

公司代码：688787

公司简称：海天瑞声

北京海天瑞声科技股份有限公司
2022 年年度报告摘要

第一节 重要提示

1 本年度报告摘要来自年度报告全文，为全面了解本公司的经营成果、财务状况及未来发展规划，投资者应当到 www.sse.com.cn 网站仔细阅读年度报告全文。

2 重大风险提示

公司已在本报告中详细描述可能存在的风险，敬请查阅“第三节管理层讨论与分析”（之四）“风险因素”部分，请投资者注意投资风险。

3 本公司董事会、监事会及董事、监事、高级管理人员保证年度报告内容的真实性、准确性、完整性，不存在虚假记载、误导性陈述或重大遗漏，并承担个别和连带的法律责任。

4 公司全体董事出席董事会会议。

5 信永中和会计师事务所（特殊普通合伙）为本公司出具了标准无保留意见的审计报告。

6 公司上市时未盈利且尚未实现盈利

是 否

7 董事会决议通过的本报告期利润分配预案或公积金转增股本预案

经公司第二届董事会第十三次会议审议，公司 2022 年度利润分配方案拟定如下：

拟以实施权益分派股权登记日登记的总股本为基数分配利润，向全体股东每 10 股派发现金股利 6 元（含税），截至 2022 年 12 月 31 日，公司总股本 42,800,000 股，以此计算合计拟分派现金红利 25,680,000 元（含税），现金分红比例为 87.19%；同时以资本公积向全体股东每 10 股转增 4 股，合计转增 17,120,000 股，转增后公司总股本增加至 59,920,000 股。

2022 年度利润分配及资本公积转增股本预案披露后至实施权益分派的股权登记日期间，若公司总股本发生变动，公司将按照每股分红金额及每股转增比例不变的原则相应调整分配总额。

上述利润分配方案尚需提交公司 2022 年年度股东大会审议通过。

8 是否存在公司治理特殊安排等重要事项

适用 不适用

第二节 公司基本情况

1 公司简介

公司股票简况

适用 不适用

公司股票简况				
股票种类	股票上市交易所及板块	股票简称	股票代码	变更前股票简称
人民币普通股（A）	上海证券交易所	海天瑞声	688787	不适用

股)	科创板			
----	-----	--	--	--

公司存托凭证简况

适用 不适用

联系人和联系方式

联系人和联系方式	董事会秘书（信息披露境内代表）	证券事务代表
姓名	吕思遥	张哲
办公地址	北京市海淀区成府路28号4-801	北京市海淀区成府路28号4-801
电话	010-62660772	010-62660772
电子信箱	ir@speechocean.com	ir@speechocean.com

2 报告期公司主要业务简介

(一) 主要业务、主要产品或服务情况

1. 主要业务情况

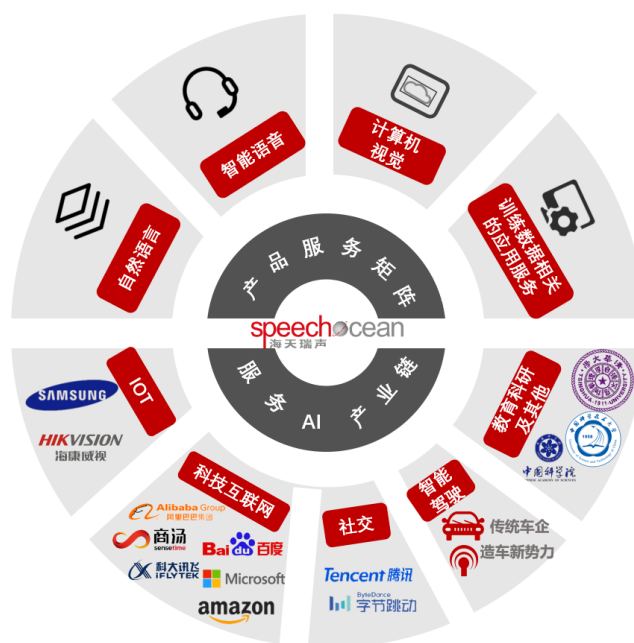
公司主要从事 AI 训练数据的研发设计、生产及销售业务。公司通过设计数据集结构、组织数据采集、对取得的原料数据进行加工，最终形成可供 AI 算法模型训练使用的专业数据集，通过软件形式向客户交付。

算法、算力、数据是人工智能技术发展的三大要素，其中训练数据是算法发展和演进的“燃料”。在当前技术发展进程中，深度学习算法是推动人工智能技术取得突破性发展的关键技术理论，而大量训练数据的训练支撑则是深度学习算法实现的基础。深度学习分为“训练”和“推断”两个环节：训练需要海量数据输入，训练出一个复杂的深度神经网络模型；推断指利用训练好的模型，去“推断”现实场景中的待判断数据，并得出各种结论。训练数据越多、越完整、质量越高，模型推断的结论越可靠。因此，要使算法模型实现从技术理论到应用实践的落地过程，就需要提供大量的训练数据，对算法模型加以训练。通常，从自然数据源简单收集取得的原料数据并不能直接用于深度学习算法的训练，必须经过专业化的采集、加工处理，形成相应的工程化数据集后才能供深度学习算法等算法、模型训练使用。

自 2005 年成立以来，公司始终致力于为 AI 产业链上的各类机构提供算法模型开发训练所需的专业数据集。经过多年发展，公司已成为人工智能基础数据服务领域具有较强国际竞争力的国内头部企业，并实现了标准化产品、定制化服务、相关应用服务全覆盖。公司所提供的训练数据涵盖智能语音（语音识别、语音合成等）、计算机视觉、自然语言等多个核心领域，全面服务于人机交互、智能家居、智能驾驶、智慧金融、智能安防等多种创新应用场景。

公司的产品和服务已获得阿里巴巴、腾讯、百度、科大讯飞、海康威视、字节跳动、微软、

亚马逊、三星、中国科学院、清华大学等国内外客户的认可，应用于其研发的个人助手、智能音箱、语音导航、搜索服务、短视频、虚拟人、智能驾驶、机器翻译等多种产品相关的算法模型训练过程中。目前公司客户累计数量为 810 家，覆盖了科技互联网、社交、IoT、智能驾驶、智慧金融等领域的主流企业，教育科研机构以及部分政企机构。



图：公司产品服务矩阵示意

2. 主要产品及服务情况

2.1 主要产品及服务按业务类型分类

公司研发、生产的训练数据覆盖了智能语音、计算机视觉及自然语言处理三大 AI 核心领域，广泛应用于算法模型的开发、训练、优化、应用场景拓展等环节。此外，公司还提供与训练数据相关的应用服务。

(1) 智能语音

人工智能在语音领域的应用技术主要包括语音识别、语音合成等。

语音识别 (Automatic Speech Recognition, ASR) 是让机器能够“听懂”人类语音的技术，它能使机器自动将语音信号转换为对应的文本信息。

语音合成 (Text to Speech, TTS) 是让机器能够“说出”人类语音的技术，它使机器能将文字信息转化为流畅的语音“朗读”出来，相当于给机器安上了人工嘴巴。

以日常生活中的情景为例，语音输入法、即时通讯软件运用了语音识别技术将用户输入的语音实时转换为文字，实现了软件“听懂”语音并“听写”出文字的效果；而地图、导航软件则运

用语音合成技术，实现了软件“发声说话”的效果，为用户提供即时语音导航。

公司通过设计（设计训练数据集结构、供发音人朗读录制的语料文本或对话场景、发音人分布、录音设备场景等）、采集（定义合适的发音人、选取录音设备及软件、组织发音人朗读录制音频）、加工（对音频文件进行切分、标注各类声音特征，形成带时间戳和特征标签的文本和标注文件等）、质检（对数据集进行质量检测，如音字一致性、标注准确率检查等）等训练数据集生产环节；或者针对客户提供的原料音频文件执行加工、质检工作，最终形成客户所需的智能语音训练数据集。

（2）计算机视觉

计算机视觉（Computer Vision, CV）是使机器具备“看”的功能的技术，它使得智能驾驶、智能家居、手机、安防设备等机器能够代替人眼对目标进行识别、跟踪和测量等。

以日常生活中的情景为例，在汽车的自动驾驶功能中，计算机视觉技术使得汽车能够“看见”并识别行车过程中的各种行人、路况场景，为后续作出相应的反应奠定基础；在机场、车站安检中，计算机视觉技术使得人脸识别设备能够识别被检验人员是否为其出示的身份证件显示的人员。

