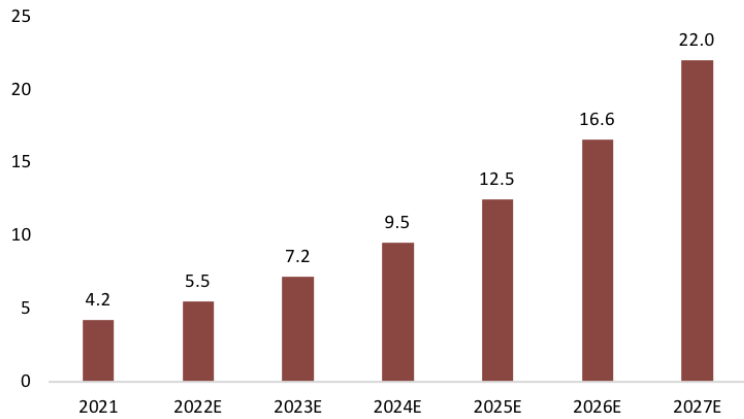
图2: 2016-2022H1 中国算力总规模 (单位: EFlops)



资料来源: 前瞻产业研究院, 浙商证券研究所

AI 模型训练需要大量高质量数据作支撑。在语音交互、智能驾驶等领域, 国内外厂商除了自有的 AI 数据团队对公司业务中产生的数据进行采集和标注, 还需要购买大量外部数据作为 AI 模型训练的养料。根据 Cognilytica 数据统计显示, 2021 年全球 AI 训练数据市场需求约为 42 亿美元, 并预计到 2027 年这一需求将增长到 220 亿美元, 2021-2027 年复合增长率达 32%。

图3： 2021-2027 全球 AI 训练数据市场需求（单位：十亿美元）



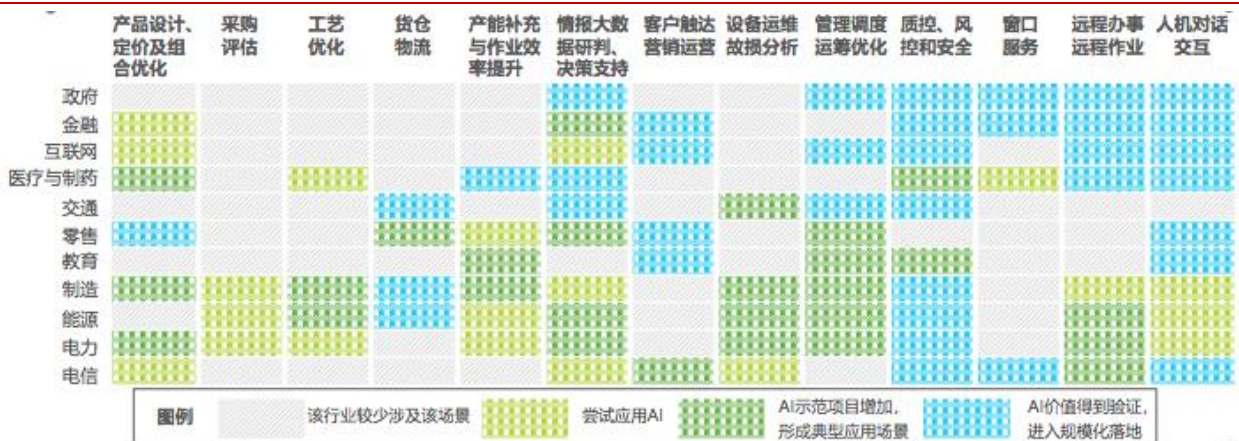
资料来源：Congnilytica, 浙商证券研究所

1.2 AI 进入商业化时代，与行业和场景的紧密结合催化高质量训练数据需求

近年来，随着新技术模型出现、各行业应用场景价值打磨与海量数据积累下的产品效果提升，人工智能应用已从消费、互联网等泛 C 端领域，向制造、能源、电力等传统行业辐射。各行业企业在设计、采购、生产、管理、营销等经济生产活动主要环节的人工智能技术与应用成熟度在不断提升，加速人工智能在各环节的落地覆盖，逐渐将其与主营业务相结合，以实现产业地位提高或经营效益优化，进一步扩大自身优势。AI 技术创新应用的大规模落地，带动了大数据智能市场的蓬勃发展，同样也为底层的数据治理服务注入了市场活力。

信息化和数字化基础较好的行业及环节将率先拥抱 AI 商业化进程。根据艾瑞咨询数据显示，目前人工智能技术已渗透进各行业的经济生产活动主要环节中，并在政府、金融、互联网等领域的质检、窗口服务、远程办公等环节进入规模化落地阶段。未来随着各垂直行业数字化信息化的持续推进，人工智能技术作为效率提升的重要手段将会加速渗透到各个环节中，作为底层模型支撑的训练数据需求将加速成长。

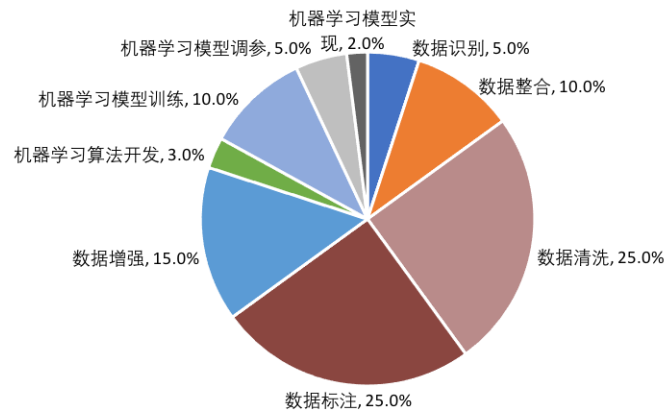
图4： 人工智能技术广泛渗透进经济生产活动主要环节



资料来源：艾瑞咨询，浙商证券研究所

人工智能模型和算法的发展带动数据标注的需求不断成长。实际业务场景中大量的数据只有在经过处理之后才能被用于部署、训练以及调试对应的机器学习或深度学习模型，这便是数据标注的价值所在，根据 AI 分析公司 Cognilytica 研究数据显示，在 AI 项目中，对数据相关的处理过程可占据超过 80% 的时间，其中数据标注环节的耗时占比可达 25%。

图5：机器学习项目中不同环节耗费时间的结构比重

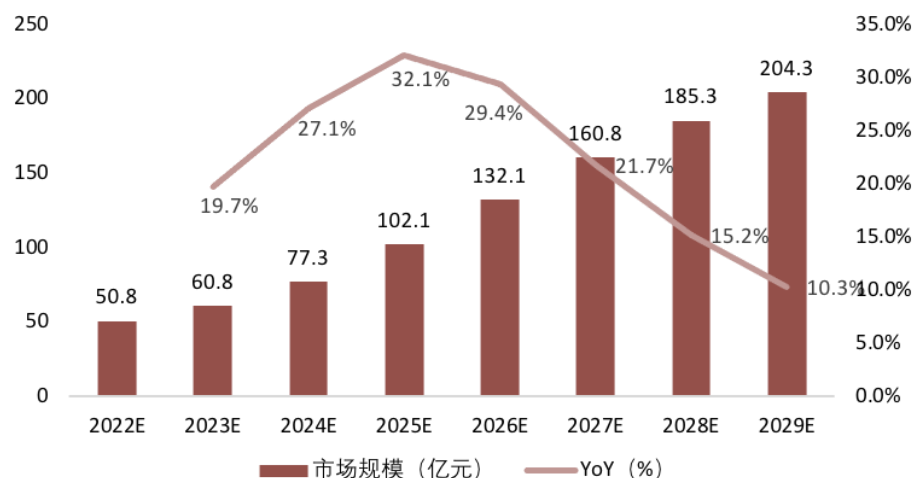


资料来源：Cognilytica，浙商证券研究所

1.3 数据标注行业发展迅速，智能化+数据安全将决定公司竞争力

AI 行业应用场景延伸及模型迭代，带来增量数据需求，AI 基础数据市场规模有望加速成长。《中国数据标注行业发展趋势与投资前景研究报告（2022-2029 年）》数据显示，2021 年我国数据标注行业市场规模达到 43.3 亿元，同比增长约 19.2%，预计到 2029 年市场规模将达到 204.3 亿元，CAGR 约为 21.4%。

