**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：2025-011**

|  |  |
| --- | --- |
| **投资者关系活动类别** | □特定对象调研 □分析师会议  □媒体采访 √业绩说明会  □新闻发布会 □路演活动  □现场参观 □电话会议  □其他（请文字说明其他活动内容） |
| **参与单位名称及人员姓名** | 参加业绩说明会的广大投资者 |
| **会议时间** | 2025年6月5日 |
| **会议地点** | 上证路演中心 <https://roadshow.sseinfo.com>  视频录播+网络文字互动 |
| **上市公司接待人员姓名** | 董事、总经理：李科  董事、副总经理、财务负责人：吕思遥  董事会秘书：张哲  独立董事：杨柳 |
| **投资者关系活动主要内容介绍** | 1、请问针对机器人行业特别是人形机器人领域，公司可以提供哪些服务？目前市场拓张情况如何？以及该行业数据来源是什么？  董事会秘书张哲答:我司提供的训练数据覆盖智能语音(语音识别、语音合成等)、计算机视觉、自然语言等多个AI核心领域，经过多年人工智能数据服务，我司已积累了较为丰富的多模态的数据获取与处理能力，可服务于包括具身智能等在内的多场景人工智能应用的开发。未来，公司也将持续关注具身智能等新兴领域的发展趋势，并根据客户的需求和技术的变化，适时推进公司产品创新、技术创新，满足更多客户、更新领域的AI训练数据需求。  2、行业以后的发展前景怎样？  董事、总经理李科答:受益于全球人工智能产业的高速发展，尤其是多模态大模型以及AI应用的持续深化，数据产业将迎来持续的高速增长。根据艾瑞咨询的数据，2024 年中国人工智能基础数据服务市场规模为 58 亿元，2028 年规模将达到 170 亿元，年复合增长率为 30.84%。  3、请问海外市场与国内的数据标注市场有什么具体不同？公司将如何开拓海外市场？  董事会秘书张哲答:海外市场更加注重专业化分工，会将更多的数据需求释放给专业的数据服务商，而非自建团队；同时，更注重数据服务商品牌以及数据交付质量，愿意支付更高的溢价，因此往往使得境外收入的毛利水平更高。 为更好把握海外市场发展机遇、进一步扩大全球客户辐射范围，公司从体系搭建、市场研究、品牌升级、营销推广等多维度提升业务、客户触达及服务能力。未来，公司持续加强境外销售团队建设，进一步织密客户服务网络；同时，通过参与全球知名行业及AI 学术顶会，以及布局搜索广告投放、社交媒体矩阵等多类宣传渠道等方式，向全球市场展示公司先进技术实力及创新产品，提升行业影响力和美誉度。  4、你们行业本期整体业绩怎么样？你们跟其他公司比如何？  董事、总经理李科答:本行业内上市公司有澳大利亚公司Appen、以及新三板公司数据堂，相关友商信息，还请您查看相关公司年度报告，谢谢。  5、公司之后的盈利有什么增长点？  董事、总经理李科答:2025年公司营收增长的核心驱动力来自AI产业的两大发展趋势。首先，多模态AI技术的快速演进催生了跨模态融合数据的增量需求。随着AI从单一文本处理扩展到视觉生成、语音交互等多元模态，市场对高质量图文对数据、细粒度标注语音数据集等高价值多模态数据服务的需求呈上升态势，这为公司业务增长提供了基础。其次，AI在垂直行业的深度应用创造了新的市场机遇。开源大模型的普及推动AI在政务、法律合规等专业领域快速落地，这些场景对专业化数据服务的需求将会显著提升。此外，公司去年在东南亚新增建设的数据交付体系已进入爬坡运营阶段，该基地可以帮助公司拓展海外定制服务市场，预计可为公司带来可观的收入增量，并有望成为海外业务扩展新的战略支点。  6、公司给客户做定制数据的过程中积累的经验和数据，能不能转化成标准化的产品来卖？  董事、总经理李科答:客户定制服务涉及的训练数据在交付给客户并完成验收后，所有权完全转移给客户，海天瑞声是不能用于自身产品建设的，这一点是公司始终遵循的知识产权要求。 