

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：2024-005

投资者关系活动类别	<input 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 checked="" 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	上银基金 王婷 东北证券 范彦君 盈骅资产 熊城 远希基金 高志强、李佳蓉 链友资本 杨晋 金鼎资本 王佳瑜 诺安基金 邓心怡
会议时间	2024年5月9日
会议地点	东北证券策略会
上市公司接待人员姓名	证券事务代表：张哲
投资者关系活动内容介绍	<p>1、请介绍一下公司在数据服务方面的能力和业务范围</p> <p>公司主要从事 AI 训练数据的研发设计、生产及销售业务。公司通过设计数据集结构、组织数据采集、对取得的原料数据进行加工，最终形成可供 AI 算法模型训练使用的专业数据集。公司的数据服务类型已经实现智能语音（语音识别、语音合成等）、计算机视</p>

觉、自然语言等多个核心领域全覆盖，可服务于智能驾驶、内容生成、机器人、智慧医疗、智慧教育、智慧金融等 22 种创新应用场景。目前，公司产品/服务线已覆盖全球超过 205 个主要语种及方言，已积累超过 1,550 个自有知识产权的 AI 训练数据产品。

2、公司在人工智能数据服务行业中的竞争优势是什么？

经过多年发展与积累，公司逐步构建起了在行业内的竞争壁垒，核心竞争力主要体现在：

（1）技术平台能力：公司历来重视技术的研发，近年来更是加大研发投入的力度，全面提升公司的算法能力、平台能力、工程化能力，加深算法辅助能力与人工工作的结合，达到更佳的人机协同效率，这样能够做大规模、提升效率、降低成本。

（2）公司的业务模式是服务产品双模式，且产品化贡献显著，是收入和毛利的主要来源，标准化数据集的研、产、销体系是公司从业多年探索出来的业务模式，其复用性为公司的规模化和高利润率提供了保障。而保持这样的能力需要具备对行业需求的强判断力和较强的资金实力。截至 2023 年 12 月末，公司已积累超过 1,550 个自有知识产权的训练数据标准化产品，数据库存量稳居全球企业前列。

（3）供应链资源管理能力：公司通过长期建设的供应链体系，保障资源的获取，未来，公司会进一步加大供应链资源平台的建设，使人员管理、采标资源分配、质量检验、远程工作等各方面的能力得到显著提升，为客群拓展提供有力支撑。

（4）数据安全及合规能力：数据安全及合规能力已经成为了衡量品牌数据服务商综合能力的重要指

标。公司在多年数据风险识别和管理实践中，已形成了较为成熟的安全、合规管理体系。

公司全方位做好数据风险管控工作，通过了业内重要的 ISO/IEC 27001 体系认证、ISO27701 个人隐私信息安全管理体系认证，形成了具有自身特色的数据安全与隐私保护整体解决方案。同时，公司拥有北京市规划和自然资源委员会行政许可（乙级测绘资质）、等保三级备案证明；目前，公司符合 GDPR、《数据安全法》、《个人信息保护法》等一系列国际通用与国内法律法规的管理规范要求，获得了业务领域合作客户的高度认可。

3、公司的商业模式是什么样的？

目前，公司的商业模式主要集中在向客户群体提供标准化数据集产品、及定制服务。同时，随着公司自研数据处理平台以及算法能力的日趋成熟、完善，公司还创新性开拓了 SaaS 等新型业务模式，可为客户提供数据标注平台服务。

4、大模型预训练阶段很大程度用的都是无监督学习，无监督学习下，公司的业务是否会受影响，或者会产生哪些新的数据机遇？

大模型是人工智能技术新的发展范式，该技术路线下将衍生出新型数据需求。例如，预训练阶段，数据服务公司需能根据客户的大模型需求，完成对海量数据清洗方案的设计以及执行，通过清洗出高质量数据，帮助客户高效完成模型预训练开发；同时，预训练想要达到更高阶的通用学习效果，需要更为丰富的高质量数据来源，例如各类版权数据等，因此，能聚集更为丰富的版权类数据、并以工程化和规模化地方式向客户提供相应服务的数据服务商将具有更高的市

场竞争力。

此外，为使大模型能力更趋近于人类，大模型还将在完成预训练的基础上，通过有监督微调以及基于人类反馈的强化学习，实现模型在特定方向上的优化迭代。上述两个阶段均需深度人工参与，同时参与上述两阶段的数据处理人员将被要求有更高的认知水平和专业能力，以更好指导机器完成高阶能力优化。

大模型时代实质上是一个海量数据的时代，随之将产生海量的数据需求，数据服务模式也将由传统的数据标注向获得海量高质量数据、以及更强的数据处理能力转变，这对数据服务商提出了新的、更高的要求。

5、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。

6、智能驾驶行业的核心竞争力是什么？

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。

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 标注、3D 点云标注、2D-3D 联合标注以及 4D 标注的供应商比较有限，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。

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。

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。

7、公司未来的发展规划如何？

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。
附件清单（如有）	
日期	2024年5月11日