

## 科创板投资风险提示

本次股票发行后拟在科创板市场上市，该市场具有较高的投资风险。科创板公司具有研发投入大、经营风险高、业绩不稳定、退市风险高等特点，投资者面临较大的市场风险。投资者应充分了解科创板市场的投资风险及本公司所披露的风险因素，审慎作出投资决定。

# 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)



## 首次公开发行股票并在科创板上市 招股说明书

保荐机构（主承销商）



(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)

## 声 明

发行人及全体董事、监事、高级管理人员承诺招股说明书及其他信息披露资料不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。

发行人控股股东、实际控制人承诺本招股说明书不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。

公司负责人和主管会计工作的负责人、会计机构负责人保证招股说明书中财务会计资料真实、完整。

发行人及全体董事、监事、高级管理人员、发行人的控股股东、实际控制人以及保荐人、承销的证券公司承诺因发行人招股说明书及其他信息披露资料有虚假记载、误导性陈述或者重大遗漏，致使投资者在证券发行和交易中遭受损失的，将依法赔偿投资者损失。

保荐人及证券服务机构承诺因其为发行人本次公开发行制作、出具的文件有虚假记载、误导性陈述或者重大遗漏，给投资者造成损失的，将依法赔偿投资者损失。

中国证监会、交易所对本次发行所作的任何决定或意见，均不表明其对注册申请文件及所披露信息的真实性、准确性、完整性作出保证，也不表明其对发行人的盈利能力、投资价值或者对投资者的收益作出实质性判断或保证。任何与之相反的声明均属虚假不实陈述。

根据《证券法》的规定，股票依法发行后，发行人经营与收益的变化，由发行人自行负责；投资者自主判断发行人的投资价值，自主作出投资决策，自行承担股票依法发行后因发行人经营与收益变化或者股票价格变动引致的投资风险。

## 发行概况

发行股票类型	人民币普通股（A股）
发行股数	本次公开发行股票数量为 1,070 万股，占发行后公司总股本的 25%；本次发行全部为新股发行，不涉及原股东公开发售股份。
每股面值	人民币 1.00 元
每股发行价格	人民币 36.94 元
发行日期	2021 年 8 月 3 日
拟上市证券交易所及板块	上海证券交易所科创板
发行后总股本	4,280 万股
保荐机构（主承销商）	华泰联合证券有限责任公司
招股说明书签署日期	2021 年 8 月 9 日

## 重大事项提示

本公司特别提请投资者注意，在作出投资决策之前，务必仔细阅读本招股说明书正文内容，并特别关注以下重要事项。

### 一、发行人业务定位特别提示

发行人主要从事训练数据研发设计、生产及销售业务，提供用于人工智能算法模型开发训练所需的专业数据集，报告期内前述业务<sup>1</sup>收入占比超过 97%。发行人的下游客户主要包括大型科技公司、人工智能企业及科研机构等，下游客户使用发行人提供的训练数据对算法模型进行开发、拓展和提升。当前发行人提供的训练数据主要是可满足多种场景应用需求的通用型训练数据，正在逐步探索更多垂直领域特征的训练数据覆盖。

发行人从属于软件和信息技术服务业，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及质检，形成训练数据集。发行人的主要产品、服务均以训练数据集为核心。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，并以软件形式向客户交付。发行人并非 AI 产业链上的算法模型研发商或提供商，提请广大投资者认真阅读本招股说明书关于发行人主营业务的具体说明。其中，对于发行人产品服务内容、作业模式、业务实质的通俗化说明，请参见本招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（一）主营业务概述/2、产品服务内容、作业模式和业务实质介绍”。

### 二、发行人数据服务采购模式特别提示

基于下游客户需求和发行人所从事的训练数据业务特点，发行人在生产训练数据过程中需要大量劳务人员提供原料数据采集、标注服务。发行人通过对接数据服务供应商的方式解决劳务服务需求，数据服务供应商主要为人力资源外包服务公司，具备寻找符合发行人需求的终端人员并组织其向发行人提供数据采集、标注服务的专业能力。在部分具体训练数据生产项目中，涉及的终端采集、标注人员数量较多，且覆盖的语种、口音区、年龄特征等要求多种多样，再加之发行人劳务服务通常具有较为迫切的及时响应需求，数据服务供应商等也采取与不同

---

<sup>1</sup> 即发行人的训练数据定制服务、训练数据产品业务。

类型终端人员的组织人员（即“小外包人员”）对接的方式，组织终端人员为发行人提供相应服务，因此发行人数据服务采购模式符合行业惯例，具备商业合理性。

发行人与数据服务供应商、小外包人员、终端人员之间的合作模式、实际业务执行过程等情况请参见本招股说明书“第六节 业务与技术/一、发行人主营业务、主要产品或服务的情况/（三）主要经营模式/4、数据服务采购情况介绍/（2）发行人与服务公司、小外包人员、终端人员之间的合作情况”。

### 三、特别风险提示

#### （一）市场竞争加剧的风险

发行人所在的细分领域为人工智能基础数据服务领域。近年来，参与该行业的公司不断增多，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。

#### （二）收入主要来自老客户，新客户收入占比较低的风险

报告期内，发行人收入主要来自老客户，新客户收入占比较低。2018年、2019年及2020年，发行人来自老客户的收入占比分别达到88.58%、89.09%及93.47%。发行人正致力于维护老客户、拓展新客户，若老客户降低对公司产品、服务的采购量或公司未能维持与老客户的合作关系，则公司的经营情况将受到不利影响；若新客户拓展情况不达预期，则公司的经营情况也将受到不利影响。

#### （三）专利被申请宣告无效的风险

截至本招股说明书签署日，根据发行人收到的由专利代理机构转送的国家知识产权局专利局出具的《无效宣告请求受理通知书》，发行人及其子公司已取得授权的9项发明专利和1项实用新型专利被提出无效宣告请求，涉及无效宣告请求的相关专利情况参见本招股说明书“第六节 业务和技术/五、发行人的主要固定资产和无形资产/（二）主要无形资产/1、专利”。截至2021年5月7日，发行人已经收到上述专利中7项发明专利和1项实用新型专利的无效宣告请求审查决定书，其中7项发明专利维持专利权有效，1项实用新型专利专利权维持部分有

效，其余 2 项专利的无效宣告请求审查事项尚处于正常审理阶段，国家知识产权局专利局尚未作出裁定。

尽管剩余 2 项涉及无效宣告请求的相关专利权被宣告无效的可能性较小，但不排除发行人的相关专利权被部分或者全部宣告无效的可能，若公司知识产权被宣告无效，公司被宣告无效的专利或其权利要求中公开的技术点可能存在被竞争对手或第三方模仿的风险，若相关方模仿上述技术从事与发行人同类型业务，参与市场竞争，可能会对发行人的业务经营和财务状况产生不利影响。

#### **（四）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，发行人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。如果发行人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，则可能导致发行人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，进而对发行人的客户拓展、产品创新和经营业绩造成不利影响。

#### **（五）业务发展受下游人工智能领域发展状况影响较大的风险**

发行人主要为 AI 产业链上的各类机构提供算法模型开发所需的训练数据，服务于下游人工智能行业发展。发行人的经营情况、业务发展与下游人工智能领域的发展状况息息相关。若未来人工智能领域景气度下降，或因技术更迭使得对训练数据的市场需求发生变动，将对公司业绩产生较大影响。

#### **（六）经营业绩受税收优惠政策影响较大的风险**

报告期内，公司享受了增值税即征即退、高新技术企业和国家规划布局内的重点软件企业所得税税率优惠、研发费用加计扣除等一系列税收优惠，报告期各期税收优惠影响金额占利润总额的比例分别为 28.84%、22.96%和 24.94%，经营业绩受税收优惠政策影响较大。如果未来公司所享受的税收优惠政策出现不利变化，或者公司不再符合税收优惠的条件，则公司将无法持续享受税收优惠或存在所享受的税收优惠减少的可能性，并进而对公司未来经营业绩和现金流水平造成不利影响。

### （七）采购成本上升风险

报告期内，发行人进行采购原料数据采集、标注所支付的数据服务费占发行人各年采购总额的比重较高，分别达到 81.22%、79.84% 和 84.69%。随着城乡生活成本的持续提升，前述采购价格可能存在持续上升的趋势，若发行人的技术发展带来的效率提升无法抵消前述采购成本上升的影响，可能会对公司的经营业绩带来一定不利影响。

### （八）数据安全相关风险

发行人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，随着发行人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，与数据安全、个人信息保护相关的法律规章体系逐步完善。如果将来公司未能根据法律规章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或可能产生诉讼纠纷，进而给公司的经营带来不利影响。

### （九）中美贸易摩擦风险

公司产品和服务的主要出口地包括美国、日韩等国家或地区。2018 年至 2020 年，公司出口美国的营业收入占当期公司营业收入的比重分别达到 12.96%、19.48% 和 19.93%。公司相关境外业务始终严格遵守中国及他国法律。但国际局势瞬息万变，近年来中美贸易摩擦不断，如果中美双方出台新的贸易保护主义措施，可能会影响美国市场相关客户将数据提供给公司用于标注，减少与公司进行训练数据方面的业务合作，公司相关业务可能会受到约束，将会对公司的生产经营产生不利影响。

## 四、新型冠状病毒肺炎疫情对公司经营情况的影响

2020 年 1 月，新型冠状病毒肺炎疫情爆发，致使全国各行各业均遭受了不同程度的影响，并已在全球范围内蔓延。因新冠疫情爆发期间实施隔离、交通管制等防疫管控措施，对发行人的采购、销售等环节造成了不利影响。由于疫情导致的延期复工，公司及主要客户、主要供应商的生产经营均受到一定程度的影响，

公司采购原料数据采集、标注服务以及下游客户对公司产品服务的验收周期均受到延期复工影响，相比正常进度有所延后。若本次新冠疫情在较长时间内持续，将可能对公司的经营造成不利影响。

#### （一）新冠疫情对公司营业收入结构的具体影响

新冠疫情对发行人营业收入结构的影响主要体现在业务类型收入结构方面。由于新冠疫情防控期间对于人员聚集和流动的限制，发行人训练数据定制服务项目的采集、标注等工作难以以现场方式大规模开展，因此发行人引导客户购买已有的训练数据产品满足其产品研发需求，从而使得训练数据产品收入占比明显上升。2020年，发行人训练数据产品收入占主营业务收入的比例为45.48%，明显高于2019年的38.63%和2018年的34.27%。

未来随着新冠疫情得到缓解和控制，发行人将继续均衡发展训练数据定制服务和训练数据产品两大业务类型，从而满足下游客户对于定制训练数据和通用训练数据的多样化需求，为客户提供全方位的训练数据服务。

#### （二）新冠疫情对公司经营业绩及其变动的具体影响

新冠疫情对发行人经营业绩增速和合同订单获取产生了一定的不利影响，具体体现在执行项目周期延长和新增合同订单额下降两个方面。

执行项目周期延长方面，由于新冠疫情防控期间原料数据现场采集、标注工作受到限制以及客户停工导致项目沟通及验收效率降低的因素，发行人部分在执行项目周期有所延长，未能在通常周期内完成。截至2020年12月31日，发行人未完工项目中，受新冠疫情对项目周期影响的项目数量约占30%，涉及的项目收入约1,900万元。

新增合同订单额下降方面，由于下游客户在新冠疫情爆发期间存在停工停产或居家办公的情形，产品开发计划、开发进度和年初预算制定均可能受到影响，且该等影响在年中逐步开始体现，并随着境外疫情的持续不稳定而对境外客户的合同订单签订形成了更为显著的影响，因此在一定程度上导致了发行人在2020年下半年期间的新增合同订单额增速放缓。

尽管新冠疫情对发行人经营业绩增速和合同订单获取产生了一定的不利影响，但2020年以来发行人总体经营情况仍保持稳定，各项业务持续正常运营，



未出现重大不利变化。

## 五、募集资金投资项目实施对发行人经营情况的影响

本次募集资金金额及投资项目综合考虑了行业和市场状况、技术水平及发展趋势、场地、设备和人员等因素，发行人对其可行性进行了充分论证，具备合理性。但由于发行人为轻资产型公司，且处于成长阶段，公司资产和收入规模相对较小，而本次募集资金金额超过公司资产和收入规模，募投项目的实施将可能导致公司未来的业务经营、管理模式发生变化，也对公司各方面经营管理能力、资产运营能力均提出了更高的要求。如果未来行业或市场环境发生难以预期的不利变化，或由于发行人管理能力、资产运营能力不足等使得募投项目的按期实施及完全达产受到不利影响，则募集资金投资项目的经济效益可能无法如期实现。

同时，本次募集资金投资项目实施过程中将新增大量的固定资产、无形资产购置支出及研发投入，各年新增折旧摊销等费用金额较大。如果募投项目的经济效益不能如期实现，且发行人主营业务收入的增长不足以缓冲募投项目实施带来的折旧摊销等费用的增加，则公司利润将可能出现一定程度的下滑。

## 六、财务报告审计截止日后的主要经营状况

公司财务报告审计截止日为 2020 年 12 月 31 日。毕马威华振会计师事务所（特殊普通合伙）对公司 2021 年第一季度财务报表，包括 2021 年 3 月 31 日合并及母公司的资产负债表，2021 年第一季度合并及母公司的利润表、合并及母公司的现金流量表以及相关财务报表附注进行了审阅，并出具了毕马威华振专字第 2101016 号《审阅报告》。

截至 2021 年 3 月 31 日，公司总资产为 48,773.15 万元，较上年末增加 2.17%；总负债为 3,342.91 万元，较上年末减少 15.14%；所有者权益为 45,430.25 万元，较上年末增加 3.73%。2021 年 1-3 月，公司实现营业收入 4,423.89 万元，同比增长 41.52%；实现归属于母公司所有者权益的净利润 1,634.17 万元，同比增长 19.69%；实现扣除非经常性损益后的归属于母公司股东的净利润 1,432.50 万元，同比增长 26.12%。

公司财务报告审计截止日至本招股说明书签署日，公司主要经营状况正常，在经营模式、采购规模及采购价格、销售规模及销售价格、主要客户及供应商构

成、税收政策以及其他可能影响投资者判断的重大事项方面未发生重大变化。

## 七、2021年1-6月的业绩预告信息

经公司初步测算，公司预计2021年1-6月实现营业收入约12,750万元至14,025万元，同比增长约25.07%至37.58%；预计实现归属于母公司股东的净利润约4,650万元至5,115万元，同比增长约0.74%至10.81%；预计实现扣除非经常性损益后归属于母公司股东的净利润约4,185万元至4,604万元，同比增长约0.54%至10.61%。上述2021年1-6月财务数据为公司初步核算数据，未经会计师审计或审阅，且不构成盈利预测。

# 目 录

声 明.....	1
发行概况 .....	2
重大事项提示 .....	3
一、发行人业务定位特别提示.....	3
二、发行人数据服务采购模式特别提示.....	3
三、特别风险提示.....	4
四、新型冠状病毒肺炎疫情对公司经营情况的影响.....	6
五、募集资金投资项目实施对发行人经营情况的影响.....	8
六、财务报告审计截止日后的主要经营状况.....	8
七、2021年1-6月的业绩预告信息 .....	9
目 录.....	10
第一节 释 义 .....	14
一、普通术语.....	14
二、专业术语.....	15
第二节 概 览 .....	17
一、发行人及本次发行的中介机构基本情况.....	17
二、本次发行概况.....	17
三、发行人主要财务数据和财务指标.....	19
四、发行人主营业务经营情况概述.....	19
五、发行人核心技术先进性及未来发展战略.....	21
六、发行人选择的具体上市标准.....	22
七、发行人公司治理特殊安排等重要事项.....	23
八、募集资金用途概述.....	23
第三节 本次发行概况 .....	24
一、本次发行的基本情况.....	24
二、与本次发行有关的机构.....	25
三、发行人与本次发行有关的中介机构的关系.....	27
四、本次发行上市的重要日期.....	27
五、本次发行战略配售情况.....	27

六、发行人高级管理人员、核心人员拟参与战略配售情况.....	28
七、保荐机构关联公司参与战略配售情况.....	29
<b>第四节 风险因素 .....</b>	<b>30</b>
一、市场竞争加剧的风险.....	30
二、收入主要来自老客户，新客户收入占比较低的风险.....	30
三、专利被申请宣告无效的风险.....	30
四、核心技术快速迭代的风险.....	31
五、核心技术人员流失风险.....	31
六、核心技术失密风险.....	31
七、业务发展受下游人工智能领域发展状况影响较大的风险.....	31
八、新业务拓展风险.....	32
九、训练数据产品无法实现授权销售的风险.....	32
十、经营业绩受税收优惠政策影响较大的风险.....	32
十一、采购成本上升风险.....	32
十二、汇率波动风险.....	33
十三、净资产收益率下降的风险.....	33
十四、数据安全相关风险.....	33
十五、实际控制人控制不当的风险.....	33
十六、中美贸易摩擦风险.....	34
十七、疫情风险.....	34
<b>第五节 发行人基本情况 .....</b>	<b>35</b>
一、发行人基本情况.....	35
二、发行人的设立情况.....	35
三、发行人设立以来的重大资产重组情况.....	41
四、公司在其他证券市场的上市或挂牌情况.....	41
五、发行人的股权结构及组织架构.....	41
六、发行人控股子公司及参股公司情况.....	42
七、持有发行人 5% 以上股份的主要股东及实际控制人情况 .....	45
八、发行人股本情况.....	52
九、董事、监事、高级管理人员与核心技术人员的简要情况.....	61
十、发行人本次公开发行申报前已经制定或实施的股权激励及相关安排... 71	71

十一、发行人员工情况.....	73
<b>第六节 业务和技术 .....</b>	<b>76</b>
一、发行人主营业务、主要产品或服务的情况.....	76
二、发行人所处行业的基本情况和竞争状况.....	121
三、销售情况和主要客户.....	158
四、采购情况和主要供应商.....	181
五、发行人的主要固定资产和无形资产.....	194
六、发行人的研发和技术.....	207
七、发行人的境外经营及境外资产情况.....	262
<b>第七节 公司治理与独立性 .....</b>	<b>264</b>
一、公司治理.....	264
二、独立性情况.....	267
三、同业竞争.....	269
四、关联方及关联交易.....	271
<b>第八节 财务会计信息与管理层分析 .....</b>	<b>278</b>
一、报告期经审计的财务报表.....	278
二、与财务会计信息相关的重大事项的判断标准.....	284
三、财务报表的编制基础、遵循企业会计准则的声明、合并财务报表范围及变化情况.....	284
四、报告期内主要采用的会计政策和会计估计.....	285
五、重要会计政策、会计估计的变更及其影响.....	329
六、非经常性损益情况.....	335
七、报告期内执行的主要税收政策及缴纳的主要税种.....	336
八、报告期的主要财务指标.....	340
九、经营成果分析.....	342
十、资产质量分析.....	390
十一、偿债能力、流动性与持续经营能力分析.....	403
十二、重大投资或资本性支出、重大资产业务重组或股权收购合并事项.....	410
十三、资产负债表日后事项、或有事项及其他重要事项.....	411
十四、财务报告审计截止日后主要财务信息及经营状况.....	412
<b>第九节 募集资金运用与未来发展规划 .....</b>	<b>417</b>

一、募集资金情况概述.....	417
二、募集资金运用情况.....	419
三、未来发展规划.....	448
<b>第十节 投资者保护 .....</b>	<b>450</b>
一、公司投资者权益保护的情况.....	450
二、股利分配事项.....	451
三、本次发行前滚存利润的分配安排.....	454
四、本次公司上市后未来三年分红回报规划.....	454
五、主要承诺.....	456
<b>第十一节 其他重要事项 .....</b>	<b>475</b>
一、重要合同.....	475
二、对外担保情况.....	477
三、重大诉讼或仲裁事项.....	477
四、董事、监事、高级管理人员和核心技术人员最近三年涉及行政处罚、被司法机关立案侦查、被中国证监会立案调查情况.....	478
五、控股股东、实际控制人报告期内涉及重大违法行为情况.....	478
<b>第十二节 有关声明 .....</b>	<b>479</b>
一、发行人全体董事、监事、高级管理人员声明.....	479
二、控股股东、实际控制人声明.....	481
三、保荐人（主承销商）声明.....	482
四、保荐机构董事长、总经理声明.....	483
五、发行人律师声明.....	484
六、会计师事务所声明.....	485
七、验资机构声明.....	486
八、验资复核机构声明.....	487
九、资产评估机构声明.....	488
<b>第十三节 附件 .....</b>	<b>490</b>
一、备查文件.....	490
二、文件查阅地址及时间.....	490

## 第一节 释 义

在本招股说明书中，除非文中另有说明，下列简称具有如下含义：

### 一、普通术语

发行人、本公司、公司、海天瑞声	指	北京海天瑞声科技股份有限公司
海天瑞声有限	指	北京海天瑞声科技有限公司
控股股东、实际控制人	指	贺琳
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
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
中瑞安	指	北京中瑞安投资中心（有限合伙），为发行人的股东
中瑞立	指	北京中瑞立投资中心（有限合伙），为发行人的股东
创世联合	指	北京创世联合投资管理有限公司，为北京中瑞安投资中心（有限合伙）普通合伙人、执行事务合伙人
创慧科瑞	指	北京创慧科瑞投资管理有限公司，为北京中瑞立投资中心（有限合伙）普通合伙人、执行事务合伙人
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
上海兴富	指	上海兴富创业投资管理中心（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
天津金星	指	天津金星创业投资有限公司，曾用名“天津金星投资有限公司”，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），为发行人的股东
杭州士兰	指	杭州士兰创业投资有限公司，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），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），为发行人的股东
中科海讯	指	北京中科海讯数字科技股份有限公司（300810.SZ），发行人实际控制人贺琳配偶蔡惠智控制的企业
海讯科技	指	北京中科海讯科技有限公司，发行人实际控制人贺琳及其配偶蔡惠智共同控制的企业
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
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
工信部	指	中华人民共和国工业和信息化部
《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》
《上市规则》	指	《上海证券交易所科创板股票上市规则》
《公司章程》	指	《北京海天瑞声科技股份有限公司公司章程》
本次发行	指	发行人根据本招股说明书所载条件首次公开发行人民币普通股（A股）的行为
报告期	指	2018年1月1日至2020年12月31日
A股	指	发行人本次发行的面值为人民币1.00元的普通股
上市	指	发行人股票在上海证券交易所科创板挂牌交易
保荐机构、主承销商、华泰联合	指	华泰联合证券有限责任公司
会计师事务所、毕马威华振事务所	指	毕马威华振会计师事务所（特殊普通合伙）
发行人律师、律师、天元律师事务所	指	北京市天元律师事务所
元、万元、亿元	指	若无特别说明，均以人民币为度量币种

