

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：2024-012

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 checked="" 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	东吴证券 黄诗涛 中信证券 朱珏琦、韩林轩、许英博 华夏财富创投 戴永毅 誉辉资本 郝彪 国信证券 闵晓平 建信养老 高月 景泰利丰资产 吕伟志 景泰利丰投资 邹因素 锦绣中和投资 方家逸 信达澳亚 马绮雯 大和住银 刘明 锦成盛资产 王懿晨 庶达资产 李淑花 长宜基金 王勃 新华基金 张帅 中邮计算机 丁子惠
会议时间	2024年8月28日

	2024年8月29日
会议地点	进门财经
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。</p> <p>3、公司在全球化业务增长的原因是什么？</p> <p>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</p>

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，增厚大模型领域的的数据储备，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、强化学习阶段的数据服务，今年有何进展或者变化？

目前来看，随着各大模型的陆续上线，强化学习环节的整体数据需求在逐渐攀升，并在具体标注任务上呈现如下趋势：

(1) 由单模态向多模态转变：23 年主要的标注需求集中在文本类标注，今年开始逐步向多模态拓展（例如，文本-视频、文本-图像等）；

(2) 逐渐向更多垂类拓展（例如，法律、金融、医疗）；

(3) 强化学习标注的评价/评分指标变得更为丰富，会要求标注人员从更多维度对模型的问答进行评判和打分。

7、请问大模型向多模态发展后，是否会对公司业务产生正向影响？

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

	<p>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。</p>
附件清单(如有)	
日期	2024年8月30日