

股票简称：海天瑞声

股票代码：688787

北京海天瑞声科技股份有限公司

(Beijing Haitian Ruisheng Science Technology Ltd.)

(北京市海淀区成府路 28 号 4-801)

海天瑞声

2023 年度向特定对象发行 A 股股票并在 科创板上市募集说明书

(申报稿)

保荐机构（主承销商）



华泰联合证券有限责任公司
HUATAI UNITED SECURITIES CO.,LTD.

(深圳市前海深港合作区南山街道桂湾五路128号前海深港基金小镇B7栋401)

公告日期：2023 年 7 月

重大事项提示

公司特别提请投资者注意，在作出投资决策之前，务必认真阅读本募集说明书正文内容，并特别关注以下重要事项及公司风险。

一、特别风险提示

本公司特别提醒投资者注意公司及本次发行的以下事项，并请投资者认真阅读本募集说明书“第七节 与本次发行相关的风险因素”的全部内容。

（一）业绩下滑或亏损风险

2022 年度，虽受境外业务复苏以及智能驾驶业务驱动，公司整体营收出现较高增长，但同期公司为更好建设长期技术优势和独具特色的标准化产品模式壁垒持续加大研发投入，且在 2022 年完成管理架构及激励体系升级、全球化营销体系建设等重要事项，以上因素共同导致公司人工投入及期间费用出现明显增长，使得公司净利润出现一定比例下滑。2023 年，受宏观经济增速放缓，市场竞争较为激烈以及数据出境相关法规落地实施所带来的影响，公司部分业务进度或将放缓。如果公司收入增长无法覆盖持续加大的人工投入及期间费用支出，公司业绩存在下滑或亏损的风险。

（二）核心技术快速迭代的风险

由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，公司进行数据开发所需的各项技术也面临着快速迭代更新的风险。如果公司无法保持对行业发展趋势的洞察能力以及对研发人才的有效组织和研发经费的经济投入，则可能导致公司在行业技术快速迭代过程中无法持续保持技术先进性和技术优势，进而对公司的客户拓展、产品创新和经营业绩造成不利影响。

（三）募投项目研发失败或研发成果无法转化的风险

由于大模型时代下存在数据规模极大、复杂性和多样性高，数据服务规则设计难度指数级提升等诸多问题，相关技术迭代速度快，未来发展趋势存在不确定性。若公司未能及时准确地把握行业发展趋势、突破技术难关，公司将面临研发失败或研发成果无法转化、前期的研发投入将难以收回的风险，并对公司的盈利

能力和市场竞争力造成不利影响。

（四）募投项目实施后效益不及预期的风险

公司本次部分募集资金将用于 AI 大模型训练数据集建设项目，该项目实施及未来商业化应用前景存在一定的不确定性，且公司对募投项目的效益测算系基于历史数据和未来公司及行业的发展趋势所做出的，但在募投项目的实施过程中，公司将可能面临着宏观经济及市场环境变化、下游需求放缓、市场竞争加剧、技术迭代更新等不确定或不可控因素的影响，募投项目有可能存在技术实施难度大、市场推广难、产品收益不及预期等情况，本次募集资金投资项目存在不能完全实现预期目标或效益的风险。

（五）募投项目新增折旧、摊销的风险

本次募投项目实施后公司的固定资产规模将增长较大，每年也将随之新增较多的固定资产折旧费用。如果市场情况发生重大不利变化，募投项目无法如期推进或不能产生预期收益，则公司可能存在因固定资产折旧、摊销费用大幅增加而导致利润下滑的风险。

二、本次向特定对象发行股票情况

1、本次向特定对象发行股票相关事项已经公司第二届董事会第十七次会议及 2023 年第一次临时股东大会审议通过。根据有关法律法规的规定，本次向特定对象发行股票尚需获得上海证券交易所审核通过并经中国证监会作出同意注册决定后方可实施。

2、本次发行的发行对象为不超过三十五名（含三十五名）符合中国证监会规定条件的特定投资者，包括符合规定条件的证券投资基金管理公司、证券公司、信托公司、财务公司、保险机构投资者、合格境外机构投资者以及其他符合相关规定的法人、自然人或其他合格的投资者等。其中，证券投资基金管理公司、证券公司、合格境外机构投资者、人民币合格境外机构投资者以其管理的二只以上产品认购的，视为一个发行对象；信托公司作为发行对象的，只能以自有资金认购。

最终发行对象将在本次发行申请获得上交所审核通过并由中国证监会作出予以注册决定后，按照相关法律法规的规定及监管部门要求，由公司董事会或董

事会授权人士在股东大会的授权范围内，与保荐机构（主承销商）根据询价结果协商确定。若国家法律、法规对向特定对象发行 A 股股票的发行对象有新的规定，公司将按新的规定进行调整。

所有发行对象均以人民币现金方式按照同一价格认购本次发行的股票。

3、本次发行的股票数量按照募集资金总额除以发行价格确定，同时本次发行股票数量不超过本次发行前公司总股本 60,325,180 股的 20%即 12,065,036 股（含本数），最终发行数量上限以中国证监会予以注册决定要求的发行数量上限为准。在前述范围内，最终发行数量将在本次发行由中国证监会作出予以注册决定后，由公司董事会或董事会授权人士根据公司董事会的授权和发行时的实际情况与本次发行的保荐机构（主承销商）协商确定。

在本次发行的董事会决议公告日至发行日期间，若公司发生送红股、资本公积金转增股本、股权激励、股票回购注销等事项引起公司股份变动，本次发行股票数量的上限将进行相应调整。

若国家法律、法规及规范性文件对本次发行的股票数量有新的规定或中国证监会予以注册的決定要求调整的，则本次发行的股票数量届时相应调整。

4、本次发行采取询价发行方式，定价基准日为发行期首日。

本次发行价格不低于定价基准日前 20 个交易日（不含定价基准日，下同）公司股票交易均价的 80%（即“发行底价”）。定价基准日前 20 个交易日公司股票交易均价=定价基准日前 20 个交易日公司股票交易总额/定价基准日前 20 个交易日公司股票交易总量。

若公司股票在定价基准日至发行日期间发生派息、送股、资本公积转增股本等除权除息事项，本次发行底价将按以下办法作相应调整。调整公式为：

派发现金股利： $P1=P0-D$

送红股或转增股本： $P1=P0/(1+N)$

两项同时进行： $P1=(P0-D)/(1+N)$

其中，P0 为调整前发行底价，P1 为调整后发行底价，每股派息/现金分红为 D，每股送红股或转增股本数为 N。

本次发行的最终发行价格将在公司本次发行申请获得上交所审核通过并经中国证监会作出予以注册决定后，由董事会或董事会授权人士根据股东大会的授权和保荐机构（主承销商）根据询价结果协商确定，但不低于前述发行底价。

5、本次发行的发行对象认购的股份自发行结束之日起 6 个月内不得转让。法律法规、规范性文件对限售期另有规定的，依其规定。

本次发行结束后，发行对象取得的本次发行的股份由于公司送红股、资本公积金转增股本等原因所衍生取得的公司股份，亦应遵守上述限售期安排。限售期结束后发行对象减持认购的本次发行的股票将按届时有效的中国证监会及上交所的有关规定执行。

6、本次发行股票募集资金总额不超过 78,989.00 万元（含本数），扣除相关发行费用后将用于投资以下项目：

单位：万元

序号	项目名称	项目投资总额	拟投入募集资金金额
1	AI 大模型训练数据集建设项目	38,337.36	38,337.36
2	数据生产垂直大模型研发项目	40,651.64	40,651.64
合计		78,989.00	78,989.00

在本次发行募集资金到位前，公司将根据募集资金投资项目的实际情况，以自筹资金先行投入，并在募集资金到位后按照相关法律、法规规定的程序予以置换。募集资金到位后，若扣除发行费用后的实际募集资金净额少于拟投入募集资金总额，在本次发行募集资金投资项目范围内，公司将根据实际募集资金数额，按照项目的轻重缓急等情况，调整并决定募集资金的具体投资项目、优先顺序及各项的具体投资金额，募集资金不足部分由公司自筹解决。

若本次发行募集资金总额因监管政策变化或发行注册文件的要求予以调整的，则届时将相应调整。

7、本次发行前公司滚存的未分配利润，由本次向特定对象发行股票完成后的新老股东按照本次发行后的股份比例共享。

8、本次发行决议的有效期为自公司股东大会审议通过之日起十二个月。

9、公司积极落实《上市公司监管指引第 3 号——上市公司现金分红》（中国

证监会公告[2022]3 号)、《关于进一步落实上市公司现金分红有关事项的通知》(证监发[2012]37 号)等相关法律法规和规范性文件的规定,结合公司实际情况,制定了《北京海天瑞声科技股份有限公司未来三年(2023 年-2025 年)股东分红回报规划》。关于利润分配和现金分红政策的详细情况,详见本募集说明书“第二节 发行人基本情况”之“十、报告期内利润分配政策、现金分红政策的制度及执行情况”。

10、本次向特定对象发行股票完成后,随着募集资金的到位,公司的总股本和净资产规模将相应增加。由于募集资金投资项目的使用及实施需要一定时间,因此本次发行存在每股收益等指标在短期内被摊薄的风险。为保障中小投资者的利益,公司就本次向特定对象发行股票事项对即期回报的影响进行了认真分析,并制定填补被摊薄即期回报的具体措施。相关情况详见《海天瑞声关于 2023 年度向特定对象发行 A 股股票摊薄即期回报与填补措施及相关主体承诺的公告》。特此提醒投资者关注本次发行摊薄股东即期回报的风险;同时,虽然公司为应对即期回报被摊薄风险而制定了填补回报措施,但所制定的填补回报措施不等于对公司未来利润做出保证。投资者不应据此进行投资决策,投资者据此进行投资决策造成损失的,公司不承担赔偿责任。提请广大投资者注意。

目 录

重大事项提示	1
一、特别风险提示.....	1
二、本次向特定对象发行股票情况.....	2
目 录.....	6
第一节 释 义	9
第二节 发行人基本情况	13
一、发行人基本信息.....	13
二、股权结构、控股股东及实际控制人情况.....	13
三、所处行业的主要特点及行业竞争情况.....	14
四、主要业务模式、产品或服务的主要内容.....	27
五、发行人主要资产状况.....	35
六、科技创新水平以及保持科技创新能力的机制或措施.....	37
七、现有业务发展安排及未来发展战略.....	41
八、截至最近一期末，不存在金额较大的财务性投资的基本情况.....	45
九、最近一年及一期业绩下滑情况及原因.....	50
十、报告期内利润分配政策、现金分红政策的制度及执行情况.....	57
十一、同业竞争情况.....	64
十二、发行人及其董事、监事、高级管理人员等相关主体的合法合规情况.....	65
第三节 本次证券发行概要	67
一、本次发行的背景和目的.....	67
二、本次向特定对象发行股票的发行对象及其与发行人的关系.....	73
三、发行证券的价格或定价方式、发行数量、限售期.....	73
四、募集资金金额及投向.....	75
五、本次发行是否构成关联交易.....	76
六、本次发行是否将导致公司控制权发生变化.....	76
七、本次发行方案取得有关主管部门批准的情况以及尚需呈报批准的程序.....	76

第四节 董事会关于本次募集资金使用的可行性分析	77
一、本次募集资金投资项目的的基本情况.....	77
二、募投项目的必要性及可行性.....	79
三、募投项目经营前景.....	88
四、募投项目与现有业务或发展战略的关系.....	88
五、募投项目的实施准备和进展情况，预计实施时间，整体进度安排，发行人的实施能力及资金缺口的解决方式.....	89
六、募投项目投向科创领域，募投项目实施促进公司科技创新水平提升的方式.....	91
七、募投项目的效益测算.....	92
八、本次募投项目非资本性支出的情况.....	94
九、募投项目的审批情况.....	94
十、募集资金用于研发投入的情况.....	95
第五节 董事会关于本次发行对公司影响的讨论与分析	98
一、本次发行完成后，上市公司的业务及资产的变动或整合计划.....	98
二、本次发行完成后，上市公司控制权结构的变化情况.....	98
三、本次发行完成后，上市公司新增同业竞争情况.....	98
四、本次发行完成后，上市公司新增关联交易情况.....	99
五、本次发行完成后，上市公司科研创新能力的变化.....	99
第六节 最近五年内募集资金运用的基本情况	100
一、前次募集资金的数额、资金到账时间以及资金在专项账户中的存放情况.....	100
二、前次募集资金实际使用情况说明.....	101
三、前次募集资金使用的其他情况.....	105
四、前次募集资金投资项目实现效益情况.....	106
五、前次募集资金中用于认购股份的资产运行情况说明.....	106
六、前次募集资金实际使用情况的信息披露对照情况.....	106
七、前次募集资金使用对发行人科技创新的作用.....	106
八、会计师对于前次募集资金使用情况的结论性意见.....	107
第七节 与本次发行相关的风险因素	108

一、业绩下滑或亏损风险.....	108
二、市场竞争加剧的风险.....	108
三、核心技术快速迭代的风险.....	108
四、核心技术人员流失风险.....	109
五、训练数据标准化产品无法实现授权销售的风险.....	109
六、应收账款回收的风险.....	109
七、数据安全相关风险.....	109
八、宏观经济波动的风险.....	110
九、募集资金投资项目相关风险.....	110
十、与本次发行有关的风险.....	111
第八节 与本次发行相关的声明	113
一、发行人及全体董事、监事、高级管理人员声明.....	113
二、发行人控股股东、实际控制人声明.....	114
三、保荐人声明.....	115
四、发行人律师声明.....	117
五、会计师事务所声明.....	118
六、董事会声明.....	119
附 件.....	123
附件一：发行人及其控股子公司拥有的商标.....	123
附件二：发行人及其控股子公司拥有的中国境内专利.....	125
附件三：发行人及其控股子公司拥有的中国境内计算机软件著作权.....	127

第一节 释 义

在本募集说明书中，除非文中另有所指，下列词语或简称具有如下特定含义：

一、基本定义		
公司/本公司/海天瑞声/发行人/上市公司	指	北京海天瑞声科技股份有限公司
本次向特定对象发行/本次发行	指	北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票的行为
A 股	指	人民币普通股
定价基准日	指	本次发行的发行期首日
股东大会	指	北京海天瑞声科技股份有限公司股东大会
董事会	指	北京海天瑞声科技股份有限公司董事会
监事会	指	北京海天瑞声科技股份有限公司监事会
中国证监会/证监会	指	中国证券监督管理委员会
上交所	指	上海证券交易所
报告期、最近三年及一期	指	2020 年、2021 年、2022 年、2023 年 1-3 月
《公司法》	指	《中华人民共和国公司法》
《证券法》	指	《中华人民共和国证券法》
《科创板上市规则》	指	《上海证券交易所科创板股票上市规则》
《管理办法》《注册管理办法》	指	《上市公司证券发行注册管理办法》
《章程指引》	指	《上市公司章程指引》
《公司章程》	指	《北京海天瑞声科技股份有限公司章程》
元、万元、元/股	指	人民币元、人民币万元、人民币元/股
控股股东、实际控制人	指	贺琳
海天瑞声有限	指	北京海天瑞声科技有限公司，为发行人的前身
中瑞智	指	北京中瑞智科技有限公司，为发行人的全资子公司
安徽瑞天数智	指	安徽瑞天数智科技有限公司，为发行人的全资子公司
山西瑞天数智	指	山西瑞天数智有限公司，为发行人的全资子公司
香港海天瑞声	指	海天瑞声（香港）科技有限公司，为发行人的全资子公司
DATAOCEAN	指	DATAOCEAN AI INC.，为香港海天瑞声的全资子公司
中瑞安	指	北京中瑞安投资中心（有限合伙），为公司股东
中瑞立	指	北京中瑞立投资中心（有限合伙），为公司股东
中移投资	指	中移投资控股有限责任公司，为发行人的股东
清德投资	指	北京清德投资中心（有限合伙），为发行人的股东

上海丰琬	指	上海丰琬投资合伙企业（有限合伙），为发行人的股东
中网投	指	中国互联网投资基金（有限合伙），为发行人的股东
上海兴富	指	上海兴富创业投资管理中心（有限合伙），曾为发行人的股东
天津金星	指	天津金星创业投资有限公司，曾为发行人的股东
杭州银杏数	指	杭州银杏数股权投资合伙企业（有限合伙），曾为发行人的股东
杭州士兰	指	杭州士兰创业投资有限公司，曾为发行人的股东
芜湖青和	指	芜湖朗姿青和一号股权投资合伙企业（有限合伙），曾为发行人的股东
芜湖博信	指	芜湖博信五号股权投资合伙企业（有限合伙），曾为发行人的股东
华泰联合证券	指	华泰联合证券有限责任公司，发行人保荐机构、主承销商
信永中和	指	信永中和会计师事务所（特殊普通合伙），发行人审计机构
天元	指	北京市天元律师事务所，发行人律师
《审计报告》	指	毕马威华振会计师事务所（特殊普通合伙）于 2021 年 4 月 30 日出具的《审计报告》（毕马威华振审字第 2103527 号）、信永中和于 2022 年 4 月 28 日出具的《北京海天瑞声科技股份有限公司 2021 年度审计报告》（XYZH/2022BJAA160966）、信永中和于 2023 年 4 月 24 日出具的《北京海天瑞声科技股份有限公司 2022 年度审计报告》（XYZH/2023BJAS2B0234）
Appen	指	Appen Limited，是一家成立于澳大利亚的训练数据产品和服务提供商
数据堂	指	数据堂（北京）科技股份有限公司
标贝科技	指	标贝（北京）科技有限公司，是一家国内同行业的训练数据产品和服务提供商
二、专业术语		
训练数据	指	通过采集与处理等步骤形成的、结构化的、可供人工智能算法模型训练使用的数据
算法	指	解题方案的准确而完整的描述，是一系列解决问题的清晰指令，算法代表着用系统的方法描述解决问题的策略机制
算力	指	计算能力
人工智能/AI	指	英文为 Artificial Intelligence，缩写为 AI，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门技术科学
生成式人工智能/生成式 AI	指	生成式人工智能（Generative AI）是使用机器学习和深度学习算法，从现有数据中学习并生成新内容的一种人工智能技术。它可以生成各种形式的数据，如图像、音频、视频和文本，并在虚拟现实、自然语言处理、艺术创作和模拟实验等领域有广泛应用
大模型	指	大模型就是 Foundation Model（基础模型），指通过在大规模宽泛的数据上进行训练后能适应一系列下游任务的模型
数据生产垂直大模型	指	海天瑞声数据生产垂直大模型，是公司利用大模型理论和算法，自主打造的面向人工智能数据基础服务的技术底座，

		实现面向多个下游数据任务的设计与处理等规则及内容的多项生成能力模型
多模态	指	多模态机器学习（Multi-modal Machine Learning）是一种机器学习方法，用于处理多种类型的数据，如图像、文本、音频等。它通过将来自不同模态的数据进行整合和联合建模，以提取和学习不同模态之间的相关信息和特征。多模态数据是指用于多模态机器学习训练使用的跨形态数据类型
深度学习（DL）	指	英文为 Deep Learning，缩写为 DL，是从机器学习中的神经网络发展而来的多层神经网络和它的训练方法是一种基于神经网络的机器学习，其中多层处理被用来从数据中逐步提取更高层次的特征
人类反馈强化学习	指	英文为 Reinforcement Learning from Human Feedback，缩写为 RLHF，是一种机器学习方法，旨在使智能系统从环境中学习，通过引入“奖励”和“惩罚”信号，让系统自行探索环境并学习最佳行为策略，以最大化某种特定目标
人工智能基础数据、训练数据、人工智能基础数据服务	指	人工智能基础数据、训练数据均指通过采集与处理等步骤形成的、结构化的、可供人工智能算法模型训练使用的数据；人工智能基础数据服务指为 AI 算法训练及优化提供的数据集设计、数据采集/获取、清洗、标注/优化、评测等服务
ChatGPT	指	美国 OpenAI（一家美国人工智能研究实验室和公司，在人工智能的各个领域进行前沿研究）研发的聊天机器人程序，ChatGPT 是人工智能技术驱动的自然语言处理工具，它能够通过理解和学习人类的语言来进行对话，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流，甚至能完成撰写邮件、视频脚本、文案、翻译、代码，写论文等任务
智能语音	指	实现人机语言的通信，包括语音识别技术（ASR）和语音合成技术（TTS）
语音识别（ASR）	指	Automatic Speech Recognition，是一种将语音信号自动转换为文本的技术，是实现智能人机交互的关键技术之一
语音合成（TTS）	指	Text to Speech，即“从文本到语音”，是一种将文字智能地转化为自然语音的技术
自然语言处理	指	英文为 Natural Language Processing，缩写为 NLP，是计算机科学领域与人工智能领域中的一个重要方向，研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法
计算机视觉	指	英文为 Computer Vision，缩写为 CV，是一门研究如何使机器“看”的科学，是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等
原料数据	指	Raw Data，即未经加工处理的原始数据
采集	指	使用软件工具、录音设备、图像拍摄设备等收集、录制和获取原料数据的过程
标注	指	训练数据的生产步骤之一，指对原料数据进行加工，进而生成符合算法模型训练要求的训练数据的过程
OCR	指	是指电子设备（例如扫描仪或数码相机）检查纸上打印的字符，通过检测暗、亮的模式确定其形状，然后用字符识别方法将形状翻译成计算机文字的过程

Prompt	指	提示文本，提供模型完成下游任务所需的上下文等信息，以帮助模型更好地理解任务需求，从而更好地完成下游任务
预训练/Pre-training	指	预训练是通过在大规模数据集上进行初始训练，学习通用特征和表示的机器学习方法。预训练模型可用于后续任务，并在多个领域应用广泛
模型微调/Fine-tune	指	微调，是一种常用于预训练模型的模型优化技术，基于已经训练好的预训练模型进行微调，使模型能够更好地适应目标任务
奖励模型/Reward Model	指	奖励模型（Reward Model）是在强化学习中使用的一种关键概念。它定义了特定任务中，智能体所追求的目标和奖励的方式。奖励模型通过指定在智能体采取行动后，所获得的奖励信号来引导智能体的学习和决策过程。奖励模型可以基于任务的目标设定，对智能体的行为进行正向奖励或负向惩罚，从而影响智能体的策略优化和决策选择
智算中心	指	人工智能计算中心，指为人工智能（AI）开发和应用提供所需算力服务、数据服务和算法服务的机构
引擎	指	是计算机平台上开发相应应用程序或软件系统的核心组件
智能驾驶	指	指的是机器帮助人进行驾驶，以及在特殊情况下完全取代人驾驶的技术
虚拟人	指	虚拟人（Virtual Human）是通过建模、动作捕捉或 AI 等科技手段，制作出具有外貌特征和行为模式的虚拟形象，并通过显示设备呈现出来。虚拟人创造的价值主要是打破物理的空间限制，提供了更多沉浸感、参与感和互动感
转写	指	通过软件加工、人工听写等手段将语音片段转换为对应的文本的加工步骤
词性	指	以词的特点划分的词类，如名词、代词、动词、形容词、数词、量词等
韵律	指	韵律通常指语音语句中各词之间的停顿时长情况
声纹	指	用电声学仪器显示的携带言语信息的声波频谱
音色	指	声音的特征，指的声音在波形方面具有的与众不同的特性
AIGC	指	AI Generated Content，是指利用人工智能技术来生成内容。

注：本募集说明书中所列数据可能因四舍五入原因，而与根据相关单项数据直接相加之和在尾数上略有差异。

第二节 发行人基本情况

一、发行人基本信息

中文名称：北京海天瑞声科技股份有限公司

英文名称：Beijing Haitian Ruisheng Science Technology Ltd.

注册地址：北京市海淀区成府路 28 号 4-801

股票简称：海天瑞声

股票代码：688787

股票上市交易所：上海证券交易所

二、股权结构、控股股东及实际控制人情况

（一）发行人股权结构

截至 2023 年 3 月 31 日，发行人股份总额为 42,800,000 股，前十名股东如下：

序号	股东名称	持股数量（股）	持股比例（%）
1	贺琳	8,669,725	20.26
2	中瑞安	4,954,128	11.58
3	中移投资	3,427,058	8.01
4	唐涤飞	3,314,846	7.74
5	清德投资	2,017,463	4.71
6	上海丰琬	1,452,374	3.39
7	中网投	862,062	2.01
8	中瑞立	503,496	1.18
9	程建群	250,121	0.58
10	万长城	184,563	0.43
总计		25,635,836	59.89

（二）控股股东及实际控制人情况

截至 2023 年 3 月 31 日，贺琳合计直接持有公司 20.26%的股份，并通过控制中瑞安间接控制发行人 11.58%的股份，合计控制公司 31.84%的股份，并担任发行人董事长，为公司控股股东、实际控制人。

贺琳，1968年7月出生，中国国籍，无境外永久居留权，毕业于北京大学。1987年至2017年7月任职于中国科学院声学研究所，从事语音识别、语音合成、汉语语言理解、语音心理测试等方面的研究工作，先后参与了“汉语人-机语音对话系统工程”、“汉语语音参数合成”等国家自然科学基金重点项目、国家八六三计划智能计算机主题项目和中国科学院“八五”重大项目，完成《计算机汉语口语输出系统的设计与实现》等论文十余篇，参与编写《汉语语音合成-原理和技术》著作。贺琳为海天瑞声创始人，报告期内，曾任公司董事长、总经理，现任公司董事长。

三、所处行业的主要特点及行业竞争情况

（一）发行人所处行业

根据国家统计局《战略性新兴产业分类(2018)》，公司所从事的业务属于“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。

（二）行业主管部门和主要法律法规及政策

1、行业主管部门及监管体制

公司所处的行业主管部门是国家工业和信息化部，其主要负责拟定产业发展战略、方针政策、总体规划和法规，并组织实施工业、通信业、信息化的发展规划，推进产业结构战略性调整和优化升级，推进信息化和工业化融合；指导行业技术创新和技术进步，以先进适用技术改造提升传统产业，组织实施有关国家科技重大专项，推进相关科研成果产业化，推动软件业、信息服务业和新兴产业发展。

行业内部组织管理机构主要是中国软件行业协会和中国人工智能产业发展联盟。中国软件行业协会受国家工业和信息化部委托，对各地软件企业认定机构的认定工作进行业务指导、监督和检查，并负责软件行业的市场研究、信息交流、行业统计、政策研究等方面的工作；中国人工智能产业发展联盟是在国家发展改

革委、科学技术部、工业和信息化部、中央网信办四部委共同指导下成立的行业协会组织，主要为促进联盟成员的研发、设计、生产、集成、服务等水平，构建我国人工智能产业生态，提升产业竞争力，强化人工智能与经济社会各领域深度融合，促进技术进步、提高生产效率，推动传统行业数字化转型，支持新技术、新产业、新业态、新模式加快发展。

2、行业主要法律法规及政策

(1) 行业总体政策情况

序号	实施时间	颁布主体	主要行业政策	相关内容
1	2023 年 7 月	中共北京市委、北京市人民政府	《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》	发展数据生产服务业，支持企业开展数据采集、清洗加工、存储计算、数据分析、数据标注、数据训练等数据生产服务，支持企业研发建设数据生产线，推进数据生产自动化。培育人工智能生成内容产业发展，发展人工智能生成语音、图像和自然语言等内容，丰富合成数据供给。 打造数据基础制度综合改革试验田，支持北京经济技术开发区等开展数据基础制度先行先试，打造政策高地、可信空间和数据工场。通过物理集中和逻辑汇通相结合的方式，导入工业、金融、能源、科研、商贸、电信、交通、医疗、教育等领域数据资源，促进数据跨行业融合应用，切实激活数据要素资源。
2	2023 年 5 月	北京市科学技术委员会、中关村科技园区管理委员会	《北京市促进通用人工智能创新发展的若干措施》	归集高质量基础训练数据集：组织有关机构整合、清洗中文预训练数据，形成安全合规的开放基础训练数据集；持续扩展多模态数据来源，建设高质量的文字、图片、音频、视频等大模型预训练语料库。谋划建设数据训练基地：加快建设数据基础制度先行先试示范区，探索打造数据训练基地，推动数据要素高水平开放，提升本市人工智能数据标注库规模和质量。
3	2023 年 2 月	中共中央、国务院	《数字中国建设整体布局规划》	指出夯实数字中国建设基础，畅通数据资源大循环。构建国家数据管理体制机制，健全各级数据统筹管理机构。推动公共数据汇聚利用，建设公共卫生、科技、教育等重要领域国家数据资源库。释放商业数据价值潜能，加快建立数据产权制度，开展数据资产计价研究，建立数据要素按价值贡献参与分配机制。
4	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》	数据作为新型生产要素，是数字化、网络化、智能化的基础，维护国家数据安全，促进数据合规高效流通使用。

序号	实施时间	颁布主体	主要行业政策	相关内容
5	2022 年 11 月	北京市经信局	《关于推进北京数据专区建设的指导意见》	针对重大领域、重点区域或特定场景，为推动政企数据融合和社会化开发利用建设的数据专区，旨在构建多层次数据要素市场，形成政务和社会数据流通融合体系，激发企业创新活力，释放数据要素价值，为加快推动首都新型智慧城市建设和打造全球数字经济标杆城市提供有力支撑。
6	2022 年 11 月	北京市人大常委会	《北京市数字经济促进条例》	从立法层面，加强数字基础设施建设，培育数据要素市场，规范公共数据的汇聚、清洗、共享、开放、应用和评估管理，开展公共数据专区授权运营。
7	2022 年 7 月	科技部等六部门	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》	加强人工智能场景创新要素供给。鼓励算力平台、共性技术平台、行业训练数据集、仿真训练平台等人工智能基础设施资源开放共享；推动城市和行业的人工智能“数据底座”建设和开放，为人工智能典型应用场景提供数据开放服务。
8	2022 年 1 月	国务院	《“十四五”数字经济发展规划》	强化高质量数据要素供给、加快数据要素市场化流通、创新数据要素开发利用机制等重点任务举措
9	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》	加强关键数字技术创新应用，建设重点行业人工智能数据集，发展算法推理训练场景。

(2) 数据安全领域主要法律法规及政策

序号	发布时间	颁布主体	主要法律法规	相关内容
1	2023 年 7 月	国家互联网信息办公室等七部门	《生成式人工智能服务管理暂行办法》	生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：（一）使用具有合法来源的数据和基础模型；（二）涉及知识产权的，不得侵害他人依法享有的知识产权；（三）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；（四）采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；（五）《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。
2	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

序号	发布时间	颁布主体	主要法律法规	相关内容
				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。
3	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。
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 8 月	国家互联网信息办公室等	《汽车数据安全 管理若干规定（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。
6	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。

