

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：2024-004

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	财通证券 杨焜、董佳男 野村资管 徐也清 国融资管 刘杰 兴华基金 高伟绚 招商资管 吴彤 建信养老 王天乐 运舟资本 邹伟 嘉合基金 丁俊枫 凯昇投资 陈昶旭 亿田智能 桂绍兵
会议时间	2024年5月8日
会议地点	财通证券策略会
上市公司接待人员姓名	证券事务代表：张哲
投资者关系活动主要内容介绍	1、影响智能驾驶数据业务市场需求空间的因素有哪些？ 智能驾驶数据业务的市场需求主要与三个要素相关：1) 车厂的车型及传感器丰富度。通常来说，不同

车型、不同传感器会有不同的硬件配置方案，继而需要不同的数据解决方案，因此车型/传感器等硬件配置的多样性程度将会直接影响所需数据解决方案的数量；2) 量产车数量。量产车的数量决定了整个的训练数据需求基数的大小；3) 智能驾驶级别的逐渐提升。智能驾驶级别和渗透率的提升决定了数据处理场景的种类和体量。

这三个要素对训练数据需求的影响是相互叠加的。公司预测，随着智能驾驶相关政策的推出以及单车成本的不断下降，智能驾驶的商业化进程将加速，在上述三个因素的共同作用下，数据处理需求将呈现指数级增长趋势。

## 2、为什么公司 2023 年度研发投入下降？

根据近几年公司的整体研发战略和投入节奏安排，公司在数据生产智能化、以及标准化数据集产品建设等方向的投入强度逐渐趋于平稳所致。在数据生产智能化投入方面，以 DOTS-AD、一体化数据处理平台等为代表的智能驾驶数据、以及综合性数据处理平台的整体能力已达到阶段性的成熟状态，因此公司在该等领域的研发投入强度较去年同期有所趋缓；在标准化数据集产品建设方面，根据公司 IPO 募投项目建设的规划，标准化数据集的集中建设期集中于前两年，并已于 2023 年 5 月顺利结项，因此报告期内公司在该领域的投入强度较去年呈现自然回落状态。

## 3、目前公司都在为大模型厂商提供哪些数据服务？

目前，公司已与国内众多知名大模型厂商开展业务合作，业务范围覆盖 RLHF 数据评分、Prompt 改写、人机多轮交互等强化学习阶段数据需求。此外，公司通过

研究数据清洗技术、模型评测数据集设计技术、大模型微调与应用等，进一步将数据服务拓宽范围至预训练以及模型评测阶段。

4、如何看待数据合成技术？是否会对公司传统业务造成影响？

在数据重要性凸显且数据需求快速增长的时代，合成数据可以认为是人工智能行业发展到一定阶段的必然产物。数据合成技术可以作为数据采集的有效辅助，但也存在较强的局限性，降低真实世界各类特征的训练效果，因此目前仅可作为数据采集的一种辅助方式。从目前数据服务行业来讲，以计算机视觉场景为例，合成数据主要应用于某些高危的、罕见的 corner case 的模拟训练当中，但合成数据毕竟是由机器生成的虚拟数据，其数据质量以及真实性仍无法替代真实场景数据，因此按照目前的技术路线，绝大多数企业仍在使用真实场景数据进行模型训练。但公司会紧密关注合成数据技术的发展，根据最新的行业动态及时调整公司业务布局。

5、版权数据持有方是否可以直接将数据卖给大模型厂商，海天瑞声在其中的价值是什么？

公司的价值主要体现在多版权数据的汇集、版权数据的清洗、以及基于客户大模型的后续服务。首先，公司可以汇聚不同版权方的数据，针对客户需求进行不同版权数据的提供。同时，可以针对客户具体定制化需求，对版权数据进行精细化清洗。虽然版权数据本身已为高质量数据，但仍无法直接用于模型训练，需经过高质量清洗后才能使用。例如，通常需将版权数据中重复数据以及不符合法律法规的相关内容进行清洗，以更好帮助大模型节约训练算力以及使大模型在训练后具备良好的

	的法律道德价值观。
附件清单（如有）	
日期	2024年5月10日