图6：中国数据标注行业市场规模发展趋势

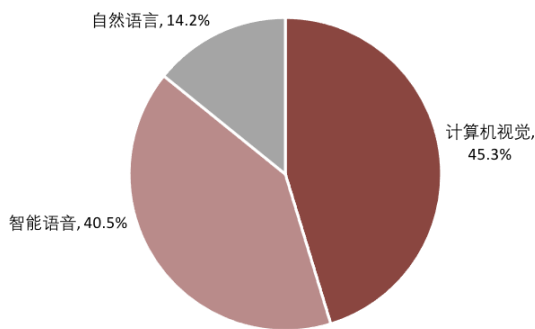


资料来源：《中国数据标注行业发展趋势与投资前景研究报告（2022-2029 年）》，浙商证券研究所

AI 训练数据标注需求以语音和计算机视觉为主，对数据质量的要求提升拉动产品服务价格不断升高。观研天下数据统计显示，2021 我国人工智能数据标注市场中，计算机视觉类、智能语音类和 NLP 类需求占比分别为 45.3%、40.5%和 14.2%，未来随着自动驾驶向 L3 及以上级别迭代，传感器数量的增加将带动数据需求量的增加，而深度人机交互的发展（虚拟人、机器人）也将拉动语音和自然语言类数据的需求量。

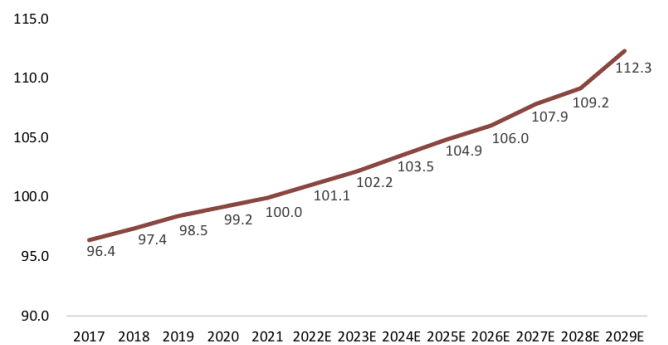
市场需求驱动 AI 训练数据质量的提升，数据价格有望持续升高。根据观研天下数据中心采用的相对价格指数表示各类数据标注产品和服务的价格水平，并以 2021 年价格为基准（100），可以发现近年来数据标注价格逐年上升，从 2017 年的 96.4 提高到 2021 年的 100，并预计在未来几年持续升高，预计到 2029 年达到 112.3。

图7： 2021 年我国数据标注行业下游需求占比



资料来源：观研天下数据中心，浙商证券研究所

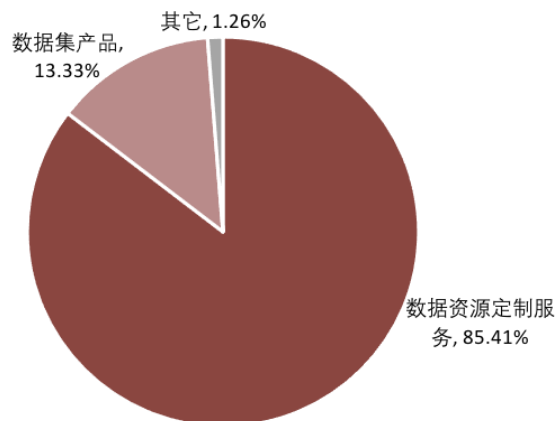
图8： 中国数据标注行业价格相对指数趋势（2021 年=100）



资料来源：观研天下数据中心，浙商证券研究所

国内数据标注行业仍以定制化服务为主，标准化产品占比较低。根据智研咨询统计数据显示，2021 年我国数据标注及审核市场中定制化服务的占比达到 85.41%，而标准化的数据集产品仅占 13.33%。我们认为随着行业的发展和 AI 商业化应用的加速，赛道内的企业能否形成智能化标注的平台能力及标准化的数据产品，将有助于提升盈利能力及行业核心竞争力。

图9： 2021 年我国数据标注与审核细分产品市场结构



资料来源：智研咨询，浙商证券研究所

数据安全合规要求提高，数据标注行业门槛提升。以智能汽车行业为例，近几年我国汽车数据处理能力增强和数据规模成长，汽车数据安全风险和隐患也日益突出。今年八月底，自然资源部发布《关于促进智能网联汽车发展维护测绘地理信息安全的通知》，对智能网联汽车的测绘主体和测绘活动的开展提出明确要求，我们认为在政策指引下，测绘资质将提升自动驾驶舱外数据采集和标注的门槛，在这方面，海天瑞声于2022年6月份获批乙级测绘资质。

表2：自然资源部测绘工作准入资质相关文件

时间	文件	内容
2022.08	《自然资源部关于促进智能网联汽车发展维护测绘地理信息安全的通知》	明确测绘地理信息数据采集和管理等相关法律法规： 第一条：“智能网联汽车安装或集成了卫星导航定位接收模块、惯性测量单元、摄像头、激光雷达等传感器后，在运行、服务和道路测试过程中对车辆及周边道路设施空间坐标、影像、点云及其属性信息等测绘地理信息数据进行采集、存储、传输和处理的行为，属于《中华人民共和国测绘法》规定的测绘活动”； 第二条：“对智能网联汽车运行、服务和道路测试过程中产生的空间坐标、影像、点云及其属性信息等测绘地理信息数据进行收集、存储、传输和处理者，是测绘活动的行为主体，应遵守相关规定并依法承担相应责任”； 第三条：“需要从事相关数据收集、存储、传输和处理的车企、服务商及智能驾驶软件提供商等，属于内资企业的，应依法取得相应测绘资质，或委托具有相应测绘资质的单位开展相应测绘活动；属于外商投资企业的，应委托具有相应测绘资质的单位开展相应测绘活动，由被委托的测绘资质单位承担收集、存储、传输和处理相关空间坐标、影像、点云及其属性信息等业务及提供地理信息服务与支持。”
2022.09	《自然资源违法行为立案查处工作规程（试行）》	明确指出，“测绘项目的招标单位让不具有相应资质等级的测绘单位中标，或者让测绘单位低于测绘成本中标的，责令改正，可以处测绘约定报酬二倍以下的罚款。招标单位的工作人员利用职务上的便利，索取他人财物，或者非法收受他人财物为他人谋取利益的，依法给予处分；构成犯罪的，依法追究刑事责任。”

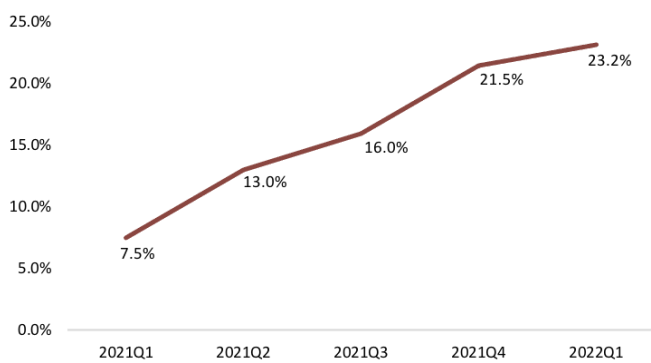
资料来源：中国政府网、自然资源部网站，浙商证券研究所

2 自动驾驶渗透率提升，公司智能驾驶数据业务有望迎来爆发

2.1 智能驾驶从 L2 向更高级别过渡，AI 数据训练需求将加速成长

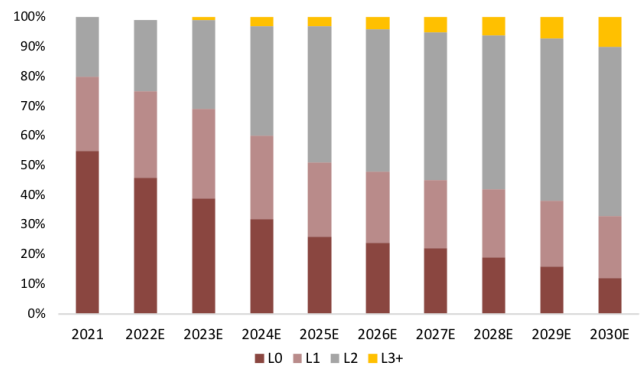
自动驾驶技术向 L3+ 发展，渗透率不断提升拉动 AI 训练数据需求。根据 IDC 数据统计显示，我国 L2 级自动驾驶乘用车渗透率从 2021 年 Q1 的 7.5 大幅提升至 2022 年 Q1 的 23.2%，未来随着自动驾驶技术向 L3+ 发展，自动驾驶渗透率将不断提升。根据华经产业研究院预测，到 2030 年全球 L2 级别自动驾驶渗透率将达到 57%，L3 及更高级别自动驾驶渗透率将达到 10%。自动驾驶技术发展将推动车辆对传感器，进而带动训练数据需求。