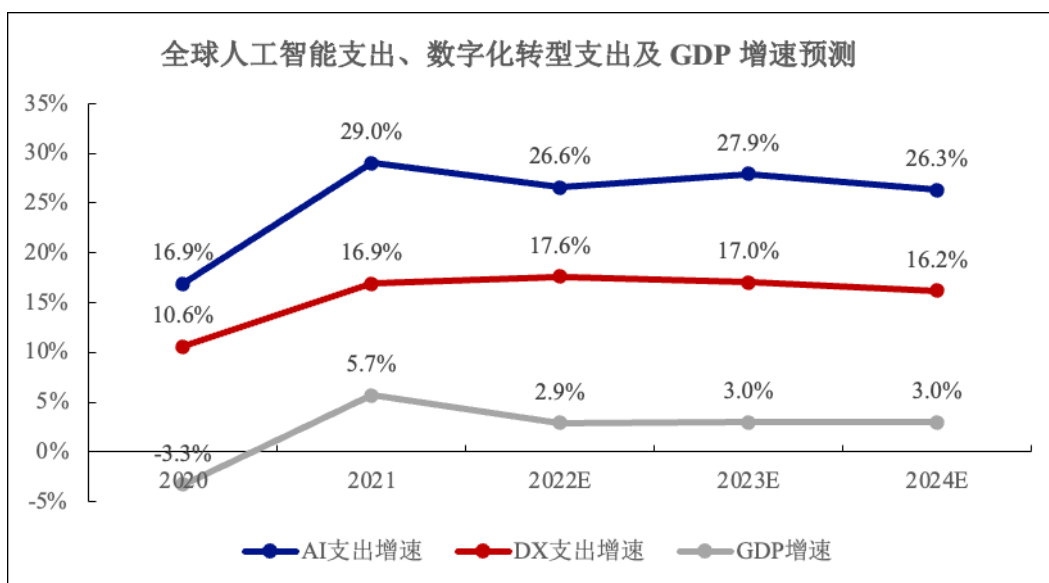
（三）行业发展现状和发展趋势

1、行业发展概况

（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。



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 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。

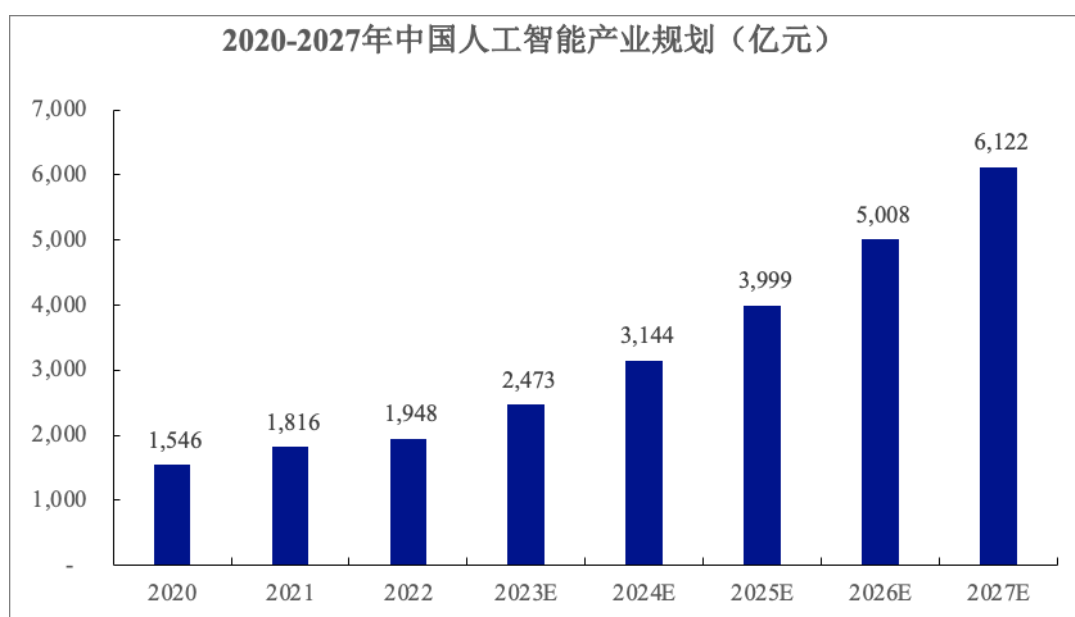
未来，受益于数字经济政策和大模型技术的双重驱动，人工智能将具备更强的产业融合能力，并将深刻影响千行百业的运行规则，以及人们的生活方式，人工智能产业的发展将随之进入快车道。

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%。



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。根据艾瑞咨询的数据显示，2022 年中国人工智能产业规模达 1,948 亿元，预计 2027 年市场规模将达到 6,122 亿元，年复合增长率为 25.6%，主要与智算中心建设以及大模型训练等需求拉动的 AI 芯片市场、无接触服务需求拉动的智能机器人及对话式 AI 市场等快速增长相关，并有望在下游制造、交通、金融医疗等多领域不断渗透，实现大规模落地应用。

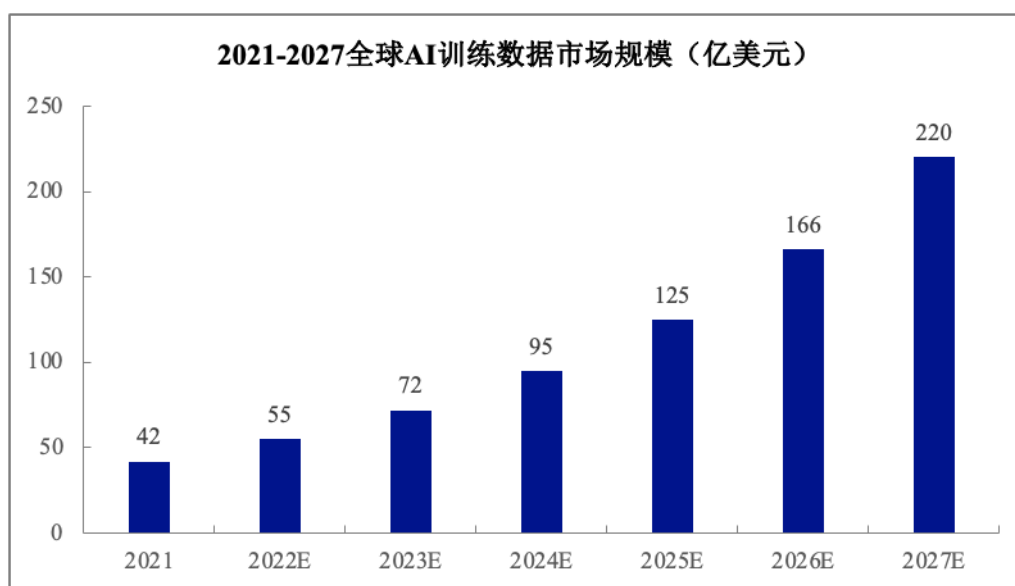


数据来源：艾瑞咨询

（2）训练数据作为 AI 算法发展和演进“燃料”的作用更加凸显

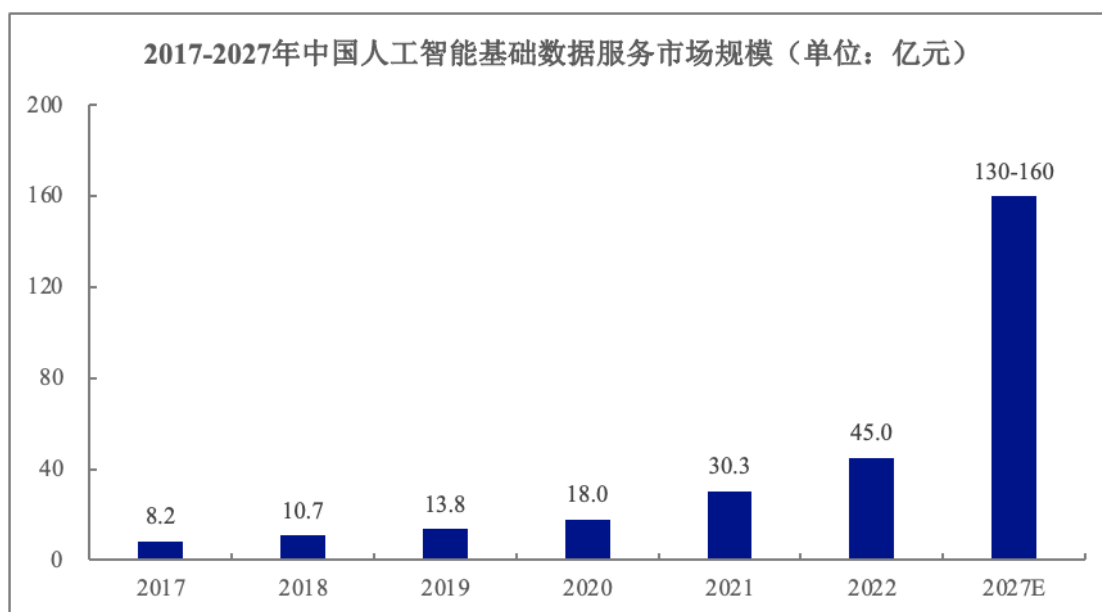
在 AI 产业链中，算法、算力和数据共同构成技术发展的三大核心要素。算法模型从技术理论到应用实践的落地过程依赖于大量的训练数据。训练数据越多、越完整、质量越高，模型推断的结论越可靠。过去十年人工智能产业是以算法为核心，通过深度学习算法的不断创新，推动人工智能产业的快速发展。但未来，随着算法发展趋于开源、算力能力大幅提升、及人工智能模型从技术理论应用到更多的垂直场景，想要更快更好提升人工智能能力，数据将发挥更重要的作用。

根据 Cognilytica 数据统计显示，2021 年全球 AI 训练数据市场需求约为 42 亿美元，并预计到 2027 年这一需求将增长到 220 亿美元，2021-2027 年复合增长率达 32%。



数据来源：Cognilytica

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。根据德勤数据，2022 年中国人工智能基础数据服务市场规模为 45 亿元，2027 年规模将达到 130-160 亿元，年复合增长率为 23.6%-28.9%。



数据来源：德勤

（3）数据要素市场蓬勃发展，数据行业迎来更为广阔的发展机遇

近年来，我国数字经济蓬勃发展，数据要素因具有基础性战略资源和关键性生产要素的双重属性，相关市场规模持续增长。尤其在《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》出台后，我国系统性布局了数据基础制度体系的“四梁八柱”，加速了数据流通交易和数据要素市场发展，进一步推动了公共数据、企业数据、个人数据合规高效流通使用。为更好响应中央号召，北京、上海、广州、深圳、杭州等地数据政策陆续出台，逐步构建了多层次、多元化数据要素市场生态体系。

以北京为例，《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》、《北京市促进通用人工智能创新发展的若干措施》和《关于推进北京市数据专区建设的指导意见》指出，北京市要加快建设“数据基础制度先行先试示范区”（以下简称“先行先试示范区”），“支持北京经济技术开发区等开展数据基础制度先行先试，打造政策高地、可信空间和数据工场”，探索打造数据训练基地，归集高质量基础训练数据集，推动数据要素高水平开放，提升本市人工智能数据标注库规模和质量，并建设针对重大领域、重点区域或特定场景建设专题数据区域，吸纳市场主体和数据、技术、资本等多元要素参与。北京市陆续出台的多项文件旨在打破数据壁垒，推动数据融合利用，加快推动公共数据开放，促进数据要素流通，激发数字市场创新活力，释放和发展数字化生产力，打造多层

级数据要素市场，成为具有竞争力和影响力的数字产业集群。按照“政府引导、市场运作、创新引领、安全可控”的原则，“先行先试示范区”有望成为国际领先的数据要素高效流通核心枢纽。

数据要素市场受政策推动，进入高速发展期，未来围绕数据的价值利用以及流通交易，将产生大量新增数据需求，为数据行业开拓了新的增长空间，提供了新的业务拓展机遇。未来，数据要素也将成为数据行业增长的重要推动力量。

2、行业发展趋势

（1）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。

首先，数据的质量以及数据清洗的工程化能力会显著拉开大模型预训练阶段的效果差距；同时，更多模型或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。对于大模型训练而言，不仅需要持续获取大规模、高质量、多模态、多场景、多垂向的数据，更须具备持续迭代的高质量数据清洗和标注策略，以不断提升包括预训练（Pre-training）、模型微调（Fine-tune）及奖励模型（Reward Model）等过程中所需数据（例如提示（Prompt）类数据）的质量，确保语言类和常识性知识之外的其他垂直领域的应用场景的能力提升，为大模型精确性、通用性及泛化能力的实现奠定坚实基础。

在以上背景下，一方面，大规模、高质量数据集重要性凸显，成为模型训练效果的核心支撑之一。另一方面，AI 发展所面对的数据前沿性及工程化技术的挑战也较为显著。长期看，只有 AI 数据处理技术的不断拓新与发展，才能及时适应甚至超前引领大模型技术和应用的发展。

（2）多模态数据受 AIGC 发展驱动，将呈现快速增长趋势

随着 AIGC 技术发展，AI 可在更多维度、更多场景辅助人类进行内容生产以及创作。例如，通过大模型等 AIGC 技术，人类仅需输入一段简单的文字指令，AI 即可按照人类描述生成一幅画、一段语音或一段视频，以此帮助人类完成内

容创作。要实现上述功能，AI 除了要具备理解人类文字指令的能力，还需要通过对齐两种独立模态关键特征的方式，建立文字与图、语音、视频等一一映射关系，这背后将依赖大量的多模态数据，AI 需要对多模态数据进行学习，以实现跨模态的创作能力。

此外，随着 AI 虚拟主播、虚拟学生、虚拟员工轮番上岗，数字人这个新兴概念逐渐走入大众视野，成为人工智能领域的热门技术赛道。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。

未来，随着以 AIGC、虚拟人为代表的 AI 技术以及应用的不断发展，多模态数据需求将呈现加速增长趋势。

(3) 人工智能企业全球化布局加速，多语种能力成为企业业务拓展核心支撑

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。

(4) 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。

（四）发行人所处行业的竞争情况

1、行业竞争格局

我国 AI 基础数据服务行业市场参与方主要包括基础数据服务商、需求方自建基础数据团队、学术机构、政府机构等几类主体，其中发行人所属的基础数据服务商，依据业务规模可以进一步分为品牌数据服务商和中小数据供应商。该类公司业务包括出售产品训练数据集的使用授权，或根据用户的具体需求提供数据处理服务（企业自行采集或用户提供原始数据、企业对数据进行转写、标注），具体业务服务形式包括且不限于提供训练数据产品、提供数据采集服务、提供数据转写标注服务等。公司作为行业内的品牌数据服务商，是我国较早专业从事训练数据产品与服务研发与及销售的主要企业之一，在行业内具备一定的领先性。

2、行业内主要企业

行业内其他主要企业包括 Appen、数据堂、标贝科技等。

（1）Appen

Appen 公司 1996 年成立于澳大利亚，2015 年 1 月于澳大利亚证券交易所上市。是全球领先的图像、文本、语音、音频、视频等 AI 训练数据服务提供商，拥有业内先进的人工智能辅助数据标注平台和一体化 AI 数据及资源管理平台及全球 100 多万名技能娴熟的众包资源，支持 200 多种语言和方言，为全球科技、

汽车、金融服务、零售、医疗和政府等行业赋能。

(2) 数据堂

数据堂成立于 2010 年，专注于为国内外人工智能技术和应用客户提供一站式数据资源服务、数据生产服务以及数据理解解决方案服务，覆盖语音识别、计算机视觉、自然语言理解等数据集，产品及服务应用于智能语音、智能驾驶、智能娱乐、智能客服、智能家居、新零售、智能医疗等领域。

(3) 标贝科技

标贝科技成立于 2016 年，是一家专注于智能语音交互和 AI 数据服务的人工智能公司。基于 AI+SaaS 开放平台，为客户提供 AI 数据服务、技术能力、智能语音交互方案赋能服务，包括通用场景的语音合成和语音识别，以及 TTS 音色定制，声音复刻，情感合成和声音转换在内的语音技术产品；AI 数据业务涵盖语音合成、语音识别、图像视觉、NLP 等采标服务和平台化自研工具能力。已与国内外百余家企业客户建立合作，涵盖汽车、教育、客服、零售、阅读、智能硬件等多个领域。

发行人行业内主要企业包括 Appen、数据堂、标贝科技等，其基本经营情况、技术研发及产品能力、综合能力、主要财务指标对比如下：

基本经营情况				
公司名称	发行人	Appen	数据堂	标贝科技
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	256 人（截至 2023 年 3 月 31 日）	超过 1,125 人	299 人	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	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、	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

	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		Samsung、Fujitsu 等企业及在华研发机构	
客户数量	800 余家	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
拥有的成品训练数据集数量	1,334 个	超过 250 个	291 个	190 个
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	33 项（31 项发明专利、1 项实用新型专利及 1 项外观设计专利）（截至 2023 年 3 月 31 日）	2 项	14 项	12 项
计算机软件著作权数量	164 项（截至 2023 年 3 月 31 日）	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、	未公开披露	乙级测绘资质；ISO27001 信息安全管理体系认证、	ISO27001 信息安全管理体系认证、ISO27701 隐私信

	ISO27701 隐私信息管理体系认证； 信息系统安全等级保护三级； 中国信通院数据安全推进计划成员单位		ISO27701 隐私信息管理体系认证	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、信息系统安全等级保护二级
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、国家专精特新“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
主要财务指标对比				
营业收入	2.63 亿元人民币	3.88 亿美元	1.69 亿元人民币	未公开披露
综合毛利率	64.73%	38.76%	45.75%	未公开披露
净利润	2,945.41 万元人民币	-2.39 亿美元	1,917.52 万元人民币	未公开披露
净利率	11.20%	-61.60%	11.34%	未公开披露

注 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。

注 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。

四、主要业务模式、产品或服务的主要内容

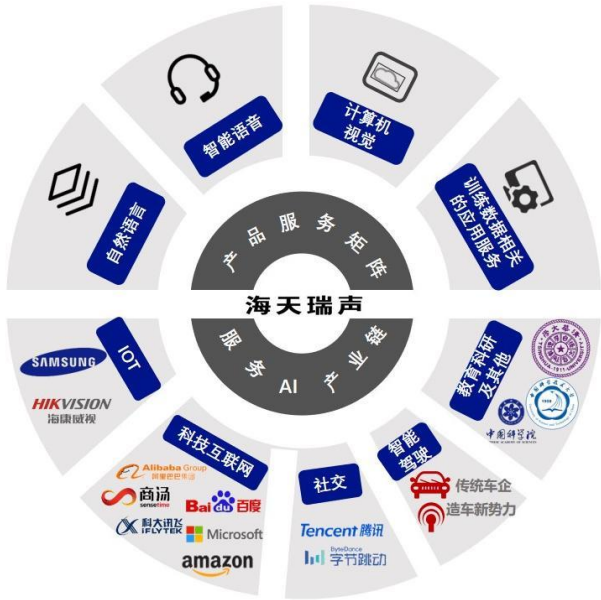
(一) 主要业务模式

1、主营业务概况

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。

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。



图：公司产品服务矩阵示意

2、盈利模式

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。

3、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（数据清洗、标注等）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、特定应用场景或特定语言的专属算法模型，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。

4、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中心负责；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。

5、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。

（二）主要产品及服务情况

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。

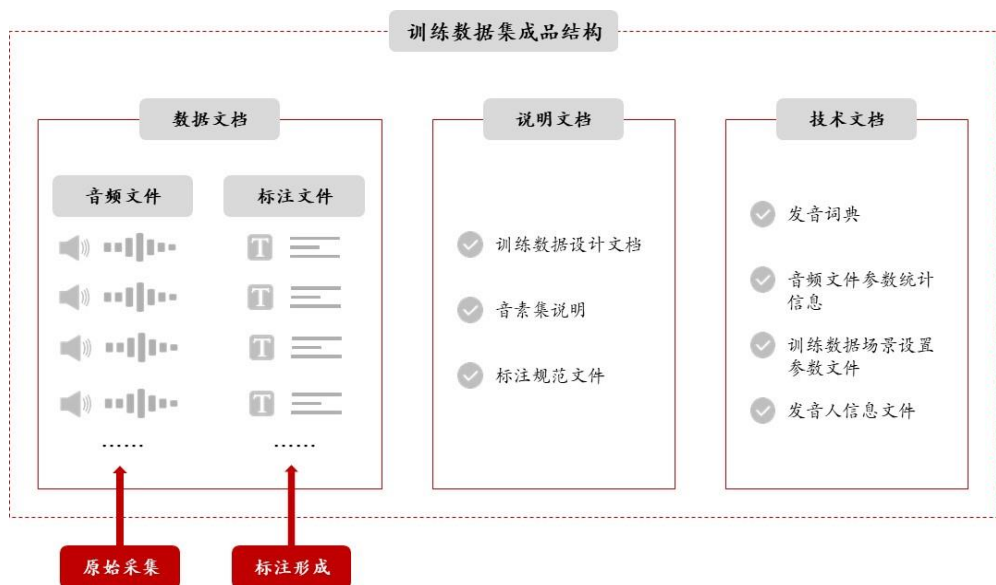
公司通过设计训练数据集结构、采集（收集或编写自然语言文本、对话等数据信息）、加工（对自然语言文本数据进行单词分割、词性标注、语义语法标注、情感属性标注等）、质检（对数据集进行质量检测，如检验文本、词性或者语义的标注结果是否准确等）；或者对客户提供的自然语言文本执行加工、质检工作，最终形成客户所需的自然语言训练数据集。

(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、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，覆盖了个人助手、语音输入、智能家居、机器人、语音导航、智能客服、智能播报、语音翻译、移动社交、虚拟人、智能驾驶、智慧交通、智慧城市、智慧金融、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

（三）境外生产经营情况

报告期内，公司境外业务情况如下：

单位：万元

项目	2023年1-3月	2022年	2021年	2020年
境外业务合计	1,013.36	11,576.20	5,729.72	8,368.56
当期公司营业收入	2,881.74	26,288.79	20,647.65	23,337.40
收入占比	35.16%	44.03%	27.75%	35.86%

从地区分布来看，报告期内公司的营业收入大部分来源于境内，收入占比分别为 64.14%、72.25%、55.97%及 64.84%，收入占比保持相对稳定。同时，公司也存在一定规模的境外收入，收入占比分别为 35.86%、27.75%、44.03%及 35.16%，其中以美国、韩国、日本等国家为主。

报告期内，发行人的主要境外经营实体如下：

境外经营实体名称	子公司类型	注册地址
香港海天瑞声	全资子公司	中国香港
DATAOCEAN AI INC.	全资孙公司	美国

报告期内，发行人中国香港子公司主要负责境外采购，设立美国子公司主要系满足公司开展境外业务、服务境外客户需求，上述子公司的情况如下：

1、香港海天瑞声

名称		海天瑞聲（香港）科技有限公司			
公司编号		2311062			
注册地		5/F, Manulife Place 348 Kwun Tong Road Kowloon Hong Kong			
法定股本		2,000,000 美元			
已发行股本		2,000,000 美元			
业务性质		计算机软件开发及技术服务			
股东构成及持股比例		股东		持股比例	
		发行人			
				100%	
控制情况		发行人的全资子公司			
成立时间		2015 年 11 月 24 日			
最近一年 一期财务 数据 （单位：万 元）	2023 年 3 月 31 日 /2023 年 1-3 月	总资产	净资产	营业收入	净利润
		534.07	65.26	144.28	-120.04
	2022 年 12 月 31 日 /2022 年度	总资产	净资产	营业收入	净利润
		523.34	55.46	1,915.38	-188.16

注 1：上述财务指标均为单体报表口径，2023 年 1-3 月数据未经审计，下同；
 注 2：总资产、净资产按照即期汇率计算，营业收入、净利润按照平均汇率计算，下同。

2、DATAOCEAN

名称		DATAOCEAN AI INC.			
UBI Number		604 843 025			
Authorized Shares		1,000,000 美元			
Issued Shares		20,000 美元			
注册地		100 N HOWARD ST STE R, SPOKANE, WA, 99201, UNITED STATES			
股东构成及持股比例		股东		持股比例	
		香港海天瑞声		100%	
控制情况		发行人的全资孙公司			
成立时间		2021 年 11 月 23 日			
最近一年 一期财务 数据（单 位：万元）	2023 年 3 月 31 日 /2023 年 1-3 月	总资产	净资产	营业收入	净利润
		541.38	295.93	85.40	11.07
	2022 年 12 月 31 日 /2022 年度	总资产	净资产	营业收入	净利润
		490.89	288.76	203.07	17.72

五、发行人主要资产状况

（一）公司固定资产基本情况

公司主要的固定资产为办公及开展经营活动所使用的房屋建筑物和办公家具，以及产品研发、生产、检测使用的仪器等电子设备和运输工具。截至 2023 年 3 月 31 日，公司的固定资产状况如下表所示：

单位：万元

类别	账面原值	累计折旧	减值准备	账面价值	成新率
房屋及建筑物	26,518.33	576.90	-	25,941.43	97.82%
运输工具	268.08	86.32	-	181.76	67.80%
电子设备	1,478.49	793.35	-	685.14	46.34%
办公家具	321.72	83.55	-	238.16	74.03%
合计	28,586.62	1,540.12	-	27,046.49	94.61%

（二）房屋所有权情况

截至本募集说明书出具日，公司拥有的房屋所有权共计 3 项，具体情况如下

表所示：

序号	权利人	房权证编号	坐落地址	面积（m ² ）	他项权利
1	发行人	京（2018）海不动产权第0066374号	海淀区知春路1号1号楼15层1501	165.55	无
2	发行人	京（2018）海不动产权第0065368号	海淀区知春路1号1号楼15层1513	178.22	无
3	发行人	京（2022）东不动产权第0012196号	东城区新中街68号8A楼1层公建一层03等21套	6,192.79	无

（三）公司主要无形资产情况

1、商标

截至2023年3月31日，发行人及其控股子公司拥有的注册商标共计36项，其中经国家知识产权局核准的中国境内注册商标30项，中国境外注册商标6项。具体情况详见“附件一：发行人及其控股子公司拥有的商标”。

2、专利

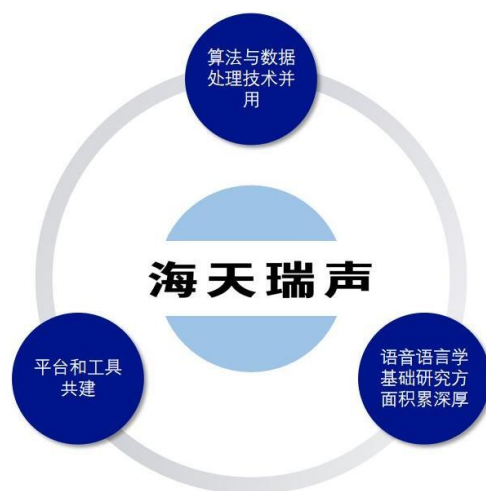
截至2023年3月31日，发行人及其控股子公司拥有的中国境内专利共计33项。具体情况详见“附件二：发行人及其控股子公司拥有的中国境内专利”。

3、软件著作权

截至2023年3月31日，发行人及其控股子公司拥有的中国境内计算机软件著作权共计164项。具体情况详见“附件三：发行人及其控股子公司拥有的中国境内计算机软件著作权”。

六、科技创新水平以及保持科技创新能力的机制或措施

（一）科技创新水平



图：海天瑞声科技创新实力示意

公司拥有对人工智能核心算法的理解和实施能力、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，以及 AI 算法自动辅助标注及精细标注能力。前述核心科技创新实力，使发行人显著区别于通过简单组织大量劳务人员进行原料数据采集、标注的企业。发行人的核心技术先进性主要体现在：

（1）算法与数据处理技术并用：通过算法与处理技术并用，可以满足算法在不同应用场景下的训练数据需求，可对大规模数据进行高效的加工或质检。截至 2022 年末，公司积累有超过 1,300 个自有知识产权的训练数据产品，并已向下游客户提供了累计约 6,000 次/个定制或标准化训练数据集，覆盖个人助手、智能家居、机器人、语音导航、移动社交、虚拟人、智能驾驶、智慧金融、机器翻译等 19 类创新应用领域，构建出独具特色的训练数据资源及服务能力集群，全面覆盖智能语音、计算机视觉、自然语言等多条业务线。

（2）平台和工具共建：公司自主开发了一体化数据处理平台，融入项目管理、质量把控、数据安全管控，嵌入生产所需的各类工具、软件模块，可实现高效、高质的训练数据生产，提升生产效率及质量控制水平。

（3）在语音语言学基础研究方面有深厚积累：公司建立了成熟的发音词典构建流程、积累了深厚的语音语言学基础研究成果。截至 2022 年末，公司已经拥有近 200 个语种/方言的覆盖能力。同时，公司已积累超过 110 个语种/方言的

发音词典，覆盖希伯来语、乌尔都语、缅甸语、阿姆哈拉语、普什图语、阿尔巴尼亚语、格鲁吉亚语等小语种，累计词条数超过 1,000 万条。

截至 2023 年 3 月 31 日，公司及其控股子公司已经取得 31 项发明专利授权、1 项实用新型专利授权及 1 项外观设计专利授权，以及 164 项计算机软件著作权。多年积累的核心技术和专业服务能力，使得公司能够更大规模、更有效率、更加精准地生产训练数据，在提升自身产出效率的同时也有效提高了训练数据对于客户算法模型的改善、优化效果。

发行人的核心技术先进性得到了业内广泛认可。自 2009 年至今，海天瑞声连续荣获“国家高新技术企业”、“中关村高新技术企业”称号。公司是“中国人工智能产业发展联盟”、“中国语音产业联盟”的理事单位，参与了行业的标准的起草工作，在训练数据集的架构设计、开发标准、质检评测等领域，以专业性和创新性获得了行业的普遍认可。2019 年，发行人入选工业和信息化部“新一代人工智能产业创新重点任务入围揭榜单位”和北京市级企业科技研究开发机构，并于 2021 年获评“新一代人工智能产业创新重点任务揭榜优胜单位”；2020 年，发行人入选工业和信息化部国家专精特新“小巨人”企业名单；2021 年获批“北京市企业技术中心”；2022 年，被选为工信部 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项资质荣誉，技术实力得到行业、主管部门的高度认可。

公司高度重视并坚持科技创新，通过持续的研发投入已积累形成了 12 项核心技术，持续应用于训练数据生产的各个环节（设计、采集、加工、质检），构成了公司生产经营的基础。上述核心技术可分为基础研究、平台工具及训练数据生产三个层次，具体如下所示：

核心技术层次	核心技术	核心技术项下的细分技术示例	核心技术在训练数据生产中的应用环节			
			设计	采集	加工	质检
基础研究	语音识别算法	语音数据库质量预估技术、语音识别系统评测技术	√	√	√	√
	语音合成算法	语音合成数据库评估技术、说话人自适应语音合成技术、语音合成系统评测技术				√

核心技术层次	核心技术	核心技术项下的细分技术示例	核心技术在训练数据生产中的应用环节			
			设计	采集	加工	质检
	计算机视觉算法	人脸检测和识别技术、物体识别技术、光学字符识别技术、场景分割技术、运动轨迹跟踪技术、3D 点云目标检测技术、3D 点云语义分割技术			√	√
	语音语言学基础研究	基于词典与模型的发音预测技术	√			
	训练数据集设计技术	多语种文本正则化技术、基于语言模型的文本易读性评测技术、多语种情感及领域分类技术	√			
平台工具	一体化数据处理平台	基于 C/S 架构和 B/S 架构的大规模语音处理平台，基于 C/S 架构的音素标注技术，多语言分布式文本处理技术，基于 C/S 架构和 B/S 架构的文本标注技术，基于 C/S 架构和 B/S 架构的图像标注技术，基于 B/S 架构的 3D 点云标注技术，大规模数据采集及标注平台	√	√	√	√
训练数据生产	多语种多模态训练数据设计技术	多语种语料清洗技术、多语种语音库设计技术、多语种音素均衡语料设计技术、混合语言语料设计技术	√			
	多模态多通道数据采集技术	多通道录音技术、移动设备上的语音采集技术、分布式图片内容分析技术、移动设备上的图片采集技术、多语言手写体数据采集技术、多模态数据采集技术		√		
	数据同步技术	音频对齐技术			√	
	实时采集质检技术	语音信号质量检测技术		√		√
	大数据驱动的高效数据处理技术	多语种拼写检查技术、视频流中特定帧定位技术、音素边界自动预测技术、基于语音特征的韵律预测技术、基于 CRF 的韵律预测技术、基于 HMM/CRF 的词性预测技术、基于 CRF 的命名实体识别技术、2D 障碍物自动检测和跟踪技术、2D 车道线自动检测技术、3D 点云自动贴合技术、3D 点云障碍物自动检测和跟踪技术			√	
	分布式高性能	音素标注正确率校验技术、相				√

