

证券代码：688787

证券简称：海天瑞声

## 北京海天瑞声科技股份有限公司

### 投资者关系活动记录表

编号：2024-013

投资者关系活动类别	<input type="checkbox"/> 特定对象调研 <input checked="" type="checkbox"/> 分析师会议 <input type="checkbox"/> 媒体采访 <input type="checkbox"/> 业绩说明会 <input type="checkbox"/> 新闻发布会 <input type="checkbox"/> 路演活动 <input type="checkbox"/> 现场参观 <input type="checkbox"/> 电话会议 <input type="checkbox"/> 其他（请文字说明其他活动内容）
参与单位名称及人员姓名	源乐晟资产管理有限公司 王雁杰 北京鸿道投资管理有限责任公司 方云龙 兴合基金管理有限公司 候吉冉
会议时间	2024年9月4日
会议地点	券商策略会
上市公司接待人员姓名	证券部总经理：张哲
投资者关系活动主要内容介绍	<p>1、公司上半年收入增长的主要驱动因素是什么？</p> <p>2024年上半年，受益于多模态大模型的发展、《促进和规范数据跨境流动规定》出台带来的政策环境改善以及公司在境外营销体系的持续投入，公司境外业务收入同比增长50.00%，拉动公司整体营业收入快速增长。</p> <p>2、公司上半年净利润增长的原因是什么？</p> <p>公司在上半年收入增长的同时，由于标准化数据集产品收入占比大幅提升，同时公司整体数据交付能力在管理能力、平台技术提升、供应链进一步整合和扩展等因素的共同推动下，使训练数据定制服务的毛利率也同比显著增长，共同驱动公司整体毛利率增加至70.34%。</p>

同时，在研发投入方面，随着首发募投项目的结项，公司在传统深度学习数据集建设等方向已达到阶段性成熟状态，整体研发投入强度呈现自然回落；同时，在销售及管理投入方面为进一步提升整体运营效率，公司进行了资源的合理配置和流程优化，使得销售费用以及管理费用有效降低，以上因素共同驱动公司上半年净利润大幅增长，实现扭亏为盈。

### 3、公司在全球化业务增长的原因是什么？

2024 年上半年，受益于大模型技术的快速发展以及应用领域不断拓展，以智能终端厂商为代表的科技巨头纷纷加大多模态数据投入，全球数据服务市场持续增长。在此背景下，公司凭借多年来在语言研究以及自然语言方面的技术积累、以及全球领先的标准化数据集储备，进一步赢得了境外客户信赖。与此同时，《促进和规范数据跨境流动规定》等法规出台，为数据出境提供了更为高效简明的路径。以上行业发展以及政策环境的改善，使得公司境外业务收入较上年同期显著增长 50.00%至 4,369.53 万元。

### 4、公司在大模型业务上有何布局和进展？

2024 年上半年，公司继续加大大模型数据方向的研发投入，增大大模型领域的数据库，已完成并持续建设包括“大语言模型中文对话预训练数据集”、“语音大模型（声音复刻、歌曲）微调数据集”、“语音大模型（多语种）预训练及微调数据集”、“视觉大模型（图像-文本）预训练及微调数据集”、“视觉大模型（视频-文本）预训练及微调数据集”等在内的多领域大模型数据集。同时，针对大模型在特定行业的应用需求，公司重点开发了医疗、金融、法律、艺术等垂直领域的标注资源，形成垂直领域专家库，为公司提供高质

量行业数据服务奠定坚实基础。与此同时，为更好理解大模型技术方向，公司通过前瞻性研究，探索大模型数据的规模化生产方式。公司已和清华大学联合启动多语种语音大模型研发计划，该项目将基于最新的语音大模型框架技术，自研多语种数据清洗技术，训练多个不同规模的语音大模型，有效提升多语种语音数据处理的效率和准确性。

