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：2024-006

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	银华基金 姚荻帆、周晶、蒋婉莹、同颖茜
会议时间	2024 年 5 月 16 日
会议地点	腾讯会议
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、2024 年第一季度，公司收入增长原因是什么？ 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、文本为代表的数据需求快速增加，整体拉动公司营业收入同比大幅提升。 2、大模型预训练会涉及到版权数据，在版权数据方面，海天的价值是什么？ 海天瑞声的价值主要体现在多版权数据的汇集、版权数据的清洗、以及基于客户大模型的后续服务。首先，海天瑞声可以汇聚不同版权方的数据，针对客户需求进行不同版权数据的提供。同时，海天瑞声可以针对客户

	具体定制化需求，对版权数据进行精细化清洗。虽然版权数据本身已为高质量数据，但仍无法直接用于模型训练，需经过高质量清洗后才能使用。例如，通常需将版权数据中重复数据以及不符合法律法规的相关内容进行清洗，以更好帮助大模型节约训练算力以及使大模型在训练后具备良好的法律道德价值观。 3、海天在数据方面的核心竞争力是什么？ 经过多年发展与积累，公司逐步构建起了在行业内的竞争壁垒，核心竞争力主要体现在： （1）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。 （2）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 2023 年 12 月末，公司已积累超过 1,550 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。 （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。 公司全方位做好数据风险管控工作，通过了业内重要的 ISO/IEC 27001 体系认证、ISO27701 个人隐私信息安全管理体系认证，形成了具有自身特色的数据安全与隐私保护整体解决方案。同时，公司拥有北京市规划和自然资源委员会行政许可（乙级测绘资质）、等保三级备案证明；目前，公司符合 GDPR、《数据安全法》、《个人信息保护法》等一系列国际通用与国内法律法规的管理规范要求，获得了业务领域合作客户的高度认可。 4、未来自动化标注对海天业务会不会产生影响？ 自动化数据标注一直以来都是数据服务行业的发展趋势，同时也是数据服务企业的核心竞争能力之一，自动化标注的核心不是完全替代人类，而是提高人机协作效率，海天瑞声近年来在研发领域持续加大投入，不断提升公司数据生产的智能化水平。 5、公司如何看待合成数据这个技术？今后是否会对公司的业务产生不利影响？ 在数据重要性凸显且数据需求快速增长的时代，合成数据可以认为是人工智能行业发展到一定阶段的必然产物。数据合成技术可以作为数据采集的有效辅助，但也存在较强的局限性，降低真实世界各类特征的训练效果，因此目前仅可作为数据采集的一种辅助方式。从目前数据服务行业来讲，以计算机视觉场景为例，合成数据主要应用于某些高危的、罕见的 corner case 的模拟训练当中，但合成数据毕竟是由机器生成的虚拟数据，其数据质量以及真实性仍无法替代真实场景数据，因此按照目前的技术路线，绝大多数企业仍在使用真实

	场景数据进行模型训练。但公司会紧密关注合成数据技术的发展，根据最新的行业动态及时调整公司业务布局。 6、请问大模型向多模态发展后，是否会对公司业务产生正向影响？ 大模型向多模态发展后，将会产生更多的新型数据需求。例如文生图的多模态大模型，通过文字输入生成对应图片，这就需要机器理解文字语义的同时将理解的关键词与图片的关键标签进行映射，通过对齐两种独立模态关键特征的方式，实现按指令的创作，以此完成学习训练过程。因此，当大模型向多模态能力维度拓展时，高质量多模态训练数据集的持续学习训练的重要性将更加凸显，多模态的发展将推动数据服务行业进入更大的增量空间。 7、是否考虑接入大模型来提升自身的数据生产效率？ 智能化标注能力是数据服务企业的核心竞争力之一，公司一直致力于不断提升数据标注的智能化水平，目前公司已通过自行研发以及 API 接入两种方式，探索将大模型接入公司一体化数据处理平台，以提高数据处理过程中的人机协作效率，辅助公司的数据生产。自研大模型方面，公司已在针对大模型预训练数据集设计与处理技术进行初步研究和规划，并基于研究成果开展了 CommonCrawl、中文书籍等适用于预训练阶段的数据的获取与清洗工作，形成了各项技术的框架方向；开展大模型评测技术调研，完成基础框架设计，形成可行性结论。另一方面，公司已将部分开源或提供 API 接口的大模型接入公司一体化数据处理平台。例如，公司已在智能驾驶平台 DOTS-AD 中接入开源的语义分割模型 SAM

	（Segment Anything Model），并基于数据预标注实际需求，对模型进行了优化升级，有效提升了 2D 语义分割项目中的降本增效能力。
附件清单（如有）	
日期	2024 年 5 月 20 日