核心技术层次	核心技术	核心技术项下的细分技术示例	核心技术在训练数据生产中的应用环节			
			设计	采集	加工	质检
	自动校验技术	似说话人自动筛查技术、音字一致校对技术				

（二）保持科技创新能力的机制或措施

1、保持科研人才的持续研发活力

一方面，发行人制定了《海天瑞声科研激励计划》，激励员工在专利、论文等知识产权方面取得原创性工作成果。对于创新成果，发行人通过季度绩效、年度创新奖等多种手段给予激励。另一方面，通过 KPI 等手段，加强对于工作成果有效性方面的要求，例如：对于研发类项目，重点考察工作成果的有效性，以及产生的商业价值；对于数据生产类项目，则以提高人效为重点考察指标，促使员工锐意进取、持续创新。

2、建设和优化公司综合研发管理体系

公司形成了以研发中心为核心，数据业务中心、语言研究中心、全球市场中心面参与的跨部门协同的研发体系，可支撑行业技术、市场动态、产品研发、质量反馈的综合协同机制。在这样的协同机制下，研发方向随行业和市场动态调整，研发成果以更新内部各类平台工具等软件的方式在第一时间应用到工作中，同时研发成果应用中出现的问题也会第一时间反馈到研发团队，并反映在接下来阶段的研发计划中。这样的机制保证了行业需求、研发创新和实际业务应用同步进行，使研发创新的成果得到最大化的应用。

3、紧密保持与行业发展趋势的一致性

（1）时刻关注行业技术、尤其是大模型领域的发展

发行人每年均派出核心技术人员和市场人员参加 Interspeech、CVPR、ICASSP 等海内外重要学术会议，关注技术与算法、尤其是大模型领域的发展，分析人工智能行业新兴数据需求痛点，指导未来创新研发训练数据产品、提供创新性的训练数据定制服务。

同时，发行人利用自有训练数据产品，多次与科研机构合作举办语种识别大赛、中英混读语音识别大赛等，吸引了众多科研院所与机构参加，一同为人工智

能行业技术水平的进步与发展做出贡献。

（2）加强与科研院所和专业人才的合作、交流、引进

在训练数据生产过程中，发行人同步开展学术研讨交流、产学研合作、企业间技术交流，把握包括大模型领域在内的行业前沿技术走向，为研发人员提供更多学习交流的机会，形成技术不断创新的坚实后盾。同时，发行人在全球范围内积极引进高端人才和语言学家，加强前沿技术研发。

（3）加强与下游客户的联系，跟踪行业发展

通过与众多下游客户建立战略合作关系，实时跟踪客户的技术发展（包括但不限于大模型领域）并持续满足客户对训练数据日新月异的需求，发行人形成了由项目人员、研发人员、市场人员共同参与的项目团队，通过规范化流程全面强化协同创新能力。

七、现有业务发展安排及未来发展战略

（一）现有业务发展安排

公司将围绕既定战略，继续发扬二次创业精神，坚守“做智能世界的数据基石”使命，成就“Changing life with AI”的愿景，不断提升核心能力，优化用户体验，按照技术+产品双轮驱动的定位持续加大研发投入，为 AI 产业链上的各类机构提供算法模型开发训练所需的高质量专业数据集。公司现有业务发展安排如下：

1、夯实传统业务，探索新兴业务

（1）全球化业务

随着以 ChatGPT 为代表的大模型技术的崛起，人工智能产业迎来了新一轮重大发展机遇，因此可以合理预期全球人工智能产业将再次进入加速期。为更好抢抓海外市场需求，公司将启动更为全面的全球化发展战略，从技术跟踪、品牌升级、体系搭建、营销推广等多维度助力全球化市场扩张。公司将搭建海外技术研究体系，通过紧密跟踪全球人工智能发展趋势，前瞻性布局适配于海外新兴技术的数据解决方案，持续提升全球市场竞争力。此外，扩大境外直营销售团队，进一步织密客户服务网络，提升客户服务体验；通过持续深度参与海外学术会议、

进行综合市场营销方案升级等方式，拓展客户触达渠道；同时，深度挖掘现有头部客户需求，进行重点突破，并进行新客户拓展。未来，公司将全面集结公司研发、交付、销售、市场力量，全方位支持海外战略扩张，力争在海外市场形成新突破。

（2）智能驾驶业务

公司将紧抓智能驾驶行业发展机遇，全方位布局和提升智能驾驶业务核心能力，通过持续跟踪前沿技术发展动态，前瞻性布局、迭代智能驾驶数据处理平台，更好服务于客户新兴需求；不断优化算法中台中枢能力，通过提高人机协作效率，实现数据处理的规模化应用提升；加大数据安全体系建设，全方位保障数据生产流程安全合规。同时，基于已经获取的测绘资质，进一步拓展数据服务环节，开展采标一体化服务，提升业务毛利水平。此外，公司也将尝试探索传统采标业务以外的新兴数据服务模式，以更好适应不同类型客户的差异化数据需求。

（3）新兴业务探索及拓展—大模型、数据要素

公司将顺应人工智能大模型的发展大势，通过建设 AI 大模型训练数据集项目，拓展训练数据产品体系，大幅提升训练数据产品的数据规模、广度、质量和精度，扩大产品覆盖的场景和领域，这是前瞻性布局大模型领域、拓展潜在高增长价值的新型业务板块的重要举措；同时，公司还将进行数据生产垂直大模型研发项目的建设，旨在以研发海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台，进一步延伸大模型领域综合数据服务能力，藉此提高公司业务能力的维度、广度、效率和水平，巩固公司的核心技术壁垒，构建长期技术实力支撑，不断提高公司核心竞争力。

同时，通过上述大模型方向的研发和业务拓新，公司也将积极参与北京市国家数据基础制度“先行先试示范区”的建设，并将充分利用“先行先试示范区”在基础制度、数据供给等方面的先行先试政策，探索以数据治理、数据交易、数据处理等为核心的数据要素领域，力争将上述方向的创新业务打造成为具有潜在高增长价值的新兴业务板块。

2、聚焦主业，进行全方位核心能力建设

(1) 核心技术

公司将继续秉承技术+产品双轮驱动的核心理念，围绕智能语音、计算机视觉、自然语言等业务领域、以及以智能驾驶为代表的新兴垂直业务领域、以及新兴的大模型数据服务领域等，在算法中台建设、业务线平台升级、新语言研究、以及多语种多模态数据布局、以及大模型数据方向研究等方面取得实质性成果，加强技术能力护城河，助力公司不断向实现训练数据生产智能化、规模化方向迈进，确保公司核心竞争力不断提升。

(2) 数据安全

近年来，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，公司所处行业的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客户选择数据服务商时的重要考量因素以及新的行业壁垒。下一步，公司将继续加大数据安全及合规能力体系建设，通过完善内部制度流程建设、根据业务发展方向提前进行资质布局、广泛参与行业交流、不断提升企业数据安全及合规技术和机制等方式，提升应对更高标准、更严格的数据监管要求的能力，提高整体数据安全保障水平，并将之逐步打造成为核心竞争力之一。

(3) 供应链

公司将对境内、境外资源群体进行持续扩充，并通过大型数据基地及资源池的精细化建设等方式对供应链体系进行优化，为实现可持续性产能扩张和成本优化奠定基础。同时，公司将持续推进供应链智能化管理及信息化建设，实现从需求申请到使用结算的全流程系统化，并对全流程业务数据进行可视化追踪展示，提高供应链的流转效率和运作速度、降低供应链的成本，更积极地支撑企业管理和决策水平提升。

(4) 市场营销

为更好支撑全球化发展战略，公司将持续加大全球市场营销体系建设，进一步升级品牌战略工程。公司将继续增加品牌营销投入，在持续拓展社交平台、直播平台、搜索引擎等多渠道营销的同时，积极参与 Interspeech、CVPR、ICASSP 等海内外重要学术会议，持续提升公司品牌认知度以及客群辐射范围。公司将加

大境外本土营销团队建设，通过进一步织密客户服务网络，提升客户服务体验。在国内，海天瑞声将持续打造自媒体矩阵，举办一系列产品发布会，积极参与泛AI行业/自动驾驶行业等行业展会，不断提升公司品牌影响力。

（5）人力资源

人才是公司发展的核心动能，公司将继续坚持以人为本，打造一支高战斗力的人才队伍。未来，公司将积极落实人才发展战略，继续通过内部培养与外部引入并重的机制，打通人才供给渠道；此外，公司将进一步完善薪酬体系改革，持续开展人才激励计划，通过将员工利益与公司业务发展及股东利益有机结合，充分调动员工积极性，增强团队稳定性；同时，优化人力资源体系，以更通畅的晋升渠道、更佳的工作环境、以及更积极的企业文化，吸引并留住人才，提高全员创效水平，为公司的可持续发展提供强劲动力。

（二）未来发展战略

1、公司发展战略

公司在保障基础数据业务稳健发展的同时，将通过可持续性的技术和商业模式创新，寻求新的业绩增长点。首先，公司将以全球化战略扩张为核心，通过全球化营销体系搭建、升级，进一步深化全球化发展策略，持续扩大公司的全球卡位优势。此外，公司将继续聚焦智能驾驶赛道，通过平台、算法研发以及数据安全能力的持续建设，不断将之打造成为公司核心竞争力之一，使之成为公司第二增长曲线。同时，公司决心通过大力度研发投入，重点聚焦探索大模型数据服务领域、与国家数字经济密切相关的数据要素领域等新兴业务领域，基于过往数据服务能力，通过前沿市场跟踪，不断拓展服务边界，挖掘潜在增量市场空间。

2、为实现战略目标已采取的措施

公司已经着手开展与上述战略规划相关的工作。在全球化拓展方面，公司已经投入研发更多应用场景下的多语种、多模态训练数据，以支撑境外客户在虚拟人、OCR及手写体、多语言扩展等研发及业务方向的发展；公司也逐渐增强境外销售团队建设、启动品牌升级，重启包括 Interspeech、ICASSP 等在内的海外研讨会及行业展会并深度参与，从而向全球市场展示公司先进技术实力及前瞻性产品积累，提升行业影响力。在智能驾驶业务领域，公司已建立独立的智能驾驶

事业部，从研发、技术、业务、销售等多维度布局本业务板块。公司已经正式推出了专为自动驾驶场景设计的全栈式数据处理平台--海天瑞声 DOTS-AD 自动驾驶平台，实现了智能驾驶应用场景全覆盖，可支持包括 4D 毫米波雷达、激光雷达、鱼眼摄像头等各类传感器的新型数据处理需求。此外，公司已获得乙级测绘资质，为公司合规开展汽车数据采集、标注业务，建立完整数据解决方案、生产标注化数据集产品，奠定了坚实的基础。

在大模型、数据要素领域，公司已经开始与国内多家头部大模型厂商开展合作，为这些客户提供大模型相关的部分数据服务，服务形态包括但不限于强化学习阶段的数据处理需求等。同时，公司受邀参与北京市开展数据基础制度先行先试规划、政策制定和生态构建，作为行业领军企业为“先行先试示范区”的建设积极建言献策，前瞻性探索以数据治理、数据交易、数据处理等为核心的数据要素领域，力争将上述方向的创新业务打造成为具有潜在高增长价值的新兴业务板块。

3、计划未来采取的措施

在未来的几年中，公司计划大规模研发多语种、多垂直场景下覆盖语音识别、语音合成、计算机视觉以及自然语言的单模态/多模态训练数据，尽可能使客户通过采购公司的训练数据产品/服务即可获得开发大模型或传统深度学习技术路线下的大规模、可商用的人工智能系统所需要的全部训练数据。为了提高完成这一战略目标的工作效率，公司计划持续升级公司自研的算法、平台及工具集群，扩建研发中心、数据业务中心，合作/聘请更多的国内外语言学家、人工智能技术专家、分布式系统技术专家、软件研发专家和数据安全技术专家，实现公司在算法研究、平台及工具研发以及数据处理技术能力等方面的全方位能力提升。同时，公司也计划与更多的境内外各类客户开展合作，开发设计更多训练数据驱动的应用场景，进一步拓展公司的业务范围。

八、截至最近一期末，不存在金额较大的财务性投资的基本情况

（一）财务性投资的认定标准

根据 2023 年 2 月证监会发布《公开发行证券的公司信息披露内容与格式准则第 61 号——上市公司向特定对象发行证券募集说明书和发行情况报告书》第

八条规定，发行人基本情况应包括：“截至最近一期末，不存在金额较大的财务性投资的基本情况”。

根据 2023 年 2 月证监会发布《<上市公司证券发行注册管理办法>第九条、第十条、第十一条、第十三条、第四十条、第五十七条、第六十条有关规定的适用意见——证券期货法律适用意见第 18 号》，对财务性投资界定如下：

“财务性投资包括但不限于：投资类金融业务；非金融企业投资金融业务（不包括投资前后持股比例未增加的对集团财务公司的投资）；与公司主营业务无关的股权投资或投资产业基金、并购基金；拆借资金；委托贷款；购买收益波动大且风险较高的金融产品等。

围绕产业链上下游以获取技术、原料或者渠道为目的的产业投资，以收购或者整合为目的的并购投资，以拓展客户、渠道为目的的拆借资金、委托贷款，如符合公司主营业务及战略发展方向，不界定为财务性投资。

上市公司及其子公司参股类金融公司的，适用本条要求；经营类金融业务的不适用本条，经营类金融业务是指将类金融业务收入纳入合并报表。

基于历史原因，通过发起设立、政策性重组等形成且短期难以清退的财务性投资，不纳入财务性投资计算口径。”

此外，根据中国证监会 2020 年 7 月发布的《监管规则适用指引——上市类第 1 号》，对上市公司募集资金投资产业基金以及其他类似基金或产品的，如同时属于以下情形的，应当认定为财务性投资：（1）上市公司为有限合伙人或其投资身份类似于有限合伙人，不具有该基金（产品）的实际管理权或控制权；（2）上市公司以获取该基金（产品）或其投资项目的投资收益为主要目的。

（二）自本次发行相关董事会决议日前六个月至本募集说明书签署日，公司已实施或拟实施的财务性投资（包括类金融投资）的具体情况

2023 年 6 月 21 日，发行人召开第二届董事会第十七次会议审议，通过了本次向特定对象发行股票相关事项。自本次发行相关董事会首次决议日前六个月（2022 年 12 月 21 日）至本募集说明书签署日，发行人存在的实施或拟实施的财务性投资及类金融业务具体如下：

1、投资类金融业务

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司未投资类金融业务。

2、非金融企业投资金融业务

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司未投资金融业务，亦不存在以超过集团持股比例向集团财务公司出资或增资的情形。

3、与公司主营业务无关的股权投资

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司未参与股权投资业务。

4、投资产业基金、并购基金

本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司不存在投资产业基金、并购基金的情形。

5、拆借资金

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司不存在拆借资金的情形。

6、委托贷款

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司不存在委托贷款的情形。

7、购买收益波动大且风险较高的金融产品

自本次发行相关董事会决议日前六个月起至本募集说明书签署日，公司不存在购买收益波动大且风险较高的金融产品。

根据上述财务性投资（包括类金融投资）的认定标准并经核查，公司本次发行相关董事会决议日（2023年6月21日）前六个月即2022年12月21日起至本募集说明书签署日，公司不存在实施或拟实施财务性投资及类金融业务的情形。

(三)最近一期末是否存在持有金额较大的交易性金融资产和可供出售的金融资产、借予他人款项、委托理财等财务性投资的情形

截至 2023 年 3 月 31 日，发行人可能涉及财务性投资（包括类金融业务）的相关会计科目情况如下：

单位：万元

序号	科目	截至 2023.3.31 账面价值	财务性投资金额	财务性投资占归属母公司所有者净资产比例
1	交易性金融资产	39,720.38	-	-
2	预付款项	552.49	-	-
3	其他应收款	268.67	-	-
4	其他流动资产	602.92	-	-
5	长期待摊费用	40.87	-	-
合计		41,185.33	-	-

注：2023 年 3 月 31 日数据未经审计。

截至报告期末，公司未持有财务性投资。

1、交易性金融资产

截至 2023 年 3 月 31 日，公司交易性金融资产明细如下：

单位：万元

项目	账面余额
以公允价值计量且其变动计入当期损益的金融资产	39,720.38
其中：理财产品	39,720.38
合计	39,720.38

截至 2023 年 3 月末，公司理财产品的明细如下：

单位：万元

产品类型	发行方	产品名称	金额	起始日	到期日
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 13035 期	800.00	2022/12/30	2023/4/1
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 13599 期	900.00	2023/2/6	2023/5/9
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 13922 期	4,800.00	2023/3/1	2023/6/1
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 13936 期	1,000.00	2023/3/6	2023/5/5
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 13937 期	2,400.00	2023/3/6	2023/6/6

产品类型	发行方	产品名称	金额	起始日	到期日
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 14051 期	800.00	2023/3/13	2023/4/13
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 14053 期	18,000.00	2023/3/13	2023/6/13
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 14269 期	430.00	2023/3/27	2023/6/27
保本	招商银行股份有限公司	招商银行点金系列看涨两层区间 90 天结构性存款	3,500.00	2023/1/13	2023/4/13
保本	招商银行股份有限公司	招商银行点金系列看涨两层区间 91 天结构性存款	7,000.00	2023/3/1	2023/5/31
合计			39,630.00	-	-

注：上述投资金额合计数与 2023 年 3 月末交易性金融资产科目余额差异为理财产品利息导致的公允价值变动。

截至 2023 年 3 月 31 日，公司购买的上述结构性存款系为提高资金使用效率而使用闲置自有资金进行现金管理，上述结构性存款为风险较低、流动性较强的理财产品，不属于“收益波动大且风险较高的金融产品”，不属于财务性投资。

2、预付款项

截至 2023 年 3 月 31 日，公司预付款项账面余额为 552.49 万元，主要为数据服务费及新办公场地设计费用等，不属于财务性投资。

3、其他应收款

截至 2023 年 3 月 31 日，公司其他应收款账面余额按性质列示如下：

单位：万元

项目	2023.3.31
房租押金	200.51
保证金	63.00
员工备用金	11.32
其他	8.51
合计	283.34

截至 2023 年 3 月 31 日，公司其他应收款账面余额为 283.34 万元，主要为房租押金、保证金、员工备用金，不属于财务性投资。

4、其他流动资产

截至 2023 年 3 月 31 日，公司其他流动资产明细情况如下：

单位：万元

项目	2023.3.31
待抵扣及待认证进项税	341.86
预缴所得税	148.41
待摊房屋租赁费	112.64
合计	602.92

截至 2023 年 3 月 31 日，公司其他流动资产账面余额为 602.92 万元，主要为待抵扣及待认证进项税、预缴所得税和待摊房屋租赁费，不属于财务性投资。

5、长期待摊费用

截至 2023 年 3 月 31 日，公司长期待摊费用明细情况如下：

单位：万元

项目	2023.3.31
办公场所装修及其他	40.87
合计	40.87

截至 2023 年 3 月 31 日，公司长期待摊费用账面余额为 40.87 万元，主要为办公场所装修等，不属于财务性投资。

综上，公司最近一期末，不存在金额较大的财务性投资。

九、最近一年及一期业绩下滑情况及原因

（一）最近一年及一期主要经营数据情况

公司 2022 年度、2023 年 1-3 月主要经营数据及其变动情况具体如下：

单位：万元

项目	2022 年		2021 年
	金额	变动幅度	金额
营业收入	26,288.79	27.32%	20,647.65
营业成本	9,271.05	24.76%	7,431.19
销售费用	1,836.16	26.43%	1,452.34
管理费用	6,140.46	74.36%	3,521.80
研发费用	9,427.34	55.77%	6,052.09
财务费用	-522.76	-558.30%	114.07
营业利润	2,148.19	-35.65%	3,338.52

利润总额	2,836.03	-16.15%	3,382.14
净利润	2,945.41	-6.81%	3,160.54
归属于母公司股东的净利润	2,945.41	-6.81%	3,160.54
扣除非经常性损益后归属于母公司股东的净利润	1,014.91	-51.83%	2,106.74
项目	2023 年 1-3 月		2022 年 1-3 月
	金额	变动幅度	金额
营业收入	2,881.74	-39.81%	4,787.64
营业成本	1,504.45	-2.96%	1,550.34
销售费用	480.04	51.01%	317.88
管理费用	1,560.22	62.70%	958.95
研发费用	1,693.31	-5.58%	1,793.47
财务费用	32.77	161.85%	12.52
营业利润	-1,903.75	-294.73%	977.62
利润总额	-1,755.25	-277.90%	986.63
净利润	-1,361.63	-244.27%	943.79
归属于母公司股东的净利润	-1,361.63	-244.27%	943.79
扣除非经常性损益后归属于母公司股东的净利润	-1,704.82	-477.02%	452.18

注：2023 年 1-3 月数据未经审计。

（二）最近一年及一期业绩下滑的原因及合理性

1、2022 年度

2022 年，公司扣除非经常性损益后归属于母公司股东的净利润为 1,014.91 万元，较上年同期下滑 51.83%，主要系：

第一，公司为加速提升智能驾驶数据服务能力、整体数据生产的智能化水平，以及推进数据集产品扩充建设计划等，持续加大研发投入；同时完成管理架构及激励体系升级、全球化营销体系建设、新办公楼购置等重要事项，共同造成期间费用大幅增长。

第二，2022 年公司收到的政府补助以及理财收益持续增加，2022 年度非经常性损益金额较 2021 年同比显著上涨。以上因素共同作用，导致公司扣除非经常性损益的净利润在报告期内出现明显下滑。

2、2023 年第一季度

2023 年第一季度，公司扣除非经常性损益后归属于母公司股东的净利润为 -1,704.82 万元，较上年同期下滑 477.02%，主要系以下原因所致：

第一，2023 年第一季度公司实现收入 2,881.74 万元，较去年同期下降 1,905.90 万元，下滑 39.81%。原因在于，一方面由于受宏观经济增速放缓，部分境外客户正在进行业务方向调整、阶段性裁员等，导致其预算释放进度放缓，叠加数据出境相关法规落地实施所带来的影响等多重因素，公司境外业务的开展受到影响，境外收入较去年同期收缩明显，一季度境外收入同比下降 56.95%；另一方面，从境内业务情况看，公司不同业务板块呈现差异化态势，智能驾驶业务需求同比显著增长约 25%；基础数据服务领域客户需求虽有所复苏，但客户投入更为谨慎，同时市场竞争较为激烈，导致公司相关收入有所下降。

第二，2023 年第一季度，公司毛利率受毛利水平较高的境外业务以及标准化产品收入占比下降影响，同比降幅明显。

第三，根据总体战略布局和年度经营计划，与去年同期相比，公司在 2023 年一季度持续升级管理架构及激励体系，加大全球营销体系建设、扩建销售团队，同时新购置房产折旧与原有房租摊销并行，使得公司 2023 年一季度销售费用及管理费用同比出现较大比例增长。

（三）是否与同行业可比公司一致

由于公司主要从事训练数据的研发设计、生产及销售业务，所在细分领域为人工智能基础数据服务领域，细分领域和业务较新，目前尚不存在业务可比或相似程度较高的同类型 A 股上市公司。发行人行业内主要企业包括 Appen、数据堂、标贝科技等，受限于可比公司数据可获取性，仅澳大利亚上市公司 Appen、新三板挂牌公司数据堂（831428）公开披露了财务数据，对比分析如下：

1、APPEN

澳大利亚上市公司 Appen 与公司业务具有一定可比性。最近三年，Appen 业绩变动情况如下表所示：

单位：千美元

项目	2023 年 1-3 月		2022 年度		2021 年度		2020 年度
	金额	变动比例	金额	变动比例	金额	变动比例	金额
主营业务收入	-	-	388,133	-13.20%	447,154	8.36%	412,638
主营业务成本	-	-	237,712	-11.43%	268,378	13.68%	236,091
主营业务毛利	-	-	150,421	-15.86%	178,776	1.26%	176,547
主营业务毛利率	-	-	38.76%	减少 1.23 个百分点	39.98%	减少 2.80 个百分点	42.78%
费用合计	-	-	389,849	172.58%	143,021	4.38%	137,026
税前利润	-	-	-239,068	-766.39%	35,875	-19.45%	44,539
净利润	-	-	-239,068	-938.28%	28,519	-19.96%	35,632

注 1: Appen 公司主营业务收入为年度报告中“Services revenue”科目。

注 2: Appen 公司主营业务成本为年度报告中“Crowd labelling services”科目。

注 3: Appen 公司的毛利率根据年度报告中“Services revenue”和“Crowd labelling services”科目计算得出，毛利率=(Services revenue - Crowd labelling services)/Services revenue。

注 4: Appen 公司的费用合计为年度报告中“Expenses”项目下除“Crowd labelling services”科目的合计数。

注 5: Appen 公司的税前利润为年度报告中“(Loss)/profit before income tax expense”科目。

注 6: Appen 公司的净利润为年度报告中“(Loss)/profit after income tax benefit/(expense) for the year attributable to the owners of Appen Limited”科目。

注 7: Appen 未公开披露 2023 年 1-3 月数据

收入方面，Appen 2021 年收入增长率为 8.36%，除中国区以外地区收入同比基本持平，较其 2016-2020 年超过 50% 的收入复合增长率出现较大幅度下降，公共卫生事件的影响是其收入增幅显著下降的主要原因之一。2022 年起，受公共卫生事件、经济疲软以及客户降低预算等因素影响，Appen 公司的业务规模出现了缩减，导致收入进一步下滑。2023 年 5 月，Appen 发布公告称，2023 年 1-4 月，公司未经审计的营收及毛利分别为 9,570 万美元和 3,580 万美元，较 2022 年同期分别下降 21.4% 和 24.7%。

利润方面，Appen 2021 年主营业务毛利率下降 2.80 个百分点，与发行人毛利率下降幅度 3.69 个百分点基本趋势一致。Appen 2021 年净利润亦出现下滑，下降幅度为 19.96%，主要系其持续加大平台产品研发投入、进行内部组织架构调整加大人员投入及支付投资并购相关费用等原因所致，但由于公司整体规模体量较 Appen 有较大差距，发行人处于业务发展和团队规模扩张的阶段，因此公司在研发、销售、管理等投入方面较 Appen 同比增幅更大，从而造成利润下降

幅度更大。2022 年，Appen 毛利率进一步下滑的同时，人员工资、折旧摊销金额的增加带来了费用支出的增长，此外公司收购的资产在 2022 年出现了大额商誉减值，以上因素共同导致了 Appen 公司 2022 年度净利润的大幅下滑。

整体上，受到宏观经济波动、内部增加投入等因素影响，报告期内发行人与 Appen 公司的净利润均出现持续下滑，但公司业绩下滑幅度相对较小，主要系一方面公司的业务收入构成中高毛利业务占比相对较高；另一方面公司持续加大研发投入，积极开拓包括智能驾驶业务在内的新业务，持续创造新的业绩增长点，使得营业收入的表现优于 Appen。

2、数据堂

数据堂是国内较早从事数据交易、数据采标的服务商之一，目前为新三板创新层挂牌企业，最近三年，数据堂业绩变动情况如下表所示：

单位：万元

项目	2023 年 1-3 月		2022 年度		2021 年度		2020 年度
	金额	变动比例	金额	变动比例	金额	变动比例	金额
营业收入	-	-	16,907.28	46.14%	11,569.42	47.19%	7,860.15
营业成本	-	-	9,172.74	39.69%	6,566.35	42.75%	4,600.02
毛利	-	-	7,734.54	54.60%	5,003.07	53.46%	3,260.14
毛利率	-	-	45.75%	增加 2.50 个百分点	43.24%	增加 1.77 个百分点	41.48%
利润总额	-	-	2,072.72	427.75%	-632.40	64.65%	-1,789.04
净利润	-	-	1,917.52	416.17%	-606.48	67.98%	-1,894.16
归属于母公司股东的净利润	-	-	1,830.16	401.29%	-607.44	69.00%	-1,959.46
扣除非经常性损益后归属于母公司股东的净利润	-	-	1,396.15	263.70%	-852.88	63.73%	-2,351.49

注 1：数据取自数据堂公开披露信息，计算变动比例时分母取绝对值；

注 2：数据堂未公开披露 2023 年 1-3 月数据。

报告期内，数据堂业务规模保持了较高的增长速度，且在 2022 年度实现盈利，主要由于一方面 AI 市场规模扩大，数据需求快速增长，数据堂由于前期规

模较小，因此收入规模增速较快；另一方面由于数据堂的产品线种类相对集中，在研发投入等方面的支出相对较小，报告期内研发费用占营业收入的比例在 10% 左右，而发行人持续加大研发投入，研发支出占比在 30% 左右，因此数据堂的净利润指标表现优于发行人。

（四）相关不利影响是否持续、是否将形成短期内不可逆转的下滑

发行人所处行业发展前景广阔，市场及技术储备良好，具备持续经营能力，发行人预计相关不利影响不具有持续性，短期内不会形成不可逆转的持续大幅下滑，具体原因如下：

1、行业需求方面，我国数字经济蓬勃发展，数据要素因具有基础性战略资源和关键性生产要素的双重属性，数据要素服务市场规模持续增长。根据国家工信安全发展研究中心数据，2022 年我国数据要素市场规模为 904 亿元，预计到 2025 年将达到 1,749 亿元左右，2020 年-2025 年年复合增长率为 26.26%，数据要素将成为赋能中国数字经济发展的主要驱动力量。因此，公司所在行业具有较高的增速空间及成长潜力。

2、公司技术积累方面，自 2005 年公司成立以来，公司始终致力于为 AI 深度学习提供算法模型开发训练所需的专业数据集，公司现已积累较为完备的综合性、一体化数据处理平台及工具体系，覆盖智能语音、计算机视觉、自然语言等全业态领域，可服务于市面上绝大多数数据处理需求，获得全球众多客户认可。截至 2023 年 3 月 31 日，公司累计客户数量已超过 800 家，公司已取得 33 项专利和 164 项计算机软件著作权。因此，发行人作为国内领先的数据服务厂商，具有深厚的技术积累和护城河优势，能够抵御政策变化、行业变化等带来的短期业绩波动风险。

3、政策变动及境外收入方面，最近一期，受宏观经济增速放缓，部分境外客户正在进行业务方向调整、阶段性裁员等，导致其预算释放进度放缓，叠加数据出境相关法规落地实施所带来的影响等多重因素，公司境外业务的开展受到影响，但长期来看，宏观经济已呈现复苏态势，境外主要客户的裁员调整已接近尾声，且相关法规的实施不会对公司持续经营能力产生实质影响，公司目前出境数据与境内个人信息关联程度较低，发行人产品的数据出境不会对国家安全产生实