#### 5、公司在智能驾驶领域有哪些最新进展？

为更好抢抓智能驾驶行业机遇，同时应对更为复杂的数据处理需求，2024年上半年，公司加大了对高级别数据标注工具的研发力度，截至6月底，DOTS-AD平台新增3D动静分离标注工具，BEV多图层4D车道线标注工具，并升级点云分割工具支持连续帧叠加标注和4D分段加载等核心能力，有效支撑了自动驾驶BEV和OCC主流算法演进对数据标注工具的需求。此外，公司不断优化算法中台中枢能力，开发了2D-3D融合的动静分离检测追踪算法，在点云连续帧融合产线实现提效30%以上，迭代优化点云分割算法和地面检测算法，在点云分割产线实现提效20%以上。此外，为进一步支撑我国车企的国际化战略，公司在全球范围内扩展了道路采集业务，可提供包括采集方案设计、采集人员组织、车队管理、车辆改装及数据清洗与脱敏等在内的全链条服务。2024年上半年，公司采集业务已覆盖包括德国、西班牙、捷克等多个国家，为公司构建智能驾驶全球化数据服务能力奠定了基础。

#### 6、目前已经有很多大模型陆续发布，未来数据需求是会持续增长还是下降？原因是什么？

随着更多大模型产品面向公众推出上线、以及不断进行版本迭代，未来的大模型相关数据需求会持续、显

著增长，这符合当 AI 进行应用侧后，对数据的需求才会规模性增长的规律。预期该类增长将主要体现在预训练阶段对高质量、大规模的版权数据、公共数据、多模态数据、垂直领域的实际场景数据等方面的需求，以及不断提升其交互精准度的强化学习阶段各类数据需求等。例如，若某大模型应用领域为文学创作，就会需要来自版权方的散文、小说、诗词等数据进行创作能力训练；若某大模型拟提升其在民生政务领域的客服能力，则来自民生政务服务平台的此类对话数据将产生重要价值。未来，随着大模型向更高级、更多维、更广泛垂直向拓展，大模型数据需求将呈现快速增长趋势。

#### 7、智能驾驶的市场竞争格局如何？

智能驾驶市场主要参与者有品牌数据服务商，客户自建团队以及一些中小服务商。从目前行业格局来看，品牌服务商占据较大比例的市场份额。海天从 21 年集中发力该领域，并将会通过持续提升平台和算法的能力、拓展客户资源，加速培育能力，在该领域实现高速增长。

#### 8、如果智能驾驶领域客户自建团队做数据业务，对海天是否会造成影响？

客户自建团队在整个 AI 训练数据市场并非新现象，而是伴随着整个 AI 行业的发展一直存在的。但客户自建团队主要解决其自身的部分数据需求，如敏感数据等，对于那些需要投入较高研发力量以及数据安全要求较高的复杂数据，客户出于成本以及规模化效益考虑，仍会大量购买数据服务提供商的数据及服务。

相较于客户自建团队，海天历来都是对接众多大型科技公司、头部人工智能企业、科研院所等，项目经验丰富，积累了大量的 know-how，对数据处理的理解更

	<p>广、更深刻，同时我们搭建了成熟的数据处理算法平台，通过更高效的人机交互实现降本增效，保证数据质量的同时能有效降低成本，为客户提供更高性价比的训练数据产品/服务。</p> <p>因此，公司判断，智能驾驶数据业务市场空间足够大，即便部分客户未来会长期维持自建团队，但出于成本以及规模化效益考虑，仍会持续大量购买第三方数据服务，不会影响海天在智能驾驶领域的整体收入体量和增长趋势。</p> <p>9、智能驾驶行业的核心竞争力是什么？</p> <p>智能驾驶数据领域的核心竞争力主要体现在三个方面，分别是平台能力、算法能力以及数据安全能力。</p> <p>平台能力是数据标注能力的基石。平台功能点覆盖的丰富度是评价平台水平的核心要素，目前同时具备2D、3D、4D标注能力的供应商比较有限，能以最快速度覆盖更多功能需求的数据服务商将能更好掌握智能驾驶数据市场的主动权以及议价能力。</p> <p>第二个核心要素是算法能力。平台的智能化程度越高，对人的依赖程度越低，在提高平台的生产效率的同时可以大幅降低生产成本。</p> <p>第三个要素是数据安全能力。智能驾驶数据不同于传统的语音类数据，由于其采集图像涉及大量的地理及个人隐私信息，为更好防范数据安全风险，国家相关法律法规要求数据流转链条上各类企业必须做好充分的数据安全保障。因此，未来不具备相关数据安全能力的供应商将逐渐被市场淘汰。</p>
附件清单（如有）	
日期	2024年9月6日