

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：2025-008

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	东方红资管 张明宇 君成投资 陈嘉欣 中信建投 张敏 华泰证券 郭雅丽 诚通证券 李天一 鸿道投资 方云龙 海富通基金 张颖
会议时间	2025 年 5 月 21 日
会议地点	现场交流
上市公司接待人员姓名	董事会秘书 张哲
投资者关系活动主要内容介绍	1、2025 年第一季度，公司收入增长的驱动因素是什么？ 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。 2、目前公司是否有在尝试新的业务或者商业模式？ 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。 3、公司与运营商的合作进展如何？ 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。 4、2025 年公司营收的核心增长点是什么？ 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。 5、数据标注行业未来会有什么样的发展趋势？ 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。 6、公司的核心竞争力主要体现在哪？ （1）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 2024 年 12 月末，公司已积累超过 1,700 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。 （2）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。 （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。 7、DeepSeek 出来后，对数据需求的影响如何？是否会降低 AI 行业对数据的需求？ （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渗透到各个行业中，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。 8、训练特定垂向领域的大模型所需的数据，主要来源于哪里？ 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。
附件清单（如有）	

日期	2025 年 5 月 23 日
----	-----------------