

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：2024-014

投资者关系活动类别	<input checked="" type="checkbox"/> 特定对象调研 <input 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	国寿安保 余舒嘉铭 华夏基金 贾静雯 万家基金 陈鹏宇、高源
会议时间	2024年9月11日 2024年9月12日
会议地点	腾讯会议
上市公司接待人员姓名	证券部总经理：张哲 投资者关系负责人：袁璐
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</p>

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%。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，增厚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、行业的竞争格局现在及未来是什么样的？未来市场集中度将如何变化？

目前来看，市场上数据服务市场主要由品牌数据服务商、客户自建团队以及一些中小数据服务商构成。

未来，公司预判整个数据服务市场将进行重新洗牌，集中度将进一步提升。市场各类主体将会通过技术研发投入、资源能力建设等主要方面的竞争，逐步淘汰掉那些研发能力弱、资源势力差的品牌服务商和中小玩家。此外，国家对于数据安全及合规要求的进一步趋严，会将那些不具备数据安全合规能力或尚未进行此方面布局的企业逐渐淘汰出局。

在客户自建团队部分，出于其自身对数据和业务的敏感性、保密性需求，可能会与品牌服务商长期共存。

8、公司主要竞争对手有哪些？

公司竞对仍是传统模式下的数据服务公司，国内的主要竞争对手是一些品牌数据提供商，如数据堂、标贝

以及一些新兴公司；国外的主要竞争对手是 Appen。

与竞争对手相比，海天瑞声自身还是存在显著的竞争优势的，如丰富的产品积累、成熟的数据处理技术和平台、全球化的供应链管理等等。另外，从公司创业历程看，由于长期与国际性科技企业合作，对数据安全和合规的重视是深入到公司运作的方方面面的。而数据安全和合规是需要投入较高的成本建设的，在日益完善的法律环境下，这方面的投入为公司带来了新的竞争壁垒，也将会为公司未来在垂直行业和政企业务拓展形成有利优势。

9、数据标注行业未来会有什么样的发展趋势？

首先是更加智能化，即通过拓展算法覆盖的场景以及算法预识别的准确率等，持续提升机器参与程度以及人机协作效率，降低数据处理成本。

其次，随着 AI 技术不断革新，应用行业以及场景不断增加，各行业、各领域数据安全规范逐渐落地将成为趋势，对于以数据生产为主营业务的数据服务企业，数据安全及合规能力将成为数据服务能力的核心评价维度，成熟的安全合规管理体系将成为重要评价标准，能持续跟踪法律环境变化，积极响应监管政策的企业将具有更强的市场竞争力。

此外，随着境内、外企业的全球化扩张成为确定性趋势以及各类客户群体扩张步伐加速，多语种能力作为支撑企业顺利出海的核心要素之一，重要意义更加凸显，具有强大语言研究能力的数据服务企业将获得更多商业机会。

另外，随着数据服务向多元化、多类型、多场景持续发展，充足、稳定且高质量的数据处理团队储备、以及更加智能化的资源配置能力，将成为数据高效、稳定

	交付的重要保障。
附件清单（如有）	
日期	2024年9月12日