公司通过设计训练数据集结构、采集（如定义合适的人脸、动作、场景作为采集对象，组织被采集人按照要求拍摄照片、录制视频等）、加工（对图像、视频文件进行打点、拉框、分割标注等）、质检（对数据集进行质量检测，如检验图片、视频文件格式是否正确，检查光照环境、物体种类的数量是否达标，打点标框的准确率是否符合要求等）；或者对客户提供的图像、视频文件执行加工、质检工作，最终形成客户所需的计算机视觉训练数据集。

（3）自然语言处理

自然语言处理（Natural Language Processing, NLP）是以机器能够像人一样理解语言意图的技术。

以日常生活中的情景为例，寄送快递时使用的“智能填写”功能即运用了自然语言处理技术，在输入框中填入整段联系信息，软件应用能够理解语义，并从中识别及提取“收件人”、“联系方式”、“地址信息”等所需信息，完成自动填写；智能客服、聊天机器人等人机交互程序也运用了自然语言处理技术，使得程序、机器能够读懂人类语言的真正意图，并相应做出反应、提供服务等。

公司通过设计训练数据集结构、采集（收集或编写自然语言文本、对话等数据信息）、加工（对自然语言文本数据进行单词分割、词性标注、语义语法标注、情感属性标注等）、质检（对数据集进行质量检测，如检验文本、词性或者语义的标注结果是否准确等）；或者对客户提供的自然语言

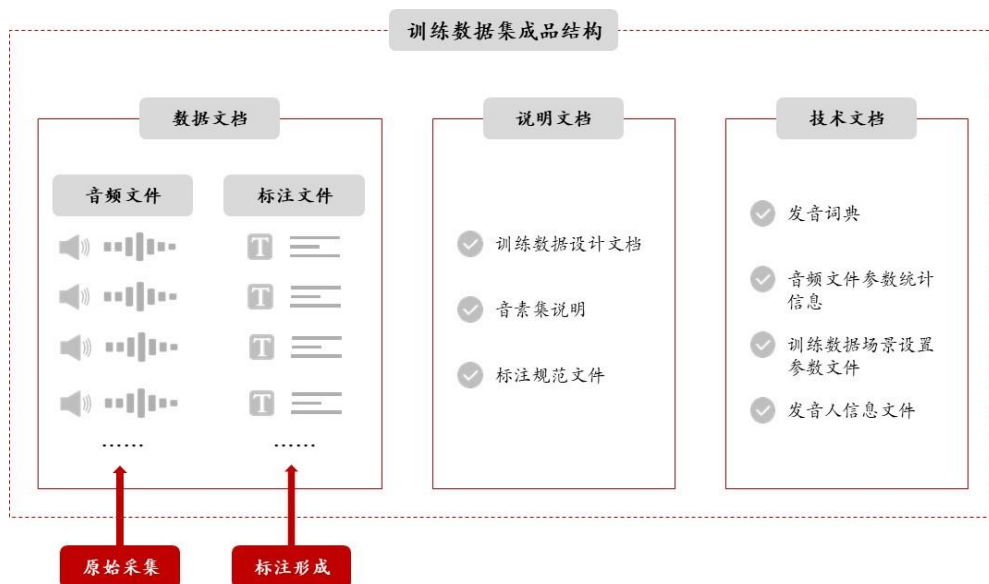
文本执行加工、质检工作，最终形成客户所需的自然语言训练数据集。

(4) 训练数据相关的应用服务

公司基于自身生产的训练数据提供算法模型相关的训练服务，运用训练数据研发能力助力下游客户完成其算法模型的语言拓展、特定算法模块拓展、垂直应用领域拓展等，为客户定制针对特定应用场景的专属算法模型，提高 AI 技术应用效果。

前述产品、服务均以公司生产的专业训练数据集为核心或基础。公司通过设计训练数据集结构、组织原料数据采集、对取得的原料数据进行加工，最终形成可供算法模型训练使用的专业数据集。

成品训练数据集主要由数据文档、说明文档、技术文档三部分构成。以智能语音训练数据集为例，成品训练数据集包含原始采集形成的音频文件、与音频文件对应的带有时间戳的标注文件，训练数据集相关的设计文档、训练数据集说明，发音词典，数据集参数信息文件等，图示如下：



图：训练数据集结构（智能语音）示例

2.2 主要产品或服务的终端应用场景

公司提供的高质量、大规模、结构化的训练数据，为算法模型的训练拓展提供了可靠的训练素材，助力 AI 技术实现实践应用及商业化落地，赋能 AI 技术与实体经济深度融合。公司提供的训练数据广泛应用于众多主流 AI 产品及终端应用的训练过程中，覆盖了智能驾驶、个人助手、语音输入、智能家居、智能客服、机器人、语音导航、智能播报、语音翻译、移动社交、虚拟人、智慧金融、智慧交通、智慧城市、机器翻译、智能问答、信息提取、情感分析、OCR 识别等多种应用场景。



图：训练数据集服务的算法模型应用场景示意

(二) 主要经营模式

1. 盈利模式

与主要产品及服务类型对应，公司的盈利模式主要包括以下三类：

(1) 定制服务：公司根据客户需求提供定制训练数据集并收取服务费。在此种模式下，公司享有服务费收入，不享有最终生成的训练数据的知识产权，不可将此类业务生产的训练数据向其他客户重复销售。

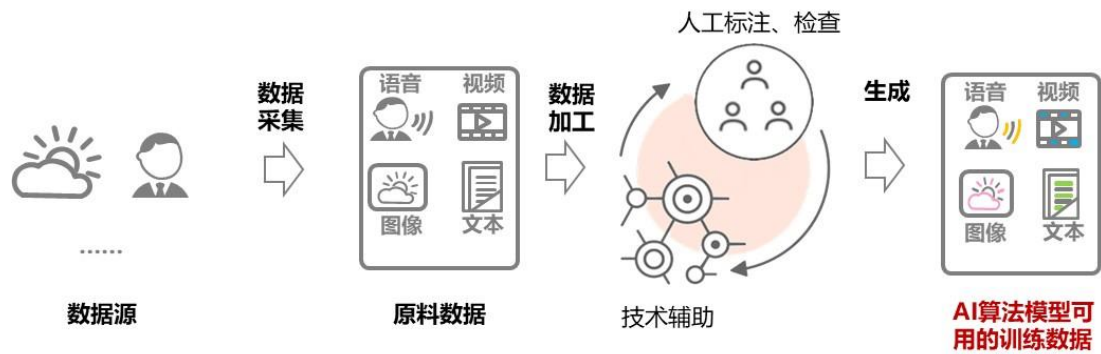
(2) 标准化产品：公司开发自有知识产权的训练数据集产品，通过销售训练数据集产品的使用授权许可，获取让渡资产使用权收入。此类训练数据集一经开发完成，可多次销售并获取授权许可收入。

(3) 训练数据相关的应用服务：公司基于生产的训练数据提供算法模型相关的模型拓展及训练服务，通常以软件授权或软硬件一体化形式交付算法模型拓展、开发成果，获取让渡资产使用权收入和技术服务收入，以及少量硬件销售收入。

2. 生产或服务模式

(1) 训练数据集生产模式

公司通过设计训练数据集结构、组织原料数据采集、对取得的原料数据进行加工，最终形成可供算法模型训练使用的专业数据集。



图：训练数据生产过程示意图

公司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）。

（2）训练数据相关的应用服务模式

公司基于其生产的训练数据提供算法模型相关训练服务，助力下游客户完成其算法模型的语言拓展、特定算法模块拓展、垂直应用领域拓展等，为客户定制针对特定行业和口音的专属算法模型，提高 AI 技术应用效果。

以某大型科技公司客户项目为例，客户研发了特定语音识别算法模型，需要根据算法模型的实际场景（如法院庭审场景）开发落地应用。公司承担了部分落地应用拓展相关的开发工作，围绕客户的算法模型和接口开发，最终协助客户算法模型实现多个麦克风收集庭审语音内容并实时转成文字记录入系统的功能。

3. 采购模式

按照采购的内容及主体划分，公司的采购包括：

数据服务采购：公司在数据采集、加工环节中，向人力资源服务等类型的公司等供应商采购的，非核心技术环节的原料数据采集、标注服务。

岗位服务采购：主要针对临时性的、不设长期岗位的业务领域的外包采购，如保洁、临时招聘服务、少量实习生招聘等。

其他采购：（1）训练数据生产所需的资产，主要包括软、硬件设备及其他需求物品采购；（2）日常运营所需的资产及物品，如办公用房、车辆、办公家具、计算机设备等；（3）日常专项服务采购等，主要包括审计服务、会议服务、差旅服务等。

