

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：2025-015

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	新华资产 兰宏阳 泰康资产 倪辰晖、颜媛、任慧峰 建信基金 刘志威 和谐健康险 朱之轩
会议时间	2025 年 9 月 4 日 2025 年 9 月 5 日
会议地点	线下交流
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 年上半年，公司收入增长的核心驱动力是什么？ 2025 年上半年公司收入增长 69.54%，主要是因为随着全球 AI 技术的快速发展和商业化落地，公司计算机视觉、自然语言 and 智能语音三大业务板块全面增长。在计算机视觉领域，视觉理解与生成技术的突破加速了在 AIGC 多模态内容生成（通用图文、3D 物体识别、高清通用场景视频、文旅图文等）、OCR 识别、数字人、多模态数据评估等场景的规模化应用，带动相关数据服

	务需求同比显著增长；自然语言方面，大模型语义理解、政务、法律合规等场景的应用落地以及海外巨头的全球化扩张，推动专业文本以及平行语料等文本数据市场持续扩容；智能语音业务则受益于科技巨头的国际化战略持续落地，高质量、多语种语音数据保持旺盛需求。此外，公司去年在东南亚新增建设的数据交付体系已进入稳定运营阶段，成功拓展海外定制服务市场并形成规模化，有望成为公司海外业务扩张的重要战略支点。 2、相比于去年同期，今年公司的收入结构是否有变化？ 2025 年上半年，公司收入结构相比去年同期有所变化，计算机视觉和自然语言业务占比提升，分别为 40%、14%，主要得益于相关领域的技术突破和市场需求增长。 在计算机视觉领域，受益于视觉大模型技术的快速迭代和应用场景的持续拓展，特别是图像/视频文本对等数据需求的爆发式增长，该业务板块实现强劲增长。同时，公司在东南亚的数据交付体系已稳定运营，进一步带动了视觉业务收入的进一步提升。在自然语言处理方面，大模型技术在政务、法律合规等专业领域的深度应用成效显著，叠加全球化扩张带来的多语言语料需求激增，共同推动了该业务的快速发展。相较而言，智能语音业务虽保持稳定增长，但增速相对平缓。因此，计算机视觉和自然语言处理这两大高增长业务在整体收入结构中的占比提升。 3、公司目前和华为合作的进展是什么？ （1）海天瑞声 x 昇腾 DeepSeek 数据飞轮智能体 在华为昇腾一体机中，整合进海天瑞声数据飞轮智能体平台、DOTS 数据标注平台及 AI 研发平台，该整合
--	--

	方案能有效解决企业数据安全问题，在确保数据安全的前提下，帮助企业以最小化的改造成本快速部署符合业务需求的智能体系统。 （2）陕西智慧文旅项目 华为与陕文投集团正合作打造陕西文旅行业示范项目，包括建设高质量文旅数据集、构建 AI 文旅垂直大模型，并开展“AI+文旅”试点应用。海天瑞声将基于其丰富的数据服务经验，深度参与该项目，持续为文旅产业提供高质量数据及场景化解决方案。 4、公司和运营商的合作情况如何？ 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，已快速成为中国移动重要的数据服务供应商。今年 7 月，公司也参与到中国移动牵头建设，聚合央企合力和产业链力量共同打造的人工智能焕新社区的发布，公司作为焕新社区唯一数据服务企业参与到项目建设中，公司将在社区中聚焦数据要素建设，贡献更多高质量数据集，为 AI 产业发展注入坚实数据动能。 5、数据要素业务的最新进展是什么？ 公司正积极探索推进数据要素相关业务，目前主要聚焦在 3 类模式：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。目前，公司已与 3 个承担数据标注基地建设任务的城市：成都、长沙、保定等地方政府成立合资公司，并与内蒙古呼和浩特签署战略合作协议，聚焦当地乳业、草种、文旅等特色产业，共建高质量数据集，有效驱动当地产业的数字化转型。此外，公司正积极拓展与上海、广州、宜宾等十余地的战略合作，进一步完善标注基地的全国性业务布局。 6、公司怎么看待具身智能的数据需求？目前公司的布局是什么？ 2025 年具身智能进入加速发展期，产业正从技术验证迈向场景应用，这对训练数据的质量、规模和场景化提出了更高要求。 为此，公司已积极布局：首先，依托自身在智能语音、计算机视觉和自然语言处理领域积累的丰富多模态数据能力，为具身智能模型训练提供支持。其次，公司携手石景山人形机器人数据训练中心共建“具身智能数据训练场”，致力于打造“数据采集-模型训练-场景验证”的全流程闭环，以解决行业数据瓶颈问题。此外，公司还发布了全国首个具身智能数据工程化服务平台，通过系统化、自动化的数据管理赋能开发者，提升研发效率。公司相信，通过上述战略布局，能够为具身智能产业提供核心数据支撑，开辟新的增长空间。 7、具身智能目前有几几种数据采集方式？ 目前行业主要依赖四种数据来源，互联网开源数据集（整合现有的公开机器人数据）、虚拟合成（利用仿
--	---