质影响，公司预计随着评估审查逐步进入常态化，公司境外业务预期将逐渐回归正常水平。另一方面，公司正在根据国家及北京市的“数据二十条”的要求，积极参与数据要素的跨境流动和业务合作，并充分利用由相关政府部门牵头建设的数字贸易港、数字贸易试验区、数据跨境实验室和数据跨境服务平台等先行先试机制，推进相关工作，提高跨境业务效率。整体上，公司与境外客户之间保持长期合作关系，未出现主要客户因上述短期影响而出现流失的情况。此外，公司也将持续加大境外业务团队的建设，促进公司境外业务实现跨越式的发展。

4、人工投入及期间费用方面，公司近年来启动了研发、销售、管理领域的加速投入计划，加大了销售团队的建设投入，新引进中高端管理人员用于加大管理团队人才储备，同时引入较多算法及平台研发人员以持续提高数据处理智能化水平、建设智能驾驶垂直领域技术平台能力，大幅新增投入标准化训练数据集产品的研发投入，提高公司产品模式壁垒。虽然上述情况带来了人工投入及期间费用的上升，使得最近一年及一期业绩下滑，但人才储备及研发投入所沉淀下来的优质积累将为公司未来 3-5 年快速扩张提供持续强劲的内在动力，虽然会使短期业绩承压，但将会为公司未来的长期发展、收入扩张、业绩释放奠定坚实的基础。

5、公司在手订单及新增业务增长点方面，公司自 2022 年度起加大了智能驾驶领域的投入，并获得了多个客户的数千万元的订单。同时，公司积极加大大模型领域布局，本次募投亦在拓展 AI 大模型训练数据集产品及服务领域，探索新型业务。

6、公司也采取一系列应对措施提高公司的持续经营能力，包括：1）立足主业，紧跟下游需求变化，拓展延伸数据服务空间；2）加强前沿技术储备，夯实公司长期竞争力；3）调整优化管理模式，推动各运营环节降本增效；4）持续管理创新，提升公司治理水平；5）提升团队建设，助力企业发展。

此外，本次向特定对象发行股票募集资金将进一步扩大公司的资产规模，随着募投项目的顺利实施，本次募集资金将会得到有效使用，公司的资金实力将显著增强，核心竞争力将全面提高，有利于增强公司未来的持续经营能力和抵御风险能力。

综上，公司目前经营情况稳定，且已采取了有效的应对措施。最近一年及一

期的业绩变动事项未改变公司的行业地位和所处行业长期向好的发展趋势，不会对公司经营产生重大不利影响，不会对本次发行构成障碍。

但未来若宏观经济形势发生变化、行业出现重大调整或市场竞争加剧等，公司将面临经营业绩下滑甚至亏损的风险。公司已在本募集说明书之“第七节 与本次发行相关的风险因素”中进行重大风险提示。

十、报告期内利润分配政策、现金分红政策的制度及执行情况

（一）公司利润分配政策

公司现行有效的《公司章程》对利润分配政策规定如下：

1、基本原则

从公司盈利情况和战略发展的实际需要出发，公司实行持续、稳定的利润分配政策，公司利润分配应重视对投资者的合理、稳定投资回报并兼顾公司的可持续发展。利润分配不得超过累计可分配利润的范围，不得损害公司持续经营能力。

2、利润分配形式

公司可以采取现金、股票或现金和股票相结合的方式分配股利。公司优先选择现金分红的利润分配方式。公司具备现金分红条件的，应当采用现金分红进行利润分配。

3、现金分红的具体条件和比例

公司当年度实现盈利且累计可供分配利润为正值，若公司无重大投资或重大支出事项（指金额占公司最近一期经审计净资产的 10%以上，且绝对金额超过 500 万元），公司每年以现金方式分配的利润不少于当年实现的可分配利润的 10%。

同时公司实施差异化的现金分红政策

（1）公司发展阶段属成熟期且无重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 80%；

（2）公司发展阶段属成熟期且有重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 40%；

(3) 公司发展阶段属成长期且有重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 20%；

公司发展阶段不易区分但有重大资金支出安排的，可以按照前项规定处理。

董事会每年综合考虑公司所处行业特点、发展阶段、自身经营模式、盈利水平以及是否有重大资金支出安排等因素，根据上述原则提出当年的利润分配方案。

4、发放股票股利的具体条件

在满足上述现金分配股利之余，在保证公司股本规模和股权结构合理的前提下，基于回报投资者和分享企业价值的考虑，从公司成长性、每股净资产的摊薄、公司股本规模和公司股票价格的匹配性等真实合理因素出发，当公司股票估值处于合理范围内，公司可以提出并实施股票股利分配方案。

5、利润分配的时间间隔

公司符合本章程规定的条件，每年度进行利润分配，也可以根据盈利状况进行中期利润分配，每年度至少进行一次现金分红。

6、公司利润分配决策程序应充分考虑独立董事、外部监事和中小股东的意见，利润分配方案应履行的审议程序具体如下

在公司实现盈利符合利润分配条件时，公司董事会应当根据公司的具体经营情况和市场环境，认真研究和论证公司现金分红的时机、条件和最低比例、调整的条件及其决策程序要求等事宜，制订中期利润分配方案（拟进行中期分配的情况下）、年度利润分配方案，利润分配方案中应说明当年未分配利润的使用计划。利润分配方案需经全体董事过半数通过。

独立董事应当对董事会制订的利润分配方案进行审核并发表明确意见，并经全体独立董事三分之二以上审核同意。独立董事可以征集中小股东的意见，提出分红提案，并直接提交董事会审议。

监事会应当对董事会制订的利润分配方案进行审核并发表审核意见，并经全体监事过半数通过；如公司有外部监事（不在公司任职的监事），则外部监事应对监事会审核意见无异议。

董事会、监事会审议通过利润分配方案后应提交股东大会审议批准。现金股利分配方案，提交股东大会审议时需经出席股东大会会议的股东（包括股东代理人）所持表决权的过半数通过；涉及股票股利分配方案的，提交股东大会审议时需经出席股东大会会议的股东（包括股东代理人）所持表决权的三分之二以上通过。

股东大会对现金分红具体方案进行审议前，公司应当通过接听投资者电话、公司公共邮箱、网络平台、召开投资者见面会等多种渠道主动与股东特别是中小股东进行沟通交流，充分听取中小股东的意见和诉求，及时答复中小股东关心的问题。

（二）公司最近三年利润分配及未分配利润使用情况

1、最近三年利润分配方案

（1）2020 年度

公司于 2021 年度上市，2020 年度未进行利润分配。

（2）2021 年度

2022 年 4 月 28 日，经公司第二届董事会第七次会议审议通过 2021 年度的利润分配方案如下：以截至 2021 年 12 月 31 日，公司的总股本 42,800,000 股为基数，拟向全体股东每 10 股派发现金红利 2.50 元（含税），以此计算合计拟派发现金红利 10,700,000.00 元（含税），不进行资本公积转增股本，不进行送股。2022 年 5 月 25 日，上述利润分配方案通过公司 2021 年年度股东大会审议批准，并于 2022 年 6 月实施完毕。

（3）2022 年度

2023 年 4 月 24 日，经公司第二届董事会第十三次会议审议通过 2022 年度的利润分配方案如下：（1）以实施权益分派股权登记日登记的总股本为基数，公司拟向全体股东每 10 股派发现金红利 6.00 元（含税），截至 2022 年年度权益分派实施的股权登记日，因 2022 年限制性股票激励计划首次授予部分第一个归属期完成归属，新增股份 289,414 股，且已办理完毕新增股份的登记手续，公司总股本增加至 43,089,414 股，以此计算合计拟派发现金红利 25,853,648.40 元（含

税)。(2)公司拟以资本公积向全体股东每 10 股转增 4 股,按照总股本 43,089,414 股计算合计拟转增 17,235,766 股,转增后公司的总股本增加至 60,325,180 股。2023 年 5 月 25 日,上述利润分配方案通过公司 2022 年年度股东大会审议批准,并于 2023 年 6 月实施完毕。

2、公司最近三年现金分红情况

公司最近三年现金分红情况如下表所示,分红情况均符合法律法规及《公司章程》的有关规定。

单位:元

年度	现金分红(含税)	合并报表中归属于母公司的净利润	现金分红占合并报表归属于母公司净利润的比率
2022 年	25,853,648.40	29,454,139.23	87.78%
2021 年	10,700,000.00	31,605,431.79	33.85%
2020 年	-	82,081,021.91	-
最近三年累计现金分红金额			36,553,648.40
最近三年年均合并报表归属于母公司的净利润			47,713,530.98
最近三年累计现金分红金额占最近三年合并报表中归属于母公司年均净利润的比例			76.61%

(三) 公司最近三年未分配利润使用安排情况

最近三年,公司滚存未分配利润主要用于公司经营,以满足公司各项业务拓展的资金需求。

(四) 公司未来三年(2023 年-2025 年)股东回报规划

根据《公司法》《上市公司监管指引第 3 号——上市公司现金分红(2022 年修订)》(中国证监会公告[2022]3 号)、《关于进一步落实上市公司现金分红有关事项的通知》(证监发[2012]37 号)及《公司章程》的有关规定,为健全公司利润分配事项的决策程序和机制,增加股利分配决策透明度和可操作性,积极回报投资者,公司在兼顾公司持续发展的基础上,制定了公司《未来三年(2023 年-2025 年)股东分红回报规划》,并经第二届董事会第十七次会议审议通过。公司未来三年(2023 年-2025 年)股东回报规划主要内容如下:

1、股东分红回报规划制定考虑因素

公司着眼于公司的长远和可持续发展,在综合分析公司经营发展实际、股东

要求和意愿、社会资金成本、外部融资环境等因素，征求和听取股东尤其是中小股东的要求和意愿，充分考虑公司目前及未来盈利规模、现金流量状况、发展所处阶段、项目投资资金需求、本次发行融资、银行信贷及债权融资环境等因素，平衡股东的短期利益和长期利益的基础上制定股东分红回报规划，建立对投资者持续、稳定、科学的回报规划与机制，对股利分配做出制度性安排，并藉此保持公司利润分配政策的连续性和稳定性。

2、股东分红回报规划制定原则

（1）从公司盈利情况和战略发展的实际需要出发，公司实行持续、稳定的利润分配政策，公司利润分配应重视对投资者的合理、稳定投资回报并兼顾公司的可持续发展。利润分配不得超过累计可分配利润的范围，不得损害公司持续经营能力；

（2）公司可以采取现金、股票或现金与股票相结合的方式分配股利，优先采取现金方式；

（3）在符合分红条件的情况下，公司原则上每年度分配一次利润，但根据公司盈利情况及资金需求情况进行中期分红。

3、股东分红回报规划制定周期和相关决策机制

公司董事会应根据股东大会制定或调整的利润分配政策，至少每三年重新审阅一次《股东分红回报规划》，根据股东（特别是公众投资者）、独立董事、监事的意见对公司正在实施的股利分配政策作出适当且必要的修改，确定该时段的股东分红回报规划，并确保调整后的股东分红回报规划不违反利润分配政策的有关规定。

董事会制定的股东分红回报规划应经全体董事过半数并经全体独立董事三分之二以上同意后提交股东大会并经出席股东大会的股东所持表决权的三分之二以上审议通过。股东分红回报规划也应经全体监事过半数审议通过。

4、公司未来三年（2023年-2025年）股东分红回报具体规划

（1）利润分配的基本原则

从公司盈利情况和战略发展的实际需要出发，公司实行持续、稳定的利润分

配政策，公司利润分配应重视对投资者的合理、稳定投资回报并兼顾公司的可持续发展。利润分配不得超过累计可分配利润的范围，不得损害公司持续经营能力。

（2）利润分配的形式

公司可以采取现金、股票或现金和股票相结合的方式分配股利。公司优先选择现金分红的利润分配方式。公司具备现金分红条件的，应当采用现金分红进行利润分配。

（3）现金分红的具体条件和比例：

公司当年度实现盈利且累计可供分配利润为正值，若公司无重大投资或重大支出事项（指金额占公司最近一期经审计净资产的 10%以上，且绝对金额超过 500 万元），公司每年以现金方式分配的利润不少于当年实现的可分配利润的 10%。

同时公司实施差异化的现金分红政策：

①公司发展阶段属成熟期且无重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 80%；

②公司发展阶段属成熟期且有重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 40%；

③公司发展阶段属成长期且有重大资金支出安排的，进行利润分配时，现金分红在本次利润分配中所占比例最低应达到 20%。

公司发展阶段不易区分但有重大资金支出安排的，可以按照前项规定处理。董事会每年综合考虑公司所处行业特点、发展阶段、自身经营模式、盈利水平以及是否有重大资金支出安排等因素，根据上述原则提出当年的利润分配方案。

（4）发放股票股利的具体条件：

在满足上述现金分配股利之余，在保证公司股本规模和股权结构合理的前提下，基于回报投资者和分享企业价值的考虑，从公司成长性、每股净资产的摊薄、公司股本规模和公司股票价格的匹配性等真实合理因素出发，当公司股票估值处于合理范围内，公司可以提出并实施股票股利分配方案。

（5）利润分配的时间间隔：

公司符合公司章程规定的条件，每年度进行利润分配，也可以根据盈利状况进行中期利润分配，每年度至少进行一次现金分红。

（6）公司利润分配方案的决策程序和机制

在公司实现盈利符合利润分配条件时，公司董事会应当根据公司的具体经营情况和市场环境，认真研究和论证公司现金分红的时机、条件和最低比例、调整的条件及其决策程序要求等事宜，制订中期利润分配方案（拟进行中期分配的情况下）、年度利润分配方案，利润分配方案中应说明当年未分配利润的使用计划。利润分配方案需经全体董事过半数通过。

独立董事应当对董事会制订的利润分配方案进行审核并发表明确意见，并经全体独立董事三分之二以上审核同意。独立董事可以征集中小股东的意见，提出分红提案，并直接提交董事会审议。

监事会应当对董事会制订的利润分配方案进行审核并发表审核意见，并经全体监事过半数通过；如公司有外部监事（不在公司任职的监事），则外部监事应对监事会审核意见无异议。

董事会、监事会审议通过利润分配方案后应提交股东大会审议批准。现金股利分配方案，提交股东大会审议时需经出席股东大会会议的股东（包括股东代理人）所持表决权的过半数通过；涉及股票股利分配方案的，提交股东大会审议时需经出席股东大会会议的股东（包括股东代理人）所持表决权的三分之二以上通过。

股东大会对现金分红具体方案进行审议前，公司应当通过接听投资者电话、公司公共邮箱、网络平台、召开投资者见面会等多种渠道主动与股东特别是中小股东进行沟通交流，充分听取中小股东的意见和诉求，及时答复中小股东关心的问题。

（7）公司调整利润分配政策的研究论证程序和决策机制

①公司调整利润分配政策时，应以股东权益保护为出发点，由董事会详细论证和说明原因，并充分听取独立董事、监事和中小股东的意见。当公司遇到战争、自然灾害等不可抗力、外部经营环境变化对公司生产经营造成重大影响时，或自身经营状况发生较大变化导致现行利润分配政策无法执行时，或有权部门颁布实

施利润分配相关新规定导致公司利润分配政策必须修改时，公司将适时调整利润分配政策。调整后的利润分配政策不得违反相关法律法规以及中国证监会、证券交易所的有关规定，董事会应在相关调整议案中详细论证和说明原因。公司应依法通过接听投资者电话、公司公共邮箱、网络平台、召开投资者见面会等多种渠道主动与独立董事、股东特别是中小股东进行沟通和交流，收集独立董事、股东对公司利润分配政策调整的意见，董事会在论证调整利润分配政策时应充分考虑中小股东的意见。

②董事会审议调整利润分配政策的议案时，应经全体董事过半数并经全体独立董事三分之二以上通过。公司监事会应当对董事会调整的利润分配政策进行审议，并且经全体监事过半数通过。

公司调整的利润分配政策应经董事会、监事会审议通过后，提请股东大会审议批准，股东大会应采取现场和网络投票相结合的方式召开。股东大会审议调整利润分配政策的议案时，需经出席股东大会会议的股东（包括股东代理人）所持表决权的三分之二以上通过。

5、本规划的生效机制

本规划未尽事宜，依照相关法律法规、规范性文件以及公司章程的相关规定执行。本规划由公司董事会负责解释，自公司股东大会审议通过之日起实施。

十一、同业竞争情况

截至本募集说明书签署日，公司不存在与控股股东、实际控制人及其控制的其他企业从事相同或相似业务的情形，不存在同业竞争情况。

（一）关于避免同业竞争的承诺

公司控股股东、实际控制人贺琳已出具《关于避免同业竞争的承诺》，作出如下承诺：

“本人及其控制的企业目前不存在且不从事与海天瑞声及其子公司主营业务相同、近似或构成竞争的业务，也未发生构成或可能构成直接或间接的同业竞争的情形；

1、将来不以任何方式从事，包括与他人合作，直接或间接从事与发行人及其子公司相同、相似或在任何方面构成竞争的业务；

2、将尽一切可能之努力使本人其他关联企业不从事与发行人及其子公司相同、类似或在任何方面构成竞争的业务；

3、不投资控股于业务与发行人及其子公司相同、类似或在任何方面构成竞争的公司、企业或其他机构、组织；

4、不向其他业务与发行人及其子公司相同、类似或在任何方面构成竞争的公司、企业或其他机构、组织或个人提供专有技术或提供销售渠道、客户信息等商业秘密；

5、如果未来本人拟从事的业务可能与发行人及其子公司存在同业竞争，承诺方将本着发行人及其子公司优先的原则与发行人协商解决。

6、承诺函自出具之日起生效，并在承诺方作为发行人关联方的整个期间持续有效。”

（二）独立董事对发行人是否存在同业竞争和避免同业竞争措施的有效性所发表的意见

公司不存在与控股股东、实际控制人及其控制的其他企业从事相同或相似业务的情形，不存在同业竞争情况。控股股东、实际控制人已经承诺采取有效措施避免同业竞争或保持与发行人之间的独立性，承诺处于正常履行中，不存在违反承诺的情形，该等承诺真实、合法、有效，能够切实维护上市公司及中小股东的利益。

十二、发行人及其董事、监事、高级管理人员等相关主体的合法合规情况

公司现任董事、监事和高级管理人员不存在最近三年受到中国证监会行政处罚，或者最近一年受到证券交易所公开谴责的情形。

公司及现任董事、监事和高级管理人员不存在因涉嫌犯罪正在被司法机关立案侦查或者涉嫌违法违规正在被中国证监会立案调查的情形。

公司控股股东、实际控制人最近三年不存在严重损害上市公司利益或者投资者合法权益的重大违法行为。

公司最近三年不存在严重损害投资者合法权益或者社会公众利益的重大违法行为。

第三节 本次证券发行概要

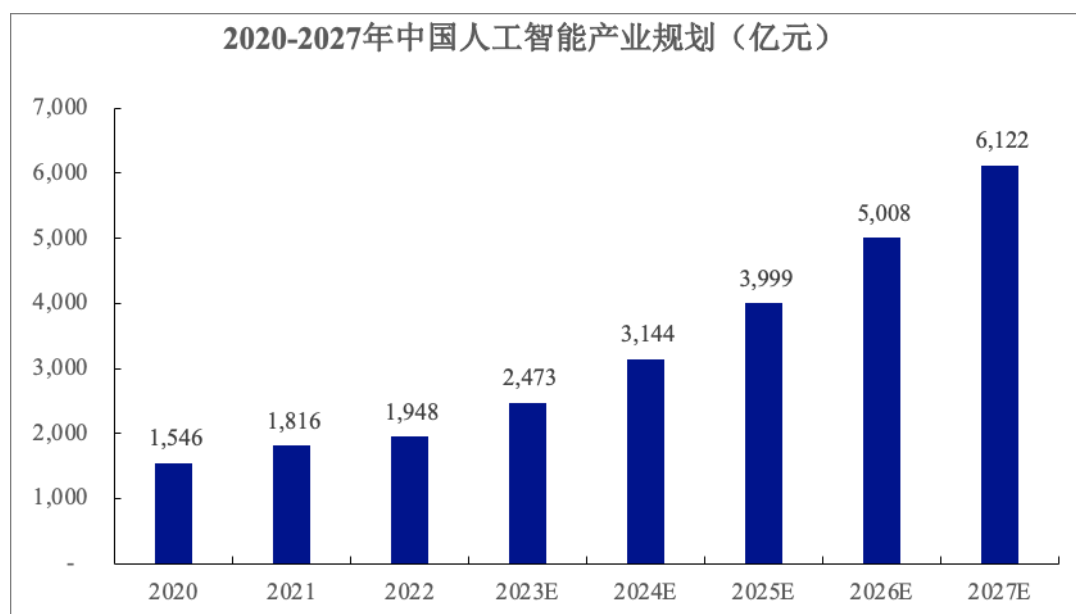
一、本次发行的背景和目的

（一）本次向特定对象发行的背景

1、全球人工智能行业快速发展，我国正加速布局 and 应对

随着互联网、云计算、物联网以及穿戴设备的发展，数据要素价值不断释放，数字化转型成为大势所趋。目前，全球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，成为数字经济发展的关键战略抓手。2022 年底 ChatGPT 的出现，掀起又一波人工智能发展热潮，以无监督学习模式为代表的预训练+人类反馈强化学习所构成的大模型技术路线的落地意味着人工智能开启发展新范式，基础模型能力通过预训练及基于人类反馈的强化学习得到不断解锁，以解决海量开放式任务。

根据艾瑞咨询数据，2022 年中国人工智能产业规模达 1,948 亿元，预计 2027 年市场规模将达到 6,122 亿元，年复合增长率为 25.6%，主要与智算中心建设以及大模型训练等需求拉动的 AI 芯片市场、无接触服务需求拉动的智能机器人及对话式 AI 市场等快速增长相关。



数据来源：艾瑞咨询

人工智能大模型因其良好的通用性、泛化性和迁移性，有助于推动人工智能

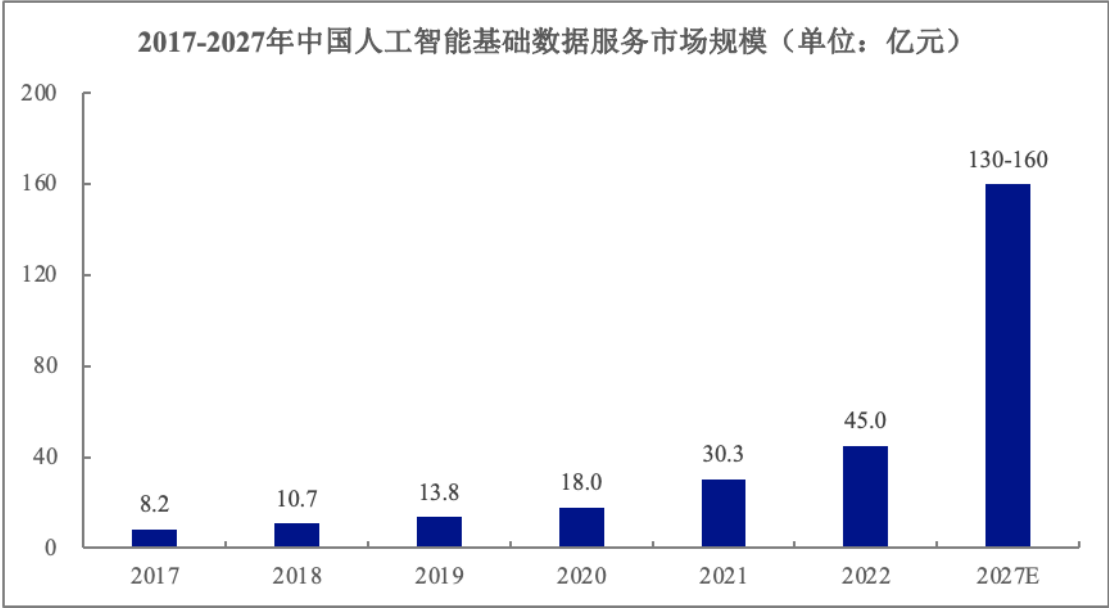
进入大规模落地应用，已成为人工智能发展新赛道。同时其强大的理解和生成能力，将驱动人工智能技术加速与实体产业融合，并深刻改变未来人类的生活和生活方式，发展大模型技术成为世界各国比拼科技实力，提升经济效率，拉升经济增长的重要动能之一。目前，国际巨头纷纷布局以大模型为核心的通用人工智能产业，产业进入加速发展期。在这一信息技术重点领域，我国与国际巨头存在一定差距，正加速布局和应对。国内众多研究机构、企业积极研究生成式 AI 大模型技术的最优路径，并进行产品发布。近期，在国内科技及投资各领域的高度关注下，百度、商汤、阿里巴巴、华为、科大讯飞、360、京东、字节跳动等企业均有所行动。

我国在“十四五”期间，针对人工智能的未来发展陆续出台了相关指导方案和激励政策，对人工智能的整体发展方向和技术发展重点做出重要规划，同时提出加强算法创新与应用、推动算力基础设施建设、完善数据基础支撑体系等关键建议，倡导未来不断夯实产业发展新基础。全国各地亦陆续出台多项数据政策，其中，《北京市促进通用人工智能创新发展的若干措施》明确提出要“系统构建大模型等通用人工智能技术体系：开展大模型创新算法及关键技术研究；加强大模型训练数据采集及治理工具研发；建设大模型评测开放服务平台；构建大模型基础软硬件体系。推动通用人工智能技术创新场景应用。”《北京市加快建设具有全球影响力的人工智能创新策源地实施方案（2023-2025 年）》提出“到 2025 年，人工智能基础理论研究取得突破；关键核心技术基本实现自主可控，其中部分技术与应用研究达到世界先进水平；人工智能高水平应用深度赋能实体经济，促进经济高质量发展”的目标，并进一步提出了“自然语言、通用视觉、多模态交互大模型等形成完整技术栈；生成式产品成为国内市场主流应用和生态平台”等具体目标。

2、人工智能基础数据服务业规模不断提升，大模型技术发展带来 AI 范式变革，正催生更多数据及其服务模式新需求

在人工智能产业链中，算法、算力和数据共同构成技术发展的三大核心要素。过去十年，人工智能产业以算法为中心，随着算法趋于开源，数据的重要性愈发凸显。在人工智能模型从技术理论到应用落地的过程中，需要依赖大量的训练数据，相较于以模型为中心的训练方法，以海量数据为中心的训练方法能够提升模

型推断结论的可靠性。万亿 GB 量级的数据随着互联网、云计算、物联网、大数据等发展源源不断地产生，但数据质量参差不齐，对海量的复杂数据进行深入挖掘、输出，进而激活和释放数据的深层价值也成为数据市场的发展重点。根据德勤数据，2022 年中国人工智能基础数据服务市场规模为 45 亿元，2027 年规模将达到 130-160 亿元，年复合增长率为 23.6%-28.9%。



数据来源：德勤

随着人工智能进入大模型时代，数据需求和数据服务模式不断提升，数据的质量以及数据清洗的工程化能力会显著拉开大模型预训练阶段的效果差距；同时，更多模型或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。对于大模型训练而言，不仅需要持续获取大规模、高质量、多模态、多场景、多垂向的数据，更须具备持续迭代的高质量数据清洗和标注策略，以不断提升包括预训练（Pre-training）、模型微调（Fine-tune）及奖励模型（Reward Model）等过程中所需数据（例如提示（Prompt）类数据）的质量，确保语言类和常识性知识之外的其他垂直领域的应用场景的能力提升，为大模型精确性、通用性及泛化能力的实现奠定坚实基础。

在以上背景下，一方面，大规模、高质量数据集重要性凸显，成为模型训练效果的核心支撑之一。除了在通用基础能力方面需要大量高质量数据训练外，大模型算法训练的需求正逐渐从通用基础能力建设向垂直领域拓展，数据需求向专业化方向发展。为加速实现人工智能产业化落地，行业将衍生出更多垂直场景的

数据需求，大模型将通过不断学习各个专业领域的行业高质量数据，实现更广阔的垂直拓展。另一方面，AI 发展所面对的数据前沿性及工程化技术的挑战也较为凸显。长期看，只有 AI 数据处理技术的不断拓新与发展，才能及时适应甚至超前引领大模型技术和应用的发展。

3、各地推动数据基础制度建设，数据要素市场迎来新发展机遇

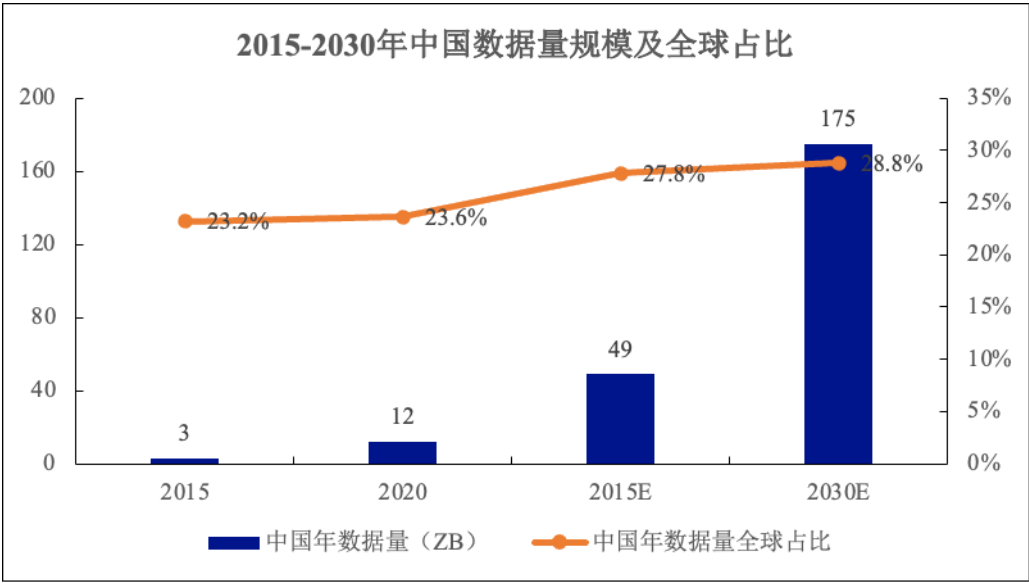
近年来，我国数字经济蓬勃发展，数据要素因具有基础性战略资源和关键性生产要素的双重属性，相关市场规模持续增长。尤其在《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》出台后，我国系统性布局了数据基础制度体系的“四梁八柱”，加速了数据流通交易和数据要素市场发展，进一步推动了公共数据、企业数据、个人数据合规高效流通使用。为更好响应中央号召，北京、上海、广州、深圳、杭州等地数据政策陆续出台，逐步构建了多层次、多元化数据要素市场生态体系。

以北京为例，《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》、《北京市促进通用人工智能创新发展的若干措施》和《关于推进北京市数据专区建设的指导意见》指出，北京市要加快建设“数据基础制度先行先试示范区”（以下简称“先行先试示范区”），“支持北京经济技术开发区等开展数据基础制度先行先试，打造政策高地、可信空间和数据工场”，探索打造数据训练基地，归集高质量基础训练数据集，推动数据要素高水平开放，提升本市人工智能数据标注库规模和质量，并建设针对重大领域、重点区域或特定场景建设专题数据区域，吸纳市场主体和数据、技术、资本等多元要素参与。北京市陆续出台的多项文件旨在打破数据壁垒，推动数据融合利用，加快推动公共数据开放，促进数据要素流通，激发数字市场创新活力，释放和发展数字化生产力，打造多层次数据要素市场，成为具有竞争力和影响力的数字产业集群。按照“政府引导、市场运作、创新引领、安全可控”的原则，“先行先试示范区”有望成为国际领先的数据要素高效流通核心枢纽。