上述原料数据采集、加工环节所涉及的数据服务采购，为公司最主要的采购类别，由集采中心负责；各部门岗位服务采购由人力资源部负责；其余日常运营相关的资产物品采购、专项服务采购等非业务采购由集采中心负责。财务中心负责参与采购供应商的遴选、监督与管理，并对采

购费用进行核算及结算。

经过多年的发展，公司已经建设有完善的《海天瑞声采购管理制度》、《海天瑞声项目资源采购管理制度》、《海天瑞声供应商管理制度》、《海天瑞声岗位服务采购管理制度》等内部规范制度，设立有完善的采购流程和体系，并与主要的供应商形成了良好稳定的长期合作关系。

4. 销售模式

公司采用直接对接并服务客户的直销模式进行营销，符合行业通行惯例。公司以高品质的训练数据集及相关服务吸引客户，并在持续服务客户的过程中提升服务价值和客户黏度。公司通过直接拜访潜在客户、口碑传播、参与学术会议和行业展会、官方网站和自媒体展示等方式建立品牌知名度、与客户建立联系，后续再通过商务谈判、招投标等形式获取具体业务机会。

(三) 所处行业情况

1. 行业的发展阶段、基本特点、主要技术门槛

根据国家统计局《战略性新兴产业分类（2018）》，公司所从事的训练数据生产业务属于“新一代信息技术产业—新兴软件和新型信息技术服务—新型信息技术服务—信息处理和存储支持服务—数据加工处理服务”行业，是国家重点支持的“新一代信息技术领域”的战略性新兴产业。公司通过设计训练数据集结构、执行数据采集、加工处理过程，生产用于算法模型开发训练用途的专业数据集，并以软件形式向客户交付，所属行业为软件和信息技术服务业。

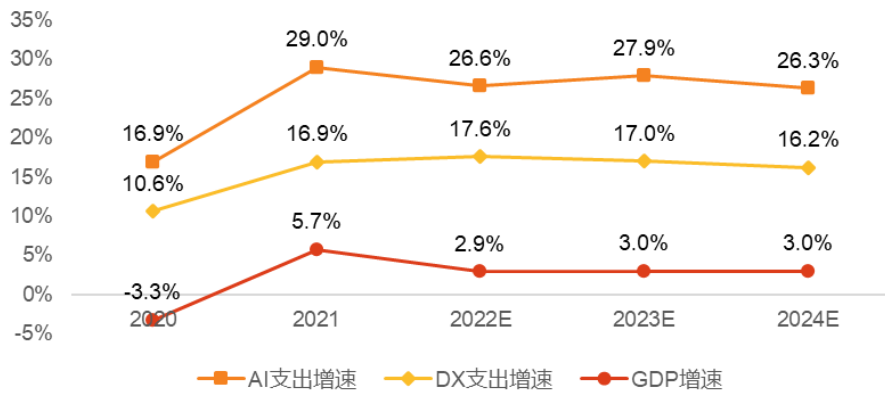
1.1 行业的发展阶段、基本特点

(1) 在全球数字经济发展以及以 ChatGPT 为代表的大模型的驱动下，全球人工智能产业即将进入新一轮加速发展期

当前，新一轮科技革命和产业变革深入发展，数字化转型成为大势所趋，世界主要国家均高度重视发展数字经济，纷纷出台战略规划，重塑数字时代的国际竞争新格局。人工智能作为数字经济发展的底层核心技术之一，正在发挥更加重要的作用。例如，随着数字经济发展的不断深入，数据体量以及复杂度均不断提升，为更好解决产业数字化中数据提取、处理、分析等工作，将会产生更多样化的人工智能需求，人工智能支出也将成为支持企业数字化转型支出的主力之一。

根据 IDC 报告，全球范围内，企业在人工智能市场的投资增速将显著高于数字化转型支出（DX）和 GDP 增速。

全球人工智能支出、数字化转型支出及 GDP 增速预测



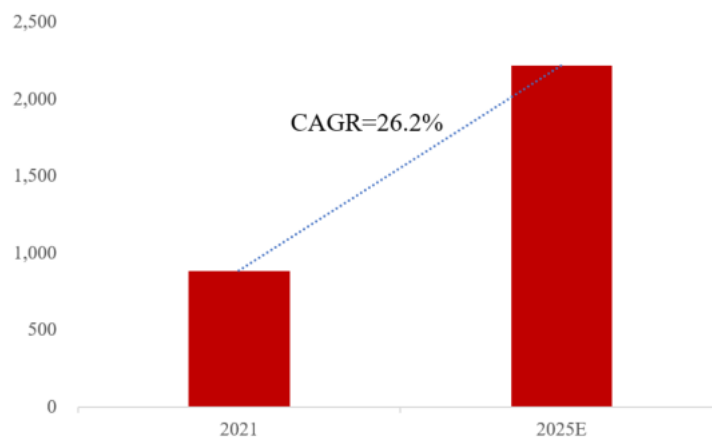
数据来源：国际数据公司（IDC）

此外，以 ChatGPT 为代表的大模型在近期的现象级智能化表现引发行业强烈关注。作为全新的人工智能聊天机器人，推出仅仅两个多月，ChatGPT 全球用户已突破 1 亿，人工智能通过极具颠覆性的能力再次走入大众视野，并获得了产业界的广泛关注。可以预见，人工智能行业将在以 ChatGPT、GPT-4 为代表的大模型技术的推动下进入新一轮产业高速发展期。

未来，受益于数字经济政策和大模型技术的双重驱动，人工智能将具备更强的产业融合能力，并将深刻影响千行百业的运行规则，以及人们的生活方式，人工智能产业的发展将随之进入快车道。

根据国际数据公司（IDC）的数据，2021 年，全球人工智能市场规模为 885.7 亿美元，预计 2025 年将达到 2,218.7 亿美元，年复合增长率达到 26.2%。

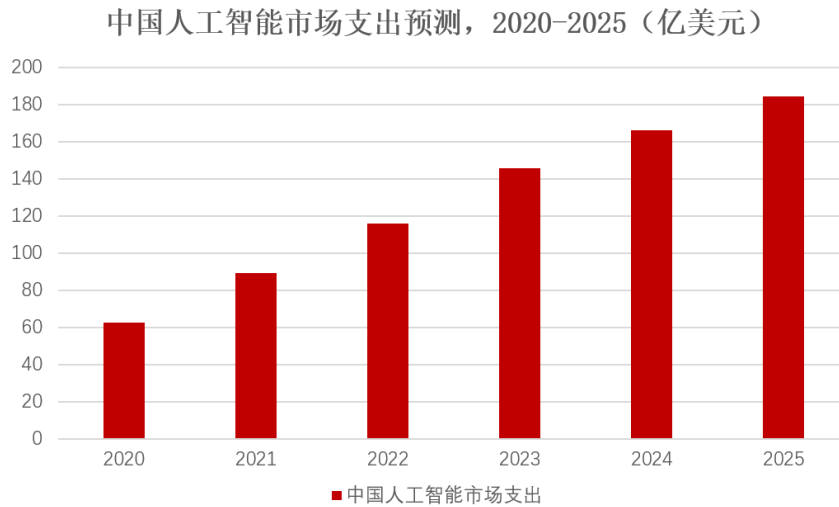
全球人工智能行业市场规模（亿美元）



数据来源：国际数据公司（IDC）

在市场需求拉动和国家政策的支持引导下，当前我国人工智能产业加速发展，已形成基础底层设施、中层技术、以及上层应用的完备的产业链生态，一批创新活跃、特色鲜明的创新企业不断涌现，并联合推动中国人工智能产业实现规模增长。根据国际数据公司（IDC）的数据，中国人工智能市场