## 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	Artificial Intelligence，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
机器学习（ML）	指	Machine Learning，是一门专门研究计算机模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身性能的学科
深度学习（DL）	指	Deep Learning，机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像、声音和文本
深度神经网络（DNN）	指	Deep Neural Network，是一个多层的人工神经网络模型，包含了：输入层，隐藏层和输出层
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/自然语言处理（NLP）	指	Natural Language Processing，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（CV）	指	Computer Vision，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
光学字符识别（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
原料数据	指	Raw Data，即未经加工处理的原始数据
语料清洗/文本清洗	指	针对文本数据审查过程中发现的错误值、缺失值、异常值、可疑数据，选用适当方法进行“清理”
文本正则化	指	将文本中的数字、符号、缩写等转换成语言文字
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
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
音素	指	根据语音的自然属性划分出来的最小语音单位，依据音节里的发音动作来分析，一个动作构成一个音素。音素分为元音与辅音两大类。如汉语音节啊(ā)只有一个音素，爱(ài)有两个音素，代(dài)有三个音素等
B/S 架构	指	Browser/Server，即浏览器/服务器模式
C/S 架构	指	Client/Server，即客户端/服务器模式
条件随机场 (CRF)	指	Conditional Random Field，是一种鉴别式机率模型，是随机场的一种，常用于标注或分析序列资料
支撑向量机 (SVM)	指	Support Vector Machine，是一类按监督学习方式对数据进行二元分类的广义线性分类器
隐马尔可夫模型 (HMM)	指	Hidden Markov Model，它用来描述一个含有隐含未知参数的马尔可夫过程

本招股说明书除特别说明外若出现总数和各分项数值之和尾数不符的情况，均为四舍五入原因造成。

## 第二节 概 览

本概览仅对招股说明书全文作扼要提示。投资者作出投资决策前，应认真阅读招股说明书全文。

### 一、发行人及本次发行的中介机构基本情况

#### （一）发行人概况

（一）发行人基本情况			
发行人名称	北京海天瑞声科技股份有限公司	成立日期	2005年05月11日
注册资本	3,210万元人民币	法定代表人	贺琳
注册地址	北京市海淀区成府路28号4-801	主要生产经营地址	北京市海淀区成府路28号4-801
控股股东	贺琳	实际控制人	贺琳
行业分类	I65软件和信息技术服务业	在其他交易场所（申请）挂牌或上市的情况	无

#### （二）中介机构概况

（二）本次发行的有关中介机构			
保荐人	华泰联合证券有限责任公司	主承销商	华泰联合证券有限责任公司
发行人律师	北京市天元律师事务所	其他承销机构	无
审计机构	毕马威华振会计师事务所（特殊普通合伙）	评估机构	蓝策亚洲（北京）资产评估有限公司

### 二、本次发行概况

（一）本次发行的基本情况	
股票种类	人民币普通股（A股）
每股面值	人民币1.00元
发行股数及占发行后总股本的比例	本次公开发行股票数量为1,070万股，占发行后公司总股本的25%；本次发行全部为新股发行，不涉及原股东公开发售股份
每股发行价格	36.94元
发行人高管、员工参与战略配售情况	发行人高级管理人员与核心员工专项资产管理计划为华泰海天瑞声家园1号科创板员工持股集合资产管理计划，发行人高级管理人员与核心员工专项资产管理计划参与战略配售的数量为1,070,000股，为本次公开发行规模的10.00%。华泰海天瑞声家园1号科创板员工持股集合资产管理计划获配股票的限售期为12个月，限售期自本次公开发行的股票在上交所上市之日起开始计算
保荐人相关子公司参与战略配售情况	保荐机构安排实际控制保荐机构的证券公司依法设立的相关子公司华泰创新投资有限公司参与本次发行战略配售，华泰创新投资有限公司最终跟投比例为本次公开发行股票数量的5.00%，

	即最终跟投数量为 535,000 股。华泰创新投资有限公司获配股票的限售期为 24 个月，限售期自本次公开发行的股票在上交所上市之日起开始计算		
发行市盈率	21.65 倍（发行价格除以发行后每股收益）		
发行前每股净资产	13.64 元/股（以 2020 年 12 月 31 日经审计的归属于母公司股东所有者权益除以本次发行前总股本计算）	发行前每股收益	2.27 元/股（以 2020 年度经审计扣除非经常性损益前后孰低的归属于公司普通股股东的净利润除以本次发行前总股本计算）
发行后每股净资产	18.09 元/股（以 2020 年 12 月 31 日经审计的归属于母公司股东所有者权益加上本次募集资金净额之和除以本次发行后总股本计算）	发行后每股收益	1.71 元/股（以 2020 年度经审计扣除非经常性损益前后孰低的归属于公司普通股股东的净利润除以本次发行后总股本计算）
发行市净率	2.04 倍（发行价格除以发行后每股净资产）		
发行方式	本次发行采用向战略投资者定向配售、网下向符合条件的投资者询价配售和网上向持有上海市场非限售 A 股股份和非限售存托凭证市值的社会公众投资者定价发行相结合的方式		
发行对象	符合资格的战略投资者、询价对象以及已开立上海证券交易所股票账户并开通科创板交易的境内自然人、法人等科创板市场投资者，但法律、法规及上海证券交易所业务规则等禁止参与者除外		
承销方式	余额包销		
拟公开发售股份股东名称	无		
募集资金总额	39,525.80 万元		
募集资金净额	33,635.69 万元		
募集资金投资项目	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以及补充流动资金		
发行费用概算	<p>本次发行费用总额 5,890.11 万元，其中：保荐费用 245.75 万元，承销费用 3,655.00 万元；审计及验资费用：733.96 万元；律师费用：716.28 万元；信息披露费用：465.09 万元；发行手续费用及其他 74.03 万元。</p> <p>注：发行费用为不含税金额，发行手续费用及其他较招股意向书的差异系包含了印花税及登记费用减免所致，除前述调整外，发行费用不存在其他调整情况。合计数与各分项数值之和尾数存在微小差异，为四舍五入造成。</p>		
<b>(二) 本次发行上市的重要日期</b>			
刊登初步询价公告日期	2021 年 7 月 26 日		

刊登发行公告日期	2021年8月2日
申购日期	2021年8月3日
缴款日期	2021年8月5日
股票上市日期	本次股票发行结束后将尽快申请在上海证券交易所科创板上市

### 三、发行人主要财务数据和财务指标

项目\年度	2020年12月31日 /2020年度	2019年12月31日 /2019年度	2018年12月31日 /2018年度
资产总额(万元)	47,735.00	40,453.94	21,232.77
归属于母公司所有者 权益(万元)	43,795.64	35,595.14	16,613.31
资产负债率(母公司)	7.92%	12.17%	21.98%
营业收入(万元)	23,337.40	23,755.81	19,265.77
净利润(万元)	8,208.10	8,158.68	6,713.07
归属于母公司所有者 的净利润(万元)	8,208.10	8,158.68	6,713.07
扣除非经常性损益后 归属于母公司所有者 的净利润(万元)	7,301.54	7,624.66	6,205.21
基本每股收益(元/股)	2.27	2.54	2.07
稀释每股收益(元/股)	2.27	2.54	2.07
加权平均净资产收益 率	18.39%	37.17%	46.82%
经营活动产生的现金 流量净额(万元)	5,117.67	8,336.33	6,427.80
现金分红(万元)	-	2,200.00	-
研发投入占营业收入 的比例	18.64%	17.55%	14.19%

注：指标计算公式如下：

资产负债率=总负债/总资产

基本每股收益、稀释每股收益和净资产收益率，依据扣除非经常性损益后的净利润计算，按照《公开发行证券的公司信息披露编报规则第9号——净资产收益率和每股收益的计算及披露》的要求进行计算。

### 四、发行人主营业务经营情况概述

海天瑞声是我国领先的训练数据专业提供商，通过设计训练数据集结构、执行数据加工处理过程，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。

自2005年成立以来，发行人始终致力于为AI产业链上的各类机构提供AI

算法模型开发训练所需的专业数据集。发行人所提供的训练数据覆盖智能语音（语音识别、语音合成等）、计算机视觉、自然语言等多个 AI 核心领域，全面服务于人机交互、智能驾驶、智慧城市等多种创新应用场景。其中智能语音为发行人的优势领域，服务/产品线已覆盖 160 余个主要语种及方言。

海天瑞声的产品和服务已获得阿里巴巴、腾讯、百度、科大讯飞、海康威视、微软、亚马逊、三星等国内外优质客户的认可，并应用于个人助手、智能音箱、语音导航、搜索服务、智能驾驶、机器翻译等多种产品相关的算法模型的训练过程中。目前，海天瑞声客户累计数量已超 500 家，覆盖了多家大型科技公司、人工智能企业及科研机构。

海天瑞声连续荣获“国家高新技术企业”及“中关村高新技术企业”称号，并于 2018 年被评为国家规划布局内重点软件企业。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。2017 年，发行人被中国信息协会、信息化观察网评选为“年度人工智能数据服务领域领军企业”；2018 年，中国语音产业联盟向发行人颁发了“智能语音创新大赛行业先锋奖”，中国信息协会、信息化和软件服务网向发行人颁发了“2018 年人工智能行业最佳产品”奖项；2020 年，发行人获评工业和信息化部国家专精特新“小巨人”企业；2021 年，发行人获评工业和信息化部“新一代人工智能产业创新重点任务揭榜优胜单位”，技术实力得到行业、主管机构的高度认可。

海天瑞声的主营业务包括训练数据定制服务、训练数据产品及训练数据相关的应用服务三大板块。报告期内，发行人的主营业务收入构成情况如下：

单位：万元

类别	2020 年		2019 年		2018 年	
	金额	占营业收入比重	金额	占营业收入比重	金额	占营业收入比重
训练数据定制服务	11,729.91	50.26%	14,232.91	59.91%	12,369.55	64.20%
训练数据产品	10,613.71	45.48%	9,176.47	38.63%	6,601.67	34.27%
训练数据相关的应用服务	993.77	4.26%	346.44	1.46%	294.55	1.53%
<b>主营业务收入</b>	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>

营业收入	23,337.40	100.00%	23,755.81	100.00%	19,265.77	100.00%
------	-----------	---------	-----------	---------	-----------	---------

## 五、发行人核心技术先进性及未来发展战略

### （一）发行人的核心技术先进性

发行人主要从事训练数据的研发设计、生产及销售业务，为人工智能产业链上的各类机构提供训练数据定制服务、训练数据产品和训练数据相关的应用服务。在多年的行业实践过程中，发行人始终注重充分挖掘和培养自身的核心技术优势，秉承基础研究与实际应用紧密结合的原则，自主开发了一体化数据处理平台，在基础研究、平台工具、训练数据生产三个维度均积累下核心技术，有机搭建起发行人的训练数据开发一站式解决方案，提升开发效率、服务质量及数据安全性，为下游客户提供高质量的训练数据产品、高效率的训练数据定制服务及高水准的训练数据相关应用服务。

海天瑞声拥有对人工智能核心算法的理解能力、前瞻性的专业数据集设计能力、丰富的语言覆盖能力及场景采集能力，以及 AI 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：公司已掌握 12 项核心技术，通过算法与数据处理技术并用，可以满足 AI 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 2020 年 12 月 31 日，公司已积累 760 余个自有知识产权的训练数据产品，可提供超过 160 余个语种/方言的训练数据。

2、工具和平台共建：公司自主开发了一体化数据处理平台，融入项目流程管理、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。截至本招股说明书签署日，公司的产品/服务已覆盖 160 余个语种/方言，已积累下超过 100 个语种/方言的发音词典，累计词条数超过 1,000 万条，可构建高质量的智能语音训练数据。

截至目前，公司已经取得 24 项发明专利、1 项实用新型专利授权及 1 项外观设计专利授权，134 项计算机软件著作权，另有 2 项专利已提交申请。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 AI 算法模型的改善、优化效果。

## （二）发行人未来发展战略

公司希望通过不断拓宽训练数据产品/服务的应用领域覆盖、扩展服务链跨度并升级服务质量，使得公司的训练数据产品/服务的发展能够匹配算法、算力发展进度，共同推动下游产业发展。

人工智能行业在未来仍将保持快速增长的趋势，随着 AI 产业落地成为主旋律、行业整体竞争愈发激烈，需求方对训练数据质量的要求不断提高，垂直场景的定制化训练数据需求成为主流，需求方市场对训练数据的要求逐渐向精细化转型。公司将会：

1、进一步扩大在训练数据领域的优势，持续提升数据采集、标注业务能力。在智能语音领域，公司将加强对“一带一路”覆盖的西亚、南亚、中东欧，及非洲语言等有文字的语言的训练数据的覆盖能力，并重点、持续支持下游客户在语言维度细化、多语种拓展方面的发展；在计算机视觉和自然语言领域，公司计划覆盖更多细分领域，并提升公司训练产品/服务对应用场景的覆盖能力；

2、将核心技术和研发生产流程体系应用延伸到行业训练数据的生产、研发中，提升公司处理垂直行业训练数据的能力，赋能人工智能技术与更多行业实体经济的深度融合；

3、紧跟下游需求变化，加深对人工智能基础技术的理解。随着人工智能算法技术逐渐成熟、广泛应用，公司将与下游算法公司合作，通过训练数据服务赋能算法、应用的落地，实现算法与数据技术的结合，实现 AI 应用领域的新拓展。

## 六、发行人选择的具体上市标准

根据毕马威华振事务所出具的《审计报告》(毕马威华振审字第 2103527 号)，发行人最近两年净利润均为正且累计净利润不低于人民币 5,000 万元。同时，考虑 A 股行业分类与发行人相同的企业在境内市场的估值情况以及发行人 2019 年 12 月份融资估值情况（增资对应发行人投后估值金额为 19.90 亿元），预计发行人发行后市值不低于人民币 10 亿元。因此，根据《上海证券交易所科创板股票

发行上市审核规则》第二十二條，发行人选择的具体上市标准为“（一）预计市值不低于人民币 10 亿元，最近两年净利润均为正且累计净利润不低于人民币 5,000 万元”。

## 七、发行人公司治理特殊安排等重要事项

截至本招股说明书签署日，发行人不存在公司治理特殊安排等重要事项。

## 八、募集资金用途概述

经第一届董事会第二十三次会议和 2020 年第二次临时股东大会审议，发行人本次拟公开发行不超过人民币普通股 1,070 万股。本次募集资金将严格执行《募集资金管理制度》的相关规定，存放于董事会确定的专项账户集中管理，并与保荐机构、存放募集资金的商业银行签订募集资金专户存储三方监管协议。公司本次募集资金运用围绕主营业务进行，扣除发行费用后，将按轻重缓急顺序投入以下项目：

单位：万元

序号	项目名称	总投资金额	拟投入募集资金金额
1	自主研发数据产品扩建项目	24,921.18	24,921.18
2	一体化数据处理技术支撑平台升级项目	25,057.72	25,057.72
3	研发中心升级建设项目	16,614.98	16,614.98
4	补充流动资金	10,000.00	10,000.00
合计		<b>76,593.88</b>	<b>76,593.88</b>

如本次募集资金不能满足项目资金的需求，公司将以自有资金、自筹资金等途径自行解决资金缺口，从而保证项目的顺利实施；如果本次募集资金超过项目资金的需要，则超募资金将用于其他与主营业务相关的业务。为充分抓住市场机遇，本次发行的募集资金到位之前，若因市场竞争或公司自身经营需要等因素导致部分投资项目必须进行先期投入的，公司可使用自有资金或者自筹资金先行投入，在募集资金到位之后予以置换。



### 第三节 本次发行概况

#### 一、本次发行的基本情况

股票种类	人民币普通股（A股）
每股面值	人民币 1.00 元
发行股数	本次公开发行股票数量1,070万股，占发行后公司总股本的25%。本次发行全部为新股发行，不涉及原股东公开发售股份。
发行后总股本	42,800,000股
每股发行价格	36.94 元
发行人高管、员工参与战略配售情况	发行人高级管理人员与核心员工专项资产管理计划为华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划，发行人高级管理人员与核心员工专项资产管理计划参与战略配售的数量为 1,070,000 股，为本次公开发行规模的 10.00%。华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划获配股票的限售期为 12 个月，限售期自本次公开发行的股票在上交所上市之日起开始计算
保荐人相关子公司参与战略配售情况	保荐机构安排实际控制保荐机构的证券公司依法设立的相关子公司华泰创新投资有限公司参与本次发行战略配售，华泰创新投资有限公司最终跟投比例为本次公开发行数量的 5.00%，即最终跟投数量为 535,000 股。华泰创新投资有限公司获配股票的限售期为 24 个月，限售期自本次公开发行的股票在上交所上市之日起开始计算
发行市盈率	21.65 倍（按发行价格除以每股收益计算，其中每股收益按照 2020 年经审计的扣除非经常性损益前后孰低的归属于公司普通股股东的净利润除以本次发行后总股本计算）
预测净利润及发行后每股收益	不适用
发行前每股净资产	13.64 元/股（根据 2020 年 12 月 31 日经审计的归属于母公司股东所有者权益除以本次发行前总股本计算）
发行后每股净资产	18.09 元/股（根据 2020 年 12 月 31 日经审计的归属于母公司股东所有者权益加上本次募集资金净额之和除以本次发行后总股本计算）
发行市净率	2.04 倍（每股发行价格除以发行后每股净资产）
发行方式	本次发行采用向战略投资者定向配售、网下向符合条件的投资者询价配售和网上向持有上海市场非限售 A 股股份和非限售存托凭证市值的社会公众投资者定价发行相结合的方式
发行对象	符合资格的战略投资者、询价对象以及已开立上海证券交易所股票账户并开通科创板交易的境内自然人、法人等科创板市场投资者，但法律、法规及上海证券交易所业务规则等禁止参与者除外
承销方式	由主承销商余额包销的方式承销
发行费用概算合计	5,890.11 万元

其中：保荐、承销费用	其中：保荐费用 245.75 万元，承销费用 3,655.00 万元
审计及验资费用	733.96 万元
律师费用	716.28 万元
用于本次发行的信息披露费用	465.09 万元
发行手续费用及其他	74.03 万元
注	发行费用为不含税金额，发行手续费用及其他较招股意向书的差异系包含了印花税及登记费用减免所致，除前述调整外，发行费用不存在其他调整情况。合计数与各分项数值之和尾数存在微小差异，为四舍五入造成。

## 二、与本次发行有关的机构

### （一）保荐人、主承销商

**华泰联合证券有限责任公司**

法定代表人：江禹

住 所：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

办公地址：北京市西城区丰盛胡同 22 号丰铭国际大厦 A 座 6 层

联系电话：010-56839300

传 真：010-56839400

保荐代表人：张鹏、葛青

项目协办人：廖蔚铭

项目经办人：杨阳、倪佳伟、答一丹

### （二）发行人律师

**北京市天元律师事务所**

负责人：朱小辉

住 所：北京市西城区丰盛胡同 28 号太平洋保险大厦 10 层

联系电话：010-57763888

传 真：010-57763777

经办律师：周世君、王韶华、顾鼎鼎

### （三）会计师事务所

**毕马威华振会计师事务所（特殊普通合伙）**

负责人：邹俊

住 所：北京市东城区东长安街1号毕马威大楼8层

联系电话：010-85085000

传 真：010-85185111

经办会计师：雷江、张杨

### （四）资产评估机构

**蓝策亚洲（北京）资产评估有限公司**

法定代表人：罗俊军

住 所：北京市海淀区西四环中路16号院2号楼3层301-3068室

联系电话：62104306

传 真：62104306

经办注册评估师：毕娇、李鹏举

### （五）股票登记机构

**中国证券登记结算有限责任公司上海分公司**

住 所：上海市浦东新区杨高南路188号

联系电话：021-68870562

传 真：021-68606910

### （六）收款银行

**中国工商银行股份有限公司深圳分行振华支行**

户 名：华泰联合证券有限责任公司

账 户：4000010209200006013

## （七）申请上市证券交易所

### 上海证券交易所

住 所：上海市浦东新区浦东南路 528 号证券大厦

联系电话：021-68808888

传 真：021-68807813

## 三、发行人与本次发行有关的中介机构的关系

截至本招股说明书签署日，发行人与本次发行有关的中介机构及其负责人、高级管理人员及经办人员之间不存在直接或间接的股权关系或其他权益关系。

## 四、本次发行上市的重要日期

1、刊登初步询价公告日期：2021 年 7 月 26 日

2、刊登发行公告日期：2021 年 8 月 2 日

3、申购日期：2021 年 8 月 3 日

4、缴款日期：2021 年 8 月 5 日

5、股票上市日期：本次股票发行结束后将尽快申请在上海证券交易所科创板上市

## 五、本次发行战略配售情况

本次发行的战略配售由保荐机构相关子公司跟投以及发行人的高级管理人员与核心员工专项资产管理计划组成，跟投机构为华泰创新投资有限公司，发行人高级管理人员与核心员工专项资产管理计划为华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划。

本次保荐机构相关子公司跟投的初始股份数量为本次公开发行股份的 5.00%，即 535,000 股；发行人高级管理人员与核心员工专项资产管理计划参与战略配售的数量为不超过本次公开发行规模的 10.00%，即 1,070,000 股，同时包含新股配售经纪佣金的总投资规模不超过 4,780 万元；战略投资者最终配售数量与初始配售数量的差额部分回拨至网下发行。

参与本次战略配售的投资者均已与发行人签署战略配售协议，不参加本次发行初步询价，并承诺按照发行人和保荐机构（主承销商）最终确定的发行价格认购其承诺认购的股票数量。

本次发行战略配售的最终情况如下：

投资者全称	获配数量 (股)	获配金额 (元, 不含佣金)	新股配售经 纪佣金 (元)	限售期限 (月)
华泰创新投资有限公司	535,000	19,762,900.00	0.00	24
华泰海天瑞声家园1号科创板 员工持股集合资产管理计划	1,070,000	39,525,800.00	197,629.00	12
合计	<b>1,605,000</b>	<b>59,288,700.00</b>	<b>197,629.00</b>	-

## 六、发行人高级管理人员、核心人员参与战略配售情况

发行人第一届董事会第三十二次会议审议通过《关于同意部分高级管理人员、核心员工设立专项资管计划参与北京海天瑞声科技股份有限公司首次公开发行人民币普通股股票并在科创板上市战略配售的议案》，同意发行人部分高级管理人员和核心员工设立专项资产管理计划参与发行人战略配售。公司高级管理人员与核心员工设立资管计划参与公司首次公开发行股票并在科创板上市战略配售已依据相关规则履行必要决策程序。具体情况如下：

具体名称：华泰海天瑞声家园1号科创板员工持股集合资产管理计划

设立时间：2021年3月1日

备案日期：2021年3月10日

备案编码：SQC329

募集资金规模：4,780万元（不含孳生利息）

管理人：华泰证券（上海）资产管理有限公司

托管人：招商银行股份有限公司南京分行

实际支配主体：华泰证券（上海）资产管理有限公司。实际支配主体非发行人高级管理人员

参与人姓名、职务与比例：

序号	姓名	职务	参与比例	实际缴款金额 (万元)
1	贺琳	董事长	47.80%	2,285.00
2	唐涤飞	董事、总经理	5.44%	260.00
3	吕思遥	董事、副总经理、财务负责人、董事会秘书	6.28%	300.00
4	李科	董事、副总经理	14.02%	670.00
5	黄宇凯	董事、副总经理、技术总监	4.60%	220.00
6	熊敏	监事、人资行政总监	4.71%	225.00
7	张蕊	全资子公司中瑞智销售总监	3.35%	160.00
8	王超	运营财务总监	3.35%	160.00
9	杨韩	核算财务经理	3.77%	180.00
10	廖晓玲	语音合成业务部总监	3.35%	160.00
11	辛晓峰	自然语言处理业务部总监	3.35%	160.00
合计			<b>100.00%</b>	<b>4,780.00</b>