4、我国拥有海量丰富的数据资源，但数据质量亟待提升

我国各行业的数据资源较为丰富，根据艾瑞咨询数据，2015 年-2030 年中国数据量规模由 3ZB 将增长至 175ZB，预计 2030 年中国数据量约占全球的 28.8%，

年复合增长率约为 31%，



数据来源：艾瑞咨询

虽然中国数据资源丰富，但由于数据挖掘不足，以及大量数据无法在市场上自由流通等原因，优质中文数据集仍然稀缺。以 ChatGPT 为例，其模型训练数据中，中文数据来源不足千分之一。目前，国内头部科技企业主要基于公开数据集以及自身特有的数据进行大模型训练，但由于中文优质数据质量以及数据资源的制约，国内大模型的能力与以 ChatGPT 为代表的国际大模型相比仍存在一定差距。

国内缺乏高质量数据集的主要原因包括当前国内数据挖掘和数据治理的力度不足、资金投入较大；数据流通与数据安全保障措施不够健全；国内市场缺乏开源意识，大量数据无法在市场上自由流通；国内相关公司成立较晚，数据积累较少；学术领域中文数据集受重视程度低以及国产数据集市场影响力及普及度较低等。从原始数据到可被应用的数据集产品，需要经历数据集结构设计、数据获取、数据处理（包括数据清洗、数据标注/优化等）等过程，以形成可供使用的优质数据集，国内数据服务市场的发展有助于缓解中文数据集数量不足和质量欠佳等问题。

（二）本次向特定对象发行的目的

1、丰富公司训练数据集产品类别，拓展 AI 大模型训练数据集产品及服务领域，探索新型业务

公司作为全球领先的人工智能训练数据提供商，一直深耕人工智能行业，在基础数据服务领域居于龙头地位，自成立以来，公司致力于为各类机构提供深度学习算法模型开发训练所需的专业数据集，所提供训练数据涵盖智能语音、计算机数据、自然语言处理等多个核心领域。

为更好实现公司业务扩张战略，公司在保障传统业务稳健发展的同时，不断探索寻求新的业绩增长点。公司将基于过往海量数据服务经验，结合以大模型发展为代表的行业前沿需求，研发生产大模型领域专业数据集，丰富公司数据集产品类别，拓展 AI 大模型训练数据领域服务能力，力争将大模型训练数据等创新业务打造成为具有潜在高增长价值的新型业务板块。并藉此提升行业内面向大模型训练数据集的类别和质量，实现基于大模型通用能力和垂直领域数据的支撑和训练学习，协助实现公共数据、社会数据等各类高价值数据资源汇聚。

2、巩固公司核心技术壁垒，构建大模型数据处理技术通用化能力，构建长期技术实力支撑并提升数据服务综合竞争力

数据集生产各项技术与能力和一体化技术支撑平台是公司核心技术的重要体现。本次募集资金投资项目的建设将基于公司在深度学习阶段数据集生产所积累的 know-how，自主研发海天瑞声数据生产垂直大模型，构建大模型数据处理技术通用化解决方案能力，实现完整、可持续迭代的大模型数据技术框架和数据策略；并以数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台中包括设计、获取、清洗、标注、安全管理、质控评测等不同的环节的功能，对大模型时代的数据处理需求形成全面支撑。以上举措均将进一步提高公司在人工智能基础数据服务领域的智能化水平，巩固公司的核心技术壁垒，形成长期技术实力支撑。

此外，本项目有助于进一步优化公司的数据处理技术，促进数据资源处理经验的进一步沉淀，长期来看，可以大幅提高公司的数据处理能力、效率，提升服务范围 and 水平，适应人工智能发展的新阶段，获得有效长期的发展动力，进一步

巩固和提升公司在数据服务领域的竞争力。

二、本次向特定对象发行股票的发行对象及其与发行人的关系

本次发行的发行对象为不超过三十五名（含三十五名）符合中国证监会规定条件的特定投资者，包括符合规定条件的证券投资基金管理公司、证券公司、信托公司、财务公司、保险机构投资者、合格境外机构投资者以及其他符合相关规定的法人、自然人或其他合格的投资者等。其中，证券投资基金管理公司、证券公司、合格境外机构投资者、人民币合格境外机构投资者以其管理的二只以上产品认购的，视为一个发行对象；信托公司作为发行对象的，只能以自有资金认购。

最终发行对象将在本次发行申请获得上交所审核通过并由中国证监会作出予以注册决定后，按照相关法律法规的规定及监管部门要求，由公司董事会或董事会授权人士在股东大会的授权范围内，与保荐机构（主承销商）根据询价结果协商确定。若国家法律、法规对向特定对象发行 A 股股票的发行对象有新的规定，公司将按新的规定进行调整。

截至本募集说明书出具日，本次向特定对象发行尚无确定的发行对象，因而无法确定发行对象与公司的关系。发行对象与公司之间的关系将在发行完成后公告的发行情况报告书中披露。

三、发行证券的价格或定价方式、发行数量、限售期

（一）定价基准日、定价原则和发行价格

本次发行采取询价发行方式，定价基准日为发行期首日。

本次发行价格不低于定价基准日前 20 个交易日（不含定价基准日，下同）公司股票交易均价的 80%（即“发行底价”）。定价基准日前 20 个交易日公司股票交易均价=定价基准日前 20 个交易日公司股票交易总额/定价基准日前 20 个交易日公司股票交易总量。

若公司股票在定价基准日至发行日期间发生派息、送股、资本公积转增股本等除权除息事项，本次发行底价将按以下办法作相应调整。调整公式为：

派发现金股利： $P1=P0-D$

送红股或转增股本： $P1=P0/(1+N)$

两项同时进行： $P1=(P0-D)/(1+N)$

其中，P0 为调整前发行底价，P1 为调整后发行底价，每股派息/现金分红为 D，每股送红股或转增股本数为 N。

本次发行的最终发行价格将在公司本次发行申请获得上交所审核通过并经中国证监会作出予以注册决定后，由董事会或董事会授权人士根据股东大会的授权和保荐机构（主承销商）根据询价结果协商确定，但不低于前述发行底价。

（二）发行数量

本次发行的股票数量按照募集资金总额除以发行价格确定，同时本次发行股票数量不超过本次发行前公司总股本 60,325,180 股的 20%即 12,065,036 股（含本数），最终发行数量上限以中国证监会予以注册决定要求的发行数量上限为准。在前述范围内，最终发行数量将在本次发行由中国证监会作出予以注册决定后，由公司董事会或董事会授权人士根据公司股东大会的授权和发行时的实际情况与本次发行的保荐机构（主承销商）协商确定。

在本次发行的董事会决议公告日至发行日期间，若公司发生送红股、资本公积金转增股本、股权激励、股票回购注销等事项引起公司股份变动，本次发行股票数量的上限将进行相应调整。

若国家法律、法规及规范性文件对本次发行的股票数量有新的规定或中国证监会予以注册的决定要求调整的，则本次发行的股票数量届时相应调整。

（三）限售期

本次发行的发行对象认购的股份自发行结束之日起 6 个月内不得转让。法律法规、规范性文件对限售期另有规定的，依其规定。

本次发行结束后，发行对象取得的本次发行的股份由于公司送红股、资本公积金转增股本等原因所衍生取得的公司股份，亦应遵守上述限售期安排。限售期结束后发行对象减持认购的本次发行的股票将按届时有效的中国证监会及上交所的有关规定执行。

（四）本次发行符合理性融资，合理确定融资规模

本次向特定对象发行股票的股票数量按照募集资金总额除以发行价格确定，

同时本次发行股票数量不超过本次发行前公司总股本的 20%，即本次发行的股票数量不超过 12,065,036 股（含本数）。

本次向特定对象发行股票的董事会于 2023 年 6 月 21 日召开，发行人前次首次公开发行股票募集资金于 2021 年 8 月 9 日到位，本次向特定对象发行股票董事会决议日距离前次募集资金到位时间已经超过 18 个月。

募投项目 AI 大模型训练数据集建设项目、数据生产垂直大模型研发项目均围绕公司主营业务开展，主要投向科技创新领域，符合公司未来发展战略方向，对丰富公司训练数据集产品类别、拓展 AI 大模型训练数据集产品及服务领域、巩固公司核心技术壁垒、构建长期技术实力支撑有重要意义。

综上所述，公司本次发行符合《注册管理办法》《上市公司证券发行注册管理办法》第九条、第十条、第十一条、第十三条、第四十条、第五十七条、第六十条有关规定的适用意见——《证券期货法律适用意见第 18 号》关于“理性融资，合理确定融资规模”的规定。

四、募集资金金额及投向

本次发行拟募集资金总额不超过 78,989.00 万元，扣除发行费用后的净额将全部用于以下项目：

单位：万元

序号	项目名称	投资总额	拟使用募集资金金额
1	AI 大模型训练数据集建设项目	38,337.36	38,337.36
2	数据生产垂直大模型研发项目	40,651.64	40,651.64
合计		78,989.00	78,989.00

在本次发行募集资金到位前，公司将根据募集资金投资项目的实际情况，以自筹资金先行投入，并在募集资金到位后按照相关法律、法规规定的程序予以置换。募集资金到位后，若扣除发行费用后的实际募集资金净额少于拟投入募集资金总额，在本次发行募集资金投资项目范围内，公司将根据实际募集资金数额，按照项目的轻重缓急等情况，调整并最终决定募集资金的具体投资项目、优先顺序及各项目的具体投资金额，募集资金不足部分由公司自筹解决。

若本次发行募集资金总额因监管政策变化或发行注册文件的要求予以调整

的，则届时将相应调整。

五、本次发行是否构成关联交易

截至本募集说明书签署日，本次发行尚未确定具体发行对象，最终是否存在因关联方认购公司本次向特定对象发行股票构成关联交易的情形，将在发行结束后公告的发行情况报告书中予以披露。

六、本次发行是否将导致公司控制权发生变化

本次发行前，公司控股股东及实际控制人为贺琳，贺琳直接持有公司 12,137,615 股股份（占总股本的 20.12%），并通过控制中瑞安间接控制公司 6,935,779 股股份（占总股本的 11.50%），合计控制公司 31.62%的股份，为公司实际控制人。

若假设本次发行股票数量为发行上限 12,065,036 股，则本次发行完成后（仅考虑本次发行导致的公司股份数量变化），公司的总股本为 72,390,216 股，贺琳、中瑞安分别占公司股本的 16.77%、9.58%，贺琳合计控制公司 26.35%的股份，仍为公司的控股股东、实际控制人。因此，本次发行不会导致公司的控制权发生变化。

七、本次发行方案取得有关主管部门批准的情况以及尚需呈报批准的程序

本次向特定对象发行股票相关事项已经 2023 年 6 月 21 日召开的公司第二届董事会第十七次会议、2023 年 7 月 7 日召开的 2023 年度第一次临时股东大会审议通过。尚需履行以下审批程序：

- 1、本次向特定对象发行尚待上交所审核通过；
- 2、本次向特定对象发行尚待中国证监会作出予以注册决定。

第四节 董事会关于本次募集资金使用的可行性分析

本次发行拟募集资金总额不超过 78,989.00 万元（含本数），扣除相关发行费用后将用于投资以下项目：

单位：万元

序号	项目名称	投资总额	拟使用募集资金金额
1	AI 大模型训练数据集建设项目	38,337.36	38,337.36
2	数据生产垂直大模型研发项目	40,651.64	40,651.64
合计		78,989.00	78,989.00

在本次发行募集资金到位前，公司将根据募集资金投资项目的实际情况，以自筹资金先行投入，并在募集资金到位后按照相关法律、法规规定的程序予以置换。募集资金到位后，若扣除发行费用后的实际募集资金净额少于拟投入募集资金总额，在本次发行募集资金投资项目范围内，公司将根据实际募集资金数额，按照项目的轻重缓急等情况，调整并最终决定募集资金的具体投资项目、优先顺序及各项目的具体投资金额，募集资金不足部分由公司自筹解决。

若本次发行募集资金总额因监管政策变化或发行注册文件的要求予以调整的，则届时将相应调整。

一、本次募集资金投资项目的的基本情况

（一）AI 大模型训练数据集建设项目

本项目的实施主体为北京海天瑞声科技股份有限公司。鉴于大模型训练数据通常具备数据规模大、数据质量高、数据类型丰富等特点，本项目拟建设 AI 大模型训练数据集，即生产用于通用型及各种垂直领域大模型训练的海量、高品质数据集。本项目拟购置办公楼作为建设大模型训练数据研发生产基地，并购置数据采集、数据处理、数据存储和办公等软硬件设备，利用海量、高质量、多样化的公共数据资源、社会数据资源和稀缺性数据源，通过数据集设计、数据采集/获取、清洗/分类/标准化、标注/优化、评测等全流程的任务执行进行高质量大模型训练数据集建设。

本项目将充分利用“先行先试示范区”在基础制度、数据供给等方面的先行先试政策，采用多元化的方式获取大规模原始数据；利用工程化的数据处理技术

进行预训练阶段的数据清洗；采用人类反馈强化学习模式，基于微调和奖励模型训练的方法，以人类撰写少量的典型问题和标准答案与深度学习阶段基础性标注相结合的模式，生产出市场适用性较强的大模型训练数据集。

本项目建成后，将提供可供大模型训练和评测的不少于 10 个品类的专业数据集，显著提升行业内面向大模型训练数据集的类别和质量，协助实现公共数据、社会数据等各类高价值数据资源汇聚，实现基于大模型通用能力和垂直领域数据的训练学习。本项目的数据集产品具体可分为三大类：

①通用及特定垂直领域的大语言模型训练数据集，包括但不限于：

A、中文大模型预训练语料数据集（含通用场景、特定场景、对话场景、指令集等）；

B、多语言大模型预训练语料数据集（含通用场景、对话场景、指令集等）。

②多模态大模型训练数据集：可应用于多语言图文大模型训练、多模态数字人训练、多语种语音大模型训练、全场景自动驾驶大模型训练等场景的跨模态数据集。

③大模型评测数据集：可应用于大模型的能力、任务、指标等方面的评测。

（二）数据生产垂直大模型研发项目

本项目建设目标为通过大模型基础研究，研发海天瑞声数据生产垂直大模型，并以海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台。本项目的实施主体为北京海天瑞声科技股份有限公司。

为应对大模型时代下数据规模量极大、复杂性和多样性高，数据服务规则设计难度指数级提升等诸多问题，且为更高效高质完成数据规则的规模化生产，公司将采用全栈自研的数据生产垂直大模型技术，辅助完成面向多个下游任务的数据设计与处理规则。同时，为更好实现数据生产垂直大模型的生成能力，公司将研发并引入多项新兴技术，夯实数据生产垂直大模型构建的基础。

此外，基于大模型的核心能力，项目还将升级海天瑞声一体化技术支撑平台，使其能够全面拥有大模型范式下的数据服务能力。通过嵌入预训练数据下载工具、预训练数据清洗工具、指令数据集筛选工具、指令数据集生成与调优工具、

大模型评测数据集评测工具、大模型评测数据集质检工具、多模态数据集生产工具等模块，完成大模型的数据获取与处理工作，打造模型训练、模型评测的能力。

图：海天瑞声新一代基于数据生产垂直大模型的数据服务技术架构图



二、募投项目的必要性及可行性

（一）AI 大模型训练数据集建设项目

1、项目建设必要性

（1）本项目建设是响应国家建立数据基础制度，落实北京建设“先行先试示范区”的必然选择

党的十八大以来，习近平总书记屡次强调建设数字中国以及构建数据要素的重要性，并明确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。为进一步推动国家数字经济发展，发挥数据要素在经济发展中的重要价值，我国推出《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》，从顶层设计角度，在数据产权、流通交易、收益分配、安全治理等方面构建了数据发展的基础制度和规划纲要，以促进数据合规高效流通使用，充分发挥中国海量数据规模和丰富应用场景优势，赋能实体经济，激活数据要素的潜能。

北京市则率先开展国家数据基础制度“先行先试示范区”建设，2023 年 5 月发布的《北京市促进通用人工智能创新发展的若干措施》指出，充分发挥政府引导作用和创新平台催化作用，整合创新资源，加强要素配置，营造创新生态，提升高质量数据要素供给能力，归集高质量基础训练数据集，2023 年 7 月发布的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》提出，“发

展数据生产服务业，支持企业开展数据采集、清洗加工、存储计算、数据分析、数据标注、数据训练等数据生产服务，支持企业研发建设数据生产线，推进数据生产自动化”。公司作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，有义务和责任积极响应北京建设“先行先试示范区”的号召，通过本项目的实施有效助力数据要素市场培育，推动数字经济创新发展，为北京市加快建设全球数字经济标杆城市提供助力。

（2）本项目建设是践行国家规范生成式人工智能产品要求的重要举措

生成式人工智能产品因其复杂性可能带来社会风险、技术伦理风险、企业商业秘密和个人信息泄露风险、虚假信息风险、知识产权侵权风险及其他潜在风险。为了更好地促进生成式人工智能技术健康发展和规范应用，国家网信办等七部门于2023年7月出台了《生成式人工智能服务管理暂行办法》，该办法从内容合规、数据来源合法性、知识产权及商业秘密保护、虚假信息防范等方面，对生成式人工智能产品提出了全方位的合规要求。该办法明确提出，“提供者应当依法开展预训练、优化训练等训练数据处理活动”、“采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性”。

根据前述规定，数据获取、数据处理的高标准意味着数据获取难度及处理成本将大幅增加，以预训练阶段为例，由于大量数据来源应合法合规，需投入大量成本完成数据获取。因此，出于成本与数据集质量的平衡性考量，在大模型训练中，大模型厂商通常会选择与专业的第三方数据集厂商合作，由专业第三方提供的合规、高质量数据集或相关解决方案将成为践行国家规范生成式人工智能产品要求的重要举措。

（3）本项目建设是支撑大模型训练，提升大模型输出能力的有效方式

随着人工智能应用场景日益丰富、产品智能化要求的不断提升，数据需求逐渐向海量、高质量、多元化方向演进。从自然数据源简单收集、获取的数据资源，通常无法直接满足大模型的训练需求，需经专业化的数据分类设计、清洗、加工处理，形成相应的工程化数据，以供大模型训练使用。一般而言，符合大模型训练标准的数据需具备质量高、规模大、样本丰富等三个特点。首先，海量具有无毒害性、公平性等高质量特征的数据集能够提高模型效果（例如，精度与可解释

性)，并且减少收敛到最优解的时间；其次，在强化学习阶段，原始数据由于存在信息量低、含有噪声或需补齐等问题，使用前需要进行数据对齐等诸多微调操作，优秀的指令数据集能够帮助大模型更好的泛化适配更多下游任务。再次，数据丰富程度能够显著提高大模型的泛化能力，减少过拟合情况的发生，达到更优的模型效果。

当前国内数据资源虽然丰富，但优质的中文大模型训练数据仍然稀缺，中文大模型训练数据数量与质量，受国内产业环境、数据积累程度、数据运营生态等因素影响，与全球领先国家仍存在一定差距，使得国内大模型难以拥有足够专业的数据资源进行训练。本项目通过提供覆盖预训练、强化学习及应用拓展阶段的海量、高质量专业数据集，更好的支撑大模型训练，提升大模型输出能力。

（4）本项目建设符合公司“夯实传统业务，探索新型业务”的战略目标

为更好实现公司业务发展战略，公司在保障人工智能基础数据业务稳健发展的同时，不断探索寻求新的业绩增长点。如前文所述，数字经济时代下，数据要素市场发展前景广阔，大模型等人工智能技术已成为国家科技发展的重要抓手，但国内数据仍存在数据质量差、各领域数据无法流通等问题制约了人工智能行业的发展。公司将基于过往的数据服务经验，结合行业前沿需求，积极拓展大模型训练数据服务领域，力争将大模型训练数据等创新业务打造成为具有潜在高增长价值的新型业务板块。

2、项目建设可行性

（1）数据要素政策红利持续释放，利好政策支撑数据服务产业发展

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，多项政策密集出台为本项目的顺利实施提供了政策保障，具体内容如下：

序号	发布时间	颁布主体	主要行业政策及法律法规	相关内容
1	2023 年 7 月	中共北京市委、北京市人民政府	关于更好发挥数据要素作用进一步加快发展数字经济的实施意见	发展数据生产服务业，支持企业开展数据采集、清洗加工、存储计算、数据分析、数据标注、数据训练等数据生产服务，支持企业研发建设数据生产线，推进数据生产自动化。培育人工智能生成内容产业发展，发展人工智能生成语音、

序号	发布时间	颁布主体	主要行业政策及法律法规	相关内容
				图像和自然语言等内容，丰富合成数据供给。 打造数据基础制度综合改革试验田，支持北京经济技术开发区等开展数据基础制度先行先试，打造政策高地、可信空间和数据工场。通过物理集中和逻辑汇通相结合的方式，导入工业、金融、能源、科研、商贸、电信、交通、医疗、教育等领域数据资源，促进数据跨行业融合应用，切实激活数据要素资源。
2	2023 年 7 月	国家互联网信息办公室等七部门	《生成式人工智能服务管理暂行办法》	生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：（一）使用具有合法来源的数据和基础模型；（二）涉及知识产权的，不得侵害他人依法享有的知识产权；（三）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其它情形；（四）采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；（五）《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律、行政法规的其它有关规定和有关主管部门的相关监管要求。
3	2023 年 5 月	北京市科学技术委员会、中关村科技园区管理委员会	北京市促进通用人工智能创新发展的若干措施	归集高质量基础训练数据集：组织有关机构整合、清洗中文预训练数据，形成安全合规的开放基础训练数据集；持续扩展多模态数据来源，建设高质量的文字、图片、音频、视频等大模型预训练语料库。谋划建设数据训练基地：加快建设数据基础制度先行先试示范区，探索打造数据训练基地，推动数据要素高水平开放，提升本市人工智能数据标注库规模和质量。
4	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》	数据作为新型生产要素，是数字化、网络化、智能化的基础，维护国家数据安全，促进数据合规高效流通使用。
5	2022 年 11 月	北京市人大常委会	《北京市数字经济促进条例》	从立法层面，加强数字基础设施建设，培育数据要素市场，规范公共数据的汇聚、清洗、共享、开放、应用和评估管理机制，开展公共数据专区授权运营。
6	2022 年 1 月	国务院	《“十四五”数字经济发展规划》	强化高质量数据要素供给、加快数据要素市场化流通、创新数据要素开发利用机制等重点任务举措
7	2021 年 3 月	十三届全国	《中华人民共	加强关键数字技术创新应用，建设重点

序号	发布时间	颁布主体	主要行业政策及法律法规	相关内容
		人大四次会议	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》	行业人工智能数据集，发展算法推理训练场景。

（2）大模型驱动人工智能发展全面提速，新型训练数据服务具备市场空间

随着人工智能大模型技术的发展，行业对数据的依赖程度逐步加深。本项目产出的大模型训练数据集拟显著改善大模型训练中，包括预训练数据获取、清洗、强化学习调优、对齐、应用阶段评测等各个阶段的数据规模与质量问题。该类数据集将有效提升行业内面向大模型训练数据集的类别和质量，并保障数据来源与处理合法合规，也将发挥规模化运营的优势，平衡数据集成本与市场效益，实现基于大模型通用能力和垂直领域数据的支撑和训练学习，协助实现公共数据、社会数据等各类高价值数据资源汇聚。本项目与公司多年发展中持续运行的商业模式相契合，市场空间广阔，具备可行性。

（3）公司具备较强的数据生产及服务等综合能力，为项目实施奠定基础

①公司拥有深度学习的技术储备，为新业务提供技术支撑

自 2005 年以来，公司始终致力于为 AI 深度学习提供算法模型开发训练所需的专业数据集，提升模型推断结论的可靠性。公司现已积累较为完备的综合性、一体化数据处理平台及工具体系，覆盖智能语音、计算机视觉、自然语言等全业态领域，可服务于市面上绝大多数数据处理需求。截至 2023 年 3 月 31 日，公司已取得 33 项专利和 164 项计算机软件著作权，覆盖平台工具开发、算法研究、产品设计等多方面。此外，公司还设置了 AI+研发部门，前瞻性挖掘和布局新兴市场需求，抢占市场先机。

公司现有的深度学习模型数据主要是通过定向采集、精细化标注实现，即通过打标签的方式将数据类别、位置、性状、结构等信息进行精细化标注，提供给深度学习模型进行学习。大模型的训练则需要以海量数据为基础，对数据的缺失值、异常值、格式等进行清洗处理，通过高效的、多元化的、专业的人类反馈不断强化和优化模型训练，提升大模型与用户交互过程中的反馈质量。公司可将现有业务的技术储备复用到大模型业务中，将深度学习数据集生产中积累的能力延

伸使用至大模型训练数据集生产。

②公司具有丰富的、多领域数据集产品生产经验，为新业务奠定经验基础

公司的标准化数据集产品是公司区别于众多竞争对手以定制化服务为主的特有商业模式，在多语种及多音色语音数据集和发音词典、动作捕捉等多模态数据集、以及多语种 OCR 和手写体数据集等方面积累了丰富的标准化产品资源。截至 2022 年 12 月 31 日，公司拥有智能语音数据集产品储备 927 个、计算机视觉数据集产品储备 125 个、自然语言数据集产品储备 282 个。经过多年积累，公司已向下游客户提供了累计约 6,000 次/个定制或标准化训练数据集，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 类创新应用领域，构建出独具特色的训练数据资源及服务能力集群，公司在标准化数据集产品的能力获得市场认可，并为后续标准化数据产品生产奠定坚实基础。

③公司已经服务全球众多科技巨头，为新业务提供客户资源基础

公司自 2005 年成立以来，始终致力于挖掘行业客户需求，解决客户痛点，通过在智能语音、计算机视觉、自然语言等领域的技术积累，获得全球众多客户认可，包括阿里巴巴、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、清华大学等全球主流企业、教育科研机构以及政企机构。截至 2022 年底，公司累计服务客户数量已超过 800 家。公司的存量客户与新业务的客户重合程度较高，且存量客户群中的部分头部企业已输出或计划输出其大模型产品与服务，为公司该项新业务拓展提供了客户资源基础。

④公司历来重视数据安全及合规体系能力建设，为新业务提供合规保障

公司一直以来非常重视数据安全能力及合规体系建设，数据安全管理工作获得市场认可。资质方面，公司拥有 ISO27001 信息安全管理体系认证、ISO27701 隐私信息管理体系认证、国家信息安全等级保护三级认证、北京市规划和自然资源委员会行政许可乙级测绘资质等。行业参与方面，公司入选中共中央网络安全和信息化委员会办公室“人工智能企业典型应用案例”，成为中国信通院数据安全推进计划成员单位，董事兼副总经理李科入选该计划数安智库专家，发表《AI

训练数据安全实践》等文章，为人工智能领域数据安全建言献策，并荣获数安智库 2022 年度优秀专家称号；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”。

公司一直坚持安全与发展并重的原则，持续进行数据安全合规能力建设，建立了较强的数据合规体系并积累了丰富的数据合规实践经验，为大模型开展合规训练提供合规保障。

（4）公司实施本项目在经济效益和社会效益上具备可行性

基于谨慎测算，本项目内部收益率高于社会基准折现率，说明项目的经济效益较好，盈利能力较强。本项目生产的产品属于国家鼓励的行业发展方向，能够带动产业链上下游各企业协同发展，具备社会效益。

综上，从经济效益和社会效益分析来看，该项目具备较强可行性。

（二）数据生产垂直大模型研发项目

1、项目建设必要性

（1）本项目建设是落实国家科技创新发展战略的重要举措

人工智能是战略性新兴产业的重要组成部分，对我国经济发展和提升国家战略安全具有重要意义。在世界政治经济格局加速重构的影响下，未来逆全球化趋势仍将延续。全球产业合作格局重构、国际分工体系全面调整，关键环节的国际竞争将加剧，我国在关键核心技术上的问题愈发突出，战略性新兴产业的产业链安全稳定存在潜在隐患。因此，我国需要进一步集中优势资源，在重点领域加快突破一批关键核心技术，助力提升我国新兴产业的产业链关键环节、关键领域、关键产品的安全保障能力，保障国家战略安全。

公司是我国人工智能数据服务领域的龙头提供商，本项目以研发数据生产垂直大模型为核心，并基于该生产垂直大模型对数据集生产的强大支撑能力，升级海天瑞声一体化技术支撑平台，持续以自主可控的技术与平台为我国人工智能技术与产业发展提供支撑。本项目的建设是公司落实国家科技创新发展战略的重要

举措。

(2) 本项目建设是巩固公司的核心技术壁垒，构建长期技术实力支撑的必然手段

随着人工智能从深度学习阶段走向大模型阶段，对训练数据服务产生了新的需求，具体可分为预训练阶段和强化学习阶段：在预训练阶段，模型所需的数据量巨大；在强化学习阶段，模型所需的数据质量较高，并需要以相关领域 know-how 作为模型输入。此外，随着多模态大模型的不断发展，跨语音、文本和视频图像数据等多种类别的数据集需求将快速增加。

数据集生产能力和一体化技术支撑平台是公司核心技术的重要体现。目前 ChatGPT 等模型执行通用生成任务的效果证明了大模型可具备数据生成能力。本项目的建设将基于公司在深度学习阶段数据集生产所积累的 know-how，自主研发数据生产垂直大模型，构建大模型数据处理技术通用化解决方案能力，实现完整、可持续迭代的大模型数据技术框架和数据策略，进一步提高公司在人工智能基础数据服务领域的智能化水平，巩固公司的核心技术壁垒，形成长期技术实力支撑。

(3) 本项目建设是优化公司数据集生产流程，提升数据服务综合竞争力的有力途径

大模型训练数据集的生产流程包括设计、获取（模型生成）、清洗、标注、安全管理、质控评测等不同的环节。系统化的开发平台和专业化的软件处理工具对应对大模型时代的数据处理需求和全流程支撑至关重要。本项目有助于进一步优化公司的数据处理技术，促进数据资源处理经验的进一步沉淀，长期来看，可以大幅提高公司的数据处理能力、效率，提升服务范围和水平，适应人工智能发展的新阶段，获得有效长期的发展动力，进一步巩固和提升公司在数据服务领域的竞争力。

2、项目建设可行性

(1) 本项目建设符合政策要求和行业发展趋势

2023 年 7 月，国家互联网信息办公室等七部门公布《生成式人工智能服务管理暂行办法》，文件明确指出，“鼓励生成式人工智能算法、框架、芯片及配

套软件平台等基础技术的自主创新，平等互利开展国际交流与合作，参与生成式人工智能相关国际规则制定”，“生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动”。该办法从政策层面对生成式人工智能的数据集提出了明确的合法、合规、合理、准确以及知识产权清晰的高要求。

但目前国内大模型的发展普遍存在数据来源不均衡、数据更新实时性弱、垂直类型数据不足、指令集质量欠佳且存在偏见等问题，由此导致大模型的效果、效率、合规性、合理性等方面亟待完善与提升，且在大模型持续发展过程中，部分问题的影响可能持续扩大。因此，建立一套完整、完善、可持续迭代的大模型训练数据技术框架和数据策略，符合生成式人工智能技术与应用合规、高效发展的趋势。

（2）公司与现有客户、科研院所联系紧密，可确保项目技术框架明确、技术路线可行有效

公司自 2005 年成立以来，始终致力于挖掘行业客户需求，解决客户痛点，通过在智能语音、计算机视觉、自然语言等领域的技术积累，获得全球众多客户认可，截至 2022 年底，公司累计客户数量已超过 800 家。公司现有客户包括阿里巴巴、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、清华大学等全球主流企业、教育科研机构以及政企机构。