规模预计 2025 年有望达 184.3 亿美元，年复合增长率达到 24.4%。



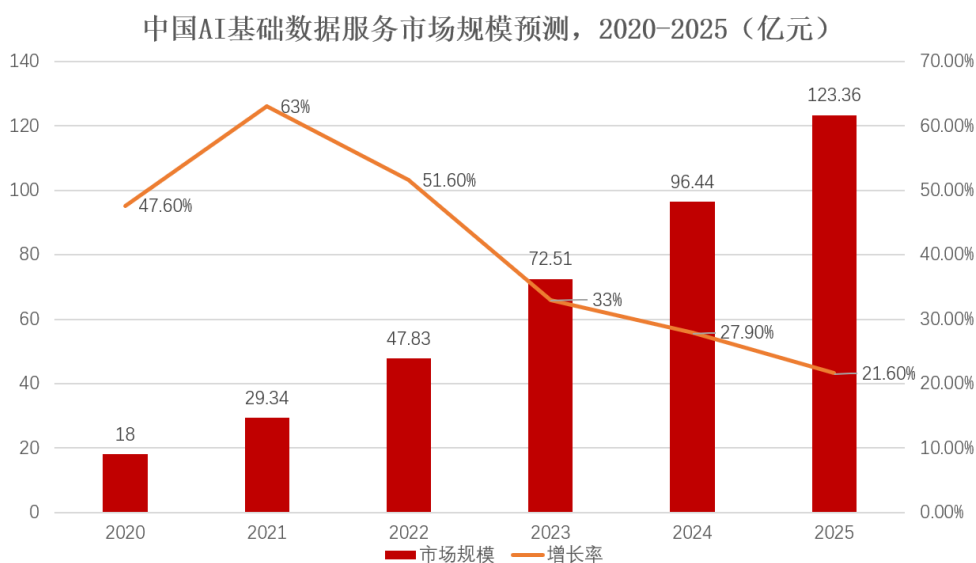
数据来源：国际数据公司（IDC）

（2）训练数据作为 AI 算法发展和演进“燃料”的作用更加凸显

在 AI 产业链中，算法、算力和数据共同构成技术发展的三大核心要素。算法模型从技术理论到应用实践的落地过程依赖于大量的训练数据。训练数据越多、越完整、质量越高，模型推断的结论越可靠。根据 Dimensional Research 的全球调研报告，72%的受访者认为至少使用超过 10 万条训练数据进行模型训练，才能保证模型的有效性和可靠性。此外，根据 Mckinsey Global Institute 研究，人工智能技术要求算法模型根据潜在的应用场景变化而持续更新，因此，算法模型所使用的训练数据亦需要定期更新。具体而言，约 1/3 的算法模型每月至少更新一次，约 1/4 的算法模型每日至少更新一次，算法模型持续更新的特点将进一步拓展各领域训练数据的需求空间。

此外，数据在人工智能三要素中的重要性也将更加凸显。过去十年人工智能产业是以算法为核心，通过深度学习算法的不断创新，推动人工智能产业的快速发展。但未来，当算法发展趋于开源，想要更快更好提升人工智能能力，数据将发挥更重要的作用。2021 年，全球人工智能和机器学习领域最权威的学者之一吴恩达教授提出二八定律：AI 研究 80%的工作应该放在数据准备上，确保数据质量是最重要的工作；业界如果更多地强调以数据为中心而不是以模型为中心，那么机器学习的发展会更快。由此可以看出，数据对于人工智能技术的长期发展的重要价值和核心作用，AI 产业的数据需求和多样性将会快速增加。根据研究机构 Cognilytica 预计，到 2027 年，全球数据标注市场规模将超过 190 亿美元。

同时，中国作为全球人工智能产业增速最快的国家之一，相关数据需求也在快速增长。根据 IDC 发布的《2021 年中国人工智能基础数据服务市场研究报告》，中国 AI 基础数据服务市场规模预期 2025 年将突破 120 亿元人民币，复合年增长率预期达到 47%，达到中国人工智能市场支出总额的约 10%。



数据来源：国际数据公司（IDC）

（3）训练数据领域的未来发展趋势

a. 人工智能技术加速向产业渗透融合，催生更多垂向领域数据需求

随着深度学习技术的不断突破，人工智能发展已经进入 2.0 时代，相关训练需求正逐渐从通用基础能力建设，向更为专业的垂向场景/行业拓展。一方面，以大模型为代表的 AI 基础技术不断取得重大突破，AI 模拟人类认知的能力飞速提升，因此从技术能力维度看，AI 已具备与垂直产业融合并规模化应用的前提条件；另一方面，受国家数字经济发展战略推动，产业数字化和智能化将进一步席卷各行各业，智能化技术与传统产业的融合将成为数字经济时代的新发展趋势，并创造出巨大的蓝海空间。

当前 AI 技术正在加速与各类产业融合，在汽车、金融、医疗、工业等传统行业的渗透率和应用场景不断提升，展现出可观的商业价值和巨大的发展潜力，而数据作为打通算法技术与行业需求的核心桥梁，作用更加凸显，可以说数据能力一定程度上决定了算法模型在对应产业的适用性以及实用性，成为加速 AI 产业化落地的关键要素。

在各类垂向场景中，智能驾驶受益于其产业本身的高速发展，以及以智能化为核心驱动的底层逻辑，相关数据需求呈现出快速增长趋势，因此成为数据服务行业首个产生规模化需求的垂向领域。未来，随着智能驾驶级别的不断提升、以及渗透率的不断提高，产业对智能驾驶的算法的精度要求会更高，一方面在硬件配置上，传感器的种类和数量会不断变化，另一方面，算法需要持续不断地学习更多场景的数据来提高罕见场景的识别能力，因此，未来智能驾驶的数据需求将向覆盖更多数据类型、更广泛应用场景以及更高质量的方向发展，数据需求也将原来单一的 2D 需求（主要集中在 L2 级别）向 3D、以及 2D-3D 融合需求拓展（L3 及以上级别）。

b.人工智能企业全球化布局加速，多语种能力成为拓展核心支撑

2013年，共建“一带一路”的倡议正式面世，十年来，随着国家“一带一路”战略的深入推进，国内一批具有较强创新能力和过硬技术实力的企业，纷纷踏出国门，积极拓展海外市场，通过不断扩大企业出海战略版图，获得高速发展机会。另一方面，境外头部企业也继续践行“全球化”战略，搭乘全球出海的快车。

随着境内、外企业的全球化扩张成为确定性趋势，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。未来，多语种训练数据将对客户侧在语音助手、智能汽车、智能家居、智能客服、机器人、多语种OCR等各领域产品/应用的全球化推广起到积极作用。因此，随着各类客户群体扩张步伐加速，多语种需求也将快速增长，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。

c.多模态数据受AIGC、元宇宙发展驱动，将呈现快速增长趋势

随着AI虚拟主播、虚拟学生、虚拟员工轮番上岗，虚拟数字人这个新兴概念逐渐走入大众视野，成为元宇宙与人工智能两大领域的热门技术赛道之一。想要让虚拟数字人实现与人类的自然交互，不仅需要发音标准自然、身体动作流畅，其表情、口型与声音也要实现细节的精准匹配，而多模态技术就是打破传统人工智能单一感官局限、让各类AI能力协同使用的重要技术。通过对高质量多模态训练数据集的持续学习，AI可实现图像、视频、音频、语义文本等多维度能力的融合，使得虚拟人在行为上更接近人类。

此外，随着AIGC技术的发展，AI将在更多维度上辅助人类创作，例如人类输入一段文字指令，AI通过理解输入的文字内容，按照人类描述生成一幅画或一段语音，以此帮助人类完成内容生产。想要实现上述功能，即需要通过对多模态数据进行学习，使文字与图画或语音形成一一映射，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。

因此，随着以虚拟人、AIGC为代表的AI技术的不断发展和应用，数据服务领域多模态需求将呈现加速增长趋势。

d.大模型技术的突破和跃升，将驱动新型数据需求持续增长

随着ChatGPT成为全球范围内的现象级应用，人工智能迎来了新的发展机遇，其背后的大模型技术也将进一步引导人工智能产业变革并带来相关数据需求的变化和增长。具体来说，为更好完成预训练阶段的通用模型训练，大量数据的质量要求将更高，高质量数据清洗的规则建立以及工程化能力将成为重点。此外，强化学习阶段的高阶数据需求将不断增加，基于人类认知的高质量的数据输入将更大程度决定模型效果。