注：上表中合计数与各明细数直接相加之和在尾数上如有差异，是由于四舍五入所造成。

华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划的份额持有人均为发行人的高级管理人员和核心员工，对于发行人生产经营具有重要影响，该资管计划中参与本次发行战略配售的人员均已与发行人或其全资子公司北京中瑞智科技有限公司签订了劳动合同。

高管和核心员工持股计划（华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划）最终实际认购数量为 1,070,000 股，为本次发行股票数量的 10.00%。华泰海天瑞声家园 1 号科创板员工持股集合资产管理计划本次获配股票的限售期为 12 个月，限售期自本次公开发行的股票在上交所上市之日起开始计算。

## 七、保荐机构关联公司参与战略配售情况

保荐机构华泰联合证券通过实际控制其的华泰证券股份有限公司依法设立的另类投资子公司华泰创新投资有限公司参与本次发行战略配售，华泰创新投资有限公司最终跟投比例为本次公开发行数量的 5.00%，最终跟投数量为 535,000 股。华泰创新投资有限公司本次跟投获配股票的限售期为 24 个月，限售期自本次公开发行的股票在上交所上市之日起开始计算。

## 第四节 风险因素

投资者在评价本公司此次发售的股票时，除本招股说明书提供的其他各项资料外，应特别认真地考虑下述各项风险因素。下列风险是根据重要性原则或可能影响投资者投资决策程度大小排序，但该排序并不表示风险因素会依次发生。

### 一、市场竞争加剧的风险

发行人所在的细分领域为人工智能基础数据服务领域。近年来，参与该行业的公司不断增多，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。

### 二、收入主要来自老客户，新客户收入占比较低的风险

报告期内，发行人收入主要来自老客户，新客户收入占比较低。2018年、2019年及2020年，发行人来自老客户的收入占比分别达到88.58%、89.09%及93.47%。发行人正致力于维护老客户、拓展新客户，若老客户降低对公司产品、服务的采购量或公司未能维持与老客户的合作关系，则公司的经营情况将受到不利影响；若新客户拓展情况不达预期，则公司的经营情况也将受到不利影响。

### 三、专利被申请宣告无效的风险

截至本招股说明书签署日，根据发行人收到的由专利代理机构转送的国家知识产权局专利局出具的《无效宣告请求受理通知书》，发行人及其子公司已取得授权的9项发明专利和1项实用新型专利被提出无效宣告请求，涉及无效宣告请求的相关专利情况参见本招股说明书“第六节 业务和技术/五、发行人的主要固定资产和无形资产/（二）主要无形资产/1、专利”。截至2021年5月7日，发行人已经收到上述专利中7项发明专利和1项实用新型专利的无效宣告请求审查决定书，其中7项发明专利维持专利权有效，1项实用新型专利专利权维持部分有效，其余2项专利的无效宣告请求审查事项尚处于正常审理阶段，国家知识产权局专利局尚未作出裁定。

尽管剩余 2 项涉及无效宣告请求的相关专利权被宣告无效的可能性较小，但不排除发行人的相关专利权被部分或者全部宣告无效的可能，若公司知识产权被宣告无效，公司被宣告无效的专利或其权利要求中公开的技术点可能存在被竞争对手或第三方模仿的风险，若相关方模仿上述技术从事与发行人同类型业务，参与市场竞争，可能会对发行人的业务经营和财务状况产生不利影响。

#### 四、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，发行人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。如果发行人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，则可能导致发行人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，进而对发行人的客户拓展、产品创新和经营业绩造成不利影响。

#### 五、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，如果发行人无法持续加强核心技术人员的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。

#### 六、核心技术失密风险

发行人数据产品和服务均以核心技术为基础。公司制定了严格的保密制度并严格执行，但上述措施仍无法完全避免公司核心技术的失密风险。如果公司相关核心技术的内控和保密机制不能得到有效执行，则可能导致公司核心技术失密的风险。

#### 七、业务发展受下游人工智能领域发展状况影响较大的风险

发行人主要为 AI 产业链上的各类机构提供算法模型开发所需的训练数据，服务于下游人工智能行业发展。发行人的经营情况、业务发展与下游人工智能领域的发展状况息息相关。若未来人工智能领域景气度下降，或因技术更迭使得对训练数据的市场需求发生变动，将对公司业绩产生较大影响。



## 八、新业务拓展风险

随着人工智能行业持续快速地增长，下游客户对高质量、结构化、大规模训练数据的需求将持续增加。公司能否紧跟下游需求变化，加深对下游行业基础技术的理解并开发出更多的业务机会，一方面受到行业发展状况、市场需求变化以及市场竞争状况等因素影响，另一方面也取决于发行人对于下游众多新领域新场景的理解认知以及合作拓展能力等诸多因素。因此，发行人新业务的开拓可能不及预期或者遇到其他不利因素，进而对发行人未来的经营业绩产生不利影响。

## 九、训练数据产品无法实现授权销售的风险

发行人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，发行人开发训练数据产品，承担开发费用，并拥有训练数据产品的知识产权。如果公司开发的训练数据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，进而对发行人未来的经营业绩产生不利影响。

## 十、经营业绩受税收优惠政策影响较大的风险

报告期内，公司享受了增值税即征即退、高新技术企业和国家规划布局内的重点软件企业所得税税率优惠、研发费用加计扣除等一系列税收优惠，报告期各期税收优惠影响金额占利润总额的比例分别为 28.84%、22.96%和 24.94%，经营业绩受税收优惠政策影响较大。如果未来公司所享受的税收优惠政策出现不利变化，或者公司不再符合税收优惠的条件，则公司将无法持续享受税收优惠或存在所享受的税收优惠减少的可能性，并进而对公司未来经营业绩和现金流水平造成不利影响。

## 十一、采购成本上升风险

报告期内，发行人进行采购原料数据采集、标注所支付的数据服务费占发行人各年采购总额的比重较高，分别达到 81.22%、79.84%和 84.69%。随着城乡生活成本的持续提升，前述采购价格可能存在持续上升的趋势，若发行人的技术发展带来的效率提升无法抵消前述采购成本上升的影响，可能会对公司的经营业绩

带来一定不利影响。

## 十二、汇率波动风险

海天瑞声作为一家面向全球的训练数据服务商，拥有数量众多的境外客户，遍布美国、韩国、日本等各地区，报告期内公司境外收入占比分别为 31.00%、49.60%和 35.86%，该等收入使用外币进行结算，并受人民币汇率水平变化的影响。人民币汇率可能受到国内外政治、经济环境等因素的影响，存在波动风险，进而对公司的经营业绩产生不利影响。

## 十三、净资产收益率下降的风险

报告期内，公司加权平均净资产收益率（按扣除非经常性损益后归属于公司普通股股东的净利润计算）分别为 46.82%、37.17%和 18.39%。2019 年及 2020 年公司加权平均净资产收益率的下降主要是由于公司进行了外部融资活动导致公司净资产增加所致。

本次发行后公司的净资产将有所提高，由于募集资金投资项目有一定的实施周期，募集资金产生经济效益需要一段时间，短期内净利润较难以实现与净资产的同步增长，因而公司在短期内净资产收益率将有所下降。

## 十四、数据安全相关风险

发行人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，随着发行人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，与数据安全、个人信息保护相关的法律规章体系逐步完善。如果将来公司未能根据法律规章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或可能产生诉讼纠纷，进而给公司的经营带来不利影响。

## 十五、实际控制人控制不当的风险

本次发行前，贺琳直接持有发行人 866.97 万股股份（占发行前总股本的 27.01%），并通过控制中瑞安间接控制发行人 495.41 万股股份（占发行前总股本

的 15.43%)，合计控制公司 42.44% 的股份，为公司实际控制人。

如果实际控制人利用其控制地位，通过行使表决权或其他方式对公司财务管理、人事、发展战略、经营决策等重大事项施加不利影响，其他股东的利益可能受到损害。

## 十六、中美贸易摩擦风险

公司产品和服务的主要出口地包括美国、日韩等国家或地区。2018 年至 2020 年，公司出口美国的营业收入占当期公司营业收入的比重分别达到 12.96%、19.48% 和 19.93%。公司相关境外业务始终严格遵守中国及他国法律。但国际局势瞬息万变，近年来中美贸易摩擦不断，如果中美双方出台新的贸易保护主义措施，可能会影响美国市场相关客户将数据提供给公司用于标注，减少与公司进行训练数据方面的业务合作，公司相关业务可能会受到约束，将会对公司的生产经营产生不利影响。

## 十七、疫情风险

2020 年 1 月，新型冠状病毒肺炎疫情爆发，致使全国各行各业均遭受了不同程度的影响，并已在全球范围内蔓延。因新冠疫情爆发期间实施隔离、交通管制等防疫管控措施，对发行人的采购、销售等环节造成了不利影响。由于疫情导致的延期复工，公司及主要客户、主要供应商的生产经营均受到一定程度的影响，公司采购原料数据采集、标注服务以及下游客户对公司产品服务的验收周期均受到延期复工影响，相比正常进度有所延后。对于公司而言，本项风险属于不可抗力风险，若本次新冠疫情在较长时间内持续，将可能对公司的经营造成不利影响。

## 第五节 发行人基本情况

### 一、发行人基本情况

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.
注册资本	3,210.00 万元
法定代表人	贺琳
有限公司成立日期	2005 年 5 月 11 日
整体变更为股份公司日期	2017 年 11 月 24 日
住所	北京市海淀区成府路 28 号 4-801
邮政编码	100083
联系电话	010-62660772
传真	010-62660892
互联网网址	<a href="http://www.speechocean.com/">http://www.speechocean.com/</a>
电子信箱	ir@speechocean.com
负责信息披露和投资者关系的部门、负责人和电话号码	董事会办公室 负责人：吕思遥，电话号码：010-62660772

### 二、发行人的设立情况

#### （一）有限责任公司设立情况

发行人由前身海天瑞声有限整体变更设立。2005 年 5 月 9 日，贺琳和俞舸签署《北京海天瑞声科技有限公司章程》，共同出资设立海天瑞声有限。根据该章程，海天瑞声有限注册资本为 30 万元，其中，贺琳以货币出资 27 万元、俞舸以货币出资 3 万元。

根据北京市工商局 2004 年 2 月颁布实施的《北京市工商局改革市场准入制度优化经济发展环境若干意见》的规定，贺琳、俞舸分别于 2005 年 5 月 9 日向海天瑞声有限于中国农业银行北京市海淀区支行开立的企业入资专用账户缴纳 27 万元、3 万元，并向北京市工商局提供了中国农业银行北京市分行出具的《交存入资资金凭证》。2005 年 5 月 11 日，北京市工商局向海天瑞声有限核发《企业法人营业执照》（注册号：1101082832381），海天瑞声有限成立。

根据中兴华会计师事务所（特殊普通合伙）出具的《北京海天瑞声科技股份有限公司出资复核报告》（中兴华核字[2020] 第 010114 号），经审验，截至 2005

年 5 月 9 日，贺琳缴纳出资款 27 万元，俞舸缴纳出资款 3 万元。

## （二）股份有限公司设立情况

发行人系海天瑞声有限以 2017 年 8 月 31 日为基准日，以经审计净资产折股，整体变更设立的股份有限公司。

2017 年 11 月 6 日，瑞华会计师事务所（特殊普通合伙）出具《审计报告》（瑞华审字[2017]02280059 号），截至 2017 年 8 月 31 日，海天瑞声有限的净资产值为 8,810.95 万元。根据中兴华会计师事务所（特殊普通合伙）出具的《审计报告专项复核报告》（中兴华核字（2020）第 010135 号），海天瑞声有限截至 2017 年 8 月 31 日净资产情况与上述情况一致。

2017 年 11 月 6 日亚洲（北京）资产评估有限公司（后更名为蓝策亚洲（北京）资产评估有限公司）出具《评估报告》（京亚评报字[2017]第 110 号），以 2017 年 8 月 31 日为评估基准日，按资产基础法海天瑞声净资产的评估值为 9,498.99 万元。

2017 年 11 月 9 日，海天瑞声有限通过股东会决议，同意公司以截至 2017 年 8 月 31 日经审计净资产 8,810.95 万元按照 1:0.3405 的比例折合 3,000 万股，整体变更为股份有限公司。

根据中兴华会计师事务所（特殊普通合伙）出具《验资报告》（中兴华验字（2020）第 010005 号），公司之全体发起人已按发起人协议、章程之规定，以其拥有的有限公司截止 2017 年 8 月 31 日经审计净资产 88,109,473.40 元，作价 88,109,473.40 元，其中 3,000 万元折合为公司的股本，股份总额为 3,000 万股，每股面值 1 元，缴纳注册资本 3,000 万元整，余额 58,109,473.40 元计入资本公积。

2017 年 11 月 24 日，北京市工商局海淀分局向发行人颁发了《营业执照》（统一社会信用代码：91110108776388751R）。

## （三）发行人 2017 年以来的股本和股东变化情况

### 1、2017 年 5 月，股权转让及增资

根据 2017 年 4 月 20 日海天瑞声有限召开的股东会决议、转让各方签署的股权转让协议以及投资协议，中瑞立将其持有的 6.07 万元出资额作价 454.905 万元

转让给上海丰琬，中瑞立将其持有的 10.13 万元出资额作价 760.095 万元转让给上海兴富，杭州银杏数以 1,300.00 万元的价格认缴公司新增注册资本 17.33 万元，杭州士兰以 1,000.00 万元的价格认缴公司新增注册资本 13.33 万元，天津金星以 1,530.095 万元的价格认缴公司新增注册资本 20.40 万元，上海兴富以 219.905 万元的价格认缴公司新增注册资本 2.94 万元。

2017 年 7 月 19 日，北京东审会计师事务所（普通合伙）出具了东审字[2017] 04-228 号《验资报告》，对上述增资情况进行了审验。根据中兴华会计师事务所（特殊普通合伙）出具的《北京海天瑞声科技股份有限公司出资复核报告》（中兴华核字[2020] 第 010115 号），北京东审会计师事务所（普通合伙）出具的东审字[2017] 04-228 号《验资报告》审验结果无误。

2017 年 5 月 31 日，海天瑞声有限完成本次工商变更并领取了新的营业执照。本次股权转让及增资完成后，海天瑞声有限股权结构如下：

序号	股东	出资额（万元）	持股比例（%）	出资方式
1	贺琳	189.00	28.90	货币
2	中瑞安	108.00	16.51	货币
3	清德投资	82.86	12.67	货币
4	唐涤飞	78.00	11.93	货币
5	上海丰琬	61.21	9.36	货币
6	上海兴富	43.07	6.59	货币
7	中瑞立	40.80	6.24	货币
8	天津金星	20.40	3.12	货币
9	杭州银杏数	17.33	2.65	货币
10	杭州士兰	13.33	2.04	货币
合计		<b>654.00</b>	<b>100.00</b>	-

## 2、2017 年 11 月，整体变更为股份有限公司

根据 2017 年 11 月 9 日海天瑞声有限的股东会决议，海天瑞声有限拟整体变更为股份有限公司。同日，原有限公司股东作为拟变更设立的股份有限公司的发起人，签订《发起人协议》并召开创立大会共同发起设立股份有限公司，整体变更后的公司名称为“北京海天瑞声科技股份有限公司”。

2017 年 11 月 6 日，瑞华会计师事务所（特殊普通合伙）出具《审计报告》

(瑞华审字[2017]02280059号),截至2017年8月31日,海天瑞声有限的净资产值为8,810.95万元。根据中兴华会计师事务所(特殊普通合伙)出具的《审计报告专项复核报告》(中兴华核字(2020)第010135号),海天瑞声有限截至2017年8月31日净资产情况与上述情况一致。

根据亚洲(北京)资产评估有限公司(后更名为蓝策亚洲(北京)资产评估有限公司)于2017年11月6日出具的京亚评报字[2017]第110号《评估报告》,以2017年8月31日为评估基准日,海天瑞声有限净资产的评估值为9,498.99万元。

海天瑞声有限以截至2017年8月31日经审计的净资产为基础,按照1:0.3405的比例折合为股本3,000万元,超出股本总额的净资产5,810.95万元计入股份有限公司的资本公积。

根据中兴华会计师事务所(特殊普通合伙)出具《验资报告》(中兴华验字(2020)第010005号),确认截至2017年11月9日止,公司之全体发起人已按发起人协议、章程之规定,以其拥有的有限公司截止2017年8月31日经审计净资产88,109,473.40元,作价88,109,473.40元,其中3,000万元折合为公司的股本,股份总额为3,000万股,每股面值1元,缴纳注册资本3,000万元整,余额58,109,473.40元计入资本公积。

2017年11月24日,北京市工商行政管理局海淀分局核准海天瑞声有限整体变更为股份有限公司,并就本次设立核发了新的营业执照。本次整体变更后,海天瑞声的股东及出资情况如下:

序号	股东	持股数(万股)	持股比例(%)	出资方式
1	贺琳	866.97	28.90	净资产折股
2	中瑞安	495.41	16.51	净资产折股
3	清德投资	380.09	12.67	净资产折股
4	唐涤飞	357.80	11.93	净资产折股
5	上海丰琬	280.78	9.36	净资产折股
6	上海兴富	197.57	6.59	净资产折股
7	中瑞立	187.16	6.24	净资产折股
8	天津金星	93.58	3.12	净资产折股

序号	股东	持股数（万股）	持股比例（%）	出资方式
9	杭州银杏数	79.50	2.65	净资产折股
10	杭州士兰	61.15	2.04	净资产折股
合计		<b>3,000.00</b>	<b>100.00</b>	

### 3、2019年12月股份转让、增加注册资本

2019年12月4日，海天瑞声召开临时股东大会，会议通过决议，同意向中移投资、中网投分别定向发行81万股、129万股股份，每股认购价格为62元；本次定向增发后，发行人注册资本增加至3,210万元，股份总数增加至3,210万股；同意就股份转让及增加注册资本事项相应修改公司章程。

2019年12月26日，中移投资与海天瑞声、贺琳签署《股份认购协议》，约定中移投资以5,022万元的价格认购公司发行的81万股新增股份，每股认购价格为62元，对应新增注册资本81万元。

2019年12月26日，中网投与海天瑞声、贺琳签署《投资协议》，约定中网投以7,998万元的价格认购公司发行的129万股新增股份，每股认购价格为62元，对应新增注册资本129万元。

上述新增入股价格系各方结合中联资产评估集团有限公司出具的评估报告对应的企业评估价值协商确定。

2019年12月26日，海天瑞声、贺琳、清德投资、上海丰琬、上海兴富、杭州银杏数、杭州士兰与中移投资签署《股份转让协议》，杭州银杏数分别与芜湖青和、芜湖博信签署《股份转让协议》，转让价格系各方结合公司的评估价值协商确定，股份转让的具体情况如下：

转让方	受让方	拟转让股份数（股）	对应转股价款（元）	每股转让价格（元）
清德投资	中移投资	1,255,454	76,059,618	60.58
上海丰琬		927,424	56,186,453	60.58
上海兴富		652,576	39,535,216	60.58
杭州士兰		201,970	12,236,000	60.58
		7,576	458,963	60.58
杭州银杏数	芜湖青和	165,062	10,000,000	60.58
	芜湖博信	89,938	5,448,750	60.58



2019年12月19日，中联资产评估集团有限公司出具《中移投资控股有限责任公司拟向北京海天瑞声科技股份有限公司增资项目资产评估报告》（中联评报字[2019]第2231号），经评估，于本次评估基准日2019年6月30日，海天瑞声净资产（股东全部权益）评估值为186,668.45万元。2019年12月26日，上述评估报告已经国务院国有资产监督管理委员会备案。根据中兴华会计师事务所（特殊普通合伙）出具的《验资报告》（中兴华验字（2020）第010006号），上述增资情况经审验无误。2019年12月26日，海天瑞声就本次股份转让及增加注册资本事宜修改了公司章程，办理了工商变更登记，领取了新的营业执照。本次股份转让及增加注册资本完成后，发行人的股本结构如下：

序号	股东	持股数（万股）	持股比例（%）	出资方式
1	贺琳	866.97	27.01	净资产折股
2	中瑞安	495.41	15.43	净资产折股
3	中移投资	304.50	9.49	净资产折股
		81.00	2.52	货币
4	唐涤飞	357.80	11.15	净资产折股
5	清德投资	254.55	7.93	净资产折股
6	上海丰琬	188.04	5.86	净资产折股
7	中瑞立	187.16	5.83	净资产折股
8	上海兴富	132.31	4.12	净资产折股
9	中网投	129.00	4.02	货币
10	天津金星	93.58	2.92	净资产折股
11	杭州银杏数	53.24	1.66	净资产折股
12	杭州士兰	40.95	1.28	净资产折股
13	芜湖青和	16.51	0.51	净资产折股
14	芜湖博信	8.99	0.28	净资产折股
合计		<b>3,210.00</b>	<b>100.00</b>	

上述参与增资的股东中移投资、中网投，参与股权转让的股东芜湖青和和芜湖博信为发行人本次申报前一年新增股东。上述股东的基本情况参见本节“八、发行人股本情况/（四）最近一年发行人新增股东的情况”。上述股东中中移投资入股发行人是为围绕中国移动“5G+”计划，推动双方优势互补、深化合作；中网投入股发行人是因为看好公司及行业发展；芜湖青和和芜湖博信入股发行人是

因为看好公司未来业务发展。经保荐机构和发行人律师核查，上述发行人申报前一年新增股东与发行人其他股东、董事、监事、高级管理人员不存在关联关系，新增股东与本次发行的中介机构及其负责人、高级管理人员、经办人员不存在关联关系，新增股东不存在股份代持情形。