公司部分现有客户是当前大模型领域的积极实践者，通过与客户的长期合作，深度交流，能够第一时间获取大模型研发中数据痛点与需求，并可在持续交流反馈中不断修正本项目的建设方案。此外，公司也与科研院所和高校等开展深入合作，可引入外部专家资源，以保证技术路线的可行性。

（3）公司拥有深厚的技术和人才储备，具有完成本项目的技术基础

公司深耕行业近 20 年，拥有一支高素质的研发团队，公司高管及核心研发人员大多毕业于清华、北大、复旦等一流院校，大部分曾在微软、阿里巴巴、英特尔、IBM、中科院等业内领先的成熟企业与研究机构担任人工智能领域技术研发与管理的领导职务。截至 2022 年 12 月 31 日，公司研发人员达到 82 人，经验丰富的技术团队为本项目的执行提供了人才保证。

截至 2022 年底，公司拥有算法模型框架 16 个、算法模型数量超过 200 个，公司自然语言理解算法支持包括语义理解、情感分析和意图识别等能力，语音识别算法支持语种 58 个，计算机视觉算法支持几十大类、上百小类的物体识别。公司在智能语音、自然语言、计算机视觉领域均有多年算法积累，该等算法模型能够全面支撑公司多个领域数据生产活动的开展。

三、募投项目经营前景

本次募投项目“AI 大模型训练数据集建设项目”投资金额 38,337.36 万元，经测算，税后内部收益率为 16.82%，税后投资回收期（含三年建设期）为 5.89 年，经济效益良好。

本次募投项目“数据生产垂直大模型研发项目”是公司落实发展战略，顺应行业发展趋势，支撑公司加速数据服务领域算法能力建设、持续构建 AI 产业核心竞争力的必要手段。本项目不直接产生效益，项目建成后将成为公司主营业务长期发展的技术底座。

本次募投项目拟投向的产品符合国家产业政策，经营发展前景良好，具体参见本募集说明书“第二节 发行人基本情况”之“三、所处行业的主要特点及行业竞争情况”。

四、募投项目与现有业务或发展战略的关系

本次募集资金拟用于“AI 大模型训练数据集建设项目”和“数据生产垂直大模型研发项目”，募集资金投资项目均围绕公司主营业务进行。

AI 大模型训练数据集建设项目将拓展训练数据产品体系，公司顺应人工智能大模型的发展趋势，大幅提升训练数据产品的数据规模、广度、质量和精度，扩大产品覆盖的场景和领域，是前瞻性布局大模型领域、拓展潜在高增长价值的新型业务板块的重要举措；数据生产垂直大模型研发项目以研发海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台，进一步延伸大模型领域综合数据服务能力，旨在提高公司业务能力的维度、广度、效率和水平，巩固公司的核心技术壁垒，构建长期技术实力支撑，不断提高公司核心竞争力。

此外，本次募投项目建设符合公司“夯实传统业务，探索新型业务”的战略

目标。为更好实现公司业务发展战略，公司在保障人工智能基础数据业务稳健发展的同时，不断探索寻求新的业绩增长点。如前文所述，数字经济时代下，数据要素市场发展前景广阔，大模型等人工智能技术已成为国家科技发展的重要抓手，但国内数据仍存在数据质量差、各领域数据无法流通等问题制约了人工智能行业的发展。公司将基于过往的数据服务经验，结合行业前沿需求，积极拓展大模型训练数据服务领域，力争将大模型训练数据等创新业务打造成为具有潜在高增长价值的新型业务板块。

五、募投项目的实施准备和进展情况，预计实施时间，整体进度安排，发行人的实施能力及资金缺口的解决方式

（一）募投项目的实施周期

1、AI 大模型训练数据集建设项目

本项目建设期 3 年，具体进度安排如下表：

序号	时间安排	Y1				Y2				Y3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	场地购置及装修												
2	设备购置												
3	硬件软件平台搭建												
4	人员引进与培训												
5	AI 大模型训练数据集生产												
6	产品销售												

注：Y1、Y2、Y3 代表建设期年份，Q1、Q2、Q3、Q4 代表季度。

2、数据生产垂直大模型研发项目

本项目建设期 3 年，具体进度安排如下表：

序号	时间安排	Y1				Y2				Y3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	场地购置及装修												
2	设备购置												
3	硬件软件平台搭建												

序号	时间安排	Y1				Y2				Y3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
4	人员引进与培训												
5	项目研发												

注：Y1、Y2、Y3 代表建设期年份，Q1、Q2、Q3、Q4 代表季度。

（二）发行人的实施能力

1、人员储备

公司拥有高素质的管理、运营及研发团队，公司高管及核心研发人员大多毕业于清华、北大、复旦等一流院校，大部分曾在微软、阿里巴巴、英特尔、IBM、中科院等业内领先成熟企业与研究机构担任人工智能领域技术研发与管理的领导职务。截至 2022 年 12 月 31 日，公司研发人员达到 82 人，经验丰富的技术团队为本项目的执行提供了人才保证。

2、技术储备

截至 2022 年底，公司拥有算法模型框架 16 个、算法模型数量超过 200 个，公司自然语言理解算法支持包括语义理解、情感分析和意图识别等能力，语音识别算法支持语种 58 个，计算机视觉算法支持几十大类、上百小类的物体识别。公司在智能语音、自然语言、计算机视觉领域均有多年算法积累，该等算法模型能够全面支撑公司多个领域数据生产活动的开展。

3、市场储备

公司自 2005 年成立以来，始终致力于挖掘行业客户需求，解决客户痛点，通过在智能语音、计算机视觉、自然语言等领域的技术积累，获得全球众多客户认可，截至 2022 年底，公司累计客户数量已超过 800 家。公司现有客户包括阿里巴巴、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、清华大学等全球主流企业、教育科研机构以及政企机构。产品广泛应用于客户所研发的个人助手、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。公司的存量客户与本次募集资金投向新业务领域的客户重合程度较高，本次募集资金投资项目具有稳定的客户资源和市场储备。

（三）项目资金缺口解决方式

本次募集资金投资项目中“AI 大模型训练数据集建设项目”总投资额为 38,337.36 万元，“数据生产垂直大模型研发项目”总投资额为 40,651.64 万元，全部使用募集资金，不涉及项目资金缺口。在本次发行募集资金到位前，公司将根据募集资金投资项目的实际情况，以自筹资金先行投入，并在募集资金到位后按照相关法律、法规规定的程序予以置换。

六、募投项目投向科创领域，募投项目实施促进公司科技创新水平提升的方式

（一）本次募集资金投向符合国家产业政策，主要投向科技创新领域

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。本次募集资金总额全部用于 AI 大模型训练数据集生产和数据生产垂直大模型的研发，系围绕公司主营业务展开。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。因此，本次募集资金投资项目投向属于科技创新领域。

（二）本次募集资金投资项目将进一步提升公司科技创新水平

在人工智能产业进入以大模型为代表的新的发展时期，通过本次募投项目的实施，公司将建设一批适用性较强的大模型训练数据集，拓展潜在高增长价值的新型业务板块，并藉此进一步扩大公司业务规模；同时，以研发海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台，研发并引入多项新兴技术，促进公司科技创新水平的不断提升，巩固公司的核心技术壁垒，构建长期技术实力支撑，从而进一步增强公司核心竞争力。

七、募投项目的效益测算

(一) AI 大模型训练数据集建设项目

1、假设条件

本项目在效益测算中主要基于如下假设：

- (1) 假定国家及项目建设所在地区的社会经济环境无重大变化；
- (2) 假定项目所处的行业领域和上游行业领域产业政策无重大变化，没有发生重大的市场突变；
- (3) 假定公司在项目建设期内各部门建设和人员招聘均按计划进行，不会发生剧烈变动；
- (4) 无其他不可抗力及不可预见因素对公司经营造成重大不利影响。

2、测算数据及测算过程

项目运营期 8 年，其中建设期 3 年，本项目相关的经济效益财务指标预计如下：

序号	全部投资税后指标	单位	数值
1	静态回收期（含建设期）	年	5.89
2	财务内部收益率	%	16.82
3	财务内部净现值(i=12%)	万元	4,131.15

结合公司实际经营情况与市场发展趋势，综合考虑本项目的收入、费用等各项指标，得到效益测算情况如下表所示：

单位：万元

序号	项目	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
1	营业收入	2,623.15	6,564.40	10,327.83	9,295.05	8,365.54	7,528.99	6,776.09	6,098.48
2	税金及附加	-	-	9.42	118.11	107.70	98.14	89.44	81.57
3	销售费用	131.16	328.22	516.39	464.75	418.28	376.45	338.80	304.92
4	管理费用	521.56	831.20	1,016.45	985.46	957.58	924.31	891.84	863.55
5	研发费用	2,550.72	5,104.44	7,385.21	2,681.32	2,086.28	1,517.41	1,077.89	813.49
6	退税收入	209.85	525.15	826.23	743.60	669.24	602.32	542.09	487.88
7	利润总额	-370.44	825.70	2,226.59	5,789.01	5,464.95	5,215.00	4,920.19	4,522.82
8	所得税费用	-	68.29	333.99	868.35	819.74	782.25	738.03	678.42
9	税后利润	-370.44	757.41	1,892.60	4,920.66	4,645.21	4,432.75	4,182.16	3,844.40

（二）数据生产垂直大模型研发项目

本项目是公司落实发展战略，顺应行业发展趋势，支撑公司加速数据服务领域算法能力建设、持续构建 AI 产业核心竞争力的必要手段。本项目不直接产生效益，项目建成后将成为公司主营业务长期发展的技术底座。

八、本次募投项目非资本性支出的情况

本次募集资金将用于“AI 大模型训练数据集建设项目”及“数据生产垂直大模型研发项目”，总投资额 78,989.00 万元，资本性支出 55,499.59 万元、非资本性支出 23,489.41 万元。本次募集资金投资项目中非资本性支出合计金额未超过募集资金总额的 30%，本次发行符合《注册管理办法》第三十条以及《证券期货法律适用意见第 18 号》的相关规定。

九、募投项目的审批情况

（一）AI 大模型训练数据集建设项目

AI 大模型训练数据集建设项目已取得北京市投资项目备案证（备案证号：京技审项（备）[2023]150 号）。本项目不同于常规生产性项目，不存在废气、废水、废渣等工业污染物，不属于根据《中华人民共和国环境影响评价法》和《建设项目环境影响评价分类管理名录》等相关法律法规需要进行环境影响评价的建设项目。因此，本项目无需进行项目环境影响评价，亦不需要取得环保主管部门对项目的审批文件。

（二）数据生产垂直大模型研发项目

数据生产垂直大模型研发项目已取得北京市投资项目备案证（备案证号：京技审项（备）[2023]149 号）。本项目不同于常规生产性项目，不存在废气、废水、废渣等工业污染物，不属于根据《中华人民共和国环境影响评价法》和《建设项目环境影响评价分类管理名录》等相关法律法规需要进行环境影响评价的建设项目。因此，本项目无需进行项目环境影响评价，亦不需要取得环保主管部门对项目的审批文件。

综上，本次募投项目已完成需履行的程序，符合国家产业政策和有关环境保护等法律、行政法规规定，不存在重大不确定性。

十、募集资金用于研发投入的情况

（一）本次募投项目研发投入的主要内容

本次募投项目中，AI 大模型训练数据集建设项目主要涉及与产品开发相关的产品研发类投入，数据生产垂直大模型研发项目为基础研发类项目，建设目标为通过大模型基础研究，研发海天瑞声数据生产垂直大模型，并以海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台。本次项目的实施主体为北京海天瑞声科技股份有限公司。

为应对大模型时代下数据规模量极大、复杂性和多样性高，数据服务规则设计难度指数级提升等诸多问题，且为更高效高质完成数据规则的规模化生产，公司将采用全栈自研的数据生产垂直大模型技术，辅助完成面向多个下游任务的数据设计与处理规则。同时，为更好实现数据生产垂直大模型的生成能力，公司将研发并引入多项新兴技术，夯实数据生产垂直大模型构建的基础。

此外，基于大模型的核心能力，项目还将升级海天瑞声一体化技术支撑平台，使其能够全面拥有大模型范式下的数据服务能力。通过嵌入预训练数据下载工具、预训练数据清洗工具、指令数据集筛选工具、指令数据集生成与调优工具、大模型评测数据集评测工具、大模型评测数据集质检工具、多模态数据集生产工具等模块，完成大模型的数据获取与处理工作，打造模型训练、模型评测的能力。

图 海天瑞声新一代基于数据生产垂直大模型的数据服务技术架构图



（二）技术可行性

公司深耕行业多年，持续加大研发投入，不断提升技术和平台工具的先进性，拥有丰富的技术储备：截至 2022 年底，公司拥有算法模型框架 16 个、算法模型

数量超过 200 个，公司自然语言理解算法支持包括语义理解、情感分析和意图识别等能力，语音识别算法支持语种 58 个，计算机视觉算法支持几十大类、上百小类的物体识别。公司在智能语音、自然语言、计算机视觉领域均有多年的算法积累，该等算法模型能够全面支撑公司多个领域数据生产活动的开展。

综上，本次募投项目在技术上具有可行性。

（三）研发预算

数据生产垂直大模型项目总投资 40,651.64 万元，分 3 年投入，具体情况如下：

单位：万元					
序号	项目	金额	第一年	第二年	第三年
1	场地购置及装修	2,346.00	2,346.00	-	-
1.1	场地购置费用	2,208.00	2,208.00	-	-
1.2	场地装修费用	138.00	138.00	-	-
2	设备购置费用	29,895.00	12,145.20	9,497.60	8,252.20
3	软件购置费用	451.89	235.77	117.89	98.24
4	研发人员费用	4,902.50	1,050.00	1,667.50	2,185.00
5	设备托管费用	3,056.25	588.75	1,042.50	1,425.00
合计		40,651.64	16,365.72	12,325.49	11,960.44

（四）时间安排

数据生产垂直大模型项目计划分 3 年进行，具体实施周期及安排如下：

第一年：购入办公场地并完成装修；购入服务器、交换机、电脑等硬件设备以及办公软件和设计软件。基于公司对于行业的分析及预测，进行人员引进与培训，研发项目开始运行。

第二年和第三年：继续根据项目研发需要，购入服务器、交换机、电脑等硬件设备以及办公软件和设计软件，并进行人员引进与培训，按照设定的研发目标，持续进行项目研发。

序号	时间安排	Y1				Y2				Y3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	场地购置及装修												

序号	时间安排	Y1				Y2				Y3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
2	设备购置												
3	硬件软件平台搭建												
4	人员引进与培训												
5	项目研发												

（五）目前研发投入及进展、预计未来研发费用资本化情况、已取得及预计取得的研发成果

数据生产垂直大模型研发项目研发投入中包括场地购置及装修、设备购置等资本性支出，其余研发投入均直接计入费用化支出，本次募集资金研发投入预计不存在资本化的情况。截至本募集说明书签署日，公司已着手开展前期研究，项目尚未正式投入建设。

第五节 董事会关于本次发行对公司影响的讨论与分析

一、本次发行完成后，上市公司的业务及资产的变动或整合计划

本次向特定对象发行股票募集资金投资项目扣除相关发行费用后将用于 AI 大模型训练数据集建设项目及数据生产垂直大模型研发项目，符合公司的业务发展方向和战略布局。其中“AI 大模型训练数据集建设项目”通过建设应用于通用和特定垂直领域的 AI 大模型训练数据集提升行业内面向大模型训练数据集的类别和质量，“数据生产垂直大模型研发项目”以研发海天瑞声数据生产垂直大模型为核心升级公司一体化技术支撑平台。本次募集资金投资项目是公司在现有主营业务基础上，结合市场需求和未来发展趋势，加大对公司核心主业重点产品及重要研究方向投资力度的体现，项目实施可以巩固和发展公司在行业中的竞争优势。本次发行完成后，公司的主营业务范围不会发生重大变化，不存在因本次发行而导致的业务及资产整合。

二、本次发行完成后，上市公司控制权结构的变化情况

本次发行前，公司控股股东及实际控制人为贺琳，贺琳直接持有公司 12,137,615 股股份（占总股本的 20.12%），并通过控制中瑞安间接控制公司 6,935,779 股股份（占总股本的 11.50%），合计控制公司 31.62%的股份，为公司实际控制人。

若假设本次发行股票数量为发行上限 12,065,036 股，则本次发行完成后（仅考虑本次发行导致的公司股份数量变化），公司的总股本为 72,390,216 股，贺琳、中瑞安分别占公司股本的 16.77%、9.58%，贺琳合计控制公司 26.35%的股份，仍为公司的控股股东、实际控制人。因此，本次发行不会导致公司的控制权发生变化。

三、本次发行完成后，上市公司新增同业竞争情况

截至本募集说明书签署日，本次向特定对象发行股票尚未确定发行对象，公司是否与发行对象及发行对象的控股股东、实际控制人从事的业务存在同业竞争或潜在同业竞争的情况，将在发行结束后公告的《发行情况报告书》中予以披露。

四、本次发行完成后，上市公司新增关联交易情况

截至本募集说明书签署日，本次向特定对象发行股票尚未确定发行对象，公司是否与发行对象及发行对象的控股股东、实际控制人存在关联交易的情况，将在发行结束后公告的《发行情况报告书》中予以披露。

五、本次发行完成后，上市公司科研创新能力的变化

人工智能产业已进入以大模型为代表的新发展时期，通过本次募投项目的实施，公司将建设一批适用性较强的大模型训练数据集，拓展潜在高增长价值的新型业务板块，并藉此进一步扩大公司业务规模；同时，以研发海天瑞声数据生产垂直大模型为核心，升级海天瑞声一体化技术支撑平台，研发并引入多项新兴技术，促进公司科研创新能力不断提升、强化公司科创属性、巩固公司核心技术壁垒，构建长期技术实力支撑，从而进一步增强公司核心竞争力。

第六节 最近五年内募集资金运用的基本情况

最近五年，公司仅通过首次公开发行股票公开募集资金，募集资金运用的基本情况如下：

一、前次募集资金的数额、资金到账时间以及资金在专项账户中的存放情况

（一）前次募集资金的数额、资金到账情况

经中国证券监督管理委员会于 2021 年 7 月 13 日出具《关于同意北京海天瑞声科技股份有限公司首次公开发行股票注册的批复》（证监许可〔2021〕2366 号），同意公司首次公开发行股票注册的申请，并经上海证券交易所同意，公司首次向社会公众公开发行人民币普通股 1,070 万股，发行价格为 36.94 元/股，募集资金总额为人民币 395,258,000.00 元，扣除各项发行费用人民币 58,901,108.97 元后，实际募集资金净额为人民币 336,356,891.03 元。上述募集资金已于 2021 年 8 月 9 日全部到账并经上会会计师事务所（特殊普通合伙）审验出具“上会师报字（2021）第 8543 号”《验资报告》。

（二）前次募集资金在专项账户中的存放情况

截至 2023 年 3 月 31 日，公司前次募集资金在银行账户的存放情况如下：

单位：人民币元

开户银行	银行账号	存款方式	初始存放金额	2023 年 3 月 31 日存放金额
招商银行股份有限公司北京清华园科技金融支行	110919851510806	活期	249,211,800.00	14,978,572.34
中信银行股份有限公司北京广安门支行	8110701012302133979	活期	106,818,250.00	9,151.52
合计			356,030,050.00	14,987,723.86

注 1：招商银行股份有限公司北京清华园支行于 2022 年 12 月 28 日更名为招商银行股份有限公司北京清华园科技金融支行。

注 2：初始存放金额与实际募集资金净额差异为尚未支付的发行费 10,718,901.33 元和置换预先支付的发行费 8,954,257.64 元，共计 19,673,158.97 元。

二、前次募集资金实际使用情况说明

（一）前次募集资金使用情况对照表

截至 2023 年 3 月 31 日，前次募集资金实际使用情况对照情况如下：

单位：人民币元

募集资金总额：			395,258,000.00			已累计使用募集资金总额：			291,036,590.99	
						各年度使用募集资金总额：				
						2021 年：			124,973,479.43	
变更用途的募集资金总额：			0.00			2022 年：			155,160,463.88	
变更用途的募集资金比例：			0.00%			2023 年 1 月 1 日-3 月 31 日：			10,902,647.68	
投资项目			募集资金投资总额			截止日募集资金累计投资额			实际投资金额 与募集后承诺 投资金额的 差额	项目达到预定可 使用状态日期（或 截止日项目完工 程度）
序 号	承诺投资 项目	实际投资 项目	募集前承诺投资 金额	募集后承诺投 资金额	实际投资 金额	募集前承诺投 资金额	募集后承诺投资 金额	实际投资金额		
1	自主研发数据产品扩建项目	自主研发数据产品扩建项目	249,211,791.03	249,211,791.03	203,891,490.99	249,211,791.03	249,211,791.03	203,891,490.99	45,320,300.04	2023 年 5 月 31 日
2	一体化数据处理技术支撑平台升级项目	一体化数据处理技术支撑平台升级项目	250,577,200.00	0.00（注释 1）	不适用	250,577,200.00	0.00	不适用	不适用	不适用
3	研发中心升级建设项目	研发中心升级建设项目	166,149,800.00	0.00（注释 1）	不适用	166,149,800.00	0.00	不适用	不适用	不适用
4	补充流动资金	补充流动资金	100,000,000.00	87,145,100.00	87,145,100.00	100,000,000.00	87,145,100.00	87,145,100.00	0.00	不适用
合计			765,938,791.03	336,356,891.03	291,036,590.99	765,938,791.03	336,356,891.03	291,036,590.99	45,320,300.04	

注释 1：公司于 2021 年 9 月 7 日召开董事会、监事会会议审议通过《关于调整部分募集资金投资项目资金来源的议案》、《关于调整募集资金投资项目拟投入金额及内部投资结构的议案》。议案决定使用募集资金进行自主研发数据产品扩建项目，并补充流动资金，使用自有资金对“一体化数据处理技术支撑平台升级项目”、“研发中心升级建设项目”继续进行投资建设，具体内容详见公司于 2021 年 9 月 9 日披露于上海证券交易所网站的《海天瑞声关于调整部分募集资金投资项目资金来源的公告》（公告编号：2021-002）、《海天瑞声关于调整募集资金投资项目拟投入金额及内部投资结构的公告》（公告编号：2021-003）。

（二）前次募集资金实际投资项目调整情况说明

公司于 2021 年 9 月 7 日召开第一届董事会第三十四次会议及第一届监事会第二十八次会议，审议通过了《关于调整募集资金投资项目拟投入金额及内部投资结构的议案》，同意公司根据实际情况，对部分募集资金投资项目拟投入金额及内部投资结构进行调整。由于前次发行募集资金净额低于拟投入的募集资金金额，为进一步提高募集资金使用效率、保障募集资金投资项目的顺利实施，结合各募集资金投资项目的实际情况，公司拟对募集资金投资项目拟投入金额进行相应调整，使用募集资金进行“自主研发数据产品扩建项目”的投资建设，并补充流动资金，使用自有资金对“一体化数据处理技术支撑平台升级项目”、“研发中心升级建设项目”继续进行投资建设。独立董事、监事会和保荐机构对该事项发表了明确的同意意见。

（三）前次募集资金投资项目的实际投资总额与承诺投资总额的差异及原因说明

截至 2023 年 3 月 31 日，前次募集资金投资项目尚在建设中，实际投资总额与承诺投资总额的差异情况参见本节“二、前次募集资金实际使用情况说明”之“（一）前次募集资金使用情况对照表”。

（四）前次募集资金投资项目已对外转让或置换情况说明

2021 年 11 月 29 日，公司召开第二届董事会第三次会议、第二届监事会第三次会议，审议通过了《关于使用募集资金置换预先投入募投项目及已支付发行费用的自筹资金的议案》，同意使用募集资金置换预先投入募投项目及已支付部分发行费用的自筹资金，置换总金额为 3,447.45 万元。公司监事会、独立董事对本事项发表了明确的同意意见，保荐机构华泰联合证券有限责任公司对本事项出具了明确的核查意见，信永中和会计师事务所（特殊普通合伙）出具了鉴证报告。2021 年 12 月 14 日，公司完成上述募集资金置换事项。

截至 2023 年 3 月 31 日止，公司不存在前次募集资金投资项目对外转让的情况。

（五）前次募集资金变更或延期情况

公司于 2022 年 4 月 28 日分别召开第二届董事会第七次会议、第二届监事会第六次会议，审议通过了《关于公司部分募投项目延期的议案》，同意公司对部分募投项目达到预定可使用状态的时间进行延长，调整后达到可使用状态日期为 2023 年 5 月 31 日。本次募投项目的延期主要是受到募集资金到账时间晚于初始预期、外部环境等因素的影响，导致项目整体进度较原有计划有所延迟。公司独立董事对本事项发表了明确同意的独立意见，保荐机构华泰联合证券有限责任公司对本事项出具了明确的核查意见。

截至 2023 年 3 月 31 日止，公司前次募投资金实际投资项目未发生变更。

（六）闲置前次募集资金进行现金管理及投资相关产品情况

2021 年 9 月 7 日，公司召开第一届董事会第三十四次会议、第一届监事会第二十八次会议，审议通过了《关于使用部分暂时闲置募集资金进行现金管理的议案》，同意公司使用额度不超过人民币 23,000.00 万元的暂时闲置募集资金在确保不影响募集资金投资项目实施、确保募集资金安全的前提下用于现金管理，自董事会审议通过之日起 12 个月内有效，在上述额度和期限范围内，资金可以滚动使用。公司独立董事、监事会对本事项发表了明确的同意意见，保荐机构华泰联合证券有限责任公司对本事项出具了明确的核查意见。

2022 年 8 月 9 日，公司召开了第二届董事会第九次会议、第二届监事会第八次会议，审议通过了《关于使用部分暂时闲置募集资金进行现金管理的议案》，同意公司使用额度不超过人民币 18,000.00 万元（含本数）的暂时闲置募集资金在确保不影响募集资金投资项目实施、确保募集资金安全的前提下用于现金管理，自董事会审议通过之日起 12 个月内有效，在上述额度和期限范围内，资金可以滚动使用。公司独立董事、监事会对本事项发表了明确的同意意见，保荐机构华泰联合证券有限责任公司对本事项出具了明确的核查意见。

截至 2023 年 3 月 31 日止，公司使用暂时闲置募集资金进行现金管理余额 3,500 万元，具体情况详见下表：

单位：万元

签约方	产品名称	收益类型	金额	起息日	到期日	年化收益率	是否到期	收益
招商银行股份有限公司	招商银行点金系列看涨两层区间90天结构性存款	保本浮动收益	3,500.00	2023-1-13	2023-4-13	1.65%-2.85%	否	不适用

（七）未使用完毕募集资金的情况

公司前次募集资金净额 33,635.69 万元，截至 2023 年 3 月 31 日，公司尚未使用前次募集资金余额为 4,998.77 万元（包括募集资金专用账户余额 1,498.77 万元和购买理财产品定期存款余额 3,500.00 万元），占前次募集资金净额的 14.86%。公司按照募集资金投资进度保障剩余募集资金的合理使用。

三、前次募集资金使用的其他情况

（一）使用自有资金方式支付募投项目所需资金并以募集资金等额置换

2021 年 9 月 7 日召开的第一届董事会第三十四次会议及第一届监事会第二十八次会议审议通过了《关于使用自有资金方式支付募投项目所需资金并以募集资金等额置换的议案》。根据中国人民银行《人民币银行结算账户管理办法》规定，为提高运营管理效率，在不影响募投项目正常进行的前提下，同意公司在募集资金投资项目实施期间，使用自有资金方式支付募投项目的技术人员工资，后续以募集资金等额置换，定期从募集资金专户划转等额资金至公司基本存款账户。

截至 2023 年 3 月 31 日，募集资金到账后因前述技术人员工资支付事项而发生的募集资金置换以自有资金支付募投项目所需资金共计 18,059,517.94 元。

（二）购买房产暨增加募投项目实施地点

2022 年 9 月 27 日，公司召开第二届董事会第十次会议、第二届监事会第九次会议；2022 年 10 月 14 日，公司召开 2022 年第三次临时股东大会，分别审议通过了《关于拟购买房产暨增加募投项目实施地点的议案》，为满足公司的战略发展及实际生产经营需要，购买位于北京市东城区新中街 68 号 8A 楼的房产用于募投项目“自主研发数据产品扩建项目”、“一体化数据处理技术支撑平台升级项目”、“研发中心升级建设项目”的实施和公司办公使用，房产建筑面积约

6,200 平方米，交易总价不超过 24,150 万元人民币，其中使用募集资金 10,674 万元，剩余价款已用自有资金补足。

（三）报告期后前次募集资金投资项目结项情况

公司于 2023 年 6 月 9 日召开第二届董事会第十六次会议及第二届监事会第十五次会议，审议通过了《关于首次公开发行股票募集资金投资项目结项并将节余募集资金永久补充流动资金的议案》，鉴于首次公开发行股票募集资金投资项目“自主研发数据产品扩建项目”已完成建设并达到预定可使用状态，同意公司将该项目结项并将节余募集资金 3,515.73 万元（截至 2023 年 5 月 31 日，含扣除手续费后的利息收入及理财收益，实际金额以资金转出当日专户余额为准）用于永久补充公司流动资金。公司独立董事对本次事项发表了明确同意的意见，保荐机构对该事项出具明确同意的核查意见。

四、前次募集资金投资项目实现效益情况

公司在首次公开发行股票并在科创板上市招股说明书中未对募集资金的使用效益做出任何承诺。因此，前次募集资金投资项目实现效益情况不适用。

五、前次募集资金中用于认购股份的资产运行情况说明

截至 2023 年 3 月 31 日，公司前次募集资金不存在以资产认购股份的情况。

六、前次募集资金实际使用情况的信息披露对照情况

截至 2023 年 3 月 31 日，上述前次募集资金实际使用情况与公司在定期报告和其他信息披露文件中披露的有关内容不存在差异。

七、前次募集资金使用对发行人科技创新的作用

前次募集资金投资项目与公司现有主营业务密切相关，具体包括自主研发数据产品扩建项目和补充流动资金。自主研发数据产品扩建项目将进一步拓展公司服务于深度学习算法训练方向上的自有知识产权训练数据产品的种类和领域覆盖，储备覆盖更大规模、更多类型、更多环境、更多情感、更多场景的训练数据产品，加快对客户需求的响应速度。目前本项目已实现新增各场景自有知识产权数据集产品。通过本项目的实施，完善了公司的产品体系，得以快速、优质地响应客户需求，进一步提升公司持续经营能力并增强企业竞争力。补充流动资金项

目增强了公司的资金实力，随着公司业务规模的持续扩张，公司需补充流动资金，以优化公司资本结构、增强公司资金实力、提升公司的抗风险实力，为公司业务的稳定增长提供财务保障。

综上，前次募集资金投资项目是公司在主营业务的基础上，按照公司未来发展的战略规划，对公司现有业务的深化和拓展。通过前次募投项目的实施，公司完善了训练数据产品体系，提升了公司的生产服务效率和技术水平，增加了公司的前瞻性技术储备，助力公司在训练数据开发及服务领域实现技术创新、提高核心竞争力。