图10：2021Q1-2022Q1 中国 L2 级自动驾驶乘用车渗透率



资料来源：IDC，浙商证券研究所

图11：2021-2030 年全球自动驾驶渗透率预测



资料来源：华经产业研究院，浙商证券研究所

车载传感器需求不断扩大，高级别自动驾驶技术对训练数据需求将显著提高。随着 L2、L3 及以上自动驾驶渗透率不断提升，无论是视觉路线还是激光雷达路线，单车搭载传感器的数量都将显著增长。奥迪威招股说明书显示，L3 级别自动驾驶单车传感器数量将达到 17-34 颗。根据英特尔测算，L3+ 级自动驾驶每辆汽车每天产生的数据高达 4000GB，作为自动驾驶技术应用落地的必要条件，数据采集、数据标注服务已成为支撑自动驾驶热潮必不可少的一环。

表3：各级别自动驾驶对传感器数量的需求

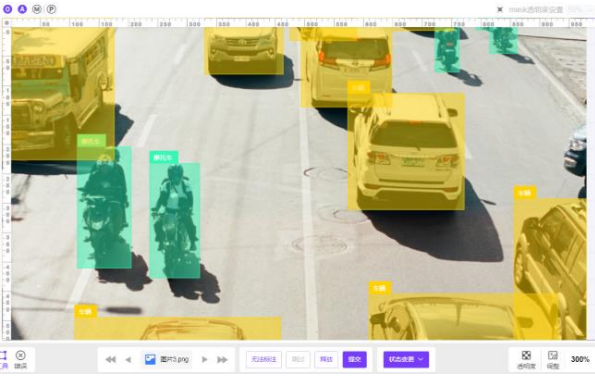
传感器类型	L0	L1	L2	L3	L4	L5
摄像头	0	1-3	3-11	3-14	3-14	3-14
毫米波传感器	0	1-3	1-3	5-7	5-7	5-7
超声波传感器	0-4	4-8	8-12	8-12	8-12	8-12
激光雷达	0	0	0	1	2	4
合计	0-4	6-14	14-26	17-34	18-35	20-37

资料来源：奥迪威招股说明书，浙商证券研究所

智能驾驶训练数据标注从 2D 点云向 3D 点云发展，激光雷达成为重要数据来源。传统的 2D 图像数据通过摄像头进行采集，涉及到的数据标注包含点标注、线标注、框标注以及语义分割。3D 标注需要实时分析大量数据，对数据标注要求更高。3D 点云数据通过激光

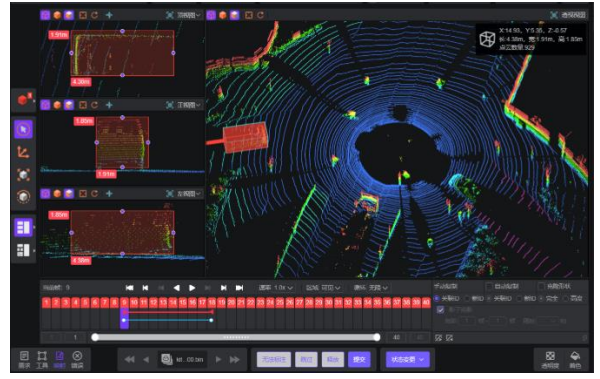
雷达进行采集，对目标进行外轮廓 3D 拉框，可精确地“描绘”出周边环境，不受外界环境、光照等因素干扰，更好的感知汽车周边道路环境，为感知系统提供决策依据。主要用于算法做 3D 目标检测以及 3D 目标跟踪，由于近年来激光雷达成本降低，3D 点云数据的量级呈现爆发式增加。

图12： 2D 点云标注示例



资料来源：CSDN，浙商证券研究所

图13： 3D 点云标注示例

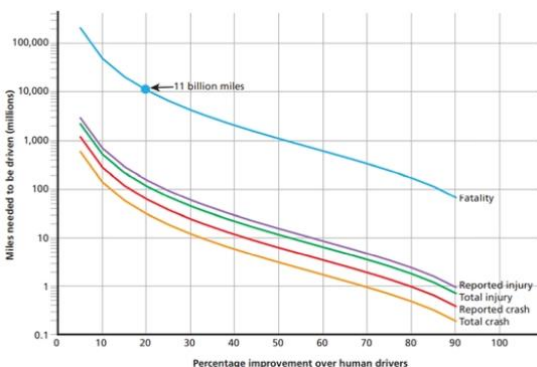


资料来源：CSDN，浙商证券研究所

高级别自动驾驶数据需求爆发式成长，驱动感知技术持续迭代，未来 4D 点云标注有望成为主流方案。根据英特尔计算，L3+级自动驾驶每辆汽车每天产生的数据高达 4000GB，而根据美国兰德公司研究，自动驾驶汽车需要在真实或者虚拟环境中至少进行 110 亿英里的里程测试，才能证明自动驾驶系统比人类驾驶员更可靠。

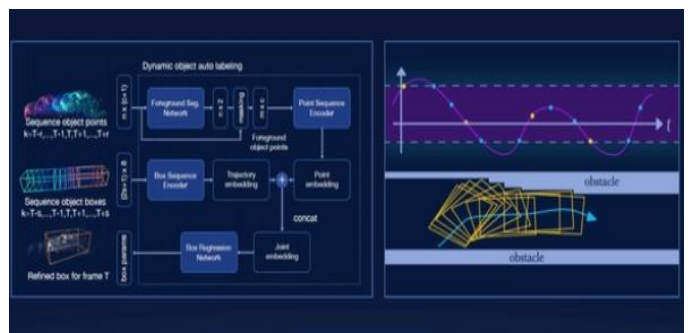
数据标注需求的成长拉动标注技术迭代，4D 点云标注技术逐渐成熟。基于 3D 点云标注的基础上，通过对感知物体进行物体的追踪与运动属性检测，并加入时序信息，由此获得 4D 点云标注信息。将当前已收集到的多维度、多角度的点云信息进行压缩，在提升现有有点云稠密度的同时，再进行物体检测可获得更精确的点云标注信息。

图14： 自动驾驶技术训练量需求



资料来源：百家号，浙商证券研究所

图15： 毫末智行 4D 点云标注示例



资料来源：毫末智行官网，浙商证券研究所

2.2 公司智能驾驶业务快速发展，智能标注平台技术优势显著

公司积极布局智能驾驶业务，产品服务矩阵持续完善。公司于2022年上半年成立智能驾驶事业部，并着重在智能驾驶数据标注平台上发力，结合业务实际需求和行业未来发展趋势，精心设计与实现了多个功能的新增和优化，在图像标注、数据管理、供应链管理等方面有大幅度提升，于6月发布了第三代智能驾驶标注平台，目前公司正在研发第四代智能驾驶标注平台，推出后有望进一步提升公司的3D功能与数据服务能力。

表4：公司智能驾驶数据解决方案

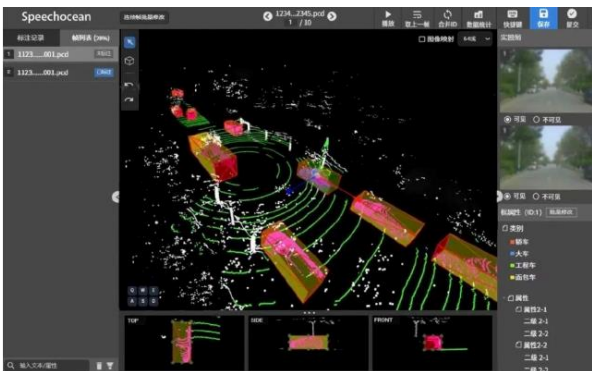
应用场景	服务模块	服务内容
智能车舱数据服务	驾乘行为检测	DMS 数据服务：对车舱内驾驶员面部表情，动作及语言进行采集和标注；驾驶员全方位监测数据处理，如是否分神、疲劳驾驶、拨打电话等；OMS 数据服务：对车舱内乘客，包括但不限于儿童、宠物的语言及行为数据进行采集标注，以用于违规行为识别或安全警示；
	车内多模交互	支持车舱内多种模态数据识别、控制和交互的数据设计及采标处理：1) 舱内不同位置、多语言、多话题内容交互的数据；2) 车内驾乘人员的手势、面部表情、唇形等交互数据；
智能驾驶数据服务	智能驾驶多场景数据设计	针对智能驾驶下的多种不同类型场景需求，例如车内多模交互、驾乘行为检测等 DMS、OMS 系统研发，不同级别 ADAS 系统研发，提供专业、覆盖样本多样性的数据设计服务；
	车外感知数据采集服务	支持智能驾驶场景下的常规道路情况数据采集，如行人、车道线、障碍物、交通标识、动态物体追踪等数据，同时包含各类场景数据如泊车、高速、园区等场景；
	车外感知数据标注服务	支持智能驾驶场景下的图片和视频的语义分割、实体分割、车道分离标注、目标追踪、障碍物等数据标注服务；
	雷达点云数据标注服务	支持智能驾驶场景下的雷达点云的语义分割、实体分割、追踪、连续帧、2D&3D 联合等数据标注服务；