	真环境生成训练数据）、动作捕捉（记录人类在特定环境中的行为模式）、以及机器人遥操（通过人工远程操控获取真实场景下的操作数据）等。其中，虚拟合成技术虽解决了初期数据短缺问题，但由于“仿真差距”的存在，其数据质量往往难以满足高精度训练需求。而开源数据集则面临着数据规模有限（通常仅百万级）、采集条件单一等挑战，无法支撑复杂场景下（需要千万级数据）的模型训练需求。 因此，行业正积极探索虚实结合的数据闭环模式，通过真实遥操数据修正仿真偏差，同时结合多模态标注，持续提升数据的多样性与有效性。 8、公司标品数据集与定制化服务的区别是什么？ 产品数据集是先于客户需求形成的模拟数据，是公司区别于其他竞争对手的一大特色，基于公司对市场的判断和通用化需求的提取能力，其属于是一次性投入、未来重复授权销售，对于公司的营收、毛利有着重要作用；而定制业务的需求来源是客户的定向化需求，有些定制业务的原始数据来源是客户提供的实网数据，公司提供纯加工的服务。 客户的 AI 产品在上线之前及初期，因为其自身尚未产生实网数据，通常需要采购模拟型数据集进行算法模型的训练，在产品上线并运行一段时间、产生大量实网数据之后，则会提供实网数据给我们进行数据加工，加工的数据反哺到客户的产品上从而促进其产品的迭代、升级。之后，客户需要进行产品功能或语种的拓展，再次需要购买模拟数据集来支撑，后续再采购数据加工服务进行迭代。 产品+服务的组合一直是公司向市场提供的综合解决方案，是一个整体，服务于不同客户的不同研发阶段
--	--

	需求，其收入贡献比例在各年间也呈现较为一致的趋势。而产品+服务带来的数据积累，也哺育了公司的数据处理平台和相关算法不断提升，努力达到数据处理场景下的行业最优。 9、训练数据产品和服务的定价模式、收费模式是什么样的？价格变动趋势如何？ 定制服务定价模式：一般采用成本加成定价法。公司根据客户的具体服务需求预估项目成本，在预估成本的基础上，参考公司制定的指导毛利率水平，结合项目技术难度、复杂程度、时限要求等进行报价，并根据市场环境与客户协商，最终确定价格。 产品定价模式：一般采用需求导向定价法。公司综合考虑训练数据集的开发支出、市场需求程度、预计未来重复销售的频率等因素，制定产品标准价格及价格区间，在销售过程中，根据客户的实际需求情况，以价格区间为基础向客户报价，经双方协商确定最终销售价格。训练数据产品通常以单个数据集为单位进行定价，定价比较灵活。 价格走势主要由市场的供需关系决定。如果某类数据为市场稀缺数据，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。但在较为成熟的细分方向，比如中文智能语音数据领域，确实存在进入者增多、价格竞争的情况。因此，未来公司将主攻有较高技术壁垒，存在较大毛利空间的细分场景，尽力避免价格竞争带来的过度消耗。 10、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。
附件清单（如有）	
日期	2025 年 9 月 8 日