

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：2024-010

| | |
|---------------|--|
| 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 | 华夏基金 贾静雯 乐正资本 白海波 银河证券 张宸 国华兴盛资产 王霆 川流资本 杨子平 国泰君安证券资产管理 范明 霁峰资本 许斐 上海锴量 钱小英 |
| 会议时间 | 2024年6月6日 |
| 会议地点 | 上海策略会 |
| 上市公司接待人员姓名 | 证券事务代表：张哲 |
| 投资者关系活动主要内容介绍 | 1、请问公司2023年收入下滑的原因是什么？ 2023年，公司营业收入为1.70亿元，较去年同期下降35.33%，主要系受境外部分客户进行阶段性裁员、业务方向及研发节奏周期性调整等影响，部分客户2023年预算释放进度放缓，同时叠加2023年上半年数据出境相关法规落地实施的阶段性影响，公司境外收入同比 |

大幅下滑。境内业务方面，虽然宏观稳经济政策已初见成效，但国内仍面临复杂严峻的内外部环境考验，部分境内客户对集中性研发投入仍持谨慎态度，基础数据服务领域客户预算及需求释放出现阶段性减缓，叠加行业内竞争加剧，综合导致境内收入同比下滑。

2、公司 2024 年第一季度收入同比增长的原因是什么？

公司 2024 年第一季度收入增长主要系在大模型技术的逐步发展和应用领域不断拓展的带动下，全球各类型科技公司对人工智能领域的研发投入呈现出复苏并增长的趋势，带动以多语言智能语音、文本为代表的的数据需求快速增加，整体拉动公司第一季度营业收入同比大幅提升。

3、公司政企类项目目前有哪些进展？

目前，公司正在与北京、安徽、山东、河北等地进行对接。若相关合作达到信息披露要求，公司将及时履行信息披露义务。

4、大模型范式下的预训练阶段数据需求和传统的数据需求有什么区别？

大模型预训练阶段的数据需求和深度学习技术路线下的传统训练数据需求，两者在形态上基本一致，都是文本、语音和图像，但在数据规模、质量、来源等方面，预训练数据较传统训练数据会存在一定差异。例如，规模上，预训练数据的 token 量普遍在万亿量级，而传统模型数量则大约在 10 亿量级。从数据来源的角度看，由于大模型所需数据规模远超传统模型，因此其数据来源将更加丰富以满足规模化、多元化数据需求。具体来说，大模型数据来源除了来自传统的定向采集外，还将涉及版权数据、公共数据等新型海量高质量数据来源。

此外,数据处理的核心技术也存在一定差异。例如,由于大模型预训练阶段的原料数据规模更大,因此大模型预训练数据更加注重数据清洗的工程化能力,在预训练阶段需要结合原料数据特点以及所涉及的主题、领域等,对海量数据完成高质量清洗,这对数据服务商的工程化数据处理能力以及过往服务经验的积累都提出了更高要求。

5、目前已经有很多大模型陆续发布,未来数据需求是会持续增长还是下降?原因是什么?

随着更多大模型产品面向公众推出上线、以及不断进行版本迭代,未来的大模型相关数据需求会持续、显著增长,这符合当 AI 进行应用侧后,对数据的需求才会规模性增长的规律。预期该类增长将主要体现在预训练阶段对高质量、大规模的版权数据、公共数据、多模态数据、垂直领域的实际场景数据等方面的需求,以及不断提升其交互精准度的强化学习阶段各类数据需求等。例如,若某大模型应用领域为文学创作,就会需要来自版权方的散文、小说、诗词等数据进行创作能力训练;若某大模型拟提升其在民生政务领域的客服能力,则来自民生政务服务平台的此类对话数据将产生重要价值。未来,随着大模型向更高级、更多维、更广泛垂向拓展,大模型数据需求将呈现快速增长趋势。

6、强化学习阶段的数据服务,今年有何进展或者变化?

目前来看,随着各大模型的陆续上线,强化学习环节的整体数据需求在逐渐攀升,并在具体标注任务上呈现如下趋势:

(1) 逐渐向更多垂类拓展(例如,法律、金融、医疗);

| | |
|----------|---|
| | <p>(2) 强化学习标注的评价/评分指标变得更为丰富，会要求标注人员从更多维度对模型的问答进行评判和打分；</p> <p>(3) 由单模态向多模态转变：23 年主要的标注需求集中在文本类标注，今年开始逐步向多模态拓展（例如，文本-视频、文本-图像等）。</p> |
| 附件清单（如有） | |
| 日期 | 2024 年 6 月 11 日 |