在定制数据集的生产过程中，积累下来的经验、know-how会帮助公司各方面能力的提升，例如工具平台因为处理了大量的定制数据集，使平台完善性有很大的增益，加强了公司的数据处理能力；再如，在一些情况下，公司在生产定制数据集时，也会根据对行业需求的判断，在保障数据权属划分清晰的前提下，利用团队管理、资源获取的便利性，同步安排额外的设计、采集和标注工作，完成产品数据集的开发。  7、请问现在行业里的竞争情况怎么样？您觉得未来这个行业是会越来越集中还是更分散？  董事、总经理李科答: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在技术研发投入、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、公司本期盈利水平如何？  董事、总经理李科答:2025年第一季度，归属于上市公司股东的净利润为37.16 万元，较去年同期增加 100.58 万元。  9、咱们的业务规模扩大后，成本是不是能相应降低？规模效应明显吗？  董事、副总经理、财务负责人吕思遥答:公司最大的成本就是原料数据采购费用，即：采集、标注成本。一方面，公司通过继续加大研发投入的力度，全面提升公司的算法能力、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，这样能够做大规模、提升效率、降低成本；另一方面是加强供应链资源管理能力，扩大资源供给，降低单位成本。  10、想请教管理层，今年一季度咱们公司收入增长不错，主要是哪些业务在带动？  董事、总经理李科答:随着多模态大模型的快速迭代及行业应用渗透提速，公司计算机视觉业务和自然语言业务分别同比实现高速增长。其中，在国家对“AI+数据要素”政策同步发力的背景下，以运营商、互联网平台公司为代表的大型客户持续加码高质量图像/视频等多模态数据采购，为其通用多模态大模型训练提供有力支撑；同时，政务、法律合规等场景应用的落地，带动场景类文本数据需求快速增加。在全球化布局方面，公司去年在东南亚新增建设的数据交付体系已进入爬坡运营阶段，通过拓展海外定制服务市场，不仅带来了可观的增量收入，并有望成为海外业务扩展新的战略支点。上述因素，共同驱动公司2025年第一季度营业收入显著增长。  11、公司最近有没有在探索什么新的业务方向或者赚钱模式？能简单说说吗？  董事、副总经理、财务负责人吕思遥答:当前，在国家大力推进“人工智能+”行动和“数据要素X”的战略指引下，公司正积极探索与实践数据产业新业务和新模式。一是按照国家推动公共数据资源的开发利用，发挥海天瑞声的技术优势，与多地政府、地方运营商等开展战略合作，共同探索数据要素市场化与产业化的创新路径，通过构建“数据可信空间”，协助地方政府打造安全、高效、合规的数据治理与流通体系，推动数据要素的价值释放。二是发挥海天瑞声的行业经验和积累，联合当地高校，培训和培养数据标注人才，提升就业率的同时夯实区域数字经济发展人才基础。三是，发挥海天瑞声的生态优势，助力地方及产业园区打造数据标注基地和构建数据标注产业新生态。  12、李总您好，海天瑞声管理层对今年全年的业绩增长怎么看？最主要的增长会来自哪些业务板块？  董事、总经理李科答:2025年公司营收增长的核心驱动力来自AI产业的两大发展趋势。首先，多模态AI技术的快速演进催生了跨模态融合数据的增量需求。随着AI从单一文本处理扩展到视觉生成、语音交互等多元模态，市场对高质量图文对数据、细粒度标注语音数据集等高价值多模态数据服务的需求呈上升态势，这为公司业务增长提供了基础。其次，AI在垂直行业的深度应用创造了新的市场机遇。开源大模型的普及推动AI在政务、法律合规等专业领域快速落地，这些场景对专业化数据服务的需求将会显著提升。此外，公司去年在东南亚新增建设的数据交付体系已进入爬坡运营阶段，该基地可以帮助公司拓展海外定制服务市场，预计可为公司带来可观的收入增量，并有望成为海外业务扩展新的战略支点。  