资料来源：公司官网，浙商证券研究所

算法迭代赋能数据标注的效率和准确度，平台化服务打开业务规模化成长空间。公司的第三代智能驾驶标注平台，覆盖了自动驾驶场景下不同类型数据的标注工具，支持3D点云标注、3D点云连续帧标注、3D连续帧与2D联合的标注和3D语义分割等，还可以根据客户的个性化需求对平台工具进行二次开发，在行业内优势显著。

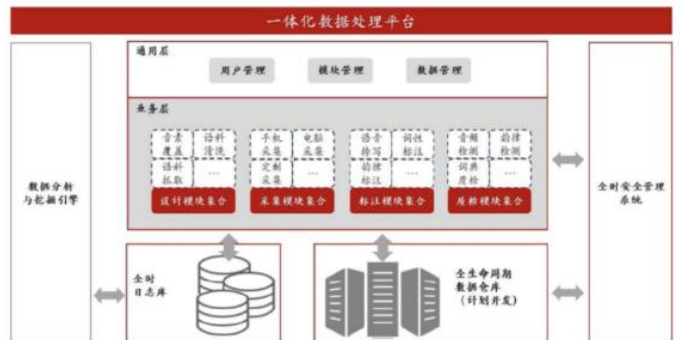
公司持续提升平台自动化、智能化水平，以人机交互为核心出发点迭代产品服务。智能标注平台在用户标注过程中会提供智能辅助以提高标注效率和准确度，例如平台包含的“辅助构建物体脑补框”工具，可自动就同一个连续帧的前后几帧内容进行预测，并给出参考。同时，一体化数据处理平台简化并统一了训练数据生产流程，模块式的项目生成和管理方式，使得生产人员可以根据实际项目需求有机结合、灵活调整，有效提升用户采集和加工数据的效率。

图16：公司3D点云标注平台



资料来源：阿里云社区，浙商证券研究所

图17：公司一体化数据处理平台示意图



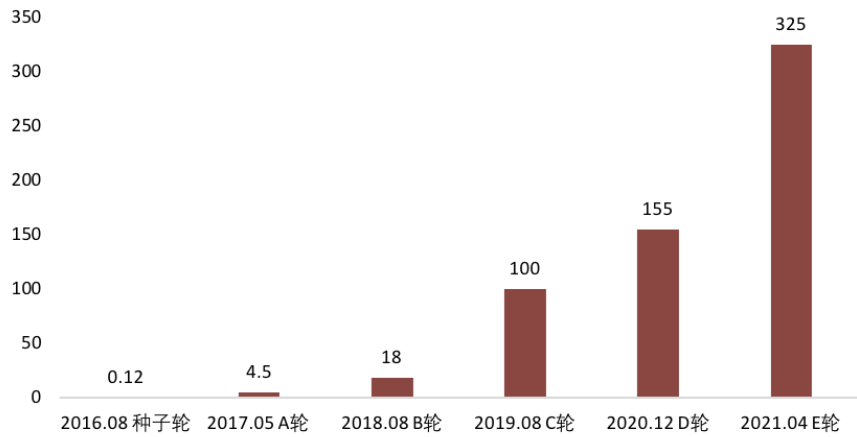
资料来源：阿里云社区，浙商证券研究所

2.3 对标海外厂商，公司发展前景广阔

2.3.1 对标 Sacle.AI，基于 AI 辅助工具大幅提升数据标注效率

Scale.AI 创立于 2016 年，先后服务于 Waymo、Toyota、Lyft 等公司，为客户公司提供用于机器学习的数据标注服务。公司于 2020 年 9 月和美国军方达成合作，拿下 9100 万美元的大单，协助美国军方开发、更新用于训练人工智能的数据集。公司成立至今已完成 7 轮融资（2021 年 8 月融资金额未披露），总计融资 6.03 亿美元，最新估值达 73 亿美元。

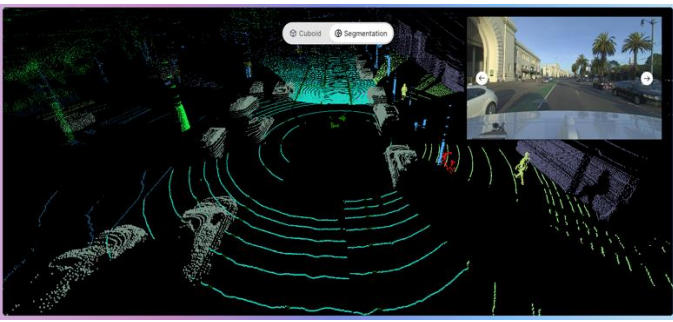
图18： Scale.AI 历轮融资情况（单位：百万美元）



资料来源：CSDN，浙商证券研究所

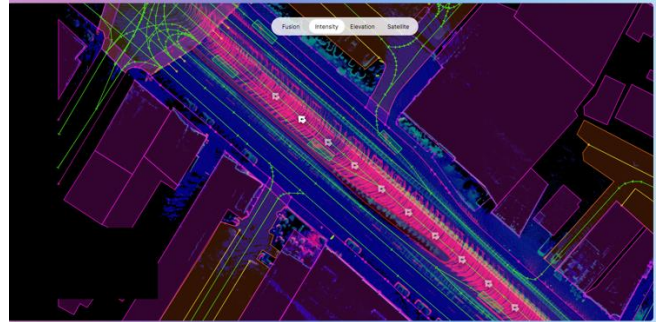
Scale.AI 借助 AI 工具，大幅提升数据标注效率，并在实际业务中不断提升标注工具的准确率和效率。Scale.AI 将 AI 应用到数据标注业务中，经过 AI 识别的任务由人工校对其中的错误，再用来训练 AI 工具，让下一次标注更精准。经过反复的实际业务训练，AI 标注工具在提升业务效率的同时，可逐步降低人工参与的比例，实现降本增效。

图19： Scale AI 3D 点云标注平台



资料来源：Scale AI 公司官网，浙商证券研究所

图20： Scale AI 稠密建图工具



资料来源：Scale AI 公司官网，浙商证券研究所

2.3.2 对标海外龙头厂商 Appen，公司盈利能力占优前景广阔

Appen（澳鹏）面向机器学习和人工智能开发提供高质量人工标注数据集。Appen 成立于 1996 年，是全球领先的图像、文本、语音、音频、视频等 AI 训练数据服务提供商，拥有业内先进的人工智能辅助数据标注平台和一体化 AI 数据及资源管理平台，支持 235 多种语言和方言，来自全球 170 多个国家和 70,000 个地区。Appen 为全球科技、汽车、金融服务、零售、医疗和政府等行业赋能，助力 AI 项目的研发及商业化。

公司通过两个运营部门提供服务：相关性（前内容相关性）部门，提供数据注释以提高搜索引擎搜索结果的相关性和准确性，包括网页、电子商务和社交参与技术。语音和影像（前语音和数据收集）部门，为计算机算法提供培训数据以开发包括语音识别器、机器翻译和语音合成器在内的产品，用于互联网连接设备、车载汽车和消费电子产品。

图21： Appen 自研 MatrixGo 数据标注平台



资料来源：美通社，浙商证券研究所

公司自研 MatrixGo 企业级 AI 数据标注平台。平台集成了丰富高效的标注工具、支持点、检测框、属性等内容的标注和自定义连线、自动顺滑及填充等多种功能，具备强大的工作流调度能力。平台可通过对自有及供应商团队及百万级众包资源进行资源整合，可支持丰富场景海量数据的各种类型标注项目。