可以预见，未来，大模型将成为 AI 时代的重要基础平台和新型基础设施，并会驱动、产生大量基于大模型的行业应用开发，AI 向各个产业的渗透将不断提升。以金融行业为例，近期工商银行发布了其自身的金融行业人工智能通用大模型，该模型与以往传统模型相比，具有更丰富的通用知识、更强的学习与推理能力，以及海量的金融专业知识，并在客户服务领域、风险防控领域、以及运营管理领域带来远超传统模型的服务效率提升。这个典型案例只是大模型应用的一个开端，未来大模型技术将通过其出色的表现，快速推广并影响更多的产业和领域，而连接行业需求和大模型技术的核心桥梁便是高质量、高精度的行业数据。因此，为加速实现 AI 产业化落地，行业将衍生出更多垂向场景的数据需求，大模型将通过不断学习各个垂直领域的行业高质量数据实现更广阔的垂向拓展。

e.随着国家法律法规更快更密集落地，对数据安全及合规会提出更高要求

近年来，数字经济规模的快速扩张，数据作为数字经济时代最核心、最具价值的生产要素，重要性更加凸显，但数据不同于传统生产要素，其承载涵盖了大量涉及个人隐私以及国家安全的重要信息，因此，为更好保障数字经济长期稳定的可持续发展，建设规范、安全、合规、高质量的数据安全体系已成为迫切需求。近年，国家陆续出台包括《数据安全法》、《个人信息保护法》、《汽车数据安全管理办法（试行）》、《数据出境安全评估办法》、《自然资源部关于促进智能网联汽车发展维护测绘地理信息安全的通知》等主流法律法规，为解决数据安全问题、净化行业快速发展中的不良乱象提供了切实可行的法律依据。

数据安全领域主要行业政策及法律法规如下：

序号	发布时间	发文机关	主要行业政策及法律法规	相关内容
1	2022年8月	自然资源部	《关于促进智能网联汽车发展维护测绘地理信息安全的通知》	《通知》进一步明确了测绘地理信息数据采集和管理等相关法律法规政策的适用与执行问题。《通知》明确规定，智能网联汽车相关测绘地理信息数据的收集、存储、传输、处理属于测绘活动，开展上述业务的企业，属于内资企业的，应依法取得相应测绘资质，或委托具有相应测绘资质的单位开展相应测绘活动；属于外商投资企业的，应委托具有相应测绘资质的单位开展相应测绘活动。

2	2022年7月	国家互联网信息办公室	《数据出境安全评估办法》	该办法明确了数据出境安全评估的适用范围，事前评估和持续监督相结合、风险自评估与安全评估相结合等原则，规定了应当申报数据出境安全评估的情形，以及数据出境安全评估的具体要求、评估程序、监督管理制度、法律责任以及合规整改要求等。
3	2021年8月	国家互联网信息办公室等	《汽车数据安全管理办法（试行）》	作为汽车数据安全领域出台的第一份有针对性的管理规定，明确了汽车数据处理者的责任和义务，规范汽车数据处理活动，对防范化解汽车数据安全风险、保障汽车数据依法合理有效利用具有重要意义。
4	2021年8月	第十三届全国人大常委会	《个人信息保护法》	进一步细化、完善个人信息保护应遵循的原则和个人信息处理规则，明确个人信息处理活动中的权利义务边界，健全个人信息保护工作体制机制。
5	2021年6月	第十三届全国人大常委会	《中华人民共和国数据安全法》	我国数据的使用和保护进入有法可依的新阶段，国家统筹发展和安全，坚持以数据开发利用和产业发展促进数据安全，以数据安全保障数据开发利用和产业发展。

未来，随着 AI 技术不断革新，应用场景不断增加，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供应商新的能力评价维度，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，牢牢把握发展与安全并重的原则的企业将具有更强的市场竞争力。

f.数据要素战略地位不断凸显，数据服务将释放更多增量空间

当前，我国开始进入由工业经济迈向数字经济的“新时代”，国家高度重视数字经济发展，而数据要素作为数字经济深化发展的核心引擎重要性更加凸显。习近平总书记在中共中央政治局就实施国家大数据战略进行第二次集体学习时曾指出：数据是新的生产要素，是基础性资源和战略性资源，也是重要生产力，要构建以数据为关键要素的数字经济。2022年1月12日，国务院印发《“十四五”数字经济发展规划》明确提出：数据要素是数字经济深化发展的核心引擎，坚持以数字化发展为导向，充

分释放要素价值，激活数据要素潜能。2022年12月20日，中共中央、国务院发布《关于构建数据基础制度更好发挥数据要素作用的意见》（简称“数据二十条”），指出“数据作为新型生产要素，是数字化、网络化、智能化的基础，数据基础制度建设事关国家发展和安全大局。要加快构建数据基础制度，充分发挥我国海量数据规模和丰富应用场景优势，激活数据要素潜能，做强做优做大数字经济，增强经济发展新动能，构筑国家竞争新优势”。

数据要素领域主要政策如下：

序号	发布时间	发文机关	主要行业政策及法律法规	相关内容
1	2022年12月	中共中央、国务院	《关于构建数据基础制度更好发挥数据要素作用的意见》（简称“数据二十条”）	提出数据作为新型生产要素，是数字化、网络化、智能化的基础。数据基础制度建设事关国家发展和安全大局，要维护国家数据安全，促进数据合规高效流通使用，以数据产权、流通交易、收益分配、安全治理为重点内容统筹推进，指出要激活数据要素潜能，加快发展数据要素市场，做大做强数据要素型企业，增强经济发展新动能。
2	2022年3月	国务院	国务院关于落实《政府工作报告》重点工作分工的意见	提出要促进数字经济发展，加强数字中国建设整体布局，建设数字信息基础设施，逐步构建全国一体化大数据中心体系，促进产业数字化转型。培育壮大人工智能等数字产业，完善数字经济治理，培育数据要素市场、释放数据要素潜力，提高应用能力。
3	2022年1月	国务院	《“十四五”数字经济发展规划》	明确指出数据要素是数字经济深化发展的核心引擎，坚持以数字化发展为导向，充分释放要素价值，激活数据要素潜能。

4	2021年3月	十三届全国人大四次会议	《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》	《十四五规划》指出要加快数字化发展，建设数字中国，同时打造数字经济新优势，充分发挥海量数据和丰富应用场景优势，促进数字技术与实体经济深度融合，赋能传统产业转型升级，催生新产业新业态新模式，壮大经济发展新引擎。同时指出要加强关键数字技术创新应用：聚焦高端芯片、操作系统、人工智能关键算法、传感器等关键领域；建设重点行业人工智能数据集，发展算法推理训练场景。
5	2020年3月	中共中央、国务院	《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》	首次将“数据”作为市场化要素写入国家顶层设计级别文件，提出要加快培育数据要素市场，发挥数据在市场化配置中的作用。

未来，随着数据要素作为国家级战略资源地位不断凸显，一系列国家引领与行业鼓励政策不断推进，数据作为当前最具时代特征的生产要素，已成为国家重点支持领域，为数据产业带来了巨大的发展机遇。具体来说，未来围绕数据采集、处理、流通、交易、应用等数据要素全生命周期，将会产生更多的数据服务形式和商业机遇。未来，数据服务商可基于过往服务经验和对数据要素市场的前瞻性判断布局，进一步拓展延伸自身数据服务空间，探索更多能力维度，迎战更广阔的数据要素市场。

1.2 行业的主要技术门槛

随着 AI 技术不断演进、产业应用不断丰富，训练数据的市场需求呈现体量、难度、复杂性、合规性持续上升的趋势，数据服务商须同时具备对人工智能核心算法的理解能力、前瞻性的专业数据集设计能力、丰富的语言覆盖能力及场景采集能力、算法辅助数据生产能力、以及数据合规管理能力，这使得行业的技术门槛持续提升，具体体现为：

(1) 在训练数据研发、生产全流程中的算法全面介入

随着 AI 技术应用落地的规模化效应凸显，客户对于数据规模和处理效率的要求不断提升，数据服务商须在研发、生产流程中全面引入算法以实现高效、合理的人机协作模式，进而实现降本增效的目标。一般而言，在训练数据研发、生产全流程中融入算法技术，可用于训练数据集的设计及训练数据生产的各个环节，例如调度不同类型的标注人员应对不同领域的任务、形成算法自

动处理能力以帮助标注人员提升效率、降低对人员的依赖（既有人员数量的降低、也有对人员标注能力要求的降低），并构建训练数据设计、加工相关的核心技术；也可用于检查训练数据集对算法模型的训练效果，进而保障训练数据集质量。

（2）平台工具链功能及适配性要求持续提升

当前，客户侧的数据采集、标注需求范围在逐渐拓宽，数据采集与标注需满足的 AI 应用场景比以往明显更加广泛、复杂，这就对数据服务商的平台工具能力提出了更高要求，平台上处理过大规模的数据、这些处理过的数据的多样性和复杂程度如何、算法引擎投票机制如何建立、置信区间如何设置、算法在平台中如何应用、数据流转的工程化程度如何等等这些因素都决定了平台的适配性和能力如何，并最终决定了数据处理的质量、效率、成本。

