

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：2024-017

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	汇添富基金 钱诗翔 东方红资管 张明宇 富国基金 蒲梦洁 西部利得 张昭君 国泰基金 韩知昂 上海淳阳基金 谢伟玉 平安资管 张昱 国盛证券 赵伟博
会议时间	2024年10月16日
会议地点	国盛证券策略会
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。

2、公司未来发展规划是什么？

公司将自身发展战略定位为以下三个方向：

（1）全球化业务：为更好把握国际市场需求，公司将推出一项更为全面的出海战略，涵盖技术创新、品牌升级、体系构建、市场推广等，全面加速全球市场的拓展。公司还将建立一个海外技术研发体系，紧跟全球 AI 的发展动态，并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。

（2）智能驾驶业务：公司将积极把握智能驾驶领域的发展良机，继续升级自动驾驶数据平台 DOTS-AD；同时，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；此外，公司会继续加强数据安全管理体系建设，确保数据处理流程的安全与合规；并进一步有效利用已获得的测绘资质，延伸数据服务范围，以提升公司智能驾驶业务的毛利水平。

（3）新兴业务探索—大模型、数据要素：公司将持续探索围绕大模型所需数据相关服务，通过前沿技术跟踪研究，开展以预训练、强化学习为代表的多元化数据获取、高阶垂向拓展等方向的数据服务能力建设；此外，还将探索以数据治理、数据交易、数据处理等为核心的数据要素领域，力争将数据要素创新业务打造成为

具有潜在高增长价值的新兴业务板块。

3、公司获得境外客户认可的核心竞争力是什么？

除了长期合作的良好口碑，公司在语音方面积累深厚，尤其是多语种方面的积累和能力，可以更好服务境外公司全球化布局。截至今年 6 月 30 日，公司已覆盖超过 200 种语种/方言，不仅包括含英、法、德、意、西、日、韩等常见语种，还包括东南亚、一带一路等国家地区的罕见多语种，尤其在亚洲多语种的服务上具备竞争优势。除核心业务能力外，公司制定了更为全面的全球化发展战略，从品牌升级、体系搭建、团队建设、营销推广等多维度升级海外市场布局，大大提升海外客户触达和服务能力，上半年境外业务增势显著，收入同比增长 50%。

4、客户对训练数据是否有持续需求？

客户对训练数据本身的需求是会长期持续的。

客户的 AI 产品在上线之前及初期，因为其自身尚未产生实网数据，通常需要采购模拟型数据集进行算法模型的训练；在产品上线并运行一段时间、产生大量实网数据之后，则会提供实网数据给到我们进行数据加工，加工的数据反哺到客户的产品上从而促进其产品的迭代、升级。之后，客户需要进行产品功能的拓展，再次需要购买模拟数据集来支撑，后续再采购数据加工服务进行迭代，如此周而复始。因此，客户对训练数据的需求是持续的，且随着应用 AI 技术的场景越来越多，各种场景的数据集需求会兴起，带来的是训练数据的需求会越来越大。

5、上半年公司在大模型业务上有哪些新的布局，以及具体进展？

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。

6、24 年大模型都在往多模态拓展，请问这个趋势对公司业务带来哪些影响和变化？

大模型向多模态发展后，将会产生更多的新型数据需求。例如文生图的多模态大模型，通过文字输入生成对应图片，这就需要机器理解文字语义的同时将理解的关键词与图片的关键标签进行映射，通过对齐两种独立模态关键特征的方式，实现按指令的创作，以此完成学习训练过程。因此，当大模型向多模态能力维度拓展时，高质量多模态训练数据集的持续学习训练的重要性将更加凸显，多模态的发展将推动数据服务行业进入更大的增量空间。

7、今年公司在智能驾驶领域的技术研发上有哪些进展？

	<p>今年上半年，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。</p>
附件清单（如有）	
日期	2024年10月21日