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：2025-010

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	国泰海通　杨林 陆家嘴国泰 张颖杰 磐稳投资　陈奕霖 华宝信托　张卿隆 国投瑞银　马柯 九方智投　吴清淳 东方红资管 张明宇 健顺投资　罗庆 巴富罗投资 周刘为 中移资本　郑逍哲
会议时间	2025 年 6 月 4 日
会议地点	现场交流
上市公司接待人员姓名	投资者关系负责人 袁璐
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、训练数据的生产过程是什么样的？ 训练数据生产过程主要包括四个环节：设计（训练数据集结构设计）、采集（获取原料数据）、加工（数据标注）及质检（各环节数据质量、加工质量检测） ① 设计——训练数据集结构设计 在设计环节中，通过考虑算法模型的具体应用领域、应用场景以及预期实现的训练效果，反过来确定训练数据集内的数据类型、数量、比例分布等，相应确定原料数据的采集要求，为后续采集工作奠定基础。以语音识别、语音合成领域的训练数据集为例，在原料数据的采集环节，发音人（被采集对象）需要朗读公司提供的基础语料，并用指定的录音设备录制以形成原料音频数据。因此，在设计阶段，公司就需要考虑如何设计基础语料，才能使得容量有限的训练

	数据集能够覆盖尽可能多的自然语言现象，如覆盖更多的发音习惯、语言特点、句长分布，达到更好的音素平衡效果等，从而使得算法模型获得更好的训练结果。 ② 采集——获取原料数据 根据此前设计好的训练数据集结构及数据量目标，制定原料数据采集方案并开展原始数据采集工作。采集过程所涉及的主要考虑因素包括： A.数据量方面：需根据成品训练数据集的目标数据量，预留少量冗余。在实际采集过程中，由于可能发生少量录音不合格的损耗情况，通常会在总采集数据量中预留少量冗余，从而略大于最终要交付的数据量，以备替换偶然出现的不合格录音数据。 B.数据属性方面：在采集环节中，根据客户算法模型应用的目标场景、领域等个性化需求，采集特定原料数据。以语音识别训练数据为例，在采集环节中，通常需要根据语音识别模型的语种/方言类别、目标应用场景（安静、噪音；家居、车载等），相应定义寻找符合要求的发音人，在合适的采集场景下由发音人朗读、或自然说出录制语音片段，生产原料音频数据。以语音合成训练数据为例，通常需要根据客户对拟合成的语音的风格（温柔、甜美、科技感等）、年龄（成人、儿童）、性别、语种、口音等方面的具体需求寻找发音人，并组织发音人按照前期设计完成的音素集、语料库等资料进行朗读，录制生成原料音频数据。此外，由于语音合成训练数据的录制对信噪比、底噪、录音棚混响时间等参数、指标和录音设备的要求很高，通常需要在专业级别的录音棚中完成录制工作。

	③加工——数据标注 通过公司 ADS 和 VDS 平台，对语音、文本、图片等原料数据进行标注，使其成为结构化可被算法识别和学习的专业训练数据集。该环节中，公司通常会应用相关算法模型，通过算法完成预识别和预标注，可以显著提高数据标注效率，降低标注成本。 ④质检——各环节数据质量检测 质检环节会渗透在整个训练数据的全生产流程，具体包括： A.在前端采集环节，公司开发的采集工具可对原始数据质量进行即时质检，不符合要求的原始数据不被计入采集数据之中； B.在中端加工环节，公司运用自动标注工具+人工校对检验的方式对数据加工情况进行检查，提升加工效率和准确度； C.在后端大规模质检环节，公司运用全自动校验技术，实现大规模训练数据集的质检需求。 7、语言学研究的具体作用和价值是什么？ 语音语言学领域的专业知识是构建高质量语音识别算法和语音合成算法的关键要素。以语音合成为例，在语音合成系统中，发音词典提供了从单词到音素之间的映射关系，将语言模型建模单位解构为声学模型的建模单元，为后续合成发音奠定基础。语音合成系统接收到文本信息后，首先运用发音词典对其进行语言处理、韵律处理，将文本（单词、字符等）转换并解构为一系列对应的发音符号（类似于国际音标）；随后，系统中的语音合成器接收到前述发音符号，运用语音库合成转换为语音对外输出，最终实现文本到语音的语音合成过程。可见，高质量的发音词

	典在语音合成系统中具备重要作用。由上述示例可以看出，要获得高准确率的语音合成算法模型，就要求智能语音训练数据结构中包含高质量的发音词典。要在大词汇量的连续语音交互中正确、合理运用智能语音相关的语言模型、语法及词法模型，则必须有效地运用计算语言学方面的基础知识和研究成果。语音语言学领域的基础研究成果和专业知识构建了发音规则、发音词典的形成基础，进而为构建高准确率的语音识别、合成训练数据提供了条件。 8、公司的业务是否存在规模效应？ 公司业务是存在规模效应的，一方面随着公司在研发方面加大投入，自研平台的能力逐步提升，可以赋能数据处理过程中的人机协作朝着更加智能化的方向前进，这就使得公司进行更大规模的数据生产成为可能。同时，数据产品的积累、平台以及工具的研发，在公司业务规模逐渐上升的情况下，相关的研发费用、管理费用将被摊薄； 从成本端看，数据生产的成本还有很大的下沉空间，对于成本控制我们会在两方面进行持续投入：一方面是继续加大技术投入，采用更为合理的人机协同比例完成数据处理任务，降低人员投入，提高处理效率；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。 此外，数据集产品一直是我们公司所坚持的重点方向，公司开发大量通用型、复卖率高的标准化产品数据集，反复给公司带来利润，也能实现训练数据产品的规模化效应。 9、公司各个业务板块的毛利率有一定差异，原因是什么？

	各版块之间的毛利差异，主要由以下两个因素综合决定： （1）产品与定制服务的收入占比 产品的毛利为100%，因此产品收入占比较高的业务版块，将具有更高的毛利水平。 （2）定制服务本身的毛利差异 客户所在区域（境内或境外）以及该区域内市场的整体供需关系将综合决定定制服务的毛利水平。通常来讲，由于境外客户更加看重数据服务商品牌和综合服务能力，因此愿意给出更高溢价；此外，若某类数据为市场稀缺产品，例如具有较高进入壁垒的多模态、虚拟人等前沿类数据需求、或传统业务里的多语种数据，都可在一定时间内维持较高的溢价水平。因此，更高的境外客户占比以及更高进入壁垒的数据需求，都将导致定制服务的综合毛利的增加。 10、我们标准数据集是如何积累的？ 公司标准数据集产品的积累方式主要为基于公司对市场需求趋势的判断和共性需求的提炼能力，先于客户需求开发数据集。数据集产品的这种商业模式在行业内往往具有较高壁垒，一方面需要公司对未来需求趋势有精准把握，另一方面由于产品开发属于先投入后产出，因此需要公司具备充足的资金保障，只有具有大量行业经验+know-how积累以及资金充足的企业，才能具备产品开发能力。因此，产品模式也成为公司区别于其他竞争对手的一大特色，目前公司产品数据集储备已处于行业头部水平，产品的积累对公司未来的收入扩张和毛利提升都将起到重要作用。
附件清单（如有）	
日期	2025年6月9日