（3）语音语言学基础研究方面须有深厚积累

伴随语音技术进一步发展落地、并向各行各业和更多垂直场景不断渗透，同时受到中国企业出海需求、国外企业区域拓展需求两方面的支撑，客户在多语种、多音色、音素集、发音规则、发音词典等方面的要求在不断抬升，这意味着只有那些在语音语言学基础研究方面投入更多、拥有深厚积累的数据服务商才能满足客户在这方面的多元化需求。

因此，市场上仅有极少数企业通过长期自主研发的方式能够达到上述核心技术门槛，成为有能力向不同客户群体提供综合、高效、合规的数据产品及服务的供应商。

2. 公司所处的行业地位分析及其变化情况

项目	海天瑞声	Appen	数据堂	标贝科技
基本经营情况				
成立年份	2005 年	1996 年	2010 年	2016 年
市场地位概述	是我国最早从事训练数据研发销售的企业之一；国内首家且是目前唯一一家 A 股上市的人工智能训练数据服务企业	较早从事数据资源开发的数据资源产品服务提供商，经营历史较长，规模、体量较大	新三板挂牌企业，是国内较早从事数据交易、数据采标的服务商之一	-
员工数量	273	超过 1,125	262	未公开披露

市场占有率	中国 AI 基础数据服务行业第二名，海天瑞声的市场占有率为 12.9% ¹	未公开披露	未公开披露	未公开披露
主要客户/合作伙伴情况	大型科技公司，如阿里巴巴、Meta、腾讯、百度、字节跳动、微软、三星等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学、中国科学技术大学等	微软、亚马逊、谷歌等大型科技公司、汽车厂商及政府	包括百度、腾讯、阿里巴巴、奇虎 360、联想、科大讯飞等国内互联网和高科技企业，微软、NEC、Canon、Intel、Samsung、Fujitsu 等企业及在华研发机构	微软、百度、阿里、腾讯、京东、滴滴、字节跳动、网易、360、三星、小鹏、美的、中科大、中电科、中国银行等
客户数量	810 家 (截至 2022 年 12 月 31 日累计)	未公开披露	未公开披露	100 余家
技术研发及产品能力				
技术实力概述	海天瑞声自主研发了覆盖智能语音、计算机视觉及自然语言等全 AI 技术领域的一体化数据处理平台，其中服务智能驾驶的第四代 VDS 平台目前已可覆盖全部舱外场景、主流传感器以及多类前沿的数据处理需求；同时，同步嵌入高质量算法，在多个服务领域实现人机协作的效率的显著提升。	Appen 拥有人工智能辅助数据注释平台，在全球 170 多个国家与 100 多万名专业承包商合作，训练数据涵盖科技、汽车、金融服务、零售、医疗健康和政府等各个领域。	拥有人工智能数据与生产服务平台，可提供数据定制服务、人工智能数据集产品、人工智能数据处理平台私有化部署服务，数据采集范围遍及全球 30 多个国家，合作伙伴遍布世界 10 多个国家。	拥有语音合成模型和算法，通过算法+专业的人工数据处理方式，为客户提供优质的语音合成服务。拥有 TOBI 标注体系，通过自主研发的 TTS 评测系统，为客户提供高质量的数据服务。
应用领域	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言、音乐
拥有的	1,334 个	超过 250 个	291 个	190 个

成品训练数据集数量	(截至2022年12月31日)			
语种/方言覆盖能力	近200个	超过235个	15余个	10余个
已取得专利授权	31项(29项发明专利、1项实用新型专利及1项外观设计专利)	2项	14项	12项
计算机软件著作权数量	163项	未公开披露	149项	30项
综合能力				
数据安全能力	乙级测绘资质； ISO27001 信息安全管理体系认证、 ISO27701 隐私信息管理体系认证； 等保三级备案证明； 中国信通院数据安全推进计划成员单位； 公司撰写的《人工智能基础数据业务之个人信息收集活动的合规审计》案例获选中国信通院、中国内审协会“全国首届数字化审计论坛”评选的“个人信息保护合规审计先锋实践案例”	未公开披露	乙级测绘资质； ISO27001 信息安全管理体系认证、 ISO27701 隐私信息管理体系认证	ISO27001 信息安全管理体系认证、 ISO27701 隐私信息管理体系认证、 ISO27017 云服务信息安全管理体系认证、 ISO27018 公有云中保护个人身份信息的信息安全管理体系认证、 国家信息安全等级保护二级
资质荣誉	国家高新技术企业、国家专精特新“小巨人”企业、“北京市企业技术中心”、工信部“新一代人工智能产业创新重点任务揭榜优胜单位”等多个国家或市级重要奖项	不适用	国家高新技术企业、北京市专精特新“小巨人”企业	国家高新技术企业、中关村高新技术企业、北京市专精特新“小巨人”企业、优秀服务机器人企业奖

主要财务指标 (数据堂为 2021 年度/2021 年末数据, 海天瑞声、Appen 为 2022 年度/2022 年末数据)				
营业收入	2.63 亿元人民币	3.88 亿美元	1.24 亿元人民币	未公开披露
综合毛利率	64.73%	38.76%	57.88%	未公开披露
净利润	2,945.41 万元人民币	-2.39 亿美元	327.50 万元人民币	未公开披露
净利率	11.20%	-61.60%	2.65%	未公开披露

数据来源及说明:

1、Appen、数据堂、标贝科技数据:截至 2022 年 12 月,前述公司官网及公开披露信息;国家知识产权局中国及多国专利审查信息查询平台(<https://www.cnipa.gov.cn/>)、中国版权保护中心 CPCC 微平台等公开信息查询渠道及第三方机构查询信息。

2、海天瑞声数据:除特别标注外,均为截至 2022 年 12 月 31 日数据。

2.1 基本经营情况

海天瑞声是我国最早从事训练数据研发销售的企业之一,深耕行业近 20 年,积累了大量的技术产品以及客户服务经验,数据服务能力覆盖个人助手、语音输入、智能家居、智能客服、机器人、语音导航、智能播报、语音翻译、移动社交、虚拟人、智能驾驶、智慧金融、智慧交通、智慧城市、机器翻译、智能问答、信息提取、情感分析、OCR 识别等 19 个领域,并获得市场认可,根据 2021 年国际数据公司(IDC)《2021 年中国人工智能基础数据服务市场研究报告》,公司的市场占有率为 12.9%。此外,在多年为全球客户提供产品、服务的过程中,公司不断受到知名科技企业的认可,截至报告期末,累计服务客户数量为 810 家,微软、Meta、百度、腾讯、阿里巴巴、字节跳动等已成为公司头部客户。

海天瑞声是国内目前唯一一家 A 股上市的人工智能训练数据服务企业,公司的综合技术实力及综合业务能力也获得了包括资本市场在内的各方认可。

2.2 技术研发及产品能力

与同行业可比公司相比,公司在如下技术能力方面具备自身特点和一定优势:

(1) 平台能力:海天瑞声自主研发的一体化数据处理平台,可实现智能语音、计算机视觉及自然语言等 AI 领域全覆盖,其中服务智能驾驶的第四代 VDS 平台目前可覆盖全部舱外场景、主流传感器以及多类前沿的数据处理需求。

(2) 算法能力:公司凭借近 20 年数据服务经验以及丰富的数据积累,在智能语音、计算机视觉、自然语言等方面储备了大量的优质算法模型,截至报告期末,公司算法模型数量已增至 200

余个。算法能力的引入极大提升了公司人机协作效率，也是行业内各企业拉开成本差距的核心驱动因素之一。

(3) 语种/方言覆盖能力：多语种能力是人工智能企业全球化发展的底层核心支撑之一，为更好配合国内外企业的全球化扩张，公司在基础语音语言学研究方面进行了持续投入，截至报告期末，公司已积累近 200 个的语种/方言，覆盖语种数量全球市场领先，在服务行业多语种语音以及多语种 OCR 和手写体业务方面具备优势。

(4) 产品能力：公司凭借多年的研发积累、对行业发展趋势的深刻洞见、以及较为充足的资金实力，已累计完成超过 1,300 个自有知识产权的训练数据标准化产品的建设，在全球企业中稳居前列，标准化数据集产品的积累意味着公司在服务响应速度、价格以及毛利水平等方面都将具有更大优势。