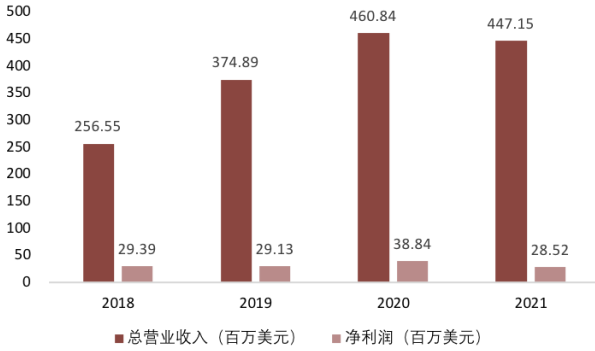
表5： MatrixGo 数据标注平台功能点

类别	细分功能	功能描述
标注内容	点标注	可分为关键点和非关键点标注，关键点即特定位置，非关键点为关键点之间沿物体轮廓均分的点。
	检测框	在关键点标注时可同时进行物体检测标注。标注时，使用检测框将每一个目标框选出来，并与点关联在同一个物体中。
	属性标注	点属性：每一个点均可设置属性，标注工具可将点属性显示为不同颜色，以方便作出区分； 物体属性：当一张图中存在多个标注目标时，可以给每一个目标设置属性。例如标注每一个人当前的动作等。
辅助功能	自定义连线	连线的定义内容包括哪些点之间需要连线以及连线颜色，通过连线快捷查看标注正确与否的目标。
	自动顺滑和自动填充	在一些关键点项目中，需要标注的点数较多（通常为人脸关键点项目）。其中，关键点为特定且可确定位置的点，比如鬓角、下巴位置等，而非关键点为关键点之间沿人脸轮廓均分的点。
	质检功能	每个标注结果都需要质检人员进行检查确认。当发现标注问题时，质检员可以对每一个点标记具体的问题；在数据被打回重做时，被标记的点会以特殊颜色显示在点的导航中。
	连续帧跟踪	Appen 关键点工具支持连续帧标注，即一次完成多帧图片的标注。在进行连续帧的标注时，可将同一个物体的前后帧关联起来，以实现跟踪的目的。

资料来源：PUDN，浙商证券研究所

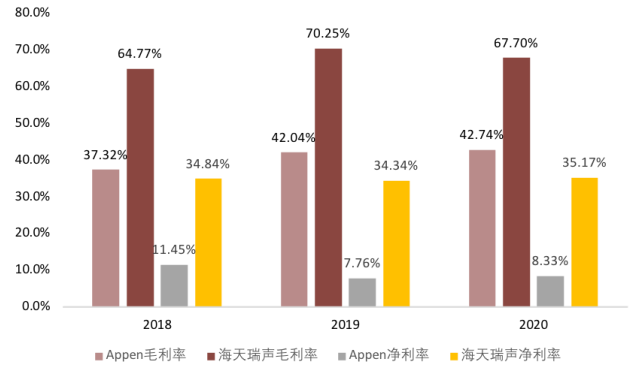
Appen 公司规模较大，营收高于海天瑞声，但毛利率、净利率指标不占优势。根据 Yahoo Finance 数据显示，Appen 在 2018 年-2020 年营业规模快速成长，2021 年受全球疫情影响，实现营业总收入 4.47 亿美元。从毛利率和净利率来看，Wind 数据统计显示，2020 年海天瑞声的毛利率、净利率分别为 67.70% 和 35.17%，而 Appen 仅为 42.74% 和 8.33%。相比于 Appen，我们认为海天瑞声在定制化的服务中不断沉淀技术经验，逐渐积累标准化产品集，使得公司盈利能力要明显优于 Appen。

图22： Appen2018-2021 年营收总收入及净利润趋势



资料来源：Yahoo Finance，浙商证券研究所

图23： Appen、海天瑞声净利率、毛利率比较



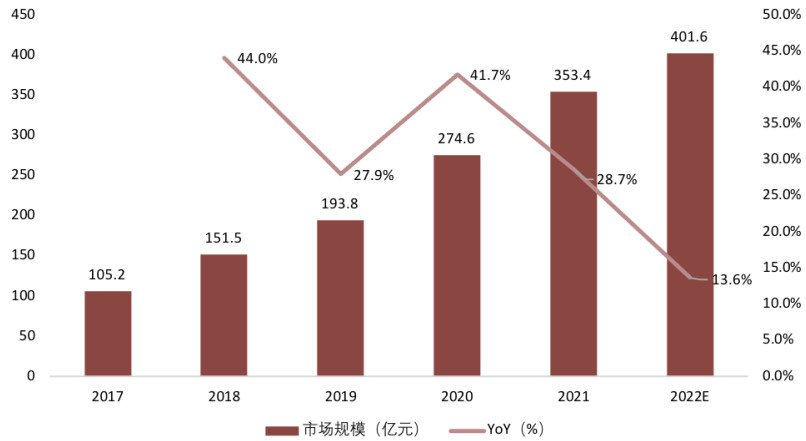
资料来源：Wind，浙商证券研究所

3 持续布局语音语义赛道，多语种为公司核心技术壁垒

3.1 公司在智能语音领域积累深厚，技术优势显著

国内智能语音市场规模持续成长。根据中商产业研究院数据显示，2022年我国智能语音市场规模预计达到401.55亿元，同比增长13.6%，2018-2022年复合增长率约27.60%，未来随着虚拟人、智能座舱、机器人等行业的发展，智能语音应用空间不断打开，将带动语音数据标注和处理的需求加速成长。

图24： 2017-2022年中国智能语音市场规模趋势及预测



资料来源：中商产业研究院，浙商证券研究所

公司智能语音服务包含语音识别、语音合成两大业务模块。公司通过设计（设计训练数据集结构、供发音人朗读录制的语料文本或对话场景、发音人分布、录音设备场景等）、采集（定义合适的发音人、选取录音设备及软件、组织发音人朗读录制音频）、加工（对音频文件进行切分、标注各类声音特征，形成带时间戳和特征标签的文本和标注文件等）、质检（对数据集进行质量检测，如音字一致性、标注准确率检查等）等训练数据集生产环节；或者针对客户提供的原料音频文件执行加工和质检工作，最终形成客户所需的智能语音训练数据。

图25： 语音识别加工

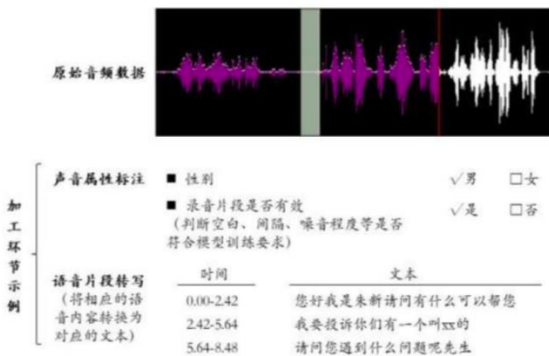
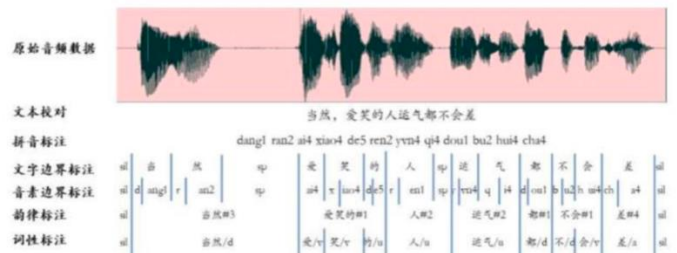


图26： 语音合成加工



资料来源：公司招股说明书，浙商证券研究所

资料来源：公司招股说明书，浙商证券研究所

公司基于多年语音识别及合成领域的技术积累，在多语种领域构筑深厚技术壁垒。公司在语音语言学研究领域储备了丰富和专家团队资源，持续研究创新发音词典构建技术以及流程技术。公司半年报显示，截至2022年上半年，公司拥有190个语种/方言的覆盖能力，除英语、法语、德语、西班牙语等常见语种外，在希伯来语、阿姆哈拉语、乌尔都语等小语种也以实现覆盖，公司累积词条数超过1000万，在高质量智能语音训练数据领域建立了深厚的技术壁垒。

表6：公司及国内外竞品公司对比

	海天瑞声	Appen	慧听科技	标贝科技
应用领域覆盖	智能语言、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言、音乐	智能语音、计算机视觉、自然语言、音乐
语言/方言覆盖能力	190个	235+个	20+个	10+个
已取得的专利授权	28项	2项	2项	6项
拥有的成品训练数据集数量	1054个	250+个	45个	192个
计算机软件著作权	159项	未公开披露	14项	30项