八、会计师对于前次募集资金使用情况的结论性意见

信永中和会计师事务所对公司前次募集资金使用情况进行了专项鉴证，并出具了《北京海天瑞声科技股份有限公司截至 2023 年 3 月 31 日前次募集资金使用情况鉴证报告》（XYZH/2023BJAS2F0439）。信永中和会计师事务所的鉴证结论如下：“我们认为，海天瑞声上述前次募集资金使用情况报告已经按照中国证券监督管理委员会、上海证券交易所相关规定编制，在所有重大方面如实反映了海天瑞声截至 2023 年 3 月 31 日前次募集资金的使用情况。”

第七节 与本次发行相关的风险因素

投资者在评价本次向特定对象发行股票时，除本募集说明书提供的其它各项资料外，应特别注意以下风险：

一、业绩下滑或亏损风险

2022 年度，虽受境外业务复苏以及智能驾驶业务驱动，公司整体营收出现较高增长，但同期公司为更好建设长期技术优势和独具特色的标准化产品模式壁垒持续加大研发投入，且在 2022 年完成管理架构及激励体系升级、全球化营销体系建设等重要事项，以上因素共同导致公司人工投入及期间费用出现明显增长，使得公司净利润出现一定比例下滑。2023 年，受宏观经济增速放缓，市场竞争较为激烈以及数据出境相关法规实施所带来的影响，公司境外业务的开展受到影响，公司部分业务进度或将放缓。如果公司收入增长无法覆盖持续加大的人工投入及期间费用支出，公司业绩存在下滑或亏损的风险。

二、市场竞争加剧的风险

公司所在的细分领域为 AI 基础数据服务领域。近年来，参与该行业的公司不断增多，由于整体行业较新，行业标准和法律监管仍处于规范阶段，因此整体行业目前处于高速发展且格局未定的状态，导致市场竞争加剧。公司如果不能持续有效地制定并实施业务发展规划，则将受到竞争者的挑战，从而面临市场竞争加剧而导致的经营状况下滑、市场地位下降和可能失去主要客户的风险，进而影响公司的盈利能力和发展潜力。

三、核心技术快速迭代的风险

由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，公司进行数据开发所需的各项技术也面临着快速迭代更新的风险。如果公司无法保持对行业发展趋势的洞察能力以及对研发人才的有效组织和研发经费的经济投入，则可能导致公司在行业技术快速迭代过程中无法持续保持技术先进性和技术优势，进而对公司的客户拓展、产品创新和经营业绩造成不利影响。

四、核心技术人员流失风险

随着行业的持续发展，行业内企业之间对于高端人才的竞争日益激烈，如果公司无法持续加强核心技术人员培养及引进并为核心技术人员提供有竞争力的激励机制和薪资待遇，则将存在核心技术人员流失的风险，公司的技术水平、研发能力也将受到不利影响。

五、训练数据标准化产品无法实现授权销售的风险

公司根据对算法模型应用领域、行业发展趋势、市场需求等的评估和研判，自行设计并开发多种训练数据集标准化产品，开发完成后根据客户实际需求进行销售，授权给客户使用。在该种业务类型下，公司开发训练数据集产品，承担开发费用，并拥有训练数据集产品的知识产权。如果公司开发的训练数据集产品不能满足下游客户对训练数据的要求，则可能存在无法实现授权销售的风险，进而对公司未来的经营业绩产生不利影响。

六、应收账款回收的风险

报告期各期末，公司的应收账款账面价值分别为 6,278.29 万元、9,019.65 万元、10,178.03 万元和 8,904.07 万元，占各期末资产总额的比例分别为 13.15%、10.73%、11.61%和 10.52%。报告期内，公司应收账款周转率分别为 4.11、2.51、2.53 和 1.11（已年化），公司应收账款周转率有所下降，倘若客户因审批流程长或经营状况出现恶化等原因拖延付款，将会造成公司应收账款逐年增加，甚至出现坏账的情况。因此，公司存在应收账款回收的风险。

七、数据安全相关风险

公司主要从事训练数据的研发设计、生产及销售业务，所提供的产品和服务主要以数据的形式体现。一方面，随着公司业务的快速发展和规模的持续扩张，原料数据采集与数据加工的数量持续增长；另外一方面，包括《数据安全法》、《个人信息保护法》、《生成式人工智能服务管理暂行办法》等在内的与数据安全、个人信息保护相关的法律规章体系正逐步完善。公司如果未能按照法律规章或客户的更新要求及时调整现行业务开展方式、公司的数据安全管理体系研发升级未能跟上业务发展的需要、或客户未能遵守训练数据产品保护相关商业约定，则公

司可能产生诉讼纠纷或面临生产经营不符合法律规章的要求、训练数据产品被泄露、盗版等数据安全相关风险。

同时,《数据出境安全评估办法》等法律法规加强了对境内收集或产生的数据的跨境行为的规制。如果将来公司未能根据《数据出境安全评估办法》的要求及时履行相应的安全评估申请程序,公司开展的属于该办法限定范围内的境外业务将受到一定程度的影响。

八、宏观经济波动的风险

公司所处的 AI 数据服务行业受下游人工智能行业的波动性影响较大,与国民经济众多领域息息相关。当期全球宏观经济状况波动,部分国家、地区出现经济增速放缓、经济复苏乏力的局面,一定程度影响了公司下游客户在相关领域的研发和投入。若未来经济大环境持续疲软、各类客户群体在人工智能领域的投资持续低迷,将对公司业务发展产生较大影响。

九、募集资金投资项目相关风险

(一) 募投项目实施后效益不及预期的风险

公司本次募集资金将用于 AI 大模型训练数据集建设项目,该项目实施及未来商业化应用前景存在一定的不确定性,且公司对募投项目的效益测算系基于历史数据和未来公司及行业的发展趋势所做出的,但在募投项目的实施过程中,公司将可能面临着宏观经济及市场环境变化、下游需求放缓、市场竞争加剧、技术迭代更新等不确定或不可控因素的影响,募投项目有可能存在技术实施难度大、市场推广难、产品收益不及预期等情况,本次募集资金投资项目存在不能完全实现预期目标或效益的风险。

(二) 募投项目研发失败或研发成果无法转化的风险

由于大模型时代下存在数据规模量极大、复杂性和多样性高,数据服务规则设计难度指数级提升等诸多问题,相关技术迭代速度快,未来发展趋势存在不确定性。若公司未能及时准确地把握行业发展趋势、突破技术难关,公司将面临研发失败或研发成果无法转化、前期的研发投入将难以收回的风险,并对公司的盈利能力和市场竞争力造成不利影响。

（三）募投项目新增折旧、摊销的风险

本次募投项目实施后公司的固定资产规模将增长较大，每年也将随之新增较多的固定资产折旧费用。如果市场情况发生重大不利变化，募投项目无法如期推进或不能产生预期收益，则公司可能存在因固定资产折旧、摊销费用大幅增加而导致利润下滑的风险。

（四）即期回报被摊薄的风险

本次向特定对象发行股票完成后，公司总股本和净资产将有所增加，而募集资金的使用和实施需要一定的时间。公司每股收益和加权平均净资产收益率等财务指标在短期内可能会出现一定幅度的下降。因此，本次向特定对象发行股票募集资金到位后，公司的即期回报存在被摊薄的风险。

（五）部分拟投入设备依赖境外采购的风险

本次募投项目所需的 GPU 服务器等设备主要向境外供应商采购，且境外供应商在相关领域占据主导位置。由于国际政治局势、全球贸易摩擦及其他不可抗力因素的影响，公司本次募投项目拟投入设备采购可能会出现延迟交货、限制供应或价格提高的情况，可能对本次募投项目的实施带来不利影响。

十、与本次发行有关的风险

（一）审核及发行风险

本次向特定对象发行股票方案尚需上交所审核通过并经中国证监会作出同意注册决定后方可实施。该等审批事项的结果及最终取得批准时间均存在不确定性。同时，本次向特定对象发行股票的发行对象为不超过三十五名（含三十五名）特定投资者。投资者的认购意向以及认购能力受到证券市场整体情况、公司股票价格走势、投资者对本次发行方案的认可程度以及市场资金面情况等多种内、外部因素的影响，可能面临募集资金不足乃至发行失败的风险。

（二）股票价格波动风险

股票价格不仅取决于公司的经营状况，同时也受国家的经济政策、经济周期、股票市场的供求状况、重大自然灾害的发生、投资者心理预期等多种因素的影响。因此，公司的股票价格存在若干不确定性，并可能因上述风险因素出现波动，直

接或间接地给投资者带来投资收益的不确定性。

第八节 与本次发行相关的声明

一、发行人及全体董事、监事、高级管理人员声明

本公司及全体董事、监事、高级管理人员承诺本募集说明书内容真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，按照诚信原则履行承诺，并承担相应的法律责任。

董事：



贺琳



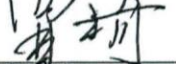
李科



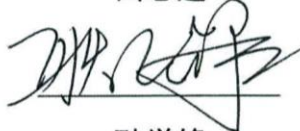
吕思遥



黄宇凯



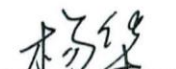
贾琦



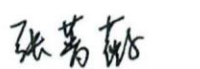
耿学锋



秦勇

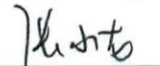


杨华

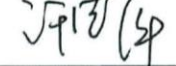


张蔷薇

监事：



张小龙



许洁萍



李娴

除董事、监事外的
高级管理人员：



王晓东



郝玉峰

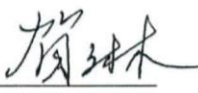
北京海天瑞声科技股份有限公司



二、发行人控股股东、实际控制人声明

本公司或本人承诺本募集说明书内容真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，按照诚信原则履行承诺，并承担相应的法律责任。

控股股东、实际控制人：


贺琳

2023年7月27日

三、保荐人声明

本公司已对募集说明书进行了核查，确认本募集说明书内容真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，并承担相应的法律责任。

项目协办人： 郭晓航
郭晓航

保荐代表人： 杨阳
杨阳

张鹏
张鹏

法定代表人（或授权代表）： 江禹
江禹



华泰联合证券有限责任公司

2023年7月27日

本人已认真阅读北京海天瑞声科技股份有限公司募集说明书的全部内容，确认募集说明书不存在虚假记载、误导性陈述或者重大遗漏，并对募集说明书真实性、准确性、完整性、及时性承担相应法律责任。

保荐人总经理：



马 骁

保荐人董事长（或授权代表）：



江 禹



华泰联合证券有限责任公司

2023年7月27日

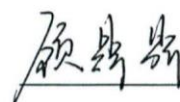
四、发行人律师声明

本所及经办律师已阅读募集说明书，确认募集说明书内容与本所出具的法律意见书不存在矛盾。本所及经办律师对发行人在募集说明书中引用的法律意见书的内容无异议，确认募集说明书不因引用上述内容而出现虚假记载、误导性陈述或重大遗漏，并承担相应的法律责任。

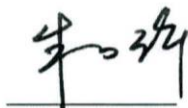
经办律师：


周世君


王韶华


顾鼎鼎

律师事务所
负责人：


朱小辉

北京市天元律师事务所

2023年7月27日



五、会计师事务所声明

本所及签字注册会计师已阅读《北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票并在科创板上市募集说明书》（以下简称“募集说明书”），确认募集说明书内容与本所出具 2021 年度、2022 年度的审计报告（报告号：XYZH/2022BJAA160966、XYZH/2023BJAS2B0234）等文件不存在矛盾。本所及签字注册会计师对发行人在募集说明书中引用的审计报告等文件的内容无异议，确认募集说明书不因引用上述内容而出现虚假记载、误导性陈述或重大遗漏，并承担相应的法律责任。

签字注册会计师：


梅秀琴


马凯飞


郭勇


左东强


庞博

会计师事务所负责人：


谭小青

信永中和会计师事务所（特殊普通合伙）



六、董事会声明

（一）未来十二个月内的其他融资计划

除本次发行外，公司在未来十二个月内暂无其他股权融资计划。若未来公司根据业务发展需要及资产负债状况安排股权融资，将按照相关法律法规履行相关审议程序和信息披露义务。

（二）关于本次向特定对象发行股票摊薄即期回报拟采取的填补措施及相关承诺

1、公司应对本次向特定对象发行摊薄即期回报采取的措施

为保护广大投资者的合法权益，降低本次向特定对象发行可能摊薄即期回报的影响，公司拟采取多种措施保证本次向特定对象发行募集资金有效使用、有效防范即期回报被摊薄的风险。公司填补即期回报的具体措施如下：

（1）强化募集资金管理，提高募集资金使用效率

根据《上市公司监管指引第2号——上市公司募集资金管理和使用的监管要求》《科创板上市规则》等法律法规的要求，结合公司实际情况，公司已制定《募集资金使用管理办法》，明确了公司对募集资金专户存储、使用、用途变更、管理和监督的规定。募集资金将存放于公司董事会决定的专项账户集中管理，做到专款专用，以保证募集资金合理规范使用。

（2）积极落实募集资金投资项目，助力公司业务发展

本次募集资金投资项目的实施，将推动公司业务发展，加强公司主营业务的研发能力及生产能力，进一步提高公司市场竞争力，为公司的战略发展带来积极影响，发行募集资金到位后，公司将积极推进募集资金投资项目，从而降低本次发行对股东即期回报摊薄的风险。

（3）不断完善公司治理，加强经营管理和内部控制

公司将严格遵循《公司法》《证券法》《科创板上市规则》等法律、法规和规范性文件的要求，不断完善公司治理结构，建立健全公司内部控制制度，促进公司规范运作并不断提高质量，维护公司整体利益，尤其是中小股东的合法权益。同时，公司将努力提高资金的使用效率，合理运用各种融资工具和渠道，控制资

金成本，提升资金使用效率，节省公司的各项费用支出，全面有效地控制经营和管控风险，保障公司持续、稳定、健康发展。

（4）严格执行利润分配政策，强化投资者回报机制

为完善公司利润分配政策，更好地维护全体股东及投资者利益，公司按照相关法律法规的要求并结合公司的实际情况，在《公司章程》中规定了有关利润分配的相关制度条款，明确了公司现金分红的具体条件和比例、股票股利发放条件以及公司利润分配的决策程序和机制。为了进一步落实利润分配政策，公司还制定了《未来三年（2023 年-2025 年）股东分红回报规划》，持续优化投资者回报机制。公司未来将严格执行分红政策，在符合利润分配条件的情况下，积极推动对股东的利润分配，切实维护投资者合法权益，努力提升股东回报水平。

（5）完善员工激励机制，加强人才队伍建设

公司将完善薪酬和激励机制，建立有市场竞争力的薪酬体系，引进市场优秀人才，建立与公司发展相匹配的人才结构，最大限度地激发员工积极性，挖掘员工的创造力和潜在动力，为公司的可持续发展提供可靠的人才保障。

（6）持续完善公司治理，为公司发展提供制度保障

公司已建立、健全了法人治理结构，具有完善的股东大会、董事会、监事会和管理层的独立运行机制，设置了与公司生产经营相适应的组织职能机构，并制定了相应的岗位职责，各职能部门之间职责明确，相互制约。公司组织机构设置合理、运行有效，股东大会、董事会、监事会和管理层之间权责分明、相互制衡、运行良好，形成了一套合理、完整、有效的公司治理与经营管理框架。

公司将继续严格遵循《公司法》《证券法》以及《上市公司治理准则》等法律、法规和规范性文件的要求，不断完善公司治理结构，切实保护投资者尤其是中小投资者权益，为公司发展提供制度保障。

公司提示投资者，上述填补回报措施不等于对公司未来利润作出保证。

2、相关主体对公司本次向特定对象发行摊薄即期回报采取填补措施出具的承诺

(1) 公司董事、高级管理人员对公司填补回报措施能够得到切实履行做出的承诺

“1、不无偿或以不公平条件向其他单位或者个人输送利益，也不采用其他方式损害公司利益。

2、对本人的职务消费行为进行约束。

3、不动用公司资产从事与本人履行职责无关的投资、消费活动。

4、由董事会或薪酬与考核委员会制定的薪酬制度与公司填补回报措施的执行情况相挂钩。

5、若公司后续推出股权激励方案，则未来股权激励方案的行权条件与公司填补回报措施的执行情况相挂钩。

6、自本承诺出具日至公司本次向特定对象发行 A 股股票实施完毕前，若中国证券监督管理委员会、上海证券交易所等证券监管机构作出关于填补回报措施及其承诺的其他新的监管规定，且上述承诺相关内容不能满足中国证券监督管理委员会、上海证券交易所等证券监管机构的该等规定时，本人承诺届时将按照中国证券监督管理委员会、上海证券交易所等证券监管机构的最新规定出具补充承诺。

7、本人承诺切实履行公司制定的有关填补回报措施以及对此作出的任何有关填补回报措施的承诺，若违反该等承诺并给公司或者投资者造成损失的，本人愿意依法承担对公司或者投资者的补偿责任。

8、作为填补回报措施相关责任主体之一，若违反上述承诺或拒不履行上述承诺，本人同意按照中国证券监督管理委员会、上海证券交易所等证券监管机构制定或发布的有关规定、规则，对本人作出相关处罚或采取相关管理措施。”

(2) 公司控股股东、实控人、董事长贺琳对公司填补回报措施能够得到切实履行作出的承诺

“1、不越权干预公司经营管理活动，不侵占公司利益，切实履行公司填补

即期回报的相关措施。

2、不无偿或以不公平条件向其他单位或者个人输送利益，也不采用其他方式损害公司利益。

3、对本人的职务消费行为进行约束。

4、不动用公司资产从事与其履行职责无关的投资、消费活动。

5、由董事会或薪酬委员会制定的薪酬制度与公司填补被摊薄即期回报措施的执行情况相挂钩。

6、若公司未来推出股权激励计划，承诺拟公布的公司股权激励计划的行权条件与公司填补被摊薄即期回报措施的执行情况相挂钩。

7、自本承诺出具日至公司本次向特定对象发行 A 股股票实施完毕前，若中国证券监督管理委员会、上海证券交易所等证券监管机构作出关于填补回报措施及其承诺的其他新的监管规定，且上述承诺相关内容不能满足中国证券监督管理委员会、上海证券交易所等证券监管机构的该等规定时，本人承诺届时将按照中国证券监督管理委员会、上海证券交易所等证券监管机构的最新规定出具补充承诺。

8、本人承诺切实履行公司制定的有关填补回报措施以及对此作出的任何有关填补回报措施的承诺，若违反该等承诺并给公司或者投资者造成损失的，本人愿意依法承担对公司或者投资者的补偿责任。

9、作为填补回报措施相关责任主体之一，若违反上述承诺或拒不履行上述承诺，本人同意按照中国证券监督管理委员会、上海证券交易所等证券监管机构制定或发布的有关规定、规则，对本人作出相关处罚或采取相关管理措施。”

北京海天瑞声科技股份有限公司董事会



附 件

附件一：发行人及其控股子公司拥有的商标

（一）中国境内商标

序号	商标	注册人	注册号	注册有效期限	类别	他项权利
1		发行人	4661751	2019 年 1 月 28 日至 2029 年 1 月 27 日	35	无
2		发行人	4661752	2018 年 12 月 28 日至 2028 年 12 月 27 日	42	无
3	speech x cean	发行人	8563300	2011 年 08 月 14 日至 2031 年 08 月 13 日	42	无
4	海天瑞声	发行人	8563299	2011 年 08 月 14 日至 2031 年 08 月 13 日	42	无
5	speech x cean	发行人	37724912	2019 年 12 月 14 日至 2029 年 12 月 13 日	9	无
6	speech x cean	发行人	37733773	2019 年 12 月 14 日至 2029 年 12 月 13 日	35	无
7	speech x cean	发行人	37711082	2019 年 12 月 14 日至 2029 年 12 月 13 日	38	无
8	speech x cean	发行人	37711106	2019 年 12 月 14 日至 2029 年 12 月 13 日	41	无
9	海天瑞声	发行人	37713178	2019 年 12 月 14 日至 2029 年 12 月 13 日	9	无
10	海天瑞声	发行人	37718092	2019 年 12 月 14 日至 2029 年 12 月 13 日	35	无
11	海天瑞声	发行人	37711077	2019 年 12 月 14 日至 2029 年 12 月 13 日	38	无
12	海天瑞声	发行人	37722765	2019 年 12 月 14 日至 2029 年 12 月 13 日	41	无
13	海天瑞声	发行人	37711715	2019 年 12 月 14 日至 2029 年 12 月 13 日	45	无
14	speech x cean	发行人	37720041	2019 年 12 月 21 日至 2029 年 12 月 20 日	45	无
15	锐听	中瑞智	36771818A	2019 年 11 月 21 日 至 2029 年 11 月 20 日	9、35、37、 38、41、 42	无
16	speech x cean	发行人	48514304	2021 年 3 月 14 日至 2031 年 3 月 13 日	9	无
17	speech x cean	发行人	48546256	2021 年 3 月 14 日至 2031 年 3 月 13 日	35	无
18	speech x cean	发行人	48533036	2021 年 3 月 14 日至 2031 年 3 月 13 日	38	无

序号	商标	注册人	注册号	注册有效期限	类别	他项权利
19	speechocean	发行人	48523324	2021 年 3 月 14 日至 2031 年 3 月 13 日	41	无
20	speechocean	发行人	48530058	2021 年 3 月 14 日至 2031 年 3 月 13 日	42	无
21	海天瑞声	发行人	48514307	2021 年 3 月 14 日至 2031 年 3 月 13 日	9	无
22	海天瑞声	发行人	48534118	2021 年 3 月 14 日至 2031 年 3 月 13 日	35	无
23	海天瑞声	发行人	48541063	2021 年 3 月 14 日至 2031 年 3 月 13 日	38	无
24	海天瑞声	发行人	48534178	2021 年 3 月 14 日至 2031 年 3 月 13 日	41	无
25	海天瑞声	发行人	48546346	2021 年 3 月 14 日至 2031 年 3 月 13 日	42	无
26	speechocean 海天瑞声	发行人	48534096	2021 年 3 月 14 日至 2031 年 3 月 13 日	9	无
27	speechocean 海天瑞声	发行人	48533021	2021 年 3 月 14 日至 2031 年 3 月 13 日	35	无
28	speechocean 海天瑞声	发行人	48534157	2021 年 3 月 14 日至 2031 年 3 月 13 日	38	无
29	speechocean 海天瑞声	发行人	48534180	2021 年 3 月 14 日至 2031 年 3 月 13 日	41	无
30	speechocean 海天瑞声	发行人	48546350	2021 年 3 月 14 日至 2031 年 3 月 13 日	42	无

(二) 中国境外商标

序号	商标	注册人	国别	注册号	注册有效期限	类别	取得方式	他项权利
1	speechocean	发行人	欧盟	018325007	2020 年 10 月 22 日至 2030 年 10 月 22 日	9、 35、 42	原始取得	无
2	speechocean	发行人	俄罗斯 印度 新加坡 以色列 英国 土耳其	1567635 (马德里商标)	2020 年 10 月 22 日至 2030 年 10 月 22 日	9、 35、 42	原始取得	无
3	speechocean	发行人	美国	6579768	2021 年 12 月 7 日至 2031 年 12 月 7 日	9、 35、 42	原始取得	无

序号	商标	注册人	国别	注册号	注册有效期限	类别	取得方式	他项权利
4	speechacean	发行人	日本	6534085	2022年3月25日至2032年3月25日	9、35、42	原始取得	无
5	speechacean	发行人	韩国	40-1837936	2022年2月25日至2032年2月25日	9、35、42	原始取得	无
6	speechacean	发行人	中国香港	305424796	2020年10月21日至2030年10月20日	9、35、42	原始取得	无

附件二：发行人及其控股子公司拥有的中国境内专利

序号	专利名称	专利权人	专利类型	专利号	申请日	授权公告日	取得方式	他项权利
1	语音对齐方法及装置	发行人	发明	201810449585.3	2018年5月11日	2020年6月23日	原始取得	无
2	语音检测方法和装置	发行人	发明	201810883930.4	2018年8月6日	2020年6月23日	原始取得	无
3	音素误标注的检测方法和装置	发行人	发明	201810827500.0	2018年7月25日	2020年1月31日	原始取得	无
4	韵律标注方法、装置和设备	发行人	发明	201810988973.9	2018年8月28日	2020年1月7日	原始取得	无
5	音频质量评估方法、装置、电子设备及存储介质	发行人	发明	201910073390.8	2019年1月25日	2019年10月29日	原始取得	无
6	中英混合语料的生成方法、装置、设备及存储介质	发行人	发明	201910022453.7	2019年1月10日	2020年3月24日	原始取得	无
7	语料选取处理方法、装置、设备及计算机可读存储介质	发行人	发明	201910072150.6	2019年1月25日	2020年1月14日	原始取得	无
8	语音库的语音获取方法、装置、电子设备及存储介质	发行人	发明	201911314504.X	2019年12月19日	2020年5月5日	原始取得	无
9	端到端的语音合成方法、装置及存储介质	发行人	发明	202010007974.8	2020年1月6日	2020年5月19日	原始取得	无
10	语音采集设备	发行人	实用新型	201922276799.8	2019年12月18日	2020年3月31日	原始取得	无
11	个性语料获取方法及个性语料获取装置	发行人	发明	201911314319.0	2019年12月19日	2020年5月5日	原始取得	无
12	短语语料获取方法及短语语料获取装置	发行人	发明	201911352915.8	2019年12月25日	2020年5月5日	原始取得	无
13	视频中音频聚类的处理方法和装置	发行人	发明	201911289077.4	2019年12月16日	2020年5月5日	原始取得	无
14	图像标注方法、图像标注装置及计算机存储介质	发行人	发明	202010096586.1	2020年2月18日	2020年8月4日	原始取得	无
15	视频抽选方法、装置及存	发行人	发明	202010100697.5	2020年2月	2020年6月	原始	无

序号	专利名称	专利权人	专利类型	专利号	申请日	授权公告日	取得方式	他项权利
	储介质				19 日	2 日	取得	
16	视频筛选方法、装置及存储介质	发行人	发明	202010096575.3	2020 年 2 月 18 日	2020 年 6 月 26 日	原始取得	无
17	图像标注方法、图像标注装置及计算机存储介质	发行人	发明	202010078873.X	2020 年 2 月 3 日	2023 年 1 月 3 日	原始取得	无
18	跨语言非标准词识别方法及装置	发行人	发明	202010122519.2	2020 年 2 月 27 日	2020 年 7 月 14 日	原始取得	无
19	语音数据库的质量评估方法、装置及计算机存储介质	发行人、清华大学	发明	202010164556.X	2020 年 3 月 11 日	2020 年 6 月 19 日	原始取得	无
20	语音转写方法、语音转写装置及计算机存储介质	发行人	发明	202010283135.9	2020 年 4 月 13 日	2020 年 7 月 28 日	原始取得	无
21	用于度量语音数据库覆盖性的无监督模型训练方法及装置	发行人、清华大学	发明	202010309303.7	2020 年 4 月 20 日	2020 年 7 月 28 日	原始取得	无
22	基于线性预测残差负熵的语音音质度量评价方法及装置	发行人、清华大学	发明	202010659644.7	2020 年 7 月 10 日	2020 年 10 月 16 日	原始取得	无
23	基于识别模型的识别方法、模型训练方法及装置	发行人、清华大学	发明	202010659647.0	2020 年 7 月 10 日	2020 年 11 月 10 日	原始取得	无
24	点云数据标注方法、点云数据标注装置及存储介质	发行人	发明	202110018470.0	2021 年 1 月 7 日	2021 年 4 月 20 日	原始取得	无
25	一种基于三元组的语音识别准确率计算方法	发行人	发明	202110026253.6	2021 年 1 月 8 日	2021 年 4 月 20 日	原始取得	无
26	在连续帧中识别物体运动状态的方法及装置	发行人	发明	202110609317.5	2021 年 6 月 1 日	2022 年 1 月 11 日	原始取得	无
27	在 3D 雷达点云连续帧数据中检测障碍物的方法及装置	发行人	发明	202110607326.0	2021 年 6 月 1 日	2022 年 2 月 18 日	原始取得	无
28	语音一致性检测方法、装置、设备及存储介质	发行人	发明	202210541377.2	2022 年 5 月 19 日	2022 年 8 月 30 日	原始取得	无
29	一种语料获取方法、装置、电子设备和存储介质	发行人	发明	202210598358.3	2022 年 5 月 30 日	2022 年 12 月 20 日	原始取得	无
30	关于激光雷达点云数据的路面检测方法、装置及介质	发行人	发明	202210671641.4	2022 年 6 月 15 日	2022 年 9 月 13 日	原始取得	无
31	文本语言识别方法、装置、设备及介质	发行人	发明	202211306400.6	2022 年 10 月 25 日	2023 年 1 月 10 日	原始取得	无
32	基于麦克风的音频处理方法和装置	中瑞智	发明	201710539306.8	2017 年 7 月 4 日	2020 年 3 月 31 日	原始取得	无
33	智能语音采集设备	中瑞智	外观设计	202030062563.X	2020 年 2 月 27 日	2020 年 8 月 18 日	原始取得	无