自 2019 年 12 月股份转让、增加注册资本以来至本招股说明书签署日，发行人未再发生股本和股东变动的情形。

### 三、发行人设立以来的重大资产重组情况

自设立以来，发行人不存在重大资产重组情况。

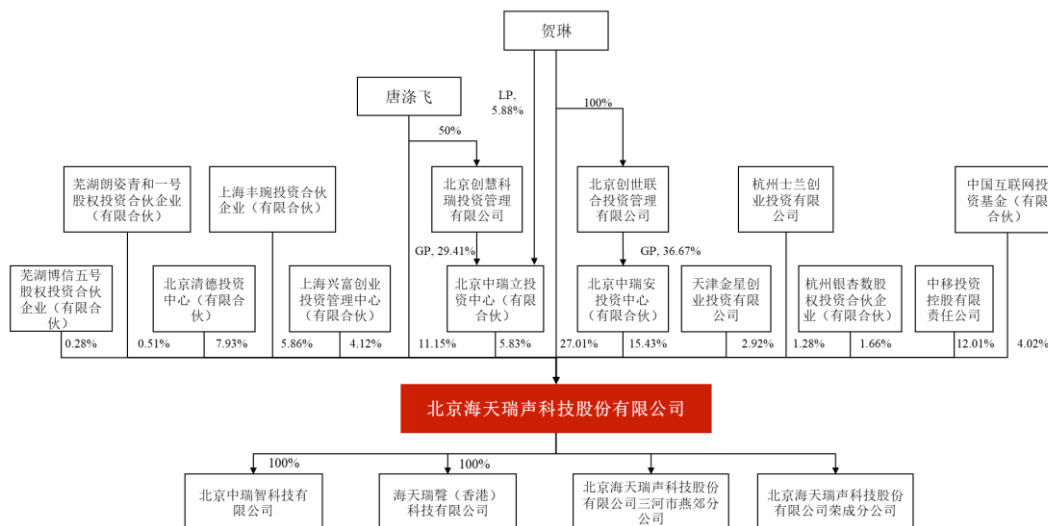
### 四、公司在其他证券市场的上市或挂牌情况

自设立以来，发行人未在其他证券市场上市或挂牌。

### 五、发行人的股权结构及组织架构

#### (一) 股权结构图

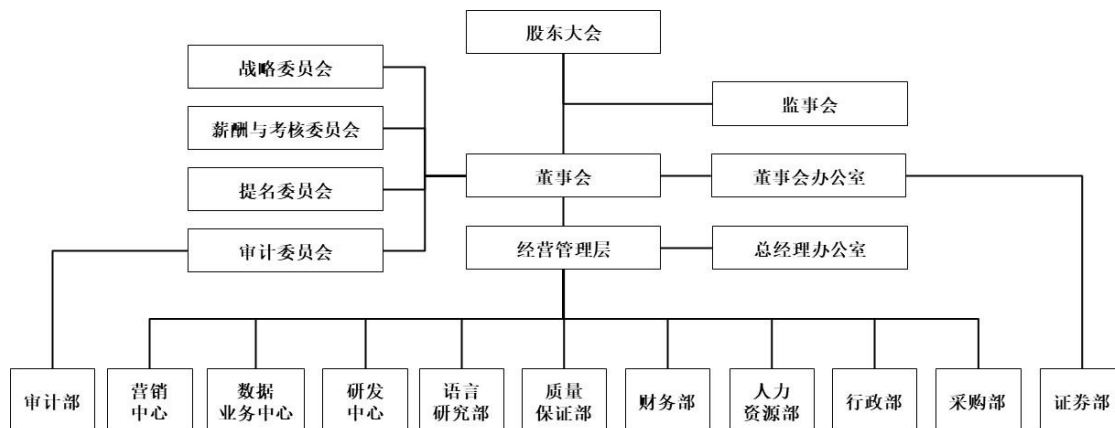
截至本招股说明书签署日，公司股权结构如下所示：



截至本招股说明书签署日，发行人的股权结构如上图所示。发行人共有 14 名股东，其中 2 名为自然人股东，2 名为员工持股平台，10 名为其他机构股东。经穿透合并计算，公司股东数量未超过 200 人。

## （二）组织架构图

截至本招股说明书签署日，公司的组织架构如下所示：



## 六、发行人控股子公司及参股公司情况

### （一）子公司

#### 1、中瑞智

##### （1）基本情况

名称	北京中瑞智科技有限公司	
统一社会信用代码	91110108MA008DL69R	
法定代表人	黄大伟	
注册资本	300 万元	
实收资本	300 万元	
经营范围	技术开发、技术服务、技术咨询、技术转让、技术推广；销售计算机、软件及辅助设备、通讯设备；企业管理咨询；货物进出口；技术进出口；计算机系统服务；软件开发。（企业依法自主选择经营项目，开展经营活动；依法须经批准的项目，经相关部门批准后依批准的内容开展经营活动；不得从事本市产业政策禁止和限制类项目的经营活动。）	
注册地	北京市海淀区知春路 1 号 1 号楼 15 层 1501	
主要生产经营地	北京市海淀区知春路 1 号 1 号楼 15 层 1501	
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	
主营业务	训练数据相关的应用服务	
主营业务与发行人的关系	属于发行人主营业务类型中的“训练数据相关的应用服务”	

成立时间	2016年9月23日
营业期限	2016年9月23日至2046年9月22日

## (2) 最近一年的主要财务数据

单位：万元

项目	2020年12月31日/2020年度
总资产	730.56
净资产	-159.81
净利润	54.26

注：中瑞智已经包含在发行人合并报表的范围内，该合并财务报表已经毕马威华振事务所审计并出具了标准无保留意见的审计报告。

## 2、香港海天瑞声

## (1) 基本情况

名称	海天瑞声（香港）科技有限公司	
公司编号	2311062	
注册地	Level 54, Hopewell Centre, 183 Queen's Road East, Hong Kong	
法定股本	800,000 美元	
已发行股本	800,000 股普通股	
主要生产经营地	香港	
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	
主营业务	数据资源开发	
主营业务与发行人的关系	承担发行人部分境外业务拓展和服务采购	
成立时间	2015年11月24日	

## (2) 最近一年的主要财务数据

单位：万元

项目	2020年12月31日/2020年度
总资产	305.65
净资产	89.98
净利润	-94.26

注：香港海天瑞声已经包含在发行人合并报表的范围内，该合并财务报表已经毕马威华振事务所审计并出具了标准无保留意见的审计报告。

香港海天瑞声设立时投资总额为 50 万美元，北京市商务委员会向发行人颁发了《企业境外投资证书》（境外投资证第 N1100201600082 号）。

2018 年 5 月，发行人向香港海天瑞声增资 30 万美元，增资完成后香港海天瑞声注册资本增加至 80 万美元。2018 年 5 月，北京市商务委员会向发行人颁发了《企业境外投资证书》（境外投资证第 N1100201800230 号）；北京市发展和改革委员会向发行人核发《项目备案通知书》（京发改[2018]1007 号），对发行人增资香港海天瑞声事项进行备案。

## （二）参股公司

截至本招股说明书签署日，发行人无参股公司。

## （三）分公司

截至本招股说明书签署日，发行人 3 家分公司基本情况如下：

名称	北京海天瑞声科技股份有限公司三河市燕郊分公司	北京海天瑞声科技股份有限公司荣成分公司	北京海天瑞声科技股份有限公司杭州分公司
营业场所	三河市燕郊开发区 102 国道南侧、迎宾南路西侧维多利亚 D 座 1805	荣成市海湾南路 86 号	浙江省杭州市萧山区宁围街道民和路 800 号宝盛世纪中心 1 幢中科宝盛科技园 5 层 -5-10 室
负责人	陈思	辛晓峰	白彬
统一社会信用代码	91131082075967275K	91371082MA3F31GH6K	91330109MA2KEJ0D31
经营范围	许可经营项目：无；一般经营项目：技术开发、技术服务、技术咨询、技术转让、技术推广；销售：计算机、软件（未取得行政许可的项目除外）。	技术开发、技术服务、技术咨询、技术转让、技术推广；销售计算机、软件（依法须经批准的项目，经相关部门批准后方可开展经营活动）。	许可项目：技术进出口；货物进出口（依法须经批准的项目，经相关部门批准后方可开展经营活动，具体经营项目以审批结果为准）。一般项目：技术服务、技术开发、技术咨询、技术交流、技术转让、技术推广；计算机软硬件及辅助设备零售；计算机软硬件及辅助设备批发（除依法须经批准的项目外，凭营业执照依法自主开展经营活动）。（在总公司经营范围内从事经营活动）
成立日期	2013 年 8 月 12 日	2017 年 6 月 19 日	2021 年 3 月 16 日

## 七、持有发行人 5% 以上股份的主要股东及实际控制人情况

### （一）控股股东、实际控制人的基本情况

贺琳直接持有发行人 866.97 万股股份（占发行前总股本的 27.01%），并通过控制中瑞安间接控制发行人 495.41 万股股份（占发行前总股本的 15.43%），合计控制公司 42.44% 的股份。因此，贺琳为发行人控股股东、实际控制人。

姓名	贺琳
性别	女
国籍	中国
身份证号	11010819680708****
住所	北京市海淀区中关村甲 13 楼
是否取得其他国家或地区的永久居留权	无

贺琳目前担任公司董事长，简历情况请参见本节“九、董事、监事、高级管理人员与核心技术人员的简要情况/（一）发行人董事、监事、高级管理人员与核心技术人员的简要情况/1、公司董事”。

### （二）其他持有本公司 5% 以上股份的主要股东的基本情况

#### 1、中瑞安

##### （1）基本情况

名称	北京中瑞安投资中心（有限合伙）
统一社会信用代码	91110108MA002G9RX4
类型	有限合伙企业
执行事务合伙人	创世联合
出资总额	1,000 万元
经营场所	北京市海淀区上地十街 1 号院 4 号楼 23 层 2304-1
经营范围/主营业务	投资管理；资产管理。
主营业务与发行人主营业务的关系	无相关性
成立时间	2015 年 12 月 15 日
合伙期限	2015 年 12 月 15 日至长期

##### （2）出资人构成和出资比例

中瑞安为发行人员工持股平台。截至本招股说明书签署日，中瑞安出资人构

成和出资比例如下：

序号	合伙人姓名或名称	在本公司及子公司任职情况	合伙人性质	出资金额（万元）	出资比例（%）
1	创世联合	-	普通合伙人	366.67	36.67
2	李科	董事、副总经理	有限合伙人	305.56	30.56
3	郝玉峰	副总经理、技术总监、首席科学家	有限合伙人	194.44	19.44
4	吕思遥	董事、副总经理、财务负责人、董事会秘书	有限合伙人	83.33	8.33
5	张蕊	子公司中瑞智销售总监	有限合伙人	50.00	5.00
合计				<b>1,000.00</b>	<b>100.00</b>

截至本招股说明书签署日，中瑞安普通合伙人创世联合的股权结构如下：

序号	股东	出资金额（万元）	持股比例（%）
1	贺琳	30.00	100.00
合计		<b>30.00</b>	<b>100.00</b>

中瑞安为发行人的员工持股平台，未以非公开方式向投资者募集资金，未委托他人管理其资产，亦未接受他人委托管理资产，不属于《中华人民共和国证券投资基金法》、《私募投资基金监督管理暂行办法》及《私募投资基金管理人登记和基金备案办法（试行）》规定的私募投资基金管理人或私募投资基金，不需办理私募投资基金管理人登记手续和私募投资基金备案手续。

## 2、中移投资

### （1）基本情况

名称	中移投资控股有限责任公司
统一社会信用代码	91440300MA5DQEEU2J
类型	有限责任公司
法定代表人	范冰
注册资本	2,000,000 万元
经营场所	深圳市前海深港合作区前湾一路1号A栋201室（入驻深圳市前海商务秘书有限公司）
经营范围/主营业务	投资管理、资本管理、资产管理（不得从事信托、金融资产管理、证券资产管理等业务）；投资咨询、

	管理咨询；计算机、通信的技术开发、技术咨询； 物业租赁；机械设备租赁（不含金融租赁）。物业管理。
主营业务与发行人主营业务的关系	无相关性
成立时间	2016年12月9日
营业期限	2016年12月9日至2076年12月9日

## （2）出资人构成和出资比例

截至本招股说明书签署日，中移投资出资人构成和出资比例如下：

序号	股东	出资金额（万元）	持股比例（%）
1	中国移动通信有限公司	2,000,000.00	100.00
	合计	<b>2,000,000.00</b>	<b>100.00</b>

中国移动通信有限公司持有中移投资 100% 股权，为中移投资的控股股东。中国移动通信有限公司为香港上市公司中国移动有限公司（HK.00941）的全资子公司，中国移动通信集团有限公司为中国移动有限公司的间接控股股东，国务院国有资产监督管理委员会持有中国移动通信集团有限公司 100% 股权，为中移投资的实际控制人。中移投资作为发行人股东，受国有独资企业中国移动通信集团有限公司控制，2020 年 7 月 15 日，发行人股东中移投资已取得国务院国有资产监督管理委员会核发的《关于北京海天瑞声科技股份有限公司国有股东标识有关事项的批复》（国资产权[2020]324 号），“如海天瑞声发行股票并上市，中移投资在证券登记结算公司设立的证券账户应标注“CS”标识。”

中移投资未以非公开方式向投资者募集资金，未委托他人管理其资产，亦未接受他人委托管理资产，不属于《中华人民共和国证券投资基金法》、《私募投资基金监督管理暂行办法》及《私募投资基金管理人登记和基金备案办法（试行）》规定的私募投资基金管理人或私募投资基金，不需办理私募投资基金管理人登记手续和私募投资基金备案手续。

## 3、唐涤飞

姓名	唐涤飞
性别	男
国籍	中国
身份证号	21040219701228****



住所	北京市朝阳区方舟苑
是否取得其他国家或地区的永久居留权	美国永久居留权

唐涤飞目前担任公司董事、总经理、核心技术人员，简历情况请参见本节“九、董事、监事、高级管理人员与核心技术人员的简要情况/（一）发行人董事、监事、高级管理人员与核心技术人员的简要情况/1、公司董事”。

#### 4、清德投资

##### （1）基本情况

名称	北京清德投资中心（有限合伙）
统一社会信用代码	91110108MA0039X16T
类型	有限合伙企业
执行事务合伙人	钟山
出资总额	3,480 万元
经营场所	北京市海淀区中关村东路 1 号院 1 号楼 A 座 14 层 1438 室
经营范围/主营业务	投资管理；资产管理。
主营业务与发行人主营业务的关系	无相关性
成立时间	2016 年 1 月 22 日
合伙期限	2016 年 1 月 22 日至长期

##### （2）出资人构成和出资比例

截至本招股说明书签署日，清德投资出资人构成和出资比例如下：

序号	合伙人姓名或名称	合伙人性质	出资金额（万元）	出资比例（%）
1	钟山	普通合伙人	305.00	8.76
2	吕柏仁	有限合伙人	800.00	22.99
3	杭世年	有限合伙人	600.00	17.24
4	刘晓波	有限合伙人	450.00	12.93
5	陈立	有限合伙人	400.00	11.49
6	孙应实	有限合伙人	300.00	8.62
7	刘三英	有限合伙人	250.00	7.18
8	张尔康	有限合伙人	200.00	5.75
9	志鹏	有限合伙人	150.00	4.31
10	罗建北	有限合伙人	25.00	0.72

序号	合伙人姓名或名称	合伙人性质	出资金额（万元）	出资比例（%）
合计			<b>3,480.00</b>	<b>100.00</b>

清德投资出资资金均为其合伙人自有资金，未以非公开方式向投资者募集资金，未委托他人管理其资产，亦未接受他人委托管理资产，不属于《中华人民共和国证券投资基金法》、《私募投资基金监督管理暂行办法》及《私募投资基金管理人登记和基金备案办法（试行）》规定的私募投资基金管理人或私募投资基金，不需办理私募投资基金管理人登记手续和私募投资基金备案手续。

## 5、上海丰琬

### （1）基本情况

名称	上海丰琬投资合伙企业（有限合伙）
统一社会信用代码	91310230MA1JX5J94W
类型	有限合伙企业
执行事务合伙人	史琰
出资总额	2,651 万元
经营场所	上海市崇明区横沙乡富民支路 58 号 A1-4411 室（上海横泰经济开发区）
经营范围/主营业务	投资管理、咨询，实业投资，资产管理，财务咨询（不得从事代理记账），商务咨询，企业管理咨询。
主营业务与发行人主营业务的关系	无相关性
成立时间	2016 年 1 月 25 日
合伙期限	2016 年 1 月 25 日至 2036 年 1 月 24 日

### （2）出资人构成和出资比例

截至本招股说明书签署日，上海丰琬出资人构成和出资比例如下：

序号	合伙人姓名或名称	合伙人性质	出资金额（万元）	出资比例（%）
1	史琰	普通合伙人	2.65	0.10
2	浙江自贸区丰航投资合伙企业（有限合伙）	有限合伙人	1,369.84	51.67
3	张桂珍	有限合伙人	555.11	20.94
4	崔宇	有限合伙人	445.70	16.81
5	史宁	有限合伙人	277.70	10.48
合计			<b>2,651.00</b>	<b>100.00</b>

根据上述有限合伙人浙江自贸区丰航投资合伙企业(有限合伙)的合伙协议,其合伙人及出资情况如下:

序号	合伙人姓名或名称	合伙人性质	出资金额(万元)	出资比例(%)
1	丰琰投资管理(上海)有限公司	普通合伙人	100.00	10.00
2	于英伦 (系史琰之女)	有限合伙人	900.00	90.00
合计			<b>1,000.00</b>	<b>100.00</b>

根据丰琰投资管理(上海)有限公司的公司章程,史琰持有丰琰投资管理(上海)有限公司 100% 股权。

上海丰琰出资资金均为其合伙人自有资金,未以非公开方式向投资者募集资金,未委托他人管理其资产,亦未接受他人委托管理资产,不属于《中华人民共和国证券投资基金法》、《私募投资基金监督管理暂行办法》及《私募投资基金管理人登记和基金备案办法(试行)》规定的私募投资基金管理人或私募投资基金,不需办理私募投资基金管理人登记手续和私募投资基金备案手续。

## 6、中瑞立

### (1) 基本情况

名称	北京中瑞立投资中心(有限合伙)
统一社会信用代码	91110108MA002L3E9M
类型	有限合伙企业
执行事务合伙人	创慧科瑞(委派唐涤飞为代表)
出资总额	1,000 万元
经营场所	北京市海淀区上地十街 1 号院 4 号楼 23 层 2304-2
经营范围/主营业务	投资管理; 资产管理。
主营业务与发行人主营业务的关系	无相关性
成立时间	2015 年 12 月 18 日
合伙期限	2015 年 12 月 18 日至长期

### (2) 出资人构成和出资比例

中瑞立为发行人员工持股平台。截至本招股说明书签署日,中瑞立出资人构成和出资比例如下:

序号	合伙人姓名或名称	在本公司及子公司任职情况	合伙人性质	出资金额(万元)	出资比例(%)
1	创慧科瑞	-	普通合伙人	294.12	29.41
2	黄大伟	子公司中瑞智总经理	有限合伙人	161.76	16.18
3	宋琼	语音识别采集业务部总监	有限合伙人	132.35	13.24
4	廖晓玲	语音合成业务部总监	有限合伙人	132.35	13.24
5	陈琪	销售经理	有限合伙人	132.35	13.24
6	贺琳	董事长	有限合伙人	58.82	5.88
7	辛晓峰	自然语言处理业务部总监	有限合伙人	44.12	4.41
8	葛星彤	语音识别转写业务部总监	有限合伙人	44.12	4.41
合计				<b>1,000.00</b>	<b>100.00</b>

截至本招股说明书签署日，中瑞立普通合伙人创慧科瑞的股权结构如下：

序号	股东	出资金额(万元)	持股比例(%)
1	唐涤飞	20.00	50.00
2	志鹏	20.00	50.00
合计		<b>40.00</b>	<b>100.00</b>

中瑞立为发行人的员工持股平台，未以非公开方式向投资者募集资金，未委托他人管理其资产，亦未接受他人委托管理资产，不属于《中华人民共和国证券投资基金法》、《私募投资基金监督管理暂行办法》及《私募投资基金管理人登记和基金备案办法（试行）》规定的私募投资基金管理人或私募投资基金，不需办理私募投资基金管理人登记手续和私募投资基金备案手续。

### （三）公司控股股东和实际控制人直接或间接持有的公司股份质押或其他有争议的情况

截至本招股说明书签署日，公司控股股东和实际控制人直接或间接持有发行人的股份不存在质押或其他有争议的情况。

### （四）公司控股股东和实际控制人控股股东、实际控制人所控制的其他企业

截至本招股说明书签署日，公司控股股东、实际控制人所控制的其他企业情况请参见本招股说明书“第七节 公司治理与独立性/三、同业竞争/（一）公司与

控股股东、实际控制人及其控制的其他企业不存在同业竞争”。

## 八、发行人股本情况

### （一）本次发行前后发行人股本情况

本公司发行前总股本为3,210万元，本次拟向社会公众公开发行1,070万股，全部为新股发行，本次发行原股东不公开发售股份。本次公开发行股份占发行后总股本的比例为25%。假设本次公开发行1,070万股，则发行后总股本预计为4,280万股。

本次发行前后，发行人的股本结构如下：

序号	类别	发行前		发行后	
		股数(万股)	比例(%)	股数(万股)	比例(%)
1	有限售条件的流通股	3,210.00	100.00	3,210.00	75.00
2	无限售条件的流通股	-	-	1,070.00	25.00
	合计	<b>3,210.00</b>	<b>100.00</b>	<b>4,280.00</b>	<b>100.00</b>

### （二）本次发行前的前十名股东情况

本次发行前，发行人前十名股东及持股情况如下：

序号	股东名称	发行前	
		股数(万股)	比例(%)
1	贺琳	866.97	27.01
2	中瑞安	495.41	15.43
3	中移投资(CS) <sup>注</sup>	385.50	12.01
4	唐涤飞	357.80	11.15
5	清德投资	254.55	7.93
6	上海丰琬	188.04	5.86
7	中瑞立	187.16	5.83
8	上海兴富	132.31	4.12
9	中网投	129.00	4.02
10	天津金星	93.58	2.92
	合计	<b>3,090.32</b>	<b>96.28</b>

注：中移投资是国有独资企业通过投资关系能够实际支配其行为的企业，符合《上市公司国有股权监督管理办法》第七十四条的规定，证券账户应被标识为“CS”。2020年7月15日，发行人股东中移投资已取得国务院国有资产监督管理委员会核发的《关于北京海天

瑞声科技股份有限公司国有股东标识有关事项的批复》(国资产权[2020]324号)，“如海天瑞声发行股票并上市，中移投资在证券登记结算公司设立的证券账户应标注“CS”标识。”除中移投资外，不存在其他需要取得国有股权批复的国有股东。

### (三) 本次发行前的前十名自然人股东及其在发行人处任职的情况

截至本招股说明书签署日，发行人共有2名自然人股东，分别为贺琳、唐涤飞，在发行人担任职务的情形如下：

序号	股东名称	直接持股数(万股)	直接持股比例(%)	在本公司任职
1	贺琳	866.97	27.01	董事长
2	唐涤飞	357.80	11.15	董事、总经理
	合计	1,224.77	38.16	

### (四) 最近一年发行人新增股东的情况

截至本招股说明书签署日，发行人本次申报前一年内公司新增股东中移投资、中网投、芜湖青和与芜湖博信，基本情况如下：

#### 1、中移投资

中移投资的具体情况详见本节之“七、持有发行人5%以上股份的主要股东及实际控制人情况”之“(二)其他持有本公司5%以上股份的主要股东的基本情况”。

#### 2、中网投

##### (1) 基本情况

名称	中国互联网投资基金(有限合伙)
统一社会信用代码	91110000MA00CXL49H
类型	有限合伙企业
执行事务合伙人	中国互联网投资基金管理有限公司(委派吴海为代表)
出资总额	3,010,000万元
经营场所	北京市大兴区大兴经济开发区科苑路18号2幢一层A032号(国家新媒体产业基地内)
经营范围/主营业务	非证券业务的投资、投资管理、咨询。
主营业务与发行人主营业务的关系	无相关性
成立时间	2017年3月23日
合伙期限	2017年3月23日至2032年3月22日

## (2) 合伙人情况

截至本招股说明书签署日，中网投合伙人情况如下：

序号	合伙人名称	合伙人性质	出资金额(万元)	出资比例(%)
1	中国互联网投资基金管理有限公司	普通合伙人	10,000.00	0.33
2	工银瑞信投资管理有限公司	有限合伙人	1,000,000.00	33.22
3	农银汇理资产管理有限公司	有限合伙人	500,000.00	16.61
4	中信国安集团有限公司	有限合伙人	450,000.00	14.95
5	中邮人寿保险股份有限公司	有限合伙人	350,000.00	11.63
6	中国移动通信集团有限公司	有限合伙人	300,000.00	9.97
7	财政部	有限合伙人	200,000.00	6.64
8	中国联合网络通信集团有限公司	有限合伙人	100,000.00	3.32
9	中国电信集团有限公司	有限合伙人	100,000.00	3.32
合计			<b>3,010,000.00</b>	<b>100.00</b>