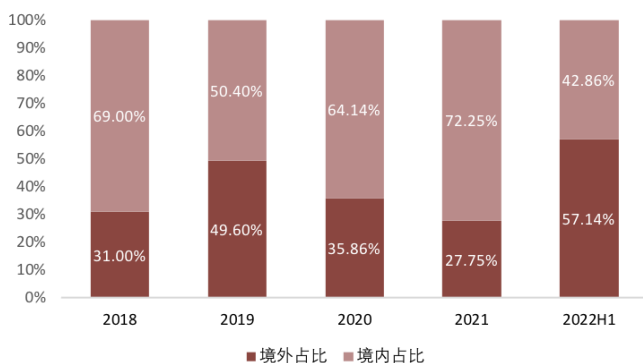
资料来源：公司半年报，浙商证券研究所

3.2 持续布局海外市场，标准化产品赋能盈利能力提升

公司持续布局海外市场，2022年境外签单数量大幅提升。近年来公司持续布局海外市场，2020-2021年受海外疫情影响，公司境外业务收入占比有所回落，分别为35.86%和27.75%。2022年以来，公司从品牌、团队、营销等多维度发力布局海外，公司公告显示，2022年上半年境外收入占总收入达57.14%，境外市场签单额同比增长近2.5倍，前三季度境外基础训练数据业务收入同比增长近140%。

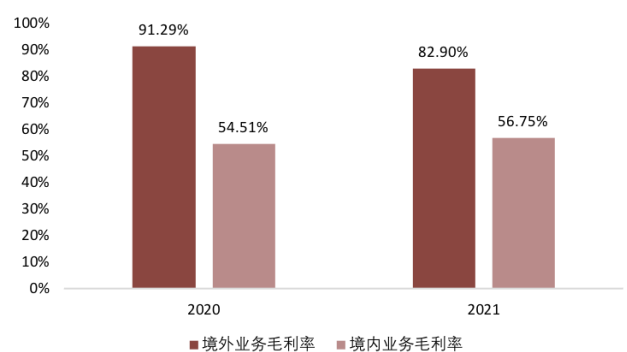
公司境外收入以标准化产品集为主，毛利率明显高于境内业务。2021年公司境外业务的毛利率达到82.90%，明显高于境内业务的56.75%。公司境外业务中标准化的数据产品销售占比较高，并且境外客户为定制化服务的付费意愿也更高。我们认为随着公司海外业务的加速布局，能够更好地发挥公司在语音和自然语言领域的多语种优势，打开公司广阔成长空间。

图27：2018-2022H1 公司境内及境外收入占比情况



资料来源：wind，浙商证券研究所

图28：公司境内外业务毛利率对比



资料来源：公司年报，浙商证券研究所

4 海天瑞声，国内 AI 训练数据提供商龙头

4.1 深耕行业近二十年，成为 AI 产业核心价值提供者

深耕 AI 数据标注行业近 20 年，发展成为全球 AI 训练数据优质提供商。公司成立于 2005 年，致力于为 AI 产业链上各类厂商提供算法模型训练所需要的高质量专业数据集及数据标注服务。公司业务涵盖智能语音、计算机视觉、自然语言理解等领域，形成了智能驾驶、智能家居、智慧金融、智能安防等多业务场景解决方案体系。截至 2022 年上半年，公司已累计为 740 家客户提供产品服务，并持续完善业务体系，不断增强综合数据服务能力竞争优势。

图29： 公司发展历程



资料来源：公司官网，公司财报，浙商证券研究所

公司产品服务覆盖 AI 多个核心领域，实现标准化产品、定制化服务、相关应用服务全覆盖。公司所提供的训练数据服务已获得字节跳动、阿里巴巴、腾讯、百度、科大讯飞、海康威视、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、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。

图30： 公司产品服务矩阵



资料来源：公司财报，浙商证券研究所

4.2 公司结构稳定，核心团队背景奠定行业竞争优势

公司核心团队行业经验丰富。公司创始人贺琳女士曾任职于中国科学院声学研究所，从事语音识别、语音合成、汉语语言理解、语音心理测试等方面的研究工作，公司核心管理团队拥有多年人工智能科研、从业经验，未来有望引领公司不断成长。

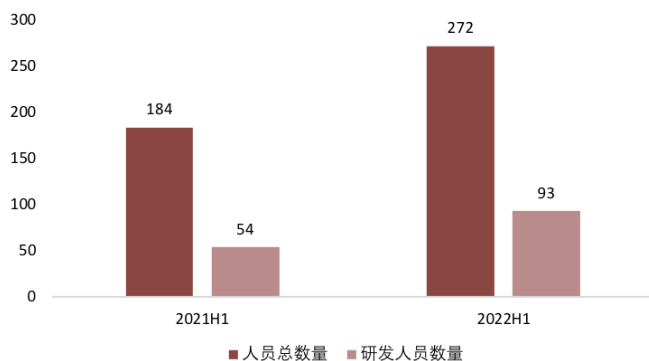
表7：公司核心管理团队

姓名	职务	任职日期	经历
贺琳	总经理	2021.10	毕业于北京大学，曾任职于中国科学院声学研究所，从事语音识别，语音合成，汉语语言理解，语音心理测试等方面的研究工作，先后参与了“汉语人-机语音对话系统工程”，“汉语语音参数合成”等国家自然科学基金重点项目，国家八六三计划智能计算机主题项目和中国科学院“八五”重大项目，完成《计算机汉语口语输出系统的设计与实现》等论文十余篇，参与编写《汉语语音合成-原理和技术》著作。公司创始人，现任公司董事长。
李科	副总经理	2017.11	硕士，毕业于清华大学，获得信息与信号处理专业硕士学位。曾任 IBM 中国技术开发中心软件工程师。
吕思遥	副总经理、财务负责人、董事会秘书	2018.1	硕士，毕业于北京大学光华管理学院，获得会计专业学士学位，中国人民大学财政与金融学院金融学硕士学位，中国注册会计师，国际注册内审师。曾任普华永道中天会计师事务所审计师，高级审计师，经理；英特尔（中国）内部控制经理；英特尔（中国）信用管理高级经理。
黄宇凯	副总经理、技术总监	2020.1	硕士，毕业于复旦大学，获得计算机科学与技术专业硕士学位。曾任微软（中国）有限公司软件研发工程师，必应网页搜索部门研发主管，地图中国团队首席研发总监等职务。
郝玉峰	副总经理、技术总监、首席科学家	2020.1	博士，毕业于东南大学，获得生物医学工程专业博士学位。曾任北京捷通华声语音技术有限公司技术研究部经理；杭州红杉树信息技术有限公司高级研发工程师。

资料来源：Wind，浙商证券研究所

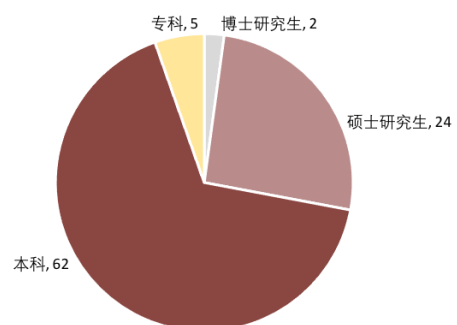
公司规模持续成长，核心技术团队加速扩充，未来有望助力业务规模拓展。公司在研发（算法、基础研究）、技术等重要岗位持续引进高级人才，优化人员结构，截至2022H1，公司人员数量达272人，同比增长47.83%，其中研发人员93人，同比增加72.22%，研发人员占比超过三分之一，公司研发人员中本科及以上占比达94.6%。未来随着业务规模成长，公司将持续扩充研发团队规模，提升技术实力，保持研发领先优势。

图31：公司研发人员数量大幅提升



资料来源：公司报告，浙商证券研究所

图32：截至2022H1公司研发人员学历构成

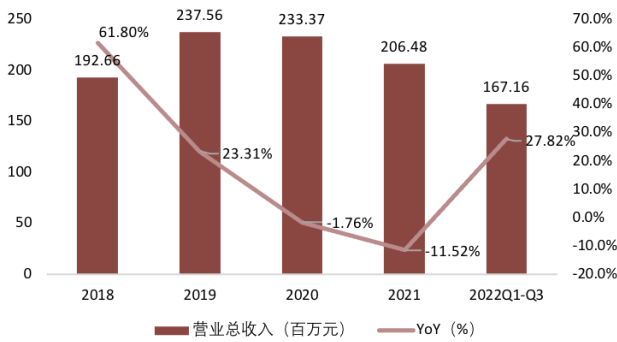


资料来源：公司半年度报告，浙商证券研究所

4.3 疫情影响下公司营收承压，智能化平台有望驱动公司快速成长

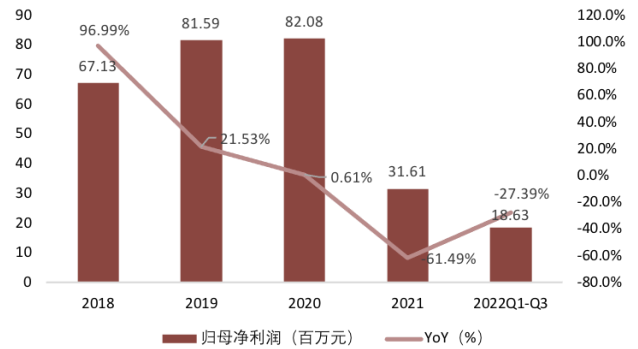
受国内外疫情影响，公司 2020-2021 年营收承压，未来有望随着疫情好转步入快速成长阶段。公司受国内外疫情影响，2020-2021 年实现营业总收入 2.33 和 2.06 亿元，同比减少 1.76% 和 11.52%，2022 年前三季度，公司在疫情反复的情况下拓展智能驾驶业务，营业总收入实现 1.67 亿元，同比增长 27.82%。公司 2020-2021 年实现归母净利润 0.82 和 0.32 亿元，同比减少 0.61% 和 61.49%，2022 年前三季度归母净利润 0.19 亿元，公司加大研发投入力度导致利润显著承压。