13、我注意到现在很多公司都在做垂直领域的大模型，想问下这些训练数据主要都是从哪里来的？  董事、副总经理、财务负责人吕思遥答:目前来看，训练垂直领域大模型的核心数据来源可分为三类：公开数据、客户自有数据和垂直场景定向采集数据。其中，公开数据（如互联网知识库、开源数据集和行业标准文档）可以为模型提供基础数据支撑；客户自有数据和定向采集数据则针对具体业务场景进行专项优化。值得注意的是，这些原始数据必须经过专业处理流程才能投入使用，主要包括：1）数据清洗与标准化；2）格式转换（如语音转文本）；3）领域专家标注与校验。以智能病历系统开发为例，数据加工流程包括：首先将门诊录音转为文本数据，再由医学专家进行专业校对并提取关键临床信息，最终生成结构化电子病历。这一过程高度依赖专业领域知识，需要大量临床医师参与质量把控。正因如此，在垂直领域大模型训练中，专业数据服务商扮演着双重角色：既是特定领域高质量数据的提供方，也是专业数据加工服务的提供商。  14、能不能介绍一下目前和三大运营商这些合作伙伴的合作情况？  董事会秘书张哲答:在国家"AI+数据要素"战略的指引下，尤其是国务院国资委连续两年开年启动部署中央企业“AI+”专项行动以来，以运营商为代表的重点央企自2024年起加速布局通用+垂向大模型研发，带动了高质量图像、视频等训练数据的规模化采购需求。公司凭借在数据领域的核心优势，已快速成为运营商类客户重要的数据服务供应商。未来，随着以运营商为代表的重点央企在多模态大模型方向的持续加码，以及其基座大模型在更多传统行业的应用落地，预计相关数据需求将进一步增长，为公司收入带来持续的增长动能。  15、DeepSeek这类大模型很火，不知道这是否会影响AI行业对数据的需求量？公司业务会不会受影响？  董事会秘书张哲答:（1）Deepseek推出了一系列模型，其中V3模型依然使用了预训练、以及SFT等训练方式，其中预训练阶段的token使用量达到了14.8T，远超GPT4等同类可比大模型预训练阶段的数据使用量，且在后训练阶段也使用了一定规模的标注数据，这也更加说明海量以及高质量数据对于基础模型能力提升的重要意义。（2）关于让大家震撼的R1模型，基于目前的公开信息来看，其部分优势体现在推理类任务上，尤其是那些具备较强的规则性、可以推导的任务类型上，确实不需要大量的人工标注，但是对于其他领域（尤其是更为广阔的垂向领域）的复杂问题，依然需要观察，我们认为高阶的数据专家的参与依然非常重要。 （3）此外，数据质量不仅影响模型获取和表达知识的能力，还决定了模型生成内容的风格和准确性，帮助DeepSeek实现了在输出端的文采能力提升。 其一，高质量数据可以提升模型表达和推理能力。优质数据包含准确、连贯且富有表现力的语言样本。例如，包含CoT数据可以引导模型在推理时进行反思，进而在生成回答时展现出清晰的逻辑和优美的语言表达。这正是DeepSeek模型能够生成既准确又具有华丽文风的关键因素之一。 其二，高质量数据可以降低噪音和确保一致性。数据中的错误、噪音或不一致信息会导致模型生成内容出现语法或逻辑问题。高质量的数据则能有效减少这些问题，使模型更好地学习到语言规律，从而提高整体生成质量。 其三，高质量数据可以提升泛化能力。数据的多样性和全面性使得模型在面对不同领域和任务时都能生成高质量的回答。丰富且准确的样本帮助模型在多种场景下自如切换风格，无论是精炼的技术解答还是文采斐然的创意写作，都能游刃有余。（4）往未来看，Deepseek模型的出现，有望进一步助推模型向产业端发展，真正让大模型技术深入滲透到各个行业中，这一过程中必将凸显专业知识的直要性，需要更多数据、以及数据专家的参与，因此我们看好并期待未来大模型在各行业百花齐放的局面。  16、作为行业从业者，您觉得数据标注这个行业未来几年会怎么发展？机会和挑战主要在哪？  独立董事杨柳答: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** | 2025年6月9日 |