中国互联网投资基金管理有限公司为中网投的普通合伙人、执行事务合伙人，其基本情况如下：

名称	中国互联网投资基金管理有限公司
统一社会信用代码	91100000MA0000015X
类型	其他有限责任公司
法定代表人	吴海
注册资本	10,000.00 万元
经营场所	北京市大兴区黄村镇工业开发区科苑路 18 号 1 幢 C1 户型 1 层 1097 室
经营范围/主营业务	投资基金管理；投资管理、资产管理。（企业依法自主选择经营项目，开展经营活动；依法须经批准的项目，经相关部门批准后依批准的内容开展经营活动；不得从事本市产业政策禁止和限制类项目的经营活动。）
主营业务与发行人主营业务的关系	无相关性
成立时间	2016 年 8 月 30 日
营业期限	2016 年 8 月 30 日至无固定期限

根据中国互联网投资基金管理有限公司的公司章程，其股东及其出资金额、出资比例如下：

序号	股东名称	出资金额（万元）	出资比例（%）
1	中网信通（北京）控股有限公司	4,000.00	40.00
2	中信国安集团有限公司	2,455.00	24.55
3	中邮人寿保险股份有限公司	1,909.00	19.09
4	中移资本控股有限责任公司	1,636.00	16.36
合计		<b>10,000.00</b>	<b>100.00</b>

根据中国互联网投资基金管理有限公司的第一大股东中网信通（北京）控股有限公司的章程，其股东及其出资金额、出资比例如下：

序号	股东	出资金额（万元）	持股比例（%）
1	国家计算机网络与信息安全管理中心	4,000.00	100.00
合计		<b>4,000.00</b>	<b>100.00</b>

中网投已在中国证券投资基金业协会进行私募基金备案，基金编号为SS8838，备案时间为2017年6月6日。其管理人为中国互联网投资基金管理有限公司，已在中国证券投资基金业协会登记为私募基金管理人，登记编号为P1060330，登记时间为2016年12月6日。

### 3、芜湖青和

#### （1）基本情况

名称	芜湖朗姿青和一号股权投资合伙企业（有限合伙）
统一社会信用代码	91340202MA2TQ6RR8P
类型	有限合伙企业
执行事务合伙人	芜湖朗姿青和投资有限公司（委派张凯为代表）
出资总额	20,100.00 万元
经营场所	安徽省芜湖市镜湖区荆西社区小微企业孵化园216-1室
经营范围/主营业务	实业投资、项目投资、股权投资、创业投资、投资咨询。
主营业务与发行人主营业务的关系	无相关性
成立时间	2019年5月17日
合伙期限	2019年5月17日至2039年5月16日

#### （2）合伙人情况

截至本招股说明书签署日，芜湖青和合伙人情况如下：



序号	合伙人名称	合伙人性质	出资金额(万元)	出资比例 (%)
1	芜湖朗姿青和投资有限公司	普通合伙人	100.00	0.50
2	江苏青和投资管理有限公司	有限合伙人	18,000.00	89.55
3	芜湖成中电子科技有限公司	有限合伙人	1,500.00	7.46
4	孙益功	有限合伙人	500.00	2.49
合计			<b>20,100.00</b>	<b>100.00</b>

芜湖朗姿青和投资有限公司为芜湖青和的普通合伙人、执行事务合伙人，其基本情况如下：

名称	芜湖朗姿青和投资有限公司
统一社会信用代码	91340202MA2TQ5BY0D
类型	其他有限责任公司
法定代表人	张凯
注册资本	1,000.00 万元
经营场所	安徽省芜湖市镜湖区荆西社区小微企业孵化园215-2号
经营范围/主营业务	实业投资、项目投资、股权投资、创业投资、投资咨询。(未经金融等监管部门的批准,不得从事向公众融资存款、融资担保、代客理财等金融服务)(依法须经批准的项目,经相关部门批准后方可开展经营活动)
主营业务与发行人主营业务的关系	无相关性
成立时间	2019年5月15日
营业期限	2019年5月15日至无固定期限

根据芜湖朗姿青和投资有限公司的公司章程,其股东及其出资金额、出资比例如下:

序号	股东名称	出资金额(万元)	出资比例 (%)
1	芜湖恒鼎投资管理有限公司	510.00	51.00
2	江苏青和投资管理有限公司	400.00	40.00
3	孙益功	90.00	9.00
合计		<b>1,000.00</b>	<b>100.00</b>

根据芜湖恒鼎投资管理有限公司的公司章程,其股东及其出资金额、出资比例如下:

序号	股东名称	出资金额（万元）	出资比例（%）
1	北京朗姿韩亚资产管理有限公司	1,000.00	100.00
合计		<b>1,000.00</b>	<b>100.00</b>

根据北京朗姿韩亚资产管理有限公司的公司章程，其股东及其出资金额、出资比例如下：

序号	股东名称	出资金额（万元）	出资比例（%）
1	芜湖德臻睿远投资合伙企业（有限合伙）	79,050.00	43.84
2	朗姿股份有限公司	52,700.00	29.23
3	株式会社韩亚银行	23,250.00	12.89
4	韩亚金融投资株式会社	16,300.00	9.04
5	三井住友信托银行股份有限公司	9,015.7895	5.00
合计		<b>180,315.7895</b>	<b>100.00</b>

根据芜湖德臻睿远投资合伙企业（有限合伙）的合伙协议，其合伙人及其出资情况如下：

序号	股东名称	出资金额（万元）	出资比例（%）
1	芜湖德臻投资有限公司	1,000	1.00
2	申东日	29,000	29.00
3	北京乐家园投资管理有限公司	20,230	20.23
4	芜湖泛游旅游产业投资合伙企业（有限合伙）	16,330	16.33
5	广东泓殿投资有限公司	15,250	15.25
6	河南瀚良企业管理咨询有限公司	14,270	14.27
7	南京衡望建材贸易有限公司	3,920	3.92
合计		<b>100,000.00</b>	<b>100.00</b>

根据芜湖德臻睿远投资合伙企业（有限合伙）普通合伙人、执行事务合伙人芜湖德臻投资有限公司的公司章程，申东日持有其 60% 股权，申今花持有其 40% 股权。申东日、申今花同时为北京朗姿韩亚资产管理有限公司股东朗姿股份有限公司（股票代码：002612）的实际控制人。

芜湖青和已在中国证券投资基金业协会进行私募基金备案，基金编号为 SGY287，备案时间为 2019 年 10 月 11 日。其管理人为北京朗姿韩亚资产管理有限公司，已在中国证券投资基金业协会登记为私募基金管理人，登记编号为

P1069524，登记时间为2019年2月1日。

#### 4、芜湖博信

##### (1) 基本情况

名称	芜湖博信五号股权投资合伙企业（有限合伙）
统一社会信用代码	91340202MA2UDGRP9E
类型	有限合伙企业
执行事务合伙人	北京朗姿韩亚资产管理有限公司（委派张凯为代表）
出资总额	660.00 万元
经营场所	芜湖市镜湖区荆西社区小微企业孵化园 230-1 号
经营范围/主营业务	企业投资、项目投资、股权投资。
主营业务与发行人主营业务的关系	无相关性
成立时间	2019 年 12 月 13 日
合伙期限	2019 年 12 月 13 日至 2039 年 12 月 12 日

##### (2) 合伙人情况

截至本招股说明书签署日，芜湖博信合伙人情况如下：

序号	合伙人名称	合伙人性质	出资金额 (万元)	出资比例 (%)
1	北京朗姿韩亚资产管理有限公司	普通合伙人	10.00	1.52
2	林鸽	有限合伙人	300.00	45.45
3	吴火忠	有限合伙人	250.00	37.88
4	赵启轩	有限合伙人	100.00	15.15
合计			<b>660.00</b>	<b>100.00</b>

北京朗姿韩亚资产管理有限公司为芜湖博信的普通合伙人、执行事务合伙人，其基本情况如下：

名称	北京朗姿韩亚资产管理有限公司
统一社会信用代码	91110113MA005JAA1G
类型	有限责任公司（中外合资）
法定代表人	袁怀中
注册资本	180,315.7895 万元
经营场所	北京市西城区后半壁街 56 号 9 号楼一层 121 号
经营范围/主营业务	资产管理。（“1、未经有关部门批准，不得以公开方

	式募集资金；2、不得公开开展证券类产品和金融衍生品交易活动；3、不得发放贷款；4、不得对所投资企业以外的其他企业提供担保；5、不得向投资者承诺投资本金不受损失或者承诺最低收益”；该公司2016年12月06日前为内资企业，于2016年12月06日变更为外商投资企业；企业依法自主选择经营项目，开展经营活动；依法须经批准的项目，经相关部门批准后依批准的内容开展经营活动；不得从事本市产业政策禁止和限制类项目的经营活动。）
主营业务与发行人主营业务的关系	无相关性
成立时间	2016年5月18日
营业期限	2016年12月6日至2036年12月5日

北京朗姿韩亚资产管理有限公司的股权结构详见本招股说明书“第五节 发行人基本情况/八、发行人股本情况/（四）最近一年发行人新增股东的情况/3、芜湖青和”。

芜湖博信已在中国证券投资基金业协会进行私募基金备案，基金编号为SJM276，备案时间为2019年12月23日。其管理人为北京朗姿韩亚资产管理有限公司，已在中国证券投资基金业协会登记为私募基金管理人，登记编号为P1069524，登记时间为2019年2月1日。

#### （五）本次发行前各股东之间的关联关系

公司控股股东、实际控制人贺琳直接持有发行人 27.01%的股权；贺琳持有100%股权的创世联合为发行人股东中瑞安（持有发行人 15.43%的股权）的普通合伙人、执行事务合伙人，持有中瑞安 36.67%的出资；贺琳为发行人股东中瑞立（持有发行人 5.83%的股权）的有限合伙人，持有中瑞立 5.88%的出资。

唐涤飞直接持有发行人 11.15%的股权；唐涤飞持有 50%股权的创慧科瑞为中瑞立（持有发行人 5.83%的股权）的普通合伙人、执行事务合伙人，持有中瑞立 29.41%的出资，唐涤飞为中瑞立执行事务合伙人的委派代表。

清德投资普通合伙人、执行事务合伙人钟山及其配偶有限合伙人志鹏分别持有清德投资 8.76%、4.31%的出资，志鹏同时持有发行人股东中瑞立普通合伙人、执行事务合伙人创慧科瑞 50%的股权。

杭州士兰为发行人股东杭州银杏数的有限合伙人，持有杭州银杏数 17.95%的出资，杭州士兰控股有限公司（持有杭州士兰 85%股权）持有杭州银杏数普通

合伙人、执行事务合伙人浙江银杏谷投资有限公司 15.74%的股权；杭州科艺投资管理有限公司持有杭州士兰 15%股权，杭州科艺投资管理有限公司及其唯一股东陈向明分别持有浙江银杏谷投资有限公司第一大股东杭州平仲投资有限公司 37.5%、25%的股权。

中国移动通信集团有限公司控制发行人的股东中移投资，为中网投的有限合伙人，持有中网投 9.97%的出资；中国移动通信集团有限公司的全资子公司中移资本控股有限责任公司持有中网投普通合伙人、执行事务合伙人中国互联网投资基金管理公司（持有中网投 0.33%的出资）16.36%的股权。

芜湖博信的普通合伙人、执行事务合伙人北京朗姿韩亚资产管理有限公司间接控制芜湖青和普通合伙人、执行事务合伙人芜湖朗姿青和投资有限公司，北京朗姿韩亚资产管理有限公司同为芜湖博信、芜湖青和的私募投资基金管理人。

除前述情形以外，本次发行前公司股东中不存在其他关联关系。

#### （六）公开发售股份对发行人的控制权、治理结构及生产经营产生的影响

本次发行不安排发行人股东公开发售股份。

#### （七）持有发行人股份的私募投资基金等金融产品纳入监管情况

截至本招股说明书签署日，发行人共有 14 名股东，其中 2 名为自然人股东，2 名为员工持股平台，10 名为其他机构股东。机构股东中，上海兴富、中网投、杭州银杏数、芜湖青和及芜湖博信为私募投资基金。中网投、芜湖青和及芜湖博信纳入监管情况参见本节“八、发行人股本情况/（四）最近一年发行人新增股东的情况”。上海兴富、杭州银杏数纳入监管的情况如下：

股东名称	私募基金备案情况	备案时间	私募基金管理人登记情况	登记时间
上海兴富	备案编码 S60283	2015-07-09	私募投资基金管理人兴富投资管理有限公司登记编号为 P1015277	2015-06-05
杭州银杏数	备案编码 SH7371	2016-11-21	私募投资基金管理人浙江银杏谷投资有限公司登记编号为 P1003808	2014-06-04

## 九、董事、监事、高级管理人员与核心技术人员的简要情况

### （一）发行人董事、监事、高级管理人员与核心技术人员的简要情况

#### 1、公司董事

截至本招股说明书签署日，公司共有 9 名董事，其中 3 名为独立董事。公司董事由股东大会选举产生，任期 3 年，任期届满可连选连任。公司现任董事基本情况如下：

姓名	在本公司任职	提名人	董事任职期间
贺琳	董事长	发行人原董事会	2017 年 11 月至 2020 年 11 月
唐涤飞	董事、总经理	发行人原董事会	2017 年 11 月至 2020 年 11 月
李科	董事、副总经理	贺琳	2019 年 9 月至 2020 年 11 月
吕思遥	董事、副总经理、财务负责人、董事会秘书	发行人原董事会	2017 年 11 月至 2020 年 11 月
黄宇凯	董事、副总经理、技术总监	贺琳	2019 年 9 月至 2020 年 11 月
志鹏	董事	发行人原董事会	2017 年 11 月至 2020 年 11 月
仪晓辉	独立董事	发行人原董事会	2017 年 11 月至 2020 年 11 月
张东晖	独立董事	发行人董事会提名委员会	2018 年 11 月至 2020 年 11 月
杨涛	独立董事	发行人原董事会	2017 年 11 月至 2020 年 11 月

注：本届董事会的任期已经届满，董事会成员将在股东大会完成换届选举前继续履行职责。

发行人现任董事简历情况如下：

贺琳，女，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，现任公司董事长。

唐涤飞，男，1970 年 12 月出生，中国国籍，有美国永久居留权，硕士，毕业于中国科学技术大学，获得通信与电子系统专业工学硕士学位。1996 年至 1997 年，任联想集团技术中心项目负责人；1997 年至 2010 年，历任微软中国研发中

心中文技术部项目经理、微软亚洲工程院资深项目经理；2010年至2012年，任阿里巴巴云计算公司数据应用与服务、无线平台的产品及项目总监。2013年3月加入海天瑞声，现任公司董事、总经理。

李科，男，1981年8月出生，中国国籍，无境外永久居留权，硕士，毕业于清华大学，获得信息与信号处理专业硕士学位。2006年至2009年任IBM中国技术开发中心软件工程师。2009年10月加入海天瑞声，现任公司董事、副总经理。

吕思遥，女，1983年8月出生，中国国籍，无境外永久居留权，硕士，毕业于北京大学光华管理学院，获得会计学专业学士学位，中国人民大学财政与金融学院金融学硕士学位，中国注册会计师、国际注册内审师。2005年至2011年，任普华永道中天会计师事务所审计师、高级审计师、经理；2011年至2012年，任英特尔（中国）内部控制经理；2012年至2014年，任北京新合作连锁超市有限公司财务总监；2015年至2016年，任英特尔（中国）信用管理高级经理。2016年8月加入海天瑞声，现任公司董事、副总经理、财务负责人、董事会秘书。

黄宇凯，男，1983年5月出生，中国国籍，无境外永久居留权，硕士，毕业于复旦大学，获得计算机科学与技术专业硕士学位。2008年至2018年，历任微软（中国）有限公司软件研发工程师、必应网页搜索部门研发主管、地图中国团队首席研发总监等职务。2018年12月加入海天瑞声，现任公司董事、副总经理、技术总监。

志鹏，女，1972年6月出生，中国国籍，无境外永久居留权，硕士，于清华大学获得计算机科学与技术专业工学学士学位，于剑桥大学获得哲学硕士学位。1996年至2004年任职于微软公司，2006年至2011年任职于TLcom Capital，2012年至今，任职昆山青石投资管理有限公司执行董事兼总经理；2013年至今任清控资产管理有限公司专务副总裁。2016年4月至今任公司董事。

仪晓辉，男，1968年10月出生，中国国籍，无境外永久居留权，硕士，毕业于北京大学无线电学系信息中心研究院，获得工学硕士学位。1993年至1997年，任方正集团销售部经理；1997年至2004年，历任IBM ThinkPad业务发展经理、IBM中国区总经理；2004年至2012年，历任联想集团中国区Think业务

部总经理、中国笔记本业务总经理、联想集团副总裁、大中华及俄罗斯区商用事业部总经理以及联想新兴市场集团下属香港、台湾和韩国区副总裁；2012年至2014年，任惠普公司全球高级副总裁、打印与信息产品集团中国区总裁；2014年至今，任成都智明达电子股份有限公司董事。2017年11月至今任公司独立董事。

张东晖，男，1967年7月出生，中国国籍，无境外永久居留权，学士，毕业于清华大学，获得工学学士学位。1993年至1995年，任北京达因电脑有限公司财务总监；1996年至2010年，任微软（中国）有限公司资深开发经理；2010年至2015年，任阿里云计算有限公司研究员，并于2014年至2016年代表阿里巴巴任恒生电子董事；2015年11月至今，为杭州几何康辰科技有限公司创始合伙人。2018年11月至今任公司独立董事。

杨涛，男，1975年2月出生，中国国籍，无境外永久居留权，学士，毕业于首都经济贸易大学，获得审计学专业学士学位，中国注册会计师。2001年至2015年，任普华永道会计师事务所审计师、高级审计师、经理、高级经理；2015年至2016年，任北京大生知行科技有限公司财务总监；2016年至2017年，任宏力医疗管理集团有限公司董事长助理；2017年至2018年3月，任张家口崇礼太舞旅游度假有限公司总经理助理；2018年3月至2020年1月，任优途天下（北京）房车有限公司副总经理；2020年2月至2020年12月，任中美医疗集团有限公司财务总监；2021年1月至今，任北京运世达医疗投资管理有限公司首席财务官。2017年11月至今任公司独立董事。

## 2、公司监事

截至本招股说明书签署日，公司共有3名监事，其中职工监事1名。监事任期3年，任期届满可连选连任。公司现任监事基本情况如下：

姓名	在本公司任职	提名人	监事任职期间
张小龙	监事会主席	发行人原监事	2017年11月至2020年11月
方彦彬	监事	发行人原监事	2017年11月至2020年11月
熊敏	监事	-	2017年11月至2020年11月

注：熊敏为职工监事，由职工代表大会选举产生。本届监事会的任期已经届满，监事会成员将在股东大会、职工代表大会完成换届选举前继续履行职责。



张小龙，男，1976年4月出生，中国国籍，无境外永久居留权，硕士，获得复旦-麻省理工国际MBA、中欧国际工商学院EMBA学位。曾任职于上海贝尔阿尔卡特移动通信有限公司、爱立信中国研究中心、艾默生、IBM、DELL等公司战略、研发和销售管理职务；2007年至2013年，任兴业创新资本执行董事；2013年至2015年，任国泰君安创新投资董事总经理；2015年至今任兴富投资管理有限公司合伙人、副总经理。2016年3月至2017年10月任公司监事，2017年11月至今任公司监事会主席。

方彦彬，男，1983年7月出生，中国国籍，无境外永久居留权，硕士，毕业于浙江大学。2006年至2015年，任职于杭州市西湖区人民政府；2015年至2019年，任浙江银杏谷投资有限公司总裁助理；2020年至今任云仰科技（杭州）有限公司总经理，2021年3月至今任该公司董事。2017年11月至今任公司监事。

熊敏，女，1987年5月出生，中国国籍，无境外永久居留权，本科，毕业于中南财经政法大学工商管理专业。历任京信通信系统（中国）有限公司、北京奥伯特石油科技有限公司人力资源专员、人事主管。2014年加入海天瑞声，先后担任人力资源主管、人资行政经理、人资行政总监。2017年11月至今担任公司监事。

### 3、公司高级管理人员

截至本招股说明书签署日，公司共有5名高级管理人员，基本情况如下：

姓名	在本公司任职	高管任职期间
唐涤飞	董事、总经理	2017年11月-2020年11月
李科	董事、副总经理	2017年11月-2020年11月
吕思遥	董事、财务负责人、董事会秘书	2017年11月-2020年11月
	副总经理	2018年1月-2020年11月
黄宇凯	董事、技术总监	2019年1月-2020年11月
	副总经理	2020年1月-2020年11月
郝玉峰	副总经理、技术总监、首席科学家	2017年11月-2020年11月

注：本届高级管理人员的任期已经届满，高级管理人员将在换届聘任前继续履行职责。

唐涤飞、吕思遥、李科、黄宇凯简历参见本节“九、董事、监事、高级管理人员与核心技术人员的简要情况/（一）发行人董事、监事、高级管理人员与核心技术人员的简要情况/1、公司董事”。

郝玉峰，男，1974年5月出生，中国国籍，无境外永久居留权，博士，毕业于东南大学，获得生物医学工程专业博士学位。2005年至2008年，任北京捷通华声语音技术有限公司技术研究部经理；2008年任杭州红杉树信息技术有限公司高级研发工程师。2008年12月加入海天瑞声，现任公司副总经理、技术总监、首席科学家。

#### 4、公司核心技术人员

截至本招股说明书签署日，公司共有6名核心技术人员。公司对核心技术人员的认定依据为：1、拥有与公司业务匹配的行业背景、科研成果；2、在研究开发、技术产业化等岗位担任重要职务，对公司核心技术创新、业务发展具有重要贡献。基本情况如下：

姓名	在本公司任职
唐涤飞	董事、总经理
李科	董事、副总经理
黄宇凯	董事、副总经理、技术总监
郝玉峰	副总经理、技术总监、首席科学家
曹琼	计算机视觉业务部总监
邵志明	高级研发工程师

唐涤飞、李科、黄宇凯、郝玉峰简历参见本节“九、董事、监事、高级管理人员与核心技术人员的简要情况/（一）发行人董事、监事、高级管理人员与核心技术人员的简要情况/1、公司董事以及3、公司高级管理人员”。

曹琼，女，1982年10月出生，中国国籍，无境外永久居留权，博士，毕业于北京航空航天大学，获得检测技术与自动化装置专业博士学位。2008年至2017年，任富士通研究开发有限公司高级研究员。2017年5月加入海天瑞声，现任公司计算机视觉业务部总监。

邵志明，男，1987年12月出生，中国国籍，无境外永久居留权，硕士，毕业于北京邮电大学，获得模式识别专业硕士学位。2014年至2015年，任百度公司语音识别研发工程师。2015年4月加入海天瑞声，现任公司高级研发工程师。