备注：发明专利权的有效期自申请日起 20 年，实用新型的有效期自申请日起 10 年，外观设计的有效期限自申请日起 10 年。

附件三：发行人及其控股子公司拥有的中国境内计算机软件著作权

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
1	软著登字第 BJ13266 号	语音自动切分系统 V1.0	2009SRBJ 2960	发行人	2009 年 3 月 31 日	2009 年 3 月 31 日	2009 年 5 月 12 日	原始取得	全部权利
2	软著登字第 BJ13373 号	外文词典信息提取软件 V1.0	2009SRBJ 3067	发行人	2009 年 4 月 13 日	2009 年 4 月 13 日	2009 年 5 月 13 日	原始取得	全部权利
3	软著登字第 BJ13369 号	递增式中文语料库抽选系统 V1.0	2009SRBJ 3063	发行人	2009 年 4 月 13 日	2009 年 4 月 13 日	2009 年 5 月 13 日	原始取得	全部权利
4	软著登字第 BJ13545 号	多通道录音软件 V1.0	2009SRBJ 3239	发行人	2009 年 4 月 15 日	2009 年 4 月 15 日	2009 年 5 月 31 日	原始取得	全部权利
5	软著登字第 BJ13537 号	音频文件及文本校对软件 V1.0	2009SRBJ 3231	发行人	2009 年 4 月 15 日	2009 年 4 月 15 日	2009 年 5 月 31 日	原始取得	全部权利
6	软著登字第 BJ13650 号	文本声音处理软件 V1.0	2009SRBJ 3344	发行人	2009 年 4 月 17 日	2009 年 4 月 17 日	2009 年 6 月 1 日	原始取得	全部权利
7	软著登字第 BJ27774 号	KDCRecorder WinMo6.1 手机录音系统[简称: KDCRecorder] V1.0	2010SRBJ 2391	发行人	2010 年 4 月 19 日	2010 年 4 月 20 日	2010 年 5 月 28 日	原始取得	全部权利
8	软著登字第 BJ27744 号	工时录入系统 V1.0	2010SRBJ 2361	发行人	2010 年 4 月 19 日	2010 年 4 月 20 日	2010 年 5 月 28 日	原始取得	全部权利
9	软著登字第 BJ27743 号	智能语音自动调幅软件[简称: 智能语音自动调幅]V1.0	2010SRBJ 2360	发行人	2010 年 4 月 19 日	2010 年 4 月 20 日	2010 年 5 月 28 日	原始取得	全部权利
10	软著登字第 0359222 号	ASIORecord 多通道录音软件 [简称: ASIORecord]V 3.4	2011SR09 5548	发行人	2011 年 10 月 31 日	2011 年 11 月 1 日	2011 年 12 月 15 日	原始取得	全部权利
11	软著登字第 0359304 号	Android 智能平台录音软件[简称: Android]V1.0	2011SR09 5630	发行人	2011 年 10 月 31 日	2011 年 11 月 1 日	2011 年 12 月 15 日	原始取得	全部权利
12	软著登字第 0360463 号	Symbian 平台语音采集软件 [简称: Symbian]V1.0	2011SR09 6789	发行人	2011 年 10 月 31 日	2011 年 11 月 1 日	2011 年 12 月 17 日	原始取得	全部权利
13	软著登字第 BJ37019 号	yTrans 语音转写标注软件[简称: yTrans]V1.0	2011SRBJ 4898	发行人	2011 年 10 月 31 日	2011 年 11 月 1 日	2011 年 12 月 28 日	原始取得	全部权利
14	软著登字第 0483542 号	合成语音评测软件 V1.0	2012SR11 5506	发行人	2012 年 9 月 10 日	2012 年 9 月 11 日	2012 年 11 月 28 日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
15	软著登字第0483700号	多通道同步视频采集器软件 V2.0	2012SR115664	发行人	2012年9月10日	2012年9月11日	2012年11月28日	原始取得	全部权利
16	软著登字第0483740号	ASIORecord 多通道录音软件 [简称: ASIORecord]V3.6	2012SR115704	发行人	2012年9月10日	2012年9月11日	2012年11月28日	原始取得	全部权利
17	软著登字第0483743号	IVR 录音平台软件 [简称: IVR] V1.0	2012SR115707	发行人	2012年9月10日	2012年9月11日	2012年11月28日	原始取得	全部权利
18	软著登字第0484348号	语料管理平台软件 V1.0	2012SR116312	发行人	2012年9月10日	2012年9月11日	2012年11月29日	原始取得	全部权利
19	软著登字第0488003号	语料提取软件 V1.0	2012SR119967	发行人	2012年9月10日	2012年9月11日	2012年12月6日	原始取得	全部权利
20	软著登字第0554117号	视频搜索与定位 GFFGUI 软件 [简称: GFFGUI]V1.0	2013SR048355	发行人	2013年3月11日	2013年3月12日	2013年5月22日	原始取得	全部权利
21	软著登字第0554348号	WindowsMobile 平台 KDCRecorder 录音工具软件 [简称: WindowsMobile KDCRecorder] V1.0	2013SR048586	发行人	2013年3月11日	2013年3月12日	2013年5月22日	原始取得	全部权利
22	软著登字第0553938号	WindowsPhone 平台 AudioRecord 录音工具软件[简称: AudioRecord]V1.0	2013SR048176	发行人	2013年3月11日	2013年3月12日	2013年5月22日	原始取得	全部权利
23	软著登字第0554449号	IOS 平台 iPAudio 录音软件 [简称: iPAudio]V1.0	2013SR048687	发行人	2013年3月11日	2013年3月12日	2013年5月22日	原始取得	全部权利
24	软著登字第0554352号	Windows8RT 平台 rtAudioRecord 录音工具软件 [简称: rtAudioRecord] V1.0	2013SR048590	发行人	2013年3月11日	2013年3月12日	2013年5月22日	原始取得	全部权利
25	软著登字第0557501号	Android 智能平台录音软件 V2.0	2013SR051739	发行人	2013年3月11日	2013年3月12日	2013年5月29日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
26	软著登字第0737811号	语音识别(ASR)数据后期处理系统[简称: ASR]V1.0	2014SR068567	发行人	2014年3月10日	2014年3月12日	2014年5月29日	原始取得	全部权利
27	软著登字第0738137号	uTrans 在线语音转写标注质检平台系统[简称: uTrans]V1.0	2014SR068893	发行人	2014年3月21日	2014年3月24日	2014年5月29日	原始取得	全部权利
28	软著登字第0738138号	SpeechOcean 语料管理系统[简称:语料管理系统]V1.0	2014SR068894	发行人	2014年3月20日	2014年3月21日	2014年5月29日	原始取得	全部权利
29	软著登字第0738332号	SMSDataFilter 软件[简称:SMSDataFilter]V1.0	2014SR069088	发行人	2014年3月20日	2014年3月21日	2014年5月29日	原始取得	全部权利
30	软著登字第0738033号	BingImgSpider 快速下载软件[简称:BingImgSpider]V1.0	2014SR068789	发行人	2014年3月10日	2014年3月12日	2014年5月29日	原始取得	全部权利
31	软著登字第0738404号	Android 平台 AudioRec 录音软件[简称:AudioRec]V2.0	2014SR069160	发行人	2014年3月10日	2014年3月12日	2014年5月29日	原始取得	全部权利
32	软著登字第0737809号	WindowsPhone 平台 AudioRecord 录音工具软件[简称: AudioRecord]V2.0	2014SR068565	发行人	2014年3月20日	2014年3月21日	2014年5月29日	原始取得	全部权利
33	软著登字第0994400号	web 平台标注与切音软件[简称: web 标注与切音]V1.1.2	2015SR107314	发行人	2015年4月22日	2015年4月22日	2015年6月16日	原始取得	全部权利
34	软著登字第0995121号	Android 平台众包录音 Sayit 软件[简称: 众包录音 Sayit]V1.0.0	2015SR108035	发行人	2015年4月25日	2015年4月25日	2015年6月16日	原始取得	全部权利
35	软著登字第0995122号	对话语音库软件[简称: 对话语音库]V1.0.0.0	2015SR108036	发行人	2015年4月22日	2015年4月22日	2015年6月16日	原始取得	全部权利
36	软著登字第0995123号	识别语音库软件[简称: 识别	2015SR108037	发行人	2015年4月22日	2015年4月22日	2015年6月16日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
		语音库]V1.0.0.0							
37	软著登字第1016654号	文本数据库软件[简称：文本库]V1.0.0.0	2015SR129568	发行人	2015年5月15日	2015年5月15日	2015年7月10日	原始取得	全部权利
38	软著登字第1017099号	词典数据库软件 V1.0.0.0	2015SR130013	发行人	2015年5月15日	2015年5月15日	2015年7月10日	原始取得	全部权利
39	软著登字第1017464号	语音合成数据库软件[简称：合成库]V1.0.0.0	2015SR130378	发行人	2015年5月15日	2015年5月15日	2015年7月10日	原始取得	全部权利
40	软著登字第1410748号	语音识别（ASR）数据后期处理软件 V2.3	2016SR232131	发行人	2016年4月29日	2016年5月3日	2016年8月24日	原始取得	全部权利
41	软著登字第1410750号	web 平台转写与校对软件[简称：web 平台转写与校对]V1.0	2016SR232133	发行人	2016年4月22日	2016年4月25日	2016年8月24日	原始取得	全部权利
42	软著登字第1400092号	图形图像数据库软件 V1.0.0.0	2016SR221475	发行人	2016年5月24日	2016年6月6日	2016年8月16日	原始取得	全部权利
43	软著登字第1439860号	《天天有赚》众包软件[简称：天天有赚]V1.0.0	2016SR261243	发行人	2016年5月5日	2016年5月5日	2016年9月14日	原始取得	全部权利
44	软著登字第1554435号	汽车图片局部特征标注软件 V1.0	2016SR375819	发行人	2016年10月7日	2016年10月7日	2016年12月15日	原始取得	全部权利
45	软著登字第1552861号	人脸图片标注软件 V1.0	2016SR374245	发行人	2016年10月6日	2016年10月7日	2016年12月15日	原始取得	全部权利
46	软著登字第1551998号	在线文字属性标注软件 V1.0	2016SR373382	发行人	2016年10月5日	2016年10月5日	2016年12月15日	原始取得	全部权利
47	软著登字第1552272号	在线图片文字标注软件 V1.0	2016SR373656	发行人	2016年10月5日	2016年10月5日	2016年12月15日	原始取得	全部权利
48	软著登字第1547781号	设备管理软件 V1.0	2016SR369165	发行人	2016年10月3日	2016年10月5日	2016年12月13日	原始取得	全部权利
49	软著登字第1900466号	行车视频标注软件 V1.0.1	2017SR315182	发行人	2017年5月1日	2017年5月1日	2017年6月27日	原始取得	全部权利
50	软著登字第1900474号	人脸五官信息标注软件 V1.0.1	2017SR315190	发行人	2017年5月3日	2017年5月3日	2017年6月27日	原始取得	全部权利
51	软著登字第1900451号	在线语句标注软件 V1.0.1	2017SR315167	发行人	2017年5月5日	2017年5月5日	2017年6月27日	原始取得	全部权利
52	软著登字第1900489号	在线语句扩展软件 V1.0.1	2017SR315205	发行人	2017年5月1日	2017年5月1日	2017年6月27日	原始取得	全部权利
53	软著登字第1900480号	图片中物体部分上色软件	2017SR315196	发行人	2017年5月5日	2017年5月5日	2017年6月27日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
		V1.0.1							
54	软著登字第2079710号	韩国语文本正则化软件 V1.5.0	2017SR494426	发行人	2017年7月10日	2017年7月10日	2017年9月7日	原始取得	全部权利
55	软著登字第2085092号	视频标注软件 V1.0.1	2017SR499808	发行人	2017年7月22日	2017年7月22日	2017年9月8日	原始取得	全部权利
56	软著登字第2228027号	视频数据库软件 V1.0.1	2017SR642743	发行人	2017年11月6日	2017年11月6日	2017年11月22日	原始取得	全部权利
57	软著登字第2192482号	天气图片标注软件 V1.0.1	2017SR607198	发行人	2017年9月13日	2017年9月13日	2017年11月6日	原始取得	全部权利
58	软著登字第2194449号	在线语句意图与属性标注软件 V1.0.1	2017SR609165	发行人	2017年9月12日	2017年9月12日	2017年11月7日	原始取得	全部权利
59	软著登字第2194452号	动漫标注软件 V1.0.1	2017SR609168	发行人	2017年9月13日	2017年9月13日	2017年11月7日	原始取得	全部权利
60	软著登字第2194458号	词典编辑软件 V1.0.1	2017SR609174	发行人	2017年9月8日	2017年9月8日	2017年11月7日	原始取得	全部权利
61	软著登字第2609455号	语音混音检查采样转换软件 V1.0.0	2018SR280360	发行人	2017年11月6日	2017年11月6日	2018年4月25日	原始取得	全部权利
62	软著登字第2609462号	分布式图片采集和分析软件 V1.0.0	2018SR280367	发行人	2017年11月20日	2017年11月20日	2018年4月25日	原始取得	全部权利
63	软著登字第2609470号	安卓图片采集软件 V1.0.3	2018SR280375	发行人	2017年12月20日	2017年12月20日	2018年4月25日	原始取得	全部权利
64	软著登字第2609477号	音频视频对齐分割软件 V1.0.0	2018SR280382	发行人	2017年11月6日	2017年11月6日	2018年4月25日	原始取得	全部权利
65	软著登字第2712455号	基于移动端手写体软件 V1.2.3	2018SR383360	发行人	2017年11月6日	2017年11月6日	2018年5月25日	原始取得	全部权利
66	软著登字第2770255号	多语言分布式文本采集和分析软件 V1.0.0	2018SR441160	发行人	2017年11月20日	2017年11月20日	2018年6月12日	原始取得	全部权利
67	软著登字第3015906号	桌面端 xml_json 通用软件 V.1.0.0.0	2018SR686811	发行人	2018年7月4日	2018年7月4日	2018年8月27日	原始取得	全部权利
68	软著登字第3015820号	桌面端 OCR 标注软件 V1.1.1.2	2018SR686725	发行人	2018年7月5日	2018年7月5日	2018年8月27日	原始取得	全部权利
69	软著登字第3095595号	AudioBook 自动切分与文本对齐软件 V1.0.0	2018SR766500	发行人	2018年7月14日	2018年7月14日	2018年9月20日	原始取得	全部权利
70	软著登字第3096532号	Kaldi 多线程并行解码软件 V1.0.0	2018SR767437	发行人	2018年7月19日	2018年7月19日	2018年9月20日	原始取得	全部权利
71	软著登字第	多语言多种类	2018SR76	发行人	2018年7	2018年7	2018年9	原始	全部

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
	3096514 号	语料设计软件 V1.0.0	7419		月 13 日	月 13 日	月 20 日	取得	权利
72	软著登字第 3095602 号	文本易读性打分软件 V1.0.0	2018SR76 6507	发行人	2018 年 7 月 14 日	2018 年 7 月 14 日	2018 年 9 月 20 日	原始取得	全部权利
73	软著登字第 3080161 号	基于 ASR 的中文语音合成音字一致性校对软件 V1.0.0	2018SR75 1066	发行人	2018 年 7 月 13 日	2018 年 7 月 13 日	2018 年 9 月 17 日	原始取得	全部权利
74	软著登字第 3075399 号	基于 hts 的语音合成软件 V1.0.0	2018SR74 6304	发行人	2018 年 7 月 20 日	2018 年 7 月 20 日	2018 年 9 月 14 日	原始取得	全部权利
75	软著登字第 3080175 号	基于 Kaldi 的语音数据库质量评测样本选取软件 V1.0.0	2018SR75 1080	发行人	2018 年 7 月 13 日	2018 年 7 月 13 日	2018 年 9 月 17 日	原始取得	全部权利
76	软著登字第 3077120 号	说话人自适应合成软件 V1.0.0	2018SR74 8025	发行人	2018 年 7 月 25 日	2018 年 7 月 25 日	2018 年 9 月 14 日	原始取得	全部权利
77	软著登字第 3772167 号	海天瑞声工作平台 V1.9.6.1	2019SR03 51410	发行人	2019 年 3 月 6 日	2019 年 3 月 6 日	2019 年 4 月 19 日	原始取得	全部权利
78	软著登字第 3936032 号	基于 PC 端视频标注软件 V1.0.1.3	2019SR51 5275	发行人	2019 年 3 月 28 日	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
79	软著登字第 3936435 号	基于 PC 端 2D 标注软件 V1.1.1.8	2019SR51 5678	发行人	2019 年 3 月 28 日	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
80	软著登字第 3935990 号	在线文本推理标注软件 V1.0.1	2019SR51 5233	发行人	2019 年 3 月 28 日	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
81	软著登字第 3936000 号	海天瑞声问卷调查系统 V1.0.0.1	2019SR51 5243	发行人	2019 年 3 月 28 日	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
82	软著登字第 3936046 号	在线滑音输入任务分配软件 V1.0.1	2019SR51 5289	发行人	2019 年 3 月 28 日	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
83	软著登字第 4175587 号	文本语音质检平台 V1.0.0.1	2019SR07 54830	发行人	2019 年 6 月 3 日	2019 年 6 月 3 日	2019 年 7 月 22 日	原始取得	全部权利
84	软著登字第 4174859 号	基于 PC 端工具插件平台 V1.0.0.5	2019SR07 54102	发行人	2019 年 6 月 3 日	2019 年 6 月 3 日	2019 年 7 月 22 日	原始取得	全部权利
85	软著登字第 4401220 号	西班牙语语音素自动切分软件 V1.0.0	2019SR09 80463	发行人	2019 年 7 月 30 日	2019 年 7 月 30 日	2019 年 9 月 23 日	原始取得	全部权利
86	软著登字第 4411334 号	基于 ios 多通道录音软件 V3.0.0.2	2019SR09 90577	发行人	2019 年 7 月 30 日	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
87	软著登字第 4411316 号	德语逆正则化软件 V1.0.0	2019SR09 90559	发行人	2019 年 7 月 30 日	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
88	软著登字第4408353号	语音数据库基频一致性评估软件 V1.0.0	2019SR0987596	发行人	2019年7月30日	2019年7月30日	2019年9月24日	原始取得	全部权利
89	软著登字第4412317号	中英数字复杂混合文本注音及韵律标注软件 V1.0.0	2019SR0991560	发行人	2019年7月30日	2019年7月30日	2019年9月25日	原始取得	全部权利
90	软著登字第4412331号	基于 window 的重音标注软件 V1.0.0.0	2019SR0991574	发行人	2019年7月30日	2019年7月30日	2019年9月25日	原始取得	全部权利
91	软著登字第4415480号	基于 Android 多通道录音软件 V3.0.0.2	2019SR0994723	发行人	2019年7月30日	2019年7月30日	2019年9月25日	原始取得	全部权利
92	软著登字第4415486号	基于 window 的新闻语料标注软件 V1.0.0.0	2019SR0994729	发行人	2019年7月30日	2019年7月30日	2019年9月25日	原始取得	全部权利
93	软著登字第4415531号	海天瑞声录音棚预约平台 V1.0.0	2019SR0994774	发行人	2019年7月30日	2019年7月30日	2019年9月25日	原始取得	全部权利
94	软著登字第4408351号	跨平台 3D 点云标注软件 V1.0.0.0	2019SR0987594	发行人	2019年7月30日	2019年7月30日	2019年9月24日	原始取得	全部权利
95	软著登字第6011193号	海天瑞声语音合成在线校对平台 V4.0.0.0	2020SR1132497	发行人	2020年4月30日	2020年4月30日	2020年9月21日	原始取得	全部权利
96	软著登字第6014182号	海天瑞声双语对齐检测软件 V1.0.0	2020SR1135486	发行人	2020年4月30日	2020年4月30日	2020年9月22日	原始取得	全部权利
97	软著登字第6014174号	海天瑞声手机录音实时检测后端 ASR 服务软件 V1.0.0	2020SR1135478	发行人	2020年4月30日	2020年4月30日	2020年9月22日	原始取得	全部权利
98	软著登字第6097955号	海天瑞声线上 MOS 评测软件 V1.3.1	2020SR1219259	发行人	2020年4月30日	2020年4月30日	2020年10月15日	原始取得	全部权利
99	软著登字第6097958号	海天瑞声安卓版 ASR 检测版在线录音软件 V3.0.0.3	2020SR1219262	发行人	2020年4月30日	2020年4月30日	2020年10月15日	原始取得	全部权利
100	软著登字第6097961号	海天瑞声移动版调查问卷软件 V1.0.0	2020SR1219265	发行人	2020年4月30日	2020年4月30日	2020年10月15日	原始取得	全部权利
101	软著登字第6097964号	中文文本易读性打分软件 V1.0.0	2020SR1219268	发行人	2020年4月30日	2020年4月30日	2020年10月15日	原始取得	全部权利
102	软著登字第6089009号	海天瑞声图形图像语义分割校对软件	2020SR1210313	发行人	2020年4月30日	2020年4月30日	2020年10月13日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
		V0.1.0.0							
103	软著登字第6088862号	海天瑞声安卓版 OCR 图片采集软件 V1.0.9	2020SR1210166	发行人	2020年4月30日	2020年4月30日	2020年10月13日	原始取得	全部权利
104	软著登字第6088857号	海天瑞声 IOS 版 ASR 检测板在线录音软件 V1.0.5	2020SR1210161	发行人	2020年4月30日	2020年4月30日	2020年10月13日	原始取得	全部权利
105	软著登字第6123616号	海天瑞声日语手写体采集标注软件 V1.0.0	2020SR1244920	发行人	2020年4月30日	2020年4月30日	2020年10月27日	原始取得	全部权利
106	软著登字第6123639号	海天瑞声中考作文打分标注软件 V1.0.0	2020SR1244943	发行人	2020年4月30日	2020年4月30日	2020年10月27日	原始取得	全部权利
107	软著登字第6123640号	海天瑞声多音字标注软件 V1.0.0	2020SR1244944	发行人	2020年4月30日	2020年4月30日	2020年10月27日	原始取得	全部权利
108	软著登字第7080694号	海天瑞声安卓版天天 AI 录音软件 V1.5.6	2021SR0358467	发行人	2020年12月1日	2021年3月3日	2021年3月9日	原始取得	全部权利
109	软著登字第7503884号	海天瑞声中文标点预测软件 V1.0.0	2021SR0781258	发行人	2021年2月21日	2021年3月16日	2021年5月27日	原始取得	全部权利
110	软著登字第7503885号	海天瑞声 OCR 采集移动端 IOS 版软件 V1.0.0	2021SR0781259	发行人	2021年3月18日	2021年3月31日	2021年5月27日	原始取得	全部权利
111	软著登字第7503883号	海天瑞声中英文混文本语种检查软件 V1.0.0	2021SR0781257	发行人	2021年2月26日	2021年3月18日	2021年5月27日	原始取得	全部权利
112	软著登字第7503882号	海天瑞声在线多语种 OCR 标注平台 V1.2.0	2021SR0781256	发行人	2021年3月20日	2021年4月7日	2021年5月27日	原始取得	全部权利
113	软著登字第7503881号	海天瑞声在线无人车多场景标注平台 V1.5.0	2021SR0781255	发行人	2021年3月1日	2021年3月16日	2021年5月27日	原始取得	全部权利
114	软著登字第7503880号	海天瑞声语音识别一体化处理平台 V1.0.0	2021SR0781254	发行人	2021年3月15日	2021年3月31日	2021年5月27日	原始取得	全部权利
115	软著登字第7503879号	海天瑞声声纹识别软件 V1.0.0	2021SR0781253	发行人	2021年3月24日	2021年4月7日	2021年5月27日	原始取得	全部权利
116	软著登字第7543504号	海天瑞声基于 PC 端多场景语义分割软件 V2.0.1	2021SR0820878	发行人	2021年3月21日	2021年4月8日	2021年6月2日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
117	软著登字第7543508号	海天瑞声安卓版手写体采集软件 V1.0.2	2021SR0820882	发行人	2021年3月18日	2021年3月31日	2021年6月2日	原始取得	全部权利
118	软著登字第7503887号	海天瑞声非标准词提取软件 V1.0.0	2021SR0781261	发行人	2021年3月18日	2021年3月31日	2021年5月27日	原始取得	全部权利
119	软著登字第7543507号	海天瑞声英文文本正则化软件 V1.0.0	2021SR0820881	发行人	2021年3月15日	2021年3月31日	2021年6月2日	原始取得	全部权利
120	软著登字第7543506号	海天瑞声手写体采集移动端 IOS 版软件 V1.0.2	2021SR0820880	发行人	2020年12月12日	2021年3月21日	2021年6月2日	原始取得	全部权利
121	软著登字第7543505号	海天瑞声语音合成一体化处理平台 V0.2.1	2021SR0820879	发行人	2020年10月22日	2021年3月10日	2021年6月2日	原始取得	全部权利
122	软著登字第7503886号	海天瑞声安卓版图片采集软件 V1.0.0	2021SR0781260	发行人	2021年3月18日	2021年3月31日	2021年5月27日	原始取得	全部权利
123	软著登字第8699718号	海天瑞声多模态标注软件[简称:多模态标注软件]V1.1.0	2021SR1977092	发行人	2021年8月25日	未发表	2021年12月2日	原始取得	全部权利
124	软著登字第8699731号	海天瑞声在线事件抽取标注系统[简称:事件抽取标注系统]V1.1.0	2021SR1977105	发行人	2021年10月15日	未发表	2021年12月2日	原始取得	全部权利
125	软著登字第8699732号	海天瑞声在线实体词标注系统[简称:在线实体词标注系统]V1.0.0	2021SR1977106	发行人	2021年10月25日	未发表	2021年12月2日	原始取得	全部权利
126	软著登字第9599011号	海天瑞声 IOS 版手部照片采集软件 V1.0	2022SR0644812	发行人	2021年12月1日	未发表	2022年5月26日	原始取得	全部权利
127	软著登字第9599012号	海天瑞声 IOS 版简笔画在线采集软件 V1.0.3	2022SR0644813	发行人	2022年3月17日	未发表	2022年5月26日	原始取得	全部权利
128	软著登字第9599010号	海天瑞声 Android 版手部照片采集软件 V1.0	2022SR0644811	发行人	2021年12月1日	未发表	2022年5月26日	原始取得	全部权利
129	软著登字第10342410号	海天瑞声语音数据预处理系统 V1.0.0	2022SR1388211	发行人	2022年8月1日	2022年8月1日	2022年10月10日	原始取得	全部权利
130	软著登字第	海天瑞声文本	2022SR13	发行人	2022年8	未发表	2022年10	原始	全部

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
	10340861 号	及拼音处理系统 V2.0.0	86662		月 1 日		月 10 日	取得	权利
131	软著登字第 10342420 号	海天瑞声巴葡文本正则化软件[简称:TN(巴葡)]V1.1.0	2022SR1388221	发行人	2022 年 7 月 19 日	未发表	2022 年 10 月 10 日	原始取得	全部权利
132	软著登字第 10342421 号	海天瑞声语音文本标注助手软件 V1.0.0	2022SR1388222	发行人	2022 年 8 月 1 日	未发表	2022 年 10 月 10 日	原始取得	全部权利
133	软著登字第 11004807 号	海天瑞声语料文本标注系统 V2.0.0	2023SR0417636	发行人	2022 年 4 月 1 日	2022 年 8 月 25 日	2023 年 3 月 30 日	原始取得	全部权利
134	软著登字第 2924705 号	电子书记员 Word 插件软件 V1.0.1	2018SR595610	中瑞智	2018 年 4 月 21 日	2018 年 4 月 21 日	2018 年 7 月 30 日	原始取得	全部权利
135	软著登字第 2482051 号	电子书记员系统服务软件 V1.0.0	2018SR152956	中瑞智	2018 年 3 月 2 日	2018 年 3 月 2 日	2018 年 3 月 8 日	原始取得	全部权利
136	软著登字第 2618122 号	电子书记员系统服务软件 V2.0.0	2018SR289027	中瑞智	2018 年 1 月 22 日	2018 年 1 月 29 日	2018 年 4 月 27 日	原始取得	全部权利
137	软著登字第 3019019 号	定制模型训练工具软件 V1.0	2018SR689924	中瑞智	2018 年 6 月 12 日	2018 年 6 月 12 日	2018 年 8 月 28 日	原始取得	全部权利
138	软著登字第 2617120 号	法官助手软件 V1.0	2018SR288025	中瑞智	2017 年 12 月 22 日	2017 年 12 月 22 日	2018 年 4 月 27 日	原始取得	全部权利
139	软著登字第 3019026 号	会议记录软件 V1.0	2018SR689931	中瑞智	2018 年 6 月 13 日	2018 年 6 月 13 日	2018 年 8 月 28 日	原始取得	全部权利
140	软著登字第 3019012 号	潘多拉魔盒软件 V1.0	2018SR689917	中瑞智	2018 年 6 月 12 日	2018 年 6 月 12 日	2018 年 8 月 28 日	原始取得	全部权利
141	软著登字第 2617499 号	实时字幕系统软件 V1.0.0	2018SR288404	中瑞智	2018 年 1 月 6 日	2018 年 1 月 8 日	2018 年 4 月 27 日	原始取得	全部权利
142	软著登字第 3019031 号	音频转写助手软件 V1.0.1	2018SR689936	中瑞智	2018 年 6 月 12 日	2018 年 6 月 12 日	2018 年 8 月 28 日	原始取得	全部权利
143	软著登字第 3204438 号	语音合成系统软件 V1.0	2018SR875343	中瑞智	2018 年 10 月 22 日	2018 年 10 月 22 日	2018 年 11 月 1 日	原始取得	全部权利
144	软著登字第 3169119 号	智控软件 V1.0	2018SR840024	中瑞智	2018 年 10 月 12 日	2018 年 10 月 12 日	2018 年 10 月 22 日	原始取得	全部权利
145	软著登字第 2710336 号	智能会议系统软件 V1.0.1	2018SR381241	中瑞智	2018 年 3 月 21 日	2018 年 3 月 21 日	2018 年 5 月 25 日	原始取得	全部权利
146	软著登字第 2481980 号	智能音频采集系统软件 V1.1.4	2018SR152885	中瑞智	2017 年 12 月 22 日	2017 年 12 月 25 日	2018 年 3 月 8 日	原始取得	全部权利
147	软著登字第 2617488 号	智能语音输入法软件 V1.0.0	2018SR288393	中瑞智	2018 年 1 月 6 日	2018 年 1 月 8 日	2018 年 4 月 27 日	原始取得	全部权利

序号	证书号	软著名称	登记号	著作权人	开发完成日期	首次发表日期	登记日期	权利取得方式	权利范围
148	软著登字第3043893号	智能语音输入法软件 V2.0.0	2018SR714798	中瑞智	2018年6月19日	2018年6月19日	2018年9月5日	原始取得	全部权利
149	软著登字第3019007号	智能语音问答记录软件 V1.0.1	2018SR689912	中瑞智	2018年6月13日	2018年6月13日	2018年8月28日	原始取得	全部权利
150	软著登字第4256134号	会议管理软件 V1.0	2019SR0835377	中瑞智	2018年12月12日	2018年12月12日	2019年8月12日	原始取得	全部权利
151	软著登字第4256238号	钉钉会议预约软件 V1.0	2019SR0835481	中瑞智	2019年1月2日	2019年1月2日	2019年8月12日	原始取得	全部权利
152	软著登字第4820054号	音频检索软件 V1.0	2019SR1399297	中瑞智	2019年10月25日	2019年10月25日	2019年12月19日	原始取得	全部权利
153	软著登字第4820060号	语音标注转写软件 V1.0	2019SR1399303	中瑞智	2019年10月25日	2019年10月25日	2019年12月19日	原始取得	全部权利
154	软著登字第4842862号	庭审智能语音助手软件 V1.0	2019SR1422105	中瑞智	2019年10月25日	2019年10月25日	2019年12月24日	原始取得	全部权利
155	软著登字第5547740号	智能监听管理软件 V1.0	2020SR0669044	中瑞智	2020年3月29日	2020年3月29日	2020年6月23日	原始取得	全部权利
156	软著登字第5547748号	产品授权软件 V1.0	2020SR0669052	中瑞智	2020年3月9日	2020年3月9日	2020年6月23日	原始取得	全部权利
157	软著登字第6355901号	自学习管理软件 V1.0	2020SR1554929	中瑞智	2020年8月28日	2020年9月16日	2020年11月9日	原始取得	全部权利
158	软著登字第6634518号	锐听管理客户端软件 V1.0	2020SR1831516	中瑞智	2020年9月8日	未发表	2020年12月16日	原始取得	全部权利
159	软著登字第6634517号	智能庭审记录系统软件 V1.0.0	2020SR1831515	中瑞智	2020年10月10日	未发表	2020年12月16日	原始取得	全部权利
160	软著登字第8915866号	电子文件管理软件 V1.0	2021SR2193240	中瑞智	2021年9月21日	2021年9月21日	2021年12月28日	原始取得	全部权利
161	软著登字第8915869号	机器人流程自动化软件 V1.0	2021SR2193243	中瑞智	2021年11月2日	2021年11月2日	2021年12月28日	原始取得	全部权利
162	软著登字第8915763号	基层卫生管理小程序软件 V1.0	2021SR2193137	中瑞智	2021年4月28日	2021年4月28日	2021年12月28日	原始取得	全部权利
163	软著登字第8915761号	智能会议管理系统软件 V1.0	2021SR2193135	中瑞智	2021年2月23日	2021年2月23日	2021年12月28日	原始取得	全部权利
164	软著登字第8915760号	多系统集中授权管理软件 V1.0	2021SR2193134	中瑞智	2021年7月8日	2021年7月8日	2021年12月28日	原始取得	全部权利

注：法人或者其他组织的软件著作权，保护期为 50 年，截止于软件首次发表后第 50 年的 12 月 31 日，但软件自开发完成之日起 50 年内未发表的，不再保护。