图33： 公司 2018-2022Q3 营业总收入



资料来源：wind，浙商证券研究所

图34： 公司 2018-2022Q3 归母净利润

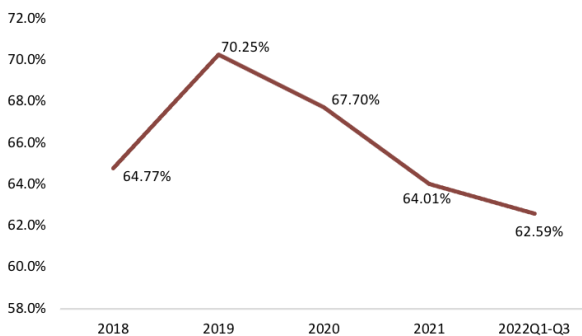


资料来源：wind，浙商证券研究所

公司毛利率短期承压，未来随着标准化产品积累有望回升。公司 2022 年前三季度综合毛利率为 62.59%，较前几年有所下降，但维持的 60% 以上的高位，主要系行业竞争态势加剧，以及公司承接部分战略性项目所致。我们认为随着公司技术不断积累和自有知识产权的训练数据集产品持续储备，以及海外高毛利收入项目的回升，公司业务盈利能力有望实现加速成长。

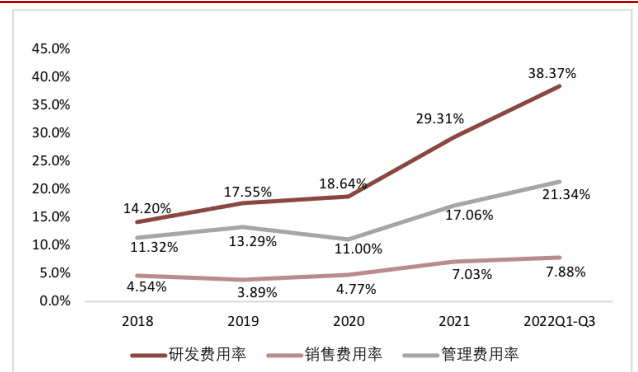
公司持续加大研发投入，销售、管理费用率稳中有升。公司 2022 年前三季度研发费用率达到 38.37%，较 2021 年有较大提升。公司在 2022 年持续加大研发投入和团队扩充，完成第三代智能驾驶标注平台并上线，同时布局第四代智能驾驶标注平台的研发工作，持续储备智能语音、计算机视觉等业务的训练数据集，未来随着相关业务规模的扩张，有望进一步释放公司技术产品储备的商业价值。公司 2022 年前三季度销售费用率和管理费用率分别为 7.88% 和 21.34%，较 2021 年小幅上升。

图35： 公司 2018-2022Q3 毛利率



资料来源：wind，浙商证券研究所

图36： 公司 2018-2022Q3 研发、销售、管理费用率



资料来源：wind，浙商证券研究所

5 盈利预测与估值

5.1 盈利预测

- **营业收入：**公司是国内 AI 训练数据优质提供商，主营业务覆盖智能语音、计算机视觉、自然语言（NLP）以及智能驾驶等核心领域，全面服务于人机交互、智能家居、智能驾驶、智慧金融、智能安防等应用场景。受国内外疫情影响，2020 年以来公司经营承压，2021 年实现营业收入 2.06 亿元。随着疫情好转叠加下游企业对 AI 训练数据需求的持续成长，以及公司对智能驾驶业务的加速布局，我们预测公司 2022-2024 年营收将持续提升，增长率分别为 29.36%、49.67%、37.00%，对应营收为 2.67 亿元、4.00 亿元、5.48 亿元。

智能语音类产品服务：根据中商产业研究院数据，2022 年我国智能语音市场规模预计达到 401.55 亿元，同比增长 13.6%，2018-2022 年复合增长率约 27.60%；同时根据观研天下数据中心数据预计，在未来几年数据标注平均价格指数将持续升高，预计到 2029 年达到 112.3（2021 年=100），2021-2029 年 CAGR 约为 1.46%。公司在多语种智能语音领域技术积累深厚，可满足国内外客户的数据需求，未来有望受益行业的持续成长，公司预计 2022-2024 年智能语音类产品服务业务收入增速为 7.15%、24.67%、27.98%。

计算机视觉类产品服务：公司在人脸检测和识别、物体识别、光学字符识别、场景分割等领域持续积累核心技术，构建竞争壁垒，数据产品服务在安防、智能家居等领域广泛应用。对标国外数据标注龙头厂商 Appen，其业务在 2018-2020 财年实现营业收入 2.57/3.75/4.61 亿美元，复合增长率达 34.03%，未来随着国内外疫情好转叠加下游客户需求成长，公司计算机视觉类产品业务有望迎来快速成长，结合数据标注产品服务平均价格指数上涨趋势，公司预计 2022-2024 年相关业务收入增速为 15.67%、46.87%、35.78%。

自然语言类产品营收：公司基于在智能语音领域的长期积累，积极探索自然语言处理领域产品及服务，公司在发音词典数据、手语语料数据、机器翻译等方向上逐渐积累核心技术，持续提升业务竞争力。根据艾瑞统计数据，2021 年，中国 NLP 核心产品和带动市场规模分别为 171 亿元和 450 亿元，预计到 2026 年规模将分别达到 459 亿元和 1061 亿元，2021-2026 年核心产品和带动市场规模 CAGR 分别为 24.8%和 22.3%。结合数据标注产品服务平均价格指数上涨趋势，我们认为公司基于智能语音方向的深厚积累，有望受益自然语言赛道的行业成长，预计 2022-2024 年自然语言类产品营收增速为 28.46%、28.49%、27.98%。

训练数据相关的应用服务：公司基于生产的训练数据提供算法模型相关的模型拓展及训练服务，以软件授权或软硬件一体化形式交付算法模型拓展、开发成果，公司该部分业务 2021 年实现营业收入 8.15 百万元，同比下降 17.97%，预计 2022-2024 年相关业务增速分别为-10.63%、-8.82%、-7.74%。

智能驾驶类产品服务：2022 年以来公司加速布局智能驾驶业务，快速拓展智能驾驶市场，公司于前三季度智能驾驶业务新增客户 14 个，截至 2022Q3 在手订单总额接近 5000 万元，同比增长超 200%。未来随着公司智能驾驶标注平台的持续升级，叠加自动驾驶技术向 L3 及以上发展推动的数据标注需求加速成长，公司智能驾驶相关业务有望迎来爆发，预计 2023-2024 年收入增速为 179.67%、59.57%。

- 毛利率：**2019-2021 年公司毛利率分别为 70.25%、67.70%和 62.59%，2021 年公司智能语音、计算机视觉、自然语言和训练数据相关应用服务各项业务毛利率分别为 65.21%、55.43%、61.34%和 75.26%，受疫情影响公司整体毛利率有所下降但仍保持高位。由于公司持续布局智能驾驶业务研发，叠加 2022 年国内外疫情影响，毛利率可能降低，未来随着公司新业务产品成熟和标准化程度提升，毛利率有望回到高位，预计公司 2022-2024 年毛利率分别为 59.18%、56.64%、57.36%。
- 费用率：**销售费用率方面，2019-2021 年公司销售费用率分别为 3.89%、4.77%和 7.03%，考虑公司业务规模成长叠加产品服务标准化程度的提升，公司有望保持销售费用率在稳定水平，预计 2022-2024 年公司销售费用率分别为 5.90%、6.47%和 6.19%；管理费用率方面，2019-2021 年公司管理费用率分别为 13.29%、11.00%和 17.06%，考虑公司短期内的限制性股票激励计划并且未来整体管理结构稳定，预计 2022-2024 年管理费用率分别为 23.66%、13.49%和 11.78%；研发费用率方面，公司 2019-2021 年研发费用率分别为 17.55%、18.64%和 29.31%，考虑公司在智能驾驶业务领域的研发投入，以及核心技术团队的持续建设，预计研发费用率将维持高位，预计 2022-2024 年研发费用率为 33.69%、29.31%、21.41%。

