

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：2025-003

投资者关系活动类别	<input type="checkbox"/> 特定对象调研 <input checked="" 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 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	东方阿尔法 梁少文 兴证全球 陈泓志 中银基金 张令泓 建信养老 王天乐 华安基金 许瀚天 风炎投资 王鹏 诺昌投资 许耀文 和信金创投资 黄庆铭 国海证券 郭义俊
会议时间	2025年3月5日 2025年3月7日
会议地点	现场交流
上市公司接待人员姓名	投资者关系 袁璐
投资者关系活动内容介绍	1、2024年收入增长的驱动因素是什么？ 受益于大模型技术的快速发展以及应用场景不断落地，以智能终端厂商、科技互联网公司为代表的国内外科技巨头纷纷加大多模态数据投入以支撑其智能

终端、内容生成等领域的 AI 能力建设。在此背景下，以多语种、多音色为代表的智能语音业务需求、以及以指令微调、偏好对齐等为代表的自然语言业务需求同比均呈现大幅增长，整体上驱动公司营业收入同比显著增加。

2、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。

3、有部分专家说可用的训练数据已经用完了，未来训练数据会不会出现瓶颈？

可用的训练数据面临枯竭主要指的是互联网上可以直接获取的数据，而在具有更大价值量、更大体量的数据并未在互联网进行公开展示，例如政务数据、公共数据、各垂类公司的私域数据、以及个人信息数据等，但上述数据对于模型训练尤其是模型在向垂类以及细分场景拓展时，起到重要作用，所以，并不会存在数据枯竭的问题。未来，随着国家政策对公共数据开放、治理、流通支持力度的持续加大，以及国家对大模型行业化拓展的持续加码，训练数据源将会持续拓展，催生更大的训练数据市场。

4、标品化的产品数据集业务与定制化服务业务的区别是什么？

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到我们进行数据加工，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。

5、训练数据产品和服务的定价模式、收费模式是什么样的？

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。

6、公司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。

7、定制数据逐渐积累，是否可以转化为自有的数据产品？

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。

8、与客户自建数据团队相比，海天的优势是什么？

相较于客户自建团队，海天瑞声历来都是对接众多大型科技公司、头部人工智能企业、科研院所等，获得的信息是广泛的，项目经验丰富，同时积累了大量的

know-how，对数据的理解更广、更深刻，同时我们搭建了成熟的数据处理算法平台，通过更高效的人机交互实现降本增效，保证数据质量的同时能有效降低成本，为客户提供更高性价比的训练数据产品/服务。

#### 9、行业里的玩家增多，会不会出现价格战？

是否存在价格竞争主要取决于该领域是否较为存在较高壁垒。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；但对于外语种领域，数据服务商则会有更高的议价空间。所以，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。

此外，公司也将通过短期内大规模专项研发投入及持续的研发升级，进一步提高自研平台能力，通过智能化促进产能提升、效率提升、成本降低实现规模效应和盈利能力的提升。

#### 10、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 6 月末，公司已积累超过 1,650 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

	<p>大规模、提升效率、降低成本。</p> <p>(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。</p> <p>(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。</p>
附件清单（如有）	
日期	2025年3月12日