(5) 知识产权储备

在专利储备方面，通过公开信息渠道可获悉的 Appen、数据堂、标贝科技的专利储备数量较少，公司在专利技术储备方面具备优势。在计算机软件著作权方面，数据堂及其子公司共拥有计算机软件著作权 149 项，标贝科技及其子公司共拥有计算机软件著作权 30 项；海天瑞声及其子公司共拥有计算机软件著作权 163 项，高于公开信息可查询的数据堂、标贝科技的计算机软件著作权数量，具备一定优势。

2.3 综合能力

(1) 数据安全能力

随着《数据安全法》、《个人信息保护法》等法律法规快速落地实施，法律环境正在快速趋严，在此背景下，数据安全、个人信息保护已成为行业用户选择数据采标服务时的重要考量因素以及新的行业壁垒。公司是行业内为数不多已获得乙级测绘资质的企业，意味着公司具备合规开展智能驾驶采集、标注等业务能力，可帮助公司更好抢占智能驾驶的数据服务市场先机。此外，2022 年公司根据实践经验总结、撰写的《人工智能基础数据业务之个人信息收集活动的合规审计》案例获选中国信通院、中国内审协会“全国首届数字化审计论坛”评选的“个人信息保护合规审计先锋实践案例”，数据安全管理工作获得了行业认可。

(2) 资质荣誉方面

公司在发展历程中获得了工信部“新一代人工智能产业创新重点任务揭榜优胜单位”、国家专精特新“小巨人”企业、国家高新技术企业、北京市企业技术中心、工信部 2022 年大数据产业发展试点示范、2022 服务贸易重点企业、2022 北京数字经济企业 100 强、北京国家人工智能创新应

用先导区示范等多个国家、北京市重要资质荣誉，行业头部地位及综合技术实力不断获得国家的充分认可。

2.4 主要财务指标

与 Appen 相比，公司整体营收规模较小，但公司整体盈利能力（包括毛利率、净利率）均明显优于 Appen；与数据堂相比，公司在营收规模、盈利能力等方面均具显现较强优势。

3. 报告期内新技术、新产业、新业态、新模式的发展情况和未来发展趋势

(1) 数据需求向海量、高质量、多元化方向演进，智能化水平成为数据服务商核心竞争力

随着 AI 应用场景日益丰富、以及产品智能化要求的不断提升，客户在数据规模、质量、多元化等方面的提出了更高的要求。以智能语音和计算机视觉领域为例，训练数据需求逐渐拓展至更多语种、更复杂场景、更多 AI 设备、更多音色、更多维的图像采集等维度，数据服务商除了要具备丰富的数据采、标经验，还需要拥有完善的多元化数据处理平台，同时，通过引入算法提升数据处理的质量和效率，降低成本，驱动行业向训练数据生产智能化的方向演进。

(2) 全球化发展的大背景下，多语种数据需求不断攀升

随着国家“一带一路”战略的进一步深入推进，我国本土头部企业纷纷走出国门，主动出海；与此同时，国外主流科技企业也在同步加速全球化布局，并呈现不断扩充、细化区域拓展策略的趋势。

在此背景下，多语种训练数据的需求迎来新一轮增长，除中、英、法、德、意、西、日、韩等常见语种外，客户还将在诸如东南亚、一带一路沿线国家地区的罕见小语种（尤其是亚洲小语种、中东欧小语种等）方向产生新的增量需求，未来或将向更多发展中国家持续拓展。因此，多语言/语种基础研究能力、以及在语言学领域的储备将成为数据服务领域的核心竞争力。

(3) 智能驾驶领域引领数据需求拓展至更多垂直场景，对行业提出更高要求

随着 AI 底层技术的持续发展创新，AI 已逐渐成为具备更强理解能力和推理能力的智能技术，极大提升了其与实体产业大规模融合和应用的可能；此外，人工智能作为国家发展数字经济以及产业数字化转型的枢纽，正在获得越来越多的政策和资本支持。在技术发展与政策推动的共同作用下，人工智能技术将向更多产业以及更广泛垂向场景渗透。

细分行业的专业知识、服务经验以及准入资质将成为衡量一家数据服务商是否具备垂直领域数据服务能力的重要考量因素。当前，以智能驾驶为代表的垂直领域已开始释放大规模训练数据需求，行业客户更加需要全栈式、闭环数据解决方案的支持，以满足智能驾驶业务的数据处理量更大、数据处理需求的迭代频次更高、合规要求更高等特点，这就要求数据服务商在专业能力（包

括但不限于对于驾驶场景、车辆传感器等要素的综合理解和实施能力)、综合能力(包括但不限于数据处理平台能力、质量管控能力、需求对接能力、项目响应能力、供应链资源管理能力等)、准入资质等方面同时满足并达到较高水准方能持续为该领域客户提供高水平支撑。

(4) 数据安全与合规能力将成为数据服务领域的新竞争壁垒

近年来,国家通过密集出台《数据安全法》、《个人信息保护法》等法律法规,加速规范数据治理体系,以保障国家数字经济的健康可持续发展。此外,随着全球化与数字经济的发展,数据在国际间的流动愈加频繁,为更好规范数据出境活动,2022年9月,国家颁布实施《数据出境安全评估管理办法》,以促进数据跨境流通的安全。作为数字经济时代里的数据服务企业,公司也深刻感受数据安全正在深刻改变着行业既往规则,数据安全及合规能力已逐渐成为数据服务商的核心竞争力之一。

在此背景下,客户在选择数据服务商时,将更加看重服务企业的数据安全及合规能力,甚至一些大型需求方在遴选数据服务商时已将此因素提升至重要级别。因此,数据服务商在此方面须紧跟国家法律法规要求的演变,相应调整、升级现行业务开展方式、数据安全及合规管理体系,及时获取合规资质(包括但不限于信息安全管理体系认证、隐私信息管理体系认证、网络安全等级保护测评等),积极开展有关数据出境风险自评、安全评估工作,并已依法向有关部门进行了安全评估申报,切实提升自身数据安全及合规能力,确保业务始终在健康、合规的环境下开展,并将自身在这方面的积累转化为竞争优势、助力业务发展。

(5) 数字经济发展催生更大增量市场空间,数据服务新业态新模式将不断涌现

百年变局加速演化,未来一个时期,在需求收缩、供给冲击、预期转弱三重压力下,发展数字经济将成为我国经济“换道超车”以及挖掘经济内生增长的重要战略举措。国家在数字经济建设方面坚定决心,通过《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》、《数字中国建设整体布局规划》等政策文件的密集发布以及组建成立国家数据局等方式,进一步统筹并加速落地数字经济发展战略,而数据要素作为深化数字经济发展的核心引擎,也将迎来新的发展机遇。未来,围绕数据确权、处理、利用和流通等环节将会产生巨大的增量市场空间,也会同期催生出数据服务领域的新业态、以及新的数据服务模式。因此,从行业需求和发展趋势来看,具备数据安全合规能力、数据智能化处理能力、以及行业资源和 know-how 的企业将能更好抢占数据要素市场竞争制高点。

(6) 大模型驱动 AI 全面提速,助推新型数据服务需求快速增长

放眼未来,从十年乃至更长的周期来看,我们认为人工智能大模型将对数据行业产生深远影

响，并带来新的数据需求和新的数据服务模式。未来，大模型发展的数据依赖程度将逐渐加深。首先，数据的质量以及数据清洗的工程化能力会极大拉开大模型预训练阶段的效果差距。其次，预期更多模型将采用类强化学习模式来进行特定领域或特定方向上的优化迭代，以使得机器能够以更加接近于人类期望的方式提供答案输出。

为实现上述目标，需首先建立基于提示（Prompt）的训练数据集的设计技术，通过建立不同数据集之间的异向性，尽可能提高有限数据集对于下游任务的覆盖能力。此外，也将更多依赖于数据集在基础模型反馈结果上的打分技术、迭代训练 Reward Model（类奖励模型）的技术、以及噪声数据过滤技术，特别是针对专业领域的知识处理，如何组建特定领域中高端标注工程团队也将成为重要课题。因此，未来数据处理将不再局限在传统的有监督学习标注，预期将向数据规模化清洗以及类强化学习等方向演进，未来具有更强的前瞻性研发能力、数据工程化能力以及更多行业资源的公司将获得更多市场青睐。