5.2 估值与投资建议

- 估值方面，**我们以行业相关性为标准，选取人工智能、智能驾驶等领域的优质企业作为可比公司，考虑到公司受疫情影响 2020-2021 年经营状况承压，随着疫情复苏有望迎来收入拐点，同时智能驾驶作为新业务成长性较高，因此我们采用 PS 估值法。人工智能领域我们选取格灵深瞳和虹软科技，智能驾驶领域我们选取四维图新和德赛西威作为可比公司进行分析。公司在智能语音、计算机视觉领域拥有多年技术积累，并积极布局智能驾驶业务，作为国内稀缺的 AI 数据标注行业上市公司，未来有望受益于人工智能行业以及数据要素生态的快速成长带来的数据需求提升，我们基于 2023 年行业平均 7.72 倍 PS，首次覆盖给予公司“买入”评级。

表8：可比公司估值分析

股票代码	公司名称	市值 (人民币)	营业收入 (百万元人民币)			PS		
			2022E	2023E	2024E	2022E	2023E	2024E
688088.SH	虹软科技	96.91 亿	552.11	682.99	932.87	17.55	14.19	10.39
688207.SH	格灵深瞳	45.91 亿	402.03	569.43	761.17	11.42	8.06	6.03
002405.SZ	四维图新	282.07 亿	3965.14	5189.62	6757.33	7.11	5.44	4.17
002920.SZ	德赛西威	604.19 亿	14004.61	19014.55	24376.06	4.31	3.18	2.48
平均值						10.10	7.72	5.77
688787.SH	海天瑞声	28.03 亿	267.09	399.75	547.65	10.49	7.01	5.12

资料来源：Wind，浙商证券研究所（注：截至 2023 年 1 月 4 日）

6 风险提示

1、技术迭代风险：由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，公司业务相关的各项技术面临快速迭代更新的压力。如果公司无法及时判断行业需求的发展趋势，合理规划研发人才及研发投入，可能导致在技术迭代进程中失去技术优势以及对新需求、新业务场景的先发优势，进而对公司经营造成不利影响。

2、汇率波动风险：公司积极拓展境外客户，2022H1 公司境外收入占总营收 57.14%，该收入采用外币结算，并受人民币汇率水平变化的影响，因此汇率波动可能对公司境外收入造成不利影响

3、新客户拓展不及预期带来的风险：2022 年上半年，公司收入主要来自老客户、占比达到 95.82%，新客户收入占比较低。若老客户降低对公司产品、服务的采购量或公司未能维持与老客户的合作关系，则公司的经营情况或将受到不利影响。

4、市场竞争加剧风险：近年来 AI 基础数据服务领域公司不断增多由于整体行业较新，行业标准和法律监管仍处于规范阶段。公司如果不能持续有效地制定并实施业务发展规划，将受到竞争者的挑战，影响公司的盈利能力和发展潜力。

表附录：三大报表预测值

资产负债表

(百万元)	2021	2022E	2023E	2024E
流动资产	809	831	897	1018
现金	278	315	298	373
交易性金融资产	418	398	408	403
应收账款	90	93	154	194
其它应收款	3	1	2	4
预付账款	2	8	14	14
存货	9	10	14	22
其他	10	5	6	7
非流动资产	32	32	33	34
金融资产类	0	0	0	0
长期投资	0	0	0	0
固定资产	25	23	21	20
无形资产	4	7	10	13
在建工程	0	0	0	0
其他	3	2	2	2
资产总计	841	863	930	1052
流动负债	35	43	65	85
短期借款	0	0	0	0
应付款项	11	14	24	31
预收账款	0	0	0	0
其他	23	30	41	55
非流动负债	0	0	0	0
长期借款	0	0	0	0
其他	0	0	0	0
负债合计	35	43	65	85
少数股东权益	0	0	0	0
归属母公司股东权益	806	820	865	967
负债和股东权益	841	863	930	1052

现金流量表

(百万元)	2021	2022E	2023E	2024E
经营活动现金流	(16)	1	(20)	59
净利润	32	14	45	102
折旧摊销	3	3	3	3
财务费用	1	(3)	(3)	(3)
投资损失	(11)	(17)	(14)	(11)
营运资金变动	(33)	5	(50)	(30)
其它	(7)	0	(1)	(1)
投资活动现金流	(84)	34	1	12
资本支出	(1)	0	0	0
长期投资	0	0	0	0
其他	(83)	34	1	13
筹资活动现金流	338	2	3	4
短期借款	0	0	0	0
长期借款	0	0	0	0
其他	338	2	3	4
现金净增加额	238	37	(17)	75

利润表

(百万元)	2021	2022E	2023E	2024E
营业收入	206	267	400	548
营业成本	74	109	173	234
营业税金及附加	1	2	3	4
营业费用	15	16	26	34
管理费用	35	63	54	65
研发费用	61	90	117	117
财务费用	1	(3)	(3)	(3)
资产减值损失	5	1	3	7
公允价值变动损益	0	0	0	0
投资净收益	11	17	14	11
其他经营收益	8	8	8	8
营业利润	33	15	49	110
营业外收支	0	0	0	0
利润总额	34	15	49	111
所得税	2	1	4	9
净利润	32	14	45	102
少数股东损益	0	0	0	0
归属母公司净利润	32	14	45	102
EBITDA	33	13	47	108
EPS (最新摊薄)	0.74	0.32	1.06	2.39

主要财务比率

	2021	2022E	2023E	2024E
成长能力				
营业收入	-11.53%	29.36%	49.67%	37.00%
营业利润	-64.18%	-56.47%	235.57%	126.53%
归属母公司净利润	-62.85%	-56.75%	230.52%	126.32%
获利能力				
毛利率	64.01%	59.18%	56.64%	57.36%
净利率	15.31%	5.12%	11.30%	18.67%
ROE	5.08%	1.68%	5.36%	11.16%
ROIC	3.43%	1.18%	4.74%	10.09%
偿债能力				
资产负债率	4.13%	5.00%	6.99%	8.10%
净负债比率	2.63%	0.71%	0.62%	0.64%
流动比率	23.28	19.26	13.80	11.95
速动比率	23.03	19.04	13.58	11.69
营运能力				
总资产周转率	0.31	0.31	0.45	0.55
应收账款周转率	2.70	2.90	3.17	3.02
应付账款周转率	8.02	8.75	9.23	8.54
每股指标(元)				
每股收益	0.74	0.32	1.06	2.39
每股经营现金	-0.36	0.03	-0.48	1.38
每股净资产	18.83	19.15	20.20	22.59
估值比率				
P/E	88.67	205.04	62.04	27.41
P/B	3.48	3.42	3.24	2.90
EV/EBITDA	98.94	159.65	44.43	18.69

资料来源：浙商证券研究所

股票投资评级说明

以报告日后的6个月内，证券相对于沪深300指数的涨跌幅为标准，定义如下：

1. 买入：相对于沪深300指数表现+20%以上；
2. 增持：相对于沪深300指数表现+10%~+20%；
3. 中性：相对于沪深300指数表现-10%~+10%之间波动；
4. 减持：相对于沪深300指数表现-10%以下。

行业的投资评级：

以报告日后的6个月内，行业指数相对于沪深300指数的涨跌幅为标准，定义如下：

1. 看好：行业指数相对于沪深300指数表现+10%以上；
2. 中性：行业指数相对于沪深300指数表现-10%~+10%以上；
3. 看淡：行业指数相对于沪深300指数表现-10%以下。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重。

建议：投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。

本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。

本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

浙商证券研究所

上海总部地址：杨高南路729号陆家嘴世纪金融广场1号楼25层

北京地址：北京市东城区朝阳门北大街8号富华大厦E座4层

深圳地址：广东省深圳市福田区广电金融中心33层

上海总部邮政编码：200127

上海总部电话：(8621) 80108518

上海总部传真：(8621) 80106010

浙商证券研究所：<https://www.stocke.com.cn>