发行人核心技术人员的  
主要经历及对公司研发的具体贡献情况详见本招股说明书“第六节 业务和技术/六、发行人的研发和技术/（九）研发人员和核心技

术人员情况”。

(二) 发行人董事、监事、高级管理人员与核心技术人员的兼职情况及与其他董事、监事、高级管理人员及核心技术人员的亲属关系

### 1、公司董事、监事、高级管理人员与核心技术人员的兼职情况

姓名	在本公司任职	兼职单位	兼职职务
贺琳	董事长	创世联合	执行董事、经理
唐涤飞	董事、总经理	创世联合	监事
		创慧科瑞	监事
		BING & DIFEI BELL-TOWER INVESTMENT LLC	董事
		BING & DIFEI YARROW POINT INVESTMENT LLC	董事
志鹏	董事	创慧科瑞	执行董事、经理
		北京海淀科技金融资本控股集团股份有限公司	董事
		清控资产管理有限公司	专务副总裁
		昆山青石投资管理有限公司	执行董事、总经理
		清源华动(北京)科技有限公司	监事
		北京清融投资管理有限公司	监事
		中关村并购发展促进会	秘书长
吕思遥	董事、副总经理、财务负责人、董事会秘书	北京贺兮科技有限公司	执行董事
仪晓辉	独立董事	成都智明达电子股份有限公司	董事
杨涛	独立董事	北京运世达医疗投资管理有限公司	首席财务官
张小龙	监事会主席	上海勤和互联网技术软件开发有限公司	董事
		上海文华财经资讯股份有限公司	董事
		上海兴富平龙创业投资有限公司	总经理
		兴富投资管理有限公司	副总经理
		深圳市博阅科技股份有限公司	董事
		北京中数智汇科技股份有限公司	监事
		南通市久正人体工学股份有限公司	董事
		广州佛朗斯股份有限公司	监事
		企查查科技有限公司	监事

姓名	在本公司任职	兼职单位	兼职职务
		杭州碧橙数字技术股份有限公司	董事
方彦彬	监事	杭州云徙科技有限公司	监事
		深圳市阿西莫夫科技有限公司	监事
		杭州迦智科技有限公司	监事会主席
		云仰科技（杭州）有限公司	董事、总经理
		飞米光年（景宁）网络科技有限公司	监事

截至本招股说明书签署日，除上述情况外，本公司董事、监事、高级管理人员及核心技术人员无在其他单位任职的情况。

## 2、董事、监事、高级管理人员及核心技术人员的兼职单位与发行人的关联关系

上述董事、监事、高级管理人员及核心技术人员的兼职单位中，部分兼职单位与发行人存在关联关系，具体情况请参见本招股说明书“第七节 公司治理与独立性/四、关联方及关联交易/（一）关联方及关联关系”；其余兼职单位与发行人不存在关联关系。

## 3、发行人董事、监事、高级管理人员及核心技术人员与其他董事、监事、高级管理人员及核心技术人员的亲属关系

发行人的董事、监事、高级管理人员及核心技术人员之间均不存在亲属关系。

## （三）发行人与董事、监事、高级管理人员及核心技术人员的重大协议及相关协议履行情况

公司董事（除志鹏和独立董事外）、职工代表监事、高级管理人员和核心技术人员均与公司签订了《劳动合同》、《竞业限制协议》以及《保密协议》。

截至本招股说明书签署日，公司董事、监事、高级管理人员及核心技术人员与公司签署的上述协议履行情况良好。

## （四）董事、监事、高级管理人员及核心技术人员所持股份发生被质押、冻结或发生诉讼纠纷等情形

截至本招股说明书签署日，公司董事、监事、高级管理人员及核心技术人员所持股份不存在被质押、冻结或发生诉讼纠纷等情形。

## （五）董事、监事、高级管理人员及核心技术人员在最近 2 年内的变动情况

### 1、董事变动情况

2018 年 11 月 30 日，发行人召开 2018 年第二次临时股东大会，由于吕一凡辞去独立董事职务，选举张东晖为独立董事。

2019 年 9 月 11 日，发行人召开 2019 年第四次临时股东大会，增选李科、黄宇凯为公司董事。

### 2、监事变动情况

发行人监事最近 2 年未发生变动。

### 3、高级管理人员变动情况

2018 年 1 月 23 日，发行人召开第一届董事会第四次会议，聘任吕思遥为副总经理。

2019 年 1 月 9 日，程显峰由于个人原因辞去副总经理职务。

2019 年 1 月 28 日，发行人召开第一届董事会第十次会议，聘任黄宇凯为技术总监。

2020 年 1 月 2 日，发行人召开第一届董事会第二十一次会议，聘任黄宇凯为副总经理。

2021 年 1 月 18 日，公司召开第一届董事会第二十八次会议，聘任郝玉峰为副总经理。

### 4、核心技术人员变动情况

截至本招股说明书签署日，发行人核心技术人员为唐涤飞、李科、黄宇凯、郝玉峰、曹琼、邵志明。除黄宇凯为 2018 年新增核心技术人员外，其他核心技术人员自报告期初即在公司任职且未发生变动。

综上，近两年内，发行人董事、监事、高级管理人员和核心技术人员未发生重大变动。

(六) 董事、监事、高级管理人员及核心技术人员与发行人业务相关的对外投资情况和以上人员及其近亲属持有发行人股份情况

1、董事、监事、高级管理人员及核心技术人员与发行人业务相关的对外投资情况

截至本招股说明书签署日，公司董事、监事、高级管理人员及核心技术人员不存在与发行人及其业务相关的其他对外投资。

2、董事、监事、高级管理人员、核心技术人员及其近亲属持有发行人股份的情况

截至本招股说明书签署日，公司董事、监事、高级管理人员及核心技术人员及其近亲属持有公司股份情况如下：

(1) 直接持股：

公司董事长贺琳直接持有公司 866.9725 万股股份，持股比例为 27.01%；董事、总经理、核心技术人员唐涤飞直接持有 357.7982 万股股份，持股比例为 11.15%。

(2) 间接持股：

姓名	任职/亲属关系	间接持股情况
贺琳	董事长	持有创世联合 100.00% 的股权，创世联合为中瑞安普通合伙人、执行事务合伙人并持有中瑞安 36.67% 的出资份额，中瑞安直接持有发行人 15.43% 的股份；作为有限合伙人持有中瑞立 5.88% 的出资份额，中瑞立直接持有发行人 5.83% 的股份。
唐涤飞	董事、总经理	持有创慧科瑞 50.00% 的股权，创慧科瑞为中瑞立普通合伙人、执行事务合伙人并持有中瑞立 29.41% 的出资份额，中瑞立直接持有发行人 5.83% 的股份。
志鹏	董事	持有创慧科瑞 50.00% 的股权，创慧科瑞为中瑞立普通合伙人、执行事务合伙人并持有中瑞立 29.41% 的出资份额，中瑞立直接持有发行人 5.83% 的股份；作为有限合伙人持有清德投资 4.31% 的出资份额，清德投资直接持有发行人 7.93% 的股份。
钟山	志鹏配偶	作为普通合伙人持有清德投资 8.76% 的出资份额，清德投资直接持有发行人 7.93% 的股份。
李科	董事、副总经理	作为有限合伙人，李科、吕思遥、郝玉峰分别持有中瑞安 30.56%、8.33%、19.44% 的出资份额，中瑞安直接持有发行人 15.43% 的股份。
吕思遥	董事、副总经理、财务负责人、董事会秘书	
郝玉峰	副总经理、技术总监、	

姓名	任职/亲属关系	间接持股情况
	首席科学家	
张小龙	监事会主席	张小龙持有宁波兴富创赢投资合伙企业（有限合伙）5.50%的出资份额，宁波兴富创赢投资合伙企业（有限合伙）持有兴富投资管理有限公司70.00%的股权，兴富投资管理有限公司为上海兴富普通合伙人、执行事务合伙人，持有上海兴富0.95%的出资份额，上海兴富直接持有发行人4.12%的股份。

截至本招股说明书签署日，除上述情况外，公司董事、监事、高级管理人员、核心技术人员及其近亲属未直接或间接持有发行人的股份；上述直接或间接持有的股份不存在质押或冻结的情况。

### （七）董事、监事、高级管理人员及其他核心人员的薪酬情况

#### 1、董事、监事、高级管理人员及核心技术人员的薪酬组成、所履行的程序

在公司担任具体经营管理职务的董事、监事、高级管理人员薪酬由基本薪酬、绩效奖金、效益奖金三部分组成，其他非高级管理人员的核心技术人员薪酬由基本薪酬、绩效奖金两部分组成。

发行人董事会薪酬与考核委员会为对上述董事、监事、高级管理人员进行考核以及确定年度薪酬分配的管理机构，其薪酬与考核以公司经济效益为出发点，由董事会薪酬与考核委员会进行综合考核，根据考核结果确定相关人员的年度薪酬分配。其他非高级管理人员的核心技术人员薪酬由公司人力资源部门按照其所在岗位的范围、职责、重要性以及绩效考核情况确定。

经股东大会审议，外部董事（未在公司任职的董事）及独立董事、外部监事享有固定数额的津贴。

#### 2、报告期内董事、监事、高级管理人员及核心技术人员薪酬总额占各期发行人利润总额的比重

报告期内，发行人董事、监事、高级管理人员及核心技术人员薪酬总额占发行人各期利润总额的比重情况如下：

项目	2020 年度	2019 年度	2018 年度
董事、监事、高级管理人员及核心技术人员薪酬总额(万元)	824.37	1,015.58	919.82
发行人利润总额（万元）	9,080.66	9,135.95	7,212.51

项目	2020 年度	2019 年度	2018 年度
占比	9.08%	11.12%	12.75%

### 3、董事、监事、高级管理人员及核心技术人员最近一年从发行人及其关联方领取收入的情况

姓名	任职	2020 年度薪酬（万元）
贺琳	董事长	66.68
唐涤飞	董事、总经理	155.44
李科	董事、副总经理	130.90
吕思遥	董事、副总经理、财务负责人、董事会秘书	127.34
黄宇凯	董事、副总经理、技术总监	94.12
志鹏	董事	6.00
仪晓辉	独立董事	6.00
张东晖	独立董事	6.00
杨涛	独立董事	6.00
张小龙	监事会主席	3.00
方彦彬	监事	3.00
熊敏	监事、人资行政总监	58.68
郝玉峰	副总经理、技术总监、首席科学家	57.80
曹琼	核心技术人员、计算机视觉业务部总监	57.50
邵志明	核心技术人员、高级研发工程师	45.91

注：外部董事志鹏、独立董事以及股东监事张小龙、方彦彬正常在其任职的其他公司领薪，享受对应待遇。

除上述情况外，公司董事、监事、高级管理人员和核心技术人员未在发行人及其关联企业享受其他待遇或退休金计划。

## 十、发行人本次公开发行申报前已经制定或实施的股权激励及相关安排

### （一）发行人员工持股平台情况

发行人于 2015 年 12 月设立中瑞安、中瑞立两家有限合伙企业作为员工持股平台，并以股权受让方式实施股权激励。截至本招股说明书签署日，中瑞安、中瑞立分别持有发行人 15.43%、5.83% 的股份。其基本情况、出资人和出资比例情



况参见本节“七、持有发行人 5%以上股份的主要股东及实际控制人情况/（二）其他持有本公司 5%以上股份的主要股东的基本情况”。

## （二）发行人股权激励实施情况

自 2015 年 12 月用于股权激励的员工持股平台设立以来,发行人股权激励有关的变动情况如下:

### 1、增加陈琪为激励对象

2016 年 3 月,贺琳与公司员工陈琪签订《出资份额转让协议》,贺琳向陈琪转让 94.7368 万元的中瑞立出资份额(对应发行人股份比例 0.90%),其中已实缴 9.4736 万元。本次出资份额转让价格为 9.4736 万元。

### 2、增加吕思遥为激励对象,调增李科出资份额,调减程显峰、张蕊出资份额

2016 年 8 月,发行人为了进一步明确管理人员职责及股权激励安排,对中瑞安员工持股情况作出调整安排:

转让方	受让方	转让出资份额 (万元)	对应发行人股 份比例 (%)	其中实缴出资 份额 (万元)	转让价格 (万元)
创世联合	李科	27.7778	0.50	4.1667	4.1667
创世联合	吕思遥	11.1111	0.20	1.6667	1.6667
程显峰	吕思遥	72.2222	1.30	10.8333	10.8333
张蕊	创世联合	88.8889	1.60	13.3333	13.3333

### 3、因程显峰离职,收回其出资份额

2019 年 1 月,程显峰由于个人原因离职。程显峰与创世联合签订了《出资份额转让协议》,程显峰向创世联合转让其所持有的 288.8889 万元出资份额(对应发行人股份比例 4.77%),其中已实缴 43.3334 万元,转让价格为 43.3334 万元。转让后,程显峰不再持有中瑞安出资份额。

除上述事项外,截至招股说明书签署日发行人无其他股权激励制定或实施安排。

## （三）股权激励实施对公司经营状况、财务状况、控制权变化等方面的影响

通过实施股权激励,发行人建立、健全了激励机制,充分调动了公司中高层

管理人员及骨干员工的工作积极性,对促进公司业务发展和人才队伍建设起到了积极作用。

发行人参考同期投资机构入股的价格,对上述员工持股调整事项在对应期间确认了股份支付费用。报告期内,公司不存在股权激励事项,也并无确认股份支付费用的情况,股权激励对于公司的财务状况影响较小。

同时公司历史上的股权激励事项对于公司的控制权认定并不产生影响。

#### (四) 上市后的股份锁定安排

关于员工持股平台股东股份锁定的承诺,参见本招股说明书“第十节 投资者保护/五、主要承诺/(一) 自愿锁定股份的承诺”。

## 十一、发行人员工情况

### (一) 员工数量及其变动

2018年末、2019年末及2020年末,发行人员工人数分别为127人、142人和151人。

### (二) 员工结构

#### 1、截至2020年12月31日,发行人员工专业结构情况如下:

类别	数量	占比
技术人员	59	39.07%
研发人员	44	29.14%
管理及行政人员	22	14.57%
销售人员	17	11.26%
财务人员	9	5.96%
<b>总计</b>	<b>151</b>	<b>100.00%</b>

#### 2、截至2020年12月31日,发行人员工学历结构情况如下:

类别	数量	占比
硕士及以上	56	37.09%
本科	87	57.62%
大专	8	5.30%

类别	数量	占比
总计	151	100%

### 3、截至 2020 年 12 月 31 日，发行人员工年龄结构情况如下：

类别	数量	占比
30 岁及以下	88	58.28%
31-40 岁	52	34.44%
41-50 岁	9	5.96%
51 岁及以上	2	1.32%
合计	151	100.00%

#### (三) 社会保险和住房公积金缴纳情况

发行人实行劳动合同制，根据《中华人民共和国劳动法》、《中华人民共和国劳动合同法》等国家及地方有关劳动法律、法规、规范性文件的规定聘用员工，与员工签订劳动合同，并按相关规定为符合条件的员工缴纳社会保险和住房公积金。报告期内，公司社会保险和住房公积金的缴纳情况如下：

项目	2020 年 12 月 31 日		2019 年 12 月 31 日		2018 年 12 月 31 日	
	已缴纳人数	占比	已缴纳人数	占比	已缴纳人数	占比
养老保险	149	98.68%	141	99.30%	122	96.06%
工伤保险	149	98.68%	141	99.30%	122	96.06%
失业保险	149	98.68%	141	99.30%	122	96.06%
医疗保险	149	98.68%	141	99.30%	122	96.06%
生育保险	149	98.68%	141	99.30%	122	96.06%
住房公积金	149	98.68%	138	97.18%	122	96.06%
员工总数	151		142		127	

截至 2020 年 12 月 31 日，公司为符合条件的员工足额缴纳了社会保险和住房公积金。社会保险和住房公积金人数不一致的主要原因为：1 名新入职员工，因已离职公司未减计其当月社保和公积金，因此当月其未于发行人处缴纳社保和公积金；1 名香港海天瑞声聘请的外籍员工，未在中国境内工作亦未在中国境内领取薪酬，无需在中国境内缴纳社保和公积金。

根据公司及各子公司和分公司所在地社会保险、住房公积金管理部门出

具的相关证明或凭证，发行人报告期内已依法足额缴纳社会保险和住房公积金，不存在因违反法律法规受到社会保险和住房公积金方面行政处罚的情形。

公司控股股东、实际控制人贺琳承诺：“北京海天瑞声科技股份有限公司及其子公司在公司首次公开发行股票并上市前如有未依法足额缴纳的任何社会保险或住房公积金等强制性社会保障费用，如果在任何时候有权机关要求公司和/或其子公司补缴，或者对公司和/或其子公司进行处罚，或者有关人员向公司和/或其子公司追索，本人将全额承担该部分补缴、被处罚或被追索的支出及费用，且在承担后不向公司和/或其子公司追偿，保证公司和/或其子公司不会因此遭受任何损失。”

## 第六节 业务和技术

### 一、发行人主营业务、主要产品或服务的情况

#### （一）主营业务概述

##### 1、概述

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。发行人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。发行人所属行业为软件和信息技术服务业。

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

海天瑞声是我国领先的训练数据专业提供商。自 2005 年成立以来，发行人始终致力于为 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。发行人所提供的训练数据覆盖智能语音（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、智能驾驶、智慧城市等多种创新应用场景。其中智能语音为发行人的优势领域，产品线已包含 160 余个主要语种及方言。

海天瑞声的产品和服务已获得阿里巴巴、腾讯、百度、科大讯飞、海康威视、微软、亚马逊、三星等国内外客户的认可，应用于其研发的个人助手、智能音箱、

语音导航、搜索服务、智能驾驶、机器翻译等多种产品相关的算法模型训练过程中。目前海天瑞声客户累计数量超 500 家，覆盖了主要的大型科技公司、人工智能企业及科研机构。



图：海天瑞声产品服务矩阵示意

在多年经营中，发行人高度重视自身核心技术优势的构建，在基础研究和产品研发两个维度持续投入。凭借多年研发积累，公司已经构筑了较为明显的核心技术先进性，主要体现在下述方面：

1、算法与数据处理技术并用：公司已掌握 12 项核心技术，通过算法与数据处理技术并用，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 2020 年 12 月 31 日，公司已积累 760 余个自有知识产权的训练数据产品，可提供超过 160 余个语种/方言的训练数据。

2、工具和平台共建：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。截至本招股说明书签署日，公司的产品/服务已覆盖 160 余个语种/方言，公司已积累下超过 100 个语种/方言的发音词典，累计词条数超过 1,000 万条，可构建高质量智能语音训练数据。

截至目前，公司已经取得 24 项发明专利、1 项实用新型专利授权及 1 项外观设计专利授权，134 项计算机软件著作权，另有 2 项专利已提交申请。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。

按照产品和服务的具体类型划分，发行人的主营业务可分为训练数据定制服务、训练数据产品和训练数据相关的应用服务。

金额单位：万元

类别	2020 年		2019 年		2018 年	
	金额	占营业收入比重	金额	占营业收入比重	金额	占营业收入比重
训练数据定制服务	11,729.91	50.26%	14,232.91	59.91%	12,369.55	64.20%
训练数据产品	10,613.71	45.48%	9,176.47	38.63%	6,601.67	34.27%
训练数据相关的应用服务	993.77	4.26%	346.44	1.46%	294.55	1.53%
<b>主营业务收入</b>	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>
<b>营业收入</b>	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>

## 2、产品服务内容、作业模式和业务实质介绍

### (1) 产品服务内容

海天瑞声是一家训练数据专业提供商，主要从事训练数据研发设计、生产及销售业务，所生产的训练数据主要运用于人工智能算法模型开发训练，产品服务形式包括训练数据定制服务、训练数据产品及相关的应用服务。

人工智能的目标是模拟人的思维和判断能力，为了实现这个目标的过程则需要对算法模型进行训练，如同学生需要通过不断的学习与训练，才能逐渐获得并提升处理判断生活中各种事件的能力，下游 AI 技术企业需要使用发行人的训练数据，对算法模型进行训练、教导。算法模型需要针对尚未学习过的模拟场景进行学习，或针对已经学习过的实际场景再进行不断迭代学习。针对下游客户算法

模型的上述需求，发行人以训练数据为核心提供定制服务、产品和相关的应用服务，充分满足客户需要。

发行人的产品和服务的交付物为成品训练数据集，内部包含数据文档、说明文档、技术文档，其中数据文档是交付物的核心，包括数据文件和标注文件。以语音识别训练数据为例，训练数据由一段段语音录音的数据文件和相关的标注文件构成，标注文件将告诉机器和算法与语音相关的多种信息，包括语音对应的文字、所代表的具体意思、说话者的情绪等等。

为了助力下游 AI 技术企业更好地使用发行人提供的训练数据，训练、教导算法模型，实现更好的训练效果，训练数据的生产过程中有很多学问：

以智能音箱为例，因为其使用者年龄跨度较大，可能从小孩覆盖到老人，使用者还可能说着各地口音，语言表达习惯也各有不同，使用的场景可能是家里的客厅或者办公室。所以发行人需要在训练数据集的设计过程中同时考虑方言的分布、不同说话人特征的分布、语言表达可能覆盖的文本分布、以及不同的声音传导环境等。因此，全面的数据集结构与内容的设计、逼真的采集环境搭建、合理的被采集人选择、准确的标注信息等，都将决定训练数据的质量，进而影响智能音箱最终的识别效果和智能音箱使用人群的体验感。

## （2）作业模式

总体而言，发行人的业务过程主要是下述几方面工作：①为了实现更好的算法模型训练效果，发行人需要对训练数据集的结构和内容进行设计；②需要采集大量数据，比如声音、文本、图像、视频；③需要对采集到的数据进行加工，把需要让算法学会的信息，通过标注的方式告诉算法；④而在采集、加工的过程中，发行人则要对质量进行把控，持续进行质检。因此形成了训练数据生产的四大环节：设计、采集、加工、质检。

## （3）业务实质

综上，发行人的业务实质是：通过执行设计、采集、加工、质检工作，生产训练数据并向下游 AI 产业链各机构提供，助力其训练算法模型。



## （二）主要产品及服务

报告期内，发行人的主要产品及服务具体情况如下：

### 1、主要产品及服务按业务类型分类

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。

根据具体业务模式差异，发行人提供的主要产品和服务又可分为下述三类：

#### （1）训练数据定制服务

发行人根据客户需求提供训练数据定制服务，具体形式包括：

①采集+加工服务：发行人根据客户需求设计训练数据集结构、组织原料数据采集、对采集到的原料数据进行加工，最终形成定制化训练数据集。该类业务主要满足算法模型对模拟场景的学习需要，数据规模越大、覆盖越广、质量越高则模拟度越强，训练效果越好。

②纯加工服务：客户提供原料数据，发行人根据客户需求进行加工，形成定制化训练数据集。该类业务主要服务于算法模型对实际场景的再学习需求，即发行人对实际场景数据进行加工，使算法模型可运用加工后的数据进行迭代学习，提升对实际场景的判断准确度。