3 公司主要会计数据和财务指标

3.1 近 3 年的主要会计数据和财务指标

单位：元 币种：人民币

	2022年	2021年	本年比上年 增减(%)	2020年
总资产	876,927,792.15	840,663,396.09	4.31	477,350,038.99
归属于上市公司股东 的净资产	829,522,849.25	805,908,403.05	2.93	437,956,372.58
营业收入	262,887,869.44	206,476,533.04	27.32	233,373,953.01
归属于上市公司股东 的净利润	29,454,139.23	31,605,431.79	-6.81	82,081,021.91
归属于上市公司股东 的扣除非经常性 损益的净利润	10,149,073.69	21,067,433.20	-51.83	73,015,355.36
经营活动产生的现 金流量净额	30,658,908.30	-15,548,319.63	不适用	51,176,659.14
加权平均净资产收 益率(%)	3.61	5.59	减少1.98个百分点	20.68
基本每股收益(元 /股)	0.69	0.89	-22.47	2.56
稀释每股收益(元 /股)	0.69	0.89	-22.47	2.56
研发投入占营业收 入的比例(%)	35.86	29.31	增加6.55个百分点	18.64

3.2 报告期分季度的主要会计数据

单位：元 币种：人民币

	第一季度 (1-3 月份)	第二季度 (4-6 月份)	第三季度 (7-9 月份)	第四季度 (10-12 月份)
营业收入	47,876,427.23	66,908,014.75	52,375,032.05	95,728,395.41
归属于上市公司股东的净利润	9,437,853.93	9,987,521.73	-794,570.29	10,823,333.86
归属于上市公司股东的扣除非经常性损益后的净利润	4,521,816.29	3,717,979.20	-3,969,136.75	5,878,414.95
经营活动产生的现金流量净额	4,768,088.45	-8,203,469.87	-228,258.05	34,322,547.77

季度数据与已披露定期报告数据差异说明

适用 不适用

4 股东情况

4.1 普通股股东总数、表决权恢复的优先股股东总数和持有特别表决权股份的股东总数及前 10 名股东情况

单位：股

截至报告期末普通股股东总数(户)								4,019
年度报告披露日前上一月末的普通股股东总数(户)								10,913
截至报告期末表决权恢复的优先股股东总数(户)								0
年度报告披露日前上一月末表决权恢复的优先股股东总数(户)								0
截至报告期末持有特别表决权股份的股东总数(户)								0
年度报告披露日前上一月末持有特别表决权股份的股东总数(户)								0
前十名股东持股情况								
股东名称 (全称)	报告期内增减	期末持股数量	比例 (%)	持有有限售条件股份数量	包含转融通借出股份的限售股份数量	质押、标记或冻结情况		股东性质
						股份状态	数量	

贺琳	0	8,669,725	20.26	8,669,725	8,669,725	无	0	境内自然人
北京中瑞安投资中心（有限合伙）	0	4,954,128	11.58	4,954,128	4,954,128	无	0	其他
中移投资控股有限责任公司	0	3,855,000	9.01	0	0	无	0	国有法人
唐涤飞	-80,000	3,497,982	8.17	0	0	无	0	境内自然人
北京清德投资中心（有限合伙）	-100,000	2,445,463	5.71	0	0	无	0	其他
上海丰琬投资合伙企业（有限合伙）	0	1,880,374	4.39	0	0	无	0	其他
中国互联网投资基金管理有限公司—中国互联网投资基金（有限合伙）	0	1,290,000	3.01	0	0	无	0	其他
北京中瑞立投资中心（有限合伙）	-805,182	1,066,378	2.49	0	0	无	0	其他
中国工商银行股份有限公司—富国新兴产业股票型证券投资基金	719,951	719,951	1.68	0	0	无	0	其他
上海兴富创业投资管理有限公司（有限合伙）	-693,272	629,840	1.47	0	0	无	0	其他

<p>上述股东关联关系或一致行动的说明</p>	<p>上述股东中，1、公司控股股东、实际控制人贺琳持有 100%股权的北京创世联合投资管理有限公司为北京中瑞安投资中心(有限合伙)的普通合伙人、执行事务合伙人，并持有北京中瑞安投资中心(有限合伙) 36.67%的出资，贺琳为北京中瑞立投资中心(有限合伙)的有限合伙人，持有北京中瑞立投资中心(有限合伙) 9.0611%的出资；2、唐涤飞持有 50%股权的北京创慧科瑞投资管理有限公司曾为北京中瑞立投资中心(有限合伙)的普通合伙人、执行事务合伙人，并持有北京中瑞立投资中心(有限合伙) 29.41%的出资，唐涤飞曾为北京中瑞立投资中心(有限合伙)的委派代表；(见注)3、北京清德投资中心(有限合伙)普通合伙人、执行事务合伙人钟山及其配偶志鹏分别持有北京清德投资中心(有限合伙) 8.76%、4.31%的出资，志鹏同时持有北京创慧科瑞投资管理有限公司 50%的股权；北京创慧科瑞投资管理有限公司曾为中瑞立投资中心(有限合伙)普通合伙人、执行事务合伙人(见注)。4、中移投资控股有限责任公司的间接控股股东中国移动通信集团有限公司为中国互联网投资基金(有限合伙)的有限合伙人，持有中国互联网投资基金(有限合伙) 9.97%的出资，中国移动通信集团有限公司的全资子公司中移资本控股有限责任公司持有中国互联网投资基金(有限合伙)普通合伙人、执行事务合伙人中国互联网投资基金管理公司(持有中国互联网投资基金(有限合伙) 0.33%的出资) 16.36%的股权。除此之外，公司未知上述其他股东之间是否存在关联关系或属于一致行动人。</p>
<p>表决权恢复的优先股股东及持股数量的说明</p>	<p>不适用</p>

存托凭证持有人情况

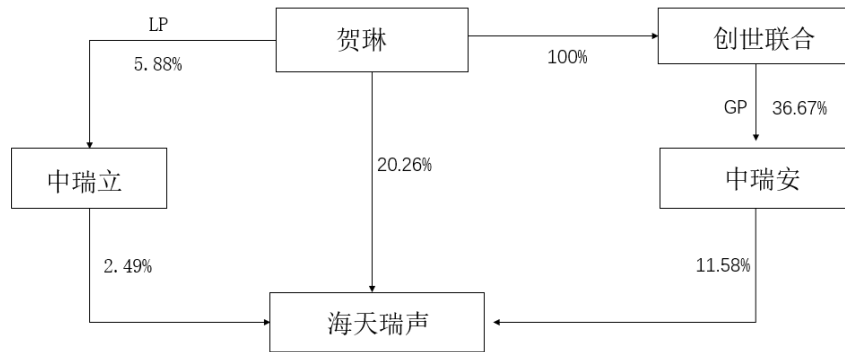
适用 不适用

截至报告期末表决权数量前十名股东情况表

适用 不适用

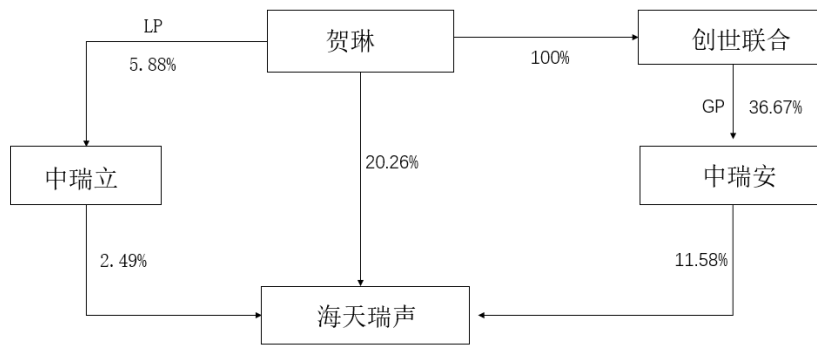
4.2 公司与控股股东之间的产权及控制关系的方框图

适用 不适用



4.3 公司与实际控制人之间的产权及控制关系的方框图

适用 不适用



4.4 报告期末公司优先股股东总数及前 10 名股东情况

适用 不适用

5 公司债券情况

适用 不适用

第三节 重要事项

1 公司应当根据重要性原则，披露报告期内公司经营情况的重大变化，以及报告期内发生的对公司经营情况有重大影响和预计未来会有重大影响的事项。

报告期内，公司实现业务收入 2.63 亿元，较上年同期增长 27.32%；实现归属于母公司所有者净利润 2,945.41 万元，较上年同期降低 6.81%；扣非后归母净利润 1,014.91 万元，较上年同期降低 51.83%。截至报告期末，公司总资产为 8.77 亿元，归属于母公司的所有者权益为 8.30 亿元，分别较上年末增加 4.31%和 2.93%。

2 公司年度报告披露后存在退市风险警示或终止上市情形的，应当披露导致退市风险警示或终

止上市情形的原因。

适用 不适用