在训练数据定制服务业务类型下，发行人为客户提供训练数据集定制服务，最终形成的训练数据集成品的知识产权由客户享有。

#### （2）训练数据产品

发行人根据市场需求，以及对算法技术应用前景、发展趋势的评估预判，开发训练数据产品，开发完成后授权客户使用。该类业务也主要满足算法模型对模拟场景的学习需要。

在训练数据产品业务类型下，发行人最终生产的训练数据集成品的知识产权由发行人享有，一次生产完成后可重复多次销售使用权。

#### （3）训练数据相关的应用服务

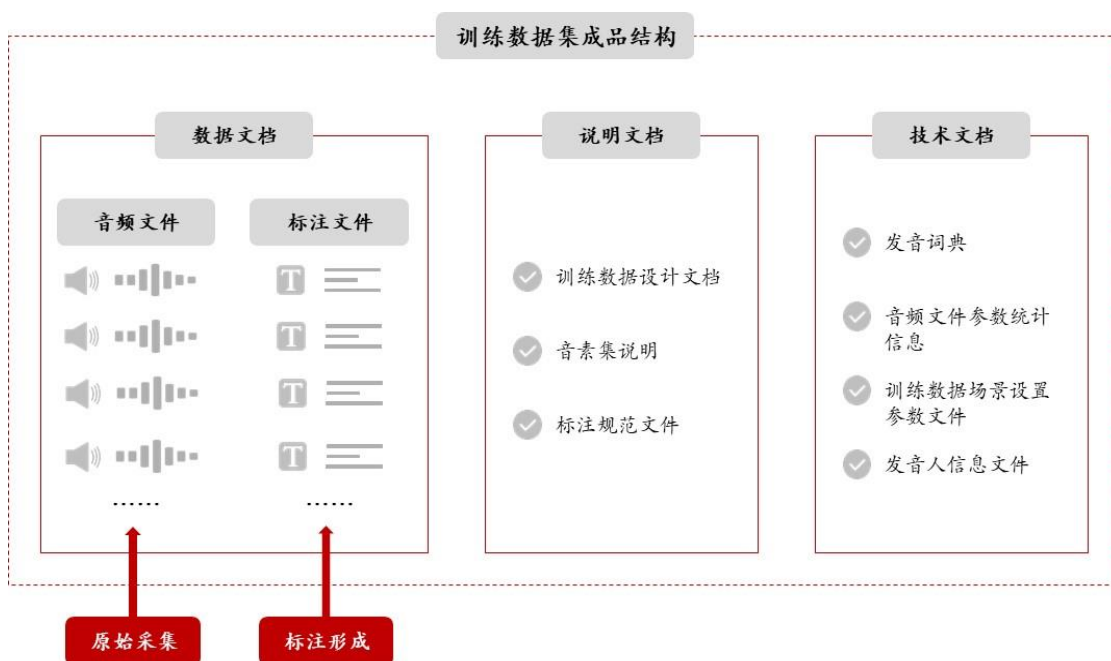
发行人基于其生产的训练数据提供算法模型相关的训练服务，运用训练数据研发能力助力下游客户完成其算法模型的语言拓展、特定算法模块拓展、垂直应

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。

训练数据相关的应用服务的具体内容包括：①依据客户需求，在客户通用算法模型的基础上，为客户定制开发特定语种/方言、应用领域的算法模型；②使用特定语种/方言、特定应用领域的训练数据，对客户的算法模型进行训练，使客户的算法模型实现语种/方言覆盖能力的拓展，满足其在特定行业、地区或者特定应用领域的运用需要；③客户负责研发/提供核心的算法模型，发行人负责开发算法模型实现后续落地运用过程所需的用户界面、软件等，及相关的系统集成工作。

如上所述，报告期内，发行人的主要产品、服务包含训练数据定制服务、训练数据产品及训练数据相关的应用服务三类。前述产品、服务均以发行人生产的专业训练数据集为核心。发行人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、主要产品或服务按下游应用领域分类

发行人研发、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。在智能语音、计算机视觉和自然语言领域，发行人可提供的细分产品/服务类别如下所示：

产品/服务类型	应用领域	产品/服务细分类别	
训练数据定制服务	智能语音	语音识别	语音识别采集
			语音识别转写
	语音合成		
	计算机视觉	计算机视觉	
	自然语言	自然语言	
训练数据产品	智能语音	语音识别	语音识别采集
			语音合成
	计算机视觉		
	自然语言		

人工智能技术的重点是让机器能够解决人脑所能解决的问题<sup>2</sup>。对于前述核心领域及发行人针对各领域提供的产品、服务简要介绍如下：

### （1）智能语音

#### ①智能语音技术简介

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。

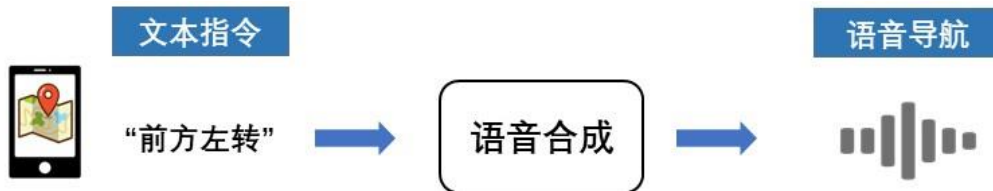
以日常生活中的情景为例，语音输入法、即时通讯软件运用了语音识别技术

<sup>2</sup> 清华大学《2018 人工智能发展报告》。

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。



图：语音识别应用示例——即时通讯软件将发送的语音信息转换为文字



图：语音合成应用示例——导航 APP 运用语音合成功能，将文本指令转换为语音输出，实现语音导航

## ②发行人在智能语音领域提供的细分产品服务

发行人提供的训练数据定制服务和训练数据产品均覆盖语音识别、语音合成两个技术领域。发行人通过设计（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）<sup>3</sup>、质检（对数据集进行质量检测，如音字一致性、标注准确率检查等）等训练数据集生产环节；或者针对客户提供的原料音频文件执行加工、质检工作，最终形成客户所需的智能语音训练数据。

特别地，语音识别领域的训练数据定制服务又可划分为语音识别采集、语音识别转写两个细分业务类别；语音识别领域的训练数据产品仅包含语音识别采集一个细分业务类别。语音识别采集、语音识别转写的具体业务内容如下：

语音识别采集指：发行人为客户提供设计、原料音频采集、加工、质检等全流程数据开发服务，形成语音识别所需的训练数据。语音识别采集业务形成的训

<sup>3</sup> 采集、加工具体内容请参见本招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（三）主要经营模式/2、生产或服务模式”中详细介绍。

练数据主要服务于语音识别模型在上线前<sup>4</sup>的训练需求，通过训练数据模拟目标应用场景下的数据情况，对算法模型进行训练。

语音识别转写是指：发行人基于客户提供的原料音频数据，执行转写、标注、质检等数据加工服务，形成与音频文件对应的、带时间戳和特征标签的文本，与音频文件一同构成语音识别算法所需的训练数据。语音识别转写业务形成的训练数据集主要服务于语音识别模型上线后的性能优化，通常是应客户需求，对其算法模型在实际运营中收集到的数据进行加工，用于算法模型上线后的迭代优化。

## （2）计算机视觉

### ①计算机视觉技术简介

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。



图：计算机视觉技术应用示例——人脸识别

### ②发行人在计算机视觉领域提供的细分产品服务

发行人提供的训练数据定制服务和训练数据产品覆盖计算机视觉领域。发行人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，拍摄自动驾驶场景视频等）、

<sup>4</sup> 上线包括初始功能上线、功能拓展上线、语言拓展上线等。

加工（对图像、视频文件进行打点、分割标注等）<sup>5</sup>、质检（对数据集进行质量检测，如检验图片、视频文件格式是否正确，检查光照环境、物体种类的数量是否达标，打点标框的准确率是否符合要求等）；或者对客户提供的图像、视频文件执行加工、质检工作，最终形成客户所需的计算机视觉训练数据。

### （3）自然语言处理

#### ①自然语言处理技术简介

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。

#### ②发行人在自然语言领域提供的细分产品服务

发行人提供的训练数据定制服务和训练数据产品覆盖自然语言领域。发行人通过设计训练数据集结构、采集（收集自然语言文本、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）<sup>6</sup>、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。

### 3、主要产品或服务的终端应用场景

发行人提供的高质量、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。发行人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、智能客服、机器

<sup>5</sup> 采集、加工具体内容请参见本招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（三）主要经营模式/2、生产或服务模式”中详细介绍。

<sup>6</sup> 采集、加工具体内容请参见本招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（三）主要经营模式/2、生产或服务模式”中详细介绍。

人、语音导航、智能播报、语音翻译、自动驾驶、智慧交通、智慧城市、机器翻译、智能问答、信息提取、情感分析、OCR 识别等多种应用场景。



图：发行人训练数据集服务的算法模型应用场景示意

#### 4、训练数据产品的数据资源泄露、盗版风险及发行人采取的保护措施情况

发行人的训练数据产品是以二进制形式存储，理论上存在被泄露或者盗版的可能，针对训练数据产品，发行人采取的保护措施如下：

(1) 外部：与客户就训练数据产品保护事项在商业层面进行条款约定

在与客户进行合作时，发行人首先会在商业合同条款中约定训练数据产品保护的相关条款，对此进行约束。其次，考虑到训练数据产品的泄露或者盗版会对发行人的客户也可能造成直接的利益损失（如增加其竞争对手等），因此发行人的客户也有动机主动保护训练数据产品不被泄露。

(2) 内部：采用技术保密手段、制定数据保密相关制度对训练数据产品加以保护

①信息安全管理体系层面——已取得了 ISO/IEC 27001 信息安全管理体系认证

发行人构建了完善的信息系统维护管理制度，确保信息系统能够对数据安全起到有力保障：通过信息系统的日常运行维护、系统更新、安全管理等手段保障服务器安全、应用系统安全、网络安全、物理安全，满足开发管理、防病毒管理、应急管理等方面要求；组织开展对信息系统的安全评估，发现危害数据、信息安全的问题及时整改；定期进行数据备份、软硬件更新，保障数据安全、降低数据

流失风险；严禁擅自修改系统软件和应用软件，确需修改需履行审批申请程序，由信息管理员负责，并在维护过程中防止数据丢失、失效。

②技术层面——构建一体化数据处理平台（以下简称“平台”），嵌入业务过程的数据安全管理需求；运用技术手段保障数据安全

对于在海天瑞声的平台上处理的数据，发行人的一体化数据处理平台可实现下述功能，在数据采集、传输、加工、存储过程中确保数据安全：

A.可实现私有化部署，加工、质检过程中数据源均保存在海天瑞声的服务器中，不允许随意复制、外泄；

B.数据加密存储：对所有存储在平台的数据进行加密存储；涉及移动设备（如移动硬盘、U 盘等）外部移动数据存储的也进行加密处理，并制定了严格的移动设备管理制度；

C.数据加密传输：对于平台之间，或平台、客户端之间的数据传输，运用行业通行的加密方式对数据进行加密传输；

D.嵌入访问权限控制、身份验证、密码管理功能，限制数据的查看、使用范围，保障数据安全。

对于在客户平台上处理的数据：发行人及其团队可依据客户需求运用客户指定的工具、平台对数据加以处理，严格防范数据外泄。

③内部控制层面——强化职责分离、权限管理

与上述信息系统安全管理相关举措配合，发行人制定、采取了有效的内控制度及举措，防止因内控弱化导致的数据被随意控制、操作的风险，具体包括：加强信息系统的职责分离、权限管理，加强与信息系统使用、授权相关的制度建设，完善授权审批管理流程等。

④保密制度层面——签署保密协议，制定保密要求、泄密惩治机制

发行人构建了有力的保密制度，对于数据、技术等各类保密信息制定了相应的保密要求，与员工签署保密协议，约定保密义务；辅以保密教育、泄密惩治机制，确保员工遵守保密要求。



综上，发行人的训练数据产品存在数据资源被泄露、盗版等的可能，发行人采取了有效的保护措施对其进行保护，防止相关泄露、盗版情形发生。但若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，则公司可能面临训练数据产品被泄露、盗版等数据安全相关风险，详见招股说明书“重大事项提示/三、特别风险提示/（八）数据安全相关风险”。

### （三）主要经营模式

#### 1、盈利模式

与发行人的主要产品及服务类型对应，发行人的盈利模式主要包括以下三类：

（1）训练数据定制服务：发行人根据客户需求提供训练数据定制服务并收取服务费。在此种模式下，发行人仅享有服务费收入，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。

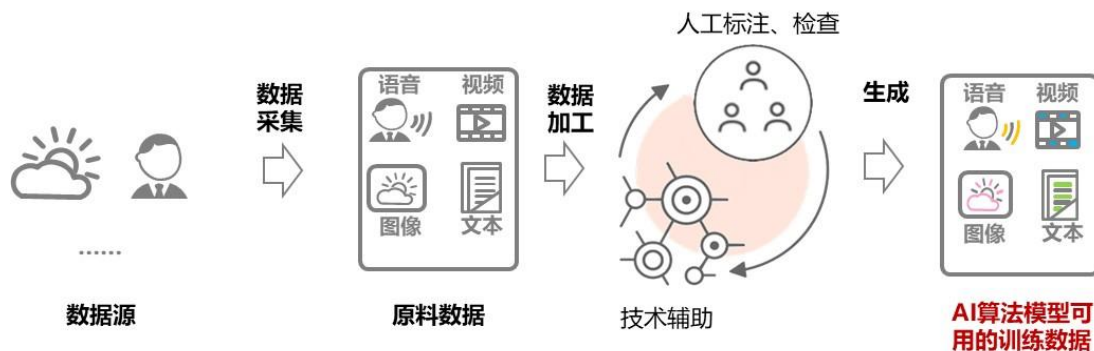
（2）训练数据产品：发行人开发自有知识产权的训练数据产品，通过销售训练数据产品的使用授权许可，获取让渡资产使用权收入。此类训练数据集一经开发完成，可多次销售并获取授权许可收入。

（3）训练数据相关的应用服务：发行人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，以及极少量硬件销售收入。

#### 2、生产或服务模式

（1）训练数据生产模式（训练数据定制服务、训练数据产品）

训练数据定制服务、训练数据产品的生产或服务流程基本相同，本质均为训练数据的研发、生产流程。在此过程中，发行人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

发行人的训练数据生产过程主要包括四个环节：设计（训练数据集结构设计）、采集（获取原料数据）、加工（数据标注）及质检（各环节数据质量、加工质量检测）。不同领域、产品服务类型所涉及的生产环节具体如下：

产品/服务类型	涉及的训练数据生产环节			
	设计 (训练数据集结构设计)	采集 (获取原料数据)	加工 (数据标注)	质检 (各环节质量检测和控制)
训练数据定制服务				
-采集+加工服务	√	√	√	√
-纯加工服务			√	√
训练数据产品	√	√	√	√

对于上述 4 个生产环节具体介绍如下：

#### ①设计——训练数据集结构设计

发行人根据客户需求以及自身专业经验研判，设计拟开发的训练数据集结构。在该环节中，发行人需要考虑算法模型的具体应用领域、应用场景以及预期实现的训练效果，从而反过来确定训练数据集内的数据类型、数量、比例分布等，相应确定原料数据的采集要求，为后续采集工作奠定基础。

以语音识别、语音合成领域的训练数据集为例，在原料数据的采集环节，发音人（被采集对象）需要朗读发行人提供的基础语料，并用指定的录音设备录制以形成原料音频数据。因此，在设计阶段，发行人就需要考虑如何设计基础语料，才能使得容量有限的训练数据集能够覆盖尽可能多的自然语言现象，如覆盖更多的发音习惯、语言特点、句长分布，达到更好的音素平衡效果等，从而使得算法模型获得更好的训练结果——使得语音识别模型能够“听懂”更多实际场景中的

语音数据；使得语音合成模型能够“说出”更多、更为自然的语音效果。

## ②采集——获取原料数据

发行人根据设计好的训练数据集结构及数据量目标，详细制定原料数据采集方案并组织开展原始数据采集工作。采集过程所涉及的主要考虑因素包括：

**A.数据量方面：**发行人根据成品训练数据集的目标数据量，预留少量冗余，确定需要采集的原始数据量

发行人的成品训练数据集结构清晰，原始采集形成的原料数据、与原料数据对应的标注文件在数据集中独立存储，最终交付的训练数据量与原始采集数据量存在对应关系。

因此，在采集环节中，发行人所需采集的数据量与成品训练数据集的目标数据量存在一定的对应关系。在实际采集过程中，由于可能发生少量录音不合格的损耗情况，发行人会在总采集数据量中预留少量冗余，从而略大于最终要交付的数据量，以备替换偶然出现的不合格录音数据。

**B.数据属性方面：**发行人根据应用场景、应用领域确定具体的采集对象

在采集环节中，发行人需要根据客户算法模型应用的目标场景、领域等个性化需求，采集特定原料数据，为后续加工生产训练数据做准备。

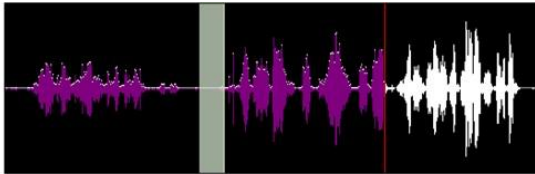
以语音识别训练数据为例，在采集环节中，发行人需要根据语音识别模型的语种/方言类别（汉语、带特定地区口音的汉语、各地方言；日语、法语等）、目标应用场景（安静、噪音；家居、车载等），相应定义并寻找符合要求的发音人，在合适的采集场景下由发音人朗读、或自然说出录制语音片段，生产原料音频数据。



以语音合成训练数据为例，发行人需要根据客户对拟合成的语音的风格（温柔、甜美、科技感等）、年龄（成人、儿童）、性别、语种、口音等方面的具体需求寻找发音人，并组织发音人按照前期设计完成的音素集、语料库等资料进行朗读，录制生成原料音频数据。此外，由于语音合成训练数据的录制对信噪比、底噪、录音棚混响时间等参数、指标和录音设备的要求很高，通常需要在专业级别的录音棚中完成录制工作。

以计算机视觉训练数据为例，发行人需要根据计算机视觉算法模型的具体应用场景（如人脸识别等）的需求，寻找合适的人脸、动作采集对象，组织采集对象按照要求进行人脸、动作的视频录制。

③加工——数据标注

发行人通过自行采集或者客户提供，获得加工所需的原料数据，并进行后续加工处理。对于不同应用领域的训练数据，具体加工操作说明如下：

训练数据类型	采集环节获得的原料数据形式	加工环节执行的主要具体操作	加工环节示例	成品训练数据集形式								
智能语音领域												
-语音识别	语音片段音频文件	<p><b>转写：</b>通过软件预标注+人工校对听写，形成与音频文件对应的、带时间戳的文本等</p> <p><b>属性标注：</b>标注原始音频的发音人性别及声音属性等</p> <p><b>截取、切分：</b>定位声音起止点、并保留一定长度的前后静音段，对音频文件做切分</p> <p><b>校对：</b>通过语音识别技术、人工听辨等核对转写、标注、切音等工作的正确性</p>	<div style="display: flex; align-items: center;"> <div style="margin-right: 10px;">原始音频数据</div>  </div> <div style="display: flex; align-items: center;"> <div style="margin-right: 10px; writing-mode: vertical-rl;">加工环节示例</div> <div> <p><b>声音属性标注</b></p> <ul style="list-style-type: none"> <li>■ 性别 <span style="float: right;"><input checked="" type="checkbox"/>男 <input type="checkbox"/>女</span></li> <li>■ 录音片段是否有效 <span style="float: right;"><input checked="" type="checkbox"/>是 <input type="checkbox"/>否</span> (判断空白、间隔、噪音程度等是否符合模型训练要求)</li> </ul> <p><b>语音片段转写</b> (将相应的语音内容转换为对应的文本)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%;">时间</th> <th style="width: 50%;">文本</th> </tr> </thead> <tbody> <tr> <td>0.00-2.42</td> <td>您好我是朱新请问有什么可以帮您</td> </tr> <tr> <td>2.42-5.64</td> <td>我要投诉你们有一个叫xx的</td> </tr> <tr> <td>5.64-8.48</td> <td>请问您遇到什么问题呢先生</td> </tr> </tbody> </table> </div> </div>	时间	文本	0.00-2.42	您好我是朱新请问有什么可以帮您	2.42-5.64	我要投诉你们有一个叫xx的	5.64-8.48	请问您遇到什么问题呢先生	原始音频文件+标注文件+技术文档+词典文件
时间	文本											
0.00-2.42	您好我是朱新请问有什么可以帮您											
2.42-5.64	我要投诉你们有一个叫xx的											
5.64-8.48	请问您遇到什么问题呢先生											

训练数据类型	采集环节获得的原料数据形式	加工环节执行的主要具体操作	加工环节示例	成品训练数据集形式
-语音合成	语音片段音频文件	<p><b>标注：</b>针对语音片段，相应标出其语音语调、音素、韵律、音节边界、词性（动词、名词）等属性；</p> <p><b>截取、切分：</b>定位声音起止点、并保留一定长度的前后静音段，对音频文件做切分</p> <p><b>校对：</b>通过语音识别技术、人工听辨等核对发音人朗读、录制的语音判断是否与所提供的语料文本等一致。</p>	<p>原始音频数据</p>  <p>文本校对 当然，爱笑的人运气都不会差</p> <p>拼音标注 dang1 ran2 ai4 xiao4 de5 ren2 yun4 qi4 dou1 bu2 hui4 cha4</p> <p>文字边界标注 sil   当   然   sp   爱   笑   的   人   sp   运   气   都   不   会   差   sil</p> <p>音素边界标注 sil   d   ang1   r   an2   sp   ai4   x   iao4   de5   r   en1   sp   r   yn4   q   i4   d   ou1   b   u2   h   ui4   ch   a4   sil</p> <p>韵律标注 sil             当然#3                   爱笑的#1                   人#2                   运气#2                   都#1                   不会#1                   差#4                   sil</p> <p>词性标注 sil             当然/d                   爱/v                   笑/v                   的/u                   人/u                   运气/u                   都/d                   不/d                   会/v                   差/a                   sil</p>	原始音频文件+音素标注文件+韵律标注文件+音素边界标注文件+词性标注文件+技术文档
计算机视觉	图像、视频等	<p><b>属性标注：</b>标注图像、视频数据的客观属性、关键特征等（如轮廓位置、人脸关键点）</p> <p><b>语义分割：</b>标出属于同一个物体、同一个类别的像素</p> <p><b>主观标注：</b>对图像、视频的内容进行主观理解标注</p>	 <p>加工环节示例：<b>图像语义分割</b>——对不规则的图片进行区域划分、标注属性，如自动驾驶场景图片分割等，可应用于自动驾驶、智能安防等算法模型训练</p>	原始数据文件+标注文件+技术文档

训练数据类型	采集环节获得的原料数据形式	加工环节执行的主要具体操作	加工环节示例	成品训练数据集形式
自然语言	对话文本等	<p><b>文本清洗:</b> 文本正则化(即字符转换, 将符号、数字等转换为正确发音对应的文字); 移除表情符号; 去除重复的字词; 去除包含敏感词的字句等</p> <p><b>词法标注:</b> 对清洗后的文本进行分词(将完整句子分割为各个单词); 词性标注(动词、名词、形容词等)标注; 实体(人名、地名、机构名等)标注</p> <p><b>语义、语法、语用相关标注:</b> 句法分析(标注短语结构、依存关系<sup>7</sup>等); 情感属性标注(标注目标属性词以及对应的情感词、情感极性)</p>	<p><b>原始文本数据</b>      当地渔民还向他们反映, 曾看到死亡江豚的尸体。</p> <p><b>加工环节示例</b> {</p> <p><b>分词</b>                      当地 渔民 还 向 他们 反映, 曾 看到 死亡 江豚 的 尸体。</p> <p><b>词性标注</b>                s   n   d   p   r   v   d   v   a   n   u   n 处所词 名词 副词 介词 代词 动词 副词 动词 形容词 名词 助词 名词</p> <p><b>情感属性标注</b>        <input type="checkbox"/>正面情绪        <input type="checkbox"/>负面情绪        <input checked="" type="checkbox"/>普通情绪</p>	原始数据文件+标注文件+技术文档

<sup>7</sup> 依存关系: 句子中各个词语在句法上的搭配关系。

#### ④质检——各环节数据质量检测

除上述环节外，质检环节也是整个训练数据生产流程中必不可少的重要环节，该环节贯穿训练数据生产的完整流程。发行人结合其专业经验，已形成了有效的质量检测和质量控制体系。发行人自主开发了一体化数据处理平台，并在该平台上整合了采集、加工、质检环节所需的软件工具和模块，将质量检测和理念嵌入到各环节工具之中：

A.在前端采集环节，发行人开发的采集工具可对原始数据质量进行即时质检，不符合要求的原始数据不被计入采集数据之中；

B.在中端加工环节，发行人运用自动标注工具+人工校对检验的方式对数据加工情况进行检查，提升加工效率和准确度；

C.在后端大规模质检环节，发行人运用全自动校验技术，实现大规模训练数据集的质检需求。

(2) 训练数据生产（训练数据定制服务、训练数据产品）过程的业务流、数据流和财务流的流转情况

##### ①发行人业务流、数据流和财务流的主要内容

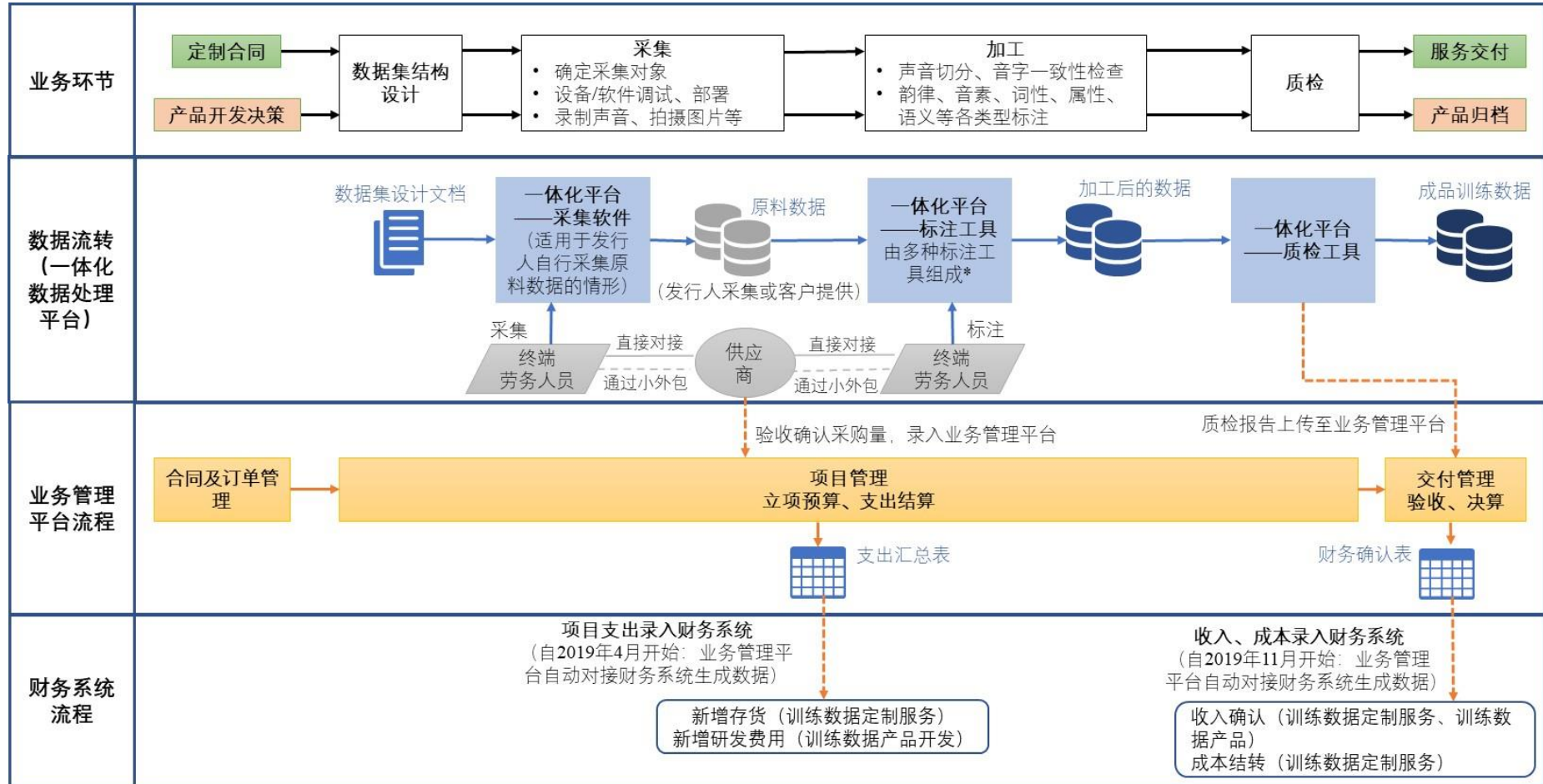
A.主要业务流包括：签署合同/订单或确定数据库产品开发决策，设计、数据采集、加工、质检，交付验收。

B.主要数据流包括：设计环节生成数据集设计文档，采集环节生成原料数据，发行人采集或客户提供的原料数据经加工环节生成加工后的数据，再通过质检环节后形成可交付的成品训练数据。上述流程涉及对外采购数据服务的为：原料数据的采集、标注环节。

C.发行人通过业务管理平台进行合同及订单管理、项目预算、采购支出申请、供应商对账、支出审批、项目验收和决算等，并与财务系统衔接。相应地，在财务系统中，在项目支出发生时确认存货（训练数据定制服务）或研发费用（训练数据产品开发），在项目验收时确认收入（训练数据定制服务、训练数据产品）并结转成本（训练数据定制服务）。



②发行人业务流、数据流和财务流的流程图



\*注：部分项目由标注人员在客户提供的标注平台上完成。

对发行人主要业务环节、数据流转、业务管理平台流程、财务流程，以列表形式进行如下进一步详细说明：

业务环节	数据流转	业务管理平台流程	财务流程
<p>对于训练数据定制服务：签订合同/获得订单</p> <p>对于训练数据产品：确定产品开发决策</p>	<ul style="list-style-type: none"> <li>◆ 无</li> </ul>	<ul style="list-style-type: none"> <li>◆ 根据合同/订单对应建立项目，并将项目信息录入业务管理平台；对于训练数据产品开发项目，根据产品开发决策建立项目并将项目信息录入业务管理平台</li> <li>◆ 核对项目与书面合同的核心内容一致性，包括客户名称、销售量、销售单价等，确保录入业务管理平台内的项目信息真实、准确</li> <li>◆ 项目立项（合同/订单与项目一一对应，确保可按项目对收入、支出进行独立的归集和核算）</li> <li>◆ 项目预算申请及审核：为成本核算完整性、准确性的重要控制节点（一般项目为三级审批制，重大项目为五级审批制）：               <ul style="list-style-type: none"> <li>✓ 核对预计采集数据量、标注数据量<sup>注</sup>，与合同/订单的交付量之间进行核对，保证预算采购量的真实、完整、准确</li> <li>✓ 对比同类项目间的采集产出比、标注产出比，保证预计采集数据量、标注工作量的合理、完整</li> <li>✓ 对比业务采购价格体系，保证预算中采集、标注单价的合理性</li> <li>✓ 测算预算毛利率，复核项目盈利能力的合理性</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>◆ 无</li> </ul>
<p>设计</p>	<ul style="list-style-type: none"> <li>◆ 生成“数据集结构设计文档”，以语音识别采集类项目为例，设计文档内容包括预设文本、发音人特征分布表（如年龄段、性别、口音区等），以及发音词典等。</li> <li>◆ 预设文本是发音人的说话素材，每个发音人的预设文本均不同，设计多少个发音人，就有多少个预设文本</li> </ul>	<ul style="list-style-type: none"> <li>◆ 无</li> </ul>	<ul style="list-style-type: none"> <li>◆ 无</li> </ul>

业务环节	数据流转	业务管理平台流程	财务流程
采集	<ul style="list-style-type: none"> <li>（以语音识别采集项目为例） 发音人根据设计的文本及场景，在发行人提供的设备及软件上进行录音，发音人的原始声音生成“原料数据”</li> <li>每个发音人对应一组声音文件</li> </ul>	<ul style="list-style-type: none"> <li>与供应商进行原料数据采集与标注工作验收，确认原料数据采集量、标注工作量<sup>注</sup>及其价格</li> <li>把实际原料数据采集量、标注工作量录入业务管理平台，进行支出报销申请</li> <li>要求终端劳务人员签字确认其提供的劳务工作量</li> <li>系统根据支出申请生成验收结算单，并发送给供应商，供应商确认原料数据采集量、标注工作量及对应价格和金额</li> <li>人工复核采集、标注量，确保系统中数据服务采购量的完整、准确，重点包括以下信息的交叉比对： <ul style="list-style-type: none"> <li>✓ 设计文档数量对比原料数据采集量</li> <li>✓ 原料数据文件个数对比原料数据采集量</li> <li>✓ 标注数据量对比原料数据采集量</li> <li>✓ 原料数据采集量、标注量对比供应商验收结算单记载数量</li> <li>✓ 原料数据采集量、标注量对比预算</li> <li>✓ 采集产出比、标注产出比对比预算</li> <li>✓ 采集、标注单价对比预算</li> </ul> </li> <li>支出申请审批通过后，系统生成“支出汇总表”</li> </ul>	<ul style="list-style-type: none"> <li>核算财务岗根据“支出汇总表”每月将项目支出录入用友财务管理系统；自2019年4月起，业务管理平台自动对接财务系统生成数据。</li> <li>全部录入完毕后，运营财务岗将财务系统中记录的项目支出导出，与“支出汇总表”进行反向核对</li> <li>根据核对无误的结果，确认存货或研发费用</li> <li>根据核对无误的结算金额向供应商付款</li> </ul>
加工	<ul style="list-style-type: none"> <li>标注人员在发行人标注工具上，对“原料数据”进行检查和标注，生成“标签数据”</li> <li>对“标签数据”进行标注质量抽查</li> </ul>	<ul style="list-style-type: none"> <li></li> </ul>	
质检	<ul style="list-style-type: none"> <li>对“标签数据”进行全面检查，形成质检结论，使“标签数据”成为通过质检的、可交付的“成品数据”</li> </ul>	<ul style="list-style-type: none"> <li>提交质检报告至业务管理平台</li> </ul>	<ul style="list-style-type: none"> <li>无</li> </ul>
交付/验收/决算	<ul style="list-style-type: none"> <li>向客户交付“成品数据”，客户对其进行验收确认</li> </ul>	<ul style="list-style-type: none"> <li>获取客户验收文件</li> <li>将验收信息录入业务管理平台</li> </ul>	<ul style="list-style-type: none"> <li>核算财务岗根据“财务确认表”将收入、成本信息录入财务管理系统</li> </ul>

业务环节	数据流转	业务管理平台流程	财务流程
		<ul style="list-style-type: none"> <li>◆ 提交项目决算，即结项申请</li> <li>◆ 复核验收、决算信息，保证验收金额的真实、准确，项目支出的准确、完整。复核内容主要包括：               <ul style="list-style-type: none"> <li>✓ 验收量对比验收文件</li> <li>✓ 验收量对比合同/订单中的销售量</li> <li>✓ 验收单价对比合同/订单中销售单价</li> <li>✓ 对比预算，整体复核项目决算中覆盖的生产环节、各环节的数量是否完整</li> <li>✓ 决算支出对比预算支出，偏离±10%需提交项目决算分析表</li> </ul> </li> <li>◆ 验收与决算审批通过后，系统生成“财务确认表”，生成收入、成本。</li> </ul>	<p>统。自 2019 年 11 月起，业务管理平台自动对接财务系统生成数据。</p> <ul style="list-style-type: none"> <li>◆ 全部录入完毕后，运营财务人员将财务系统中记录的收入、成本导出，与“财务确认表”进行反向核对</li> <li>◆ 根据核对无误的结果，确认营业收入、营业成本</li> </ul>

注：标注数据量指需要进行标注的数据量；标注工作量指对数据执行标注工作所花费的工作小时。

### （3）训练数据相关的应用服务模式

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

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

## 3、采购模式

### （1）采购模式

按照采购的内容及主体划分，发行人的采购包括：

①数据服务采购：发行人在数据采集、加工环节中，向人力资源服务公司等采购的，非核心技术环节的原料数据采集、标注服务。

②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。

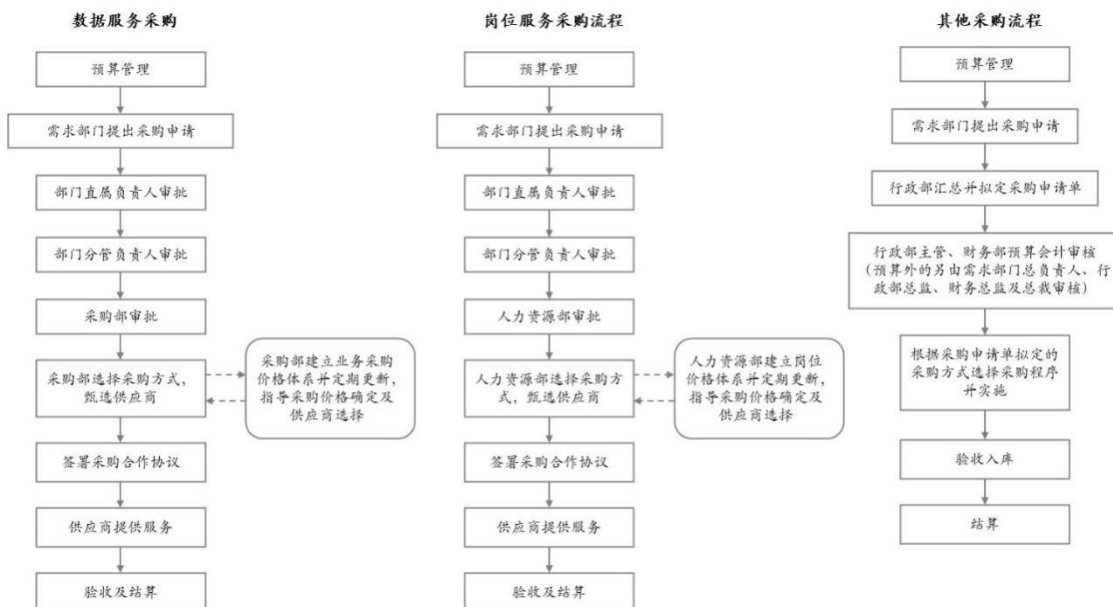
③其他采购：（1）训练数据生产所需的资产，主要包括软、硬件设备及其他需求物品采购；（2）日常运营所需的资产及物品，如办公用房、车辆、办公家具、计算机设备等；（3）日常专项服务采购等，主要包括审计服务、会议服务、差旅服务等。

上述原料数据采集、加工环节所涉及的数据服务采购，为发行人最主要的采购类别，由采购部负责；各部门岗位服务采购由人力资源部负责；其余日常运营相关的资产物品采购、专项服务采购等非业务采购由行政部负责。财务部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《采购管理制度》、《业务采购实施细则》、《岗位服务采购实施细则》等，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。

### （2）采购流程

发行人的采购流程图具体如下：



图：采购流程图

(3) 定价模式

①数据服务采购

该类采购主要采取成本导向定价法，由供应商根据发行人采购的原料数据采集、标注相关服务发生的成本，在成本基础上收取一定比例的服务管理费加成确定。训练数据定制服务、训练数据产品的数据服务采购模式无差异。按照采集、加工两个环节，采购业务的定价方式和过程如下：

A.采集

定价方式：采集环节采购的主要是音频、图像、文本等原料数据的采集服务，通过此类采购，发行人可获得后续加工所需的基础原料数据。发行人向供应商采购采集服务的采购总价=采集数据量×采集单价。采集单价根据市场上类似服务的一般价格，并参考采集对象资源的稀缺性来确定。例如，外语种、儿童或老人、行车环境等要求较为特殊的采集项目，采集资源有一定稀缺性，则采集单价高于一般项目。采集数据量根据具体原料数据的来源和类型，单位也有所不同，常见单位包括人、字/词、张等。

定价过程：确定采集方案→向供应商发出采集需求→综合考虑市场价格和同类采购的过往采购价格→确定采集价格。

B.加工

定价方式：加工环节采购的是为原料数据进行人工标注的服务。发行人向供应商采购标注服务的采购总价=有效工时×标注单价。标注单价根据市场上类似服务价格，并参考标注任务的难度来确定。有效工时根据发行人对标产比<sup>8</sup>的测速结果，与采购对象达成一致后确定。有效工时=标注数据量×标产比。

定价过程：确定加工方案→向供应商发出标注需求→综合考虑市场价格和同类采购的过往采购价格，确定项目标产比→确定标注价格。

### C.定价公允性

发行人数据服务采购定价由两部分组成，一部分是供应商支付给终端服务人员的费用，一部分是供应商就其提供的组织和管理服务收取的管理服务费。

第一部分的定价是随着整体市场人力成本的变动而变动，发行人按人或按工作小时等支付的单价整体上符合市场上对类似服务价格的定位。发行人在业务实施过程中，未发生在现行价格体系下，因价格过低造成服务资源不足的情况。根据发行人内部管理制度《海天瑞声业务采购实施细则》中，对数据服务采购价格的公允性做出了规定：采购部对业务价格体系中的价格进行定期公允性检查，以确保供应商价格在合理范围内，检查方式包括但不限于同类供应商询价、供应商及其内部人员访谈、发布模拟招聘信息询价、波动原因解析等。

第二部分的定价是根据供应商提供服务的能力不同而有所差异，目前的管理服务费率是提供类似服务的人力资源公司或技术服务公司通行的收费水平，主要涵盖其项目人员薪酬、人员招募与管理、公司税费等成本，符合该类公司提供服务的技术含量和定位。

②岗位服务采购及其他采购：该部分采购主要涉及业务过程及日常经营过程所需的软、硬件设备等其他物品及少量岗位服务等，相关领域供应商竞争充分，发行人采购的价格主要依据市场通行价格确定。

#### (4) 数据服务采购量确定机制

发行人的数据服务采购主要涉及采集、加工两个业务环节。其中采集环节采购的主要是音频、图像、文本等原料数据的采集服务，通过此类采购，发行人可

---

<sup>8</sup> 标产比：即标注产出比，指完成单位数据量的标注所需要的时间。以语音识别训练数据为例，如对某录音时长为1小时的原料音频文件进行标注，平均约需要3小时才可完成，则该项目的标产比为3。

获得后续加工所需的基础原料数据；加工环节采购的是为原料数据进行人工标注的标注服务。对于采集服务，发行人以采集数据量为单位与供应商结算；对于标注服务，发行人以有效工时为单位与供应商结算。

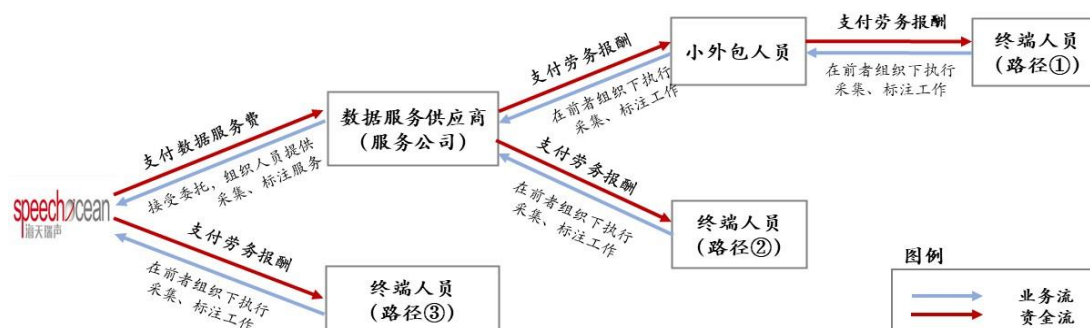
发行人训练数据定制服务和训练数据产品采购环节的数据量与交付给客户的数据量之间有密切联系。以采集环节为例，发行人会根据经验判断，在满足客户交付数据量要求的基础上留出少量冗余，例如某“200人新加坡英语手机3通道语音识别项目”，客户订单要求录音人数为200人，实际采集人数为206人。

#### 4、数据服务采购情况介绍

数据服务类采购为发行人业务经营中最主要的采购类别，对此类别采购详细说明如下：

(1) 数据服务采购模式的合理性——主要涉及非核心技术环节的采购、符合行业惯例，具备合理性

基于下游客户需求和发行人所从事的训练数据业务特点，发行人在生产训练数据过程中需要大量劳务人员提供原料数据采集、标注服务。发行人通过对接数据服务供应商的方式解决劳务用工需求，数据服务供应商主要为人力资源外包服务公司，具备寻找符合发行人用工需求的终端人员并组织其向发行人提供数据采集、标注服务的专业能力。在部分具体训练数据生产项目中，数据服务供应商等也采取与不同类型终端人员的组织人员（即“小外包人员”）对接的方式，组织终端人员为发行人提供相应服务。发行人与数据服务提供商、小外包人员、终端人员的采购模式如下图所示：



图：发行人数据服务采购模式示意图

发行人负责设计训练数据集结构，确定训练数据集开发标准，制定原料数据



采集方案，确定数据采集、标注标准和规程并提供数据采集、标注过程中所需的技术、平台和工具，组织采集、标注工作的具体开展，对采集、标注形成的数据质量进行检测、验证；人力资源服务公司等供应商负责按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员，组织其使用发行人研发的技术、平台、工具等，为发行人提供合乎标准的数据采集、标注服务。在部分具体训练数据生产项目中，涉及的终端采集、标注人员数量较多，且覆盖的语种、口音区、年龄特征等要求多种多样，再加之发行人劳务用工通常具有较为迫切的及时响应需求，人力资源服务公司等也采取与不同类型终端人员的组织人员（即小外包人员）对接的方式，来组织终端人员为发行人提供相应服务。

报告期内，在个别业务中，发行人存在向个人供应商采购原料数据采集、标注服务的情形。该等向个人供应商采购的金额较低，占比很小，并非发行人主要采用的采购模式，而是作为发行人在个别业务中的少量补充性采购，满足发行人的零散数据服务采购需求，具备商业合理性，并且不对发行人业务经营构成重大影响。

上述采购模式符合发行人的业务特点和行业惯例，具备合理性，分析如下：

①上述采购不涉及发行人业务的核心技术环节：发行人业务的核心技术环节主要体现在训练数据集结构设计，训练数据集开发标准制定，数据采集加工过程中的技术研究、工具开发，训练数据集质量标准制定、质量检测和控制在。原料数据的具体采集和标注等过程为非核心技术环节，主要为基础性、重复性的人工操作。

②上述采购符合行业惯例：原料数据采集、标注工作的需求量随业务类型结构、训练数据集所需的数据量等有所波动，且所需人工较多。行业内同类企业在业务经营中均普遍专注于训练数据设计、技术工具开发、项目流程管控、产品服务质检等重点方面。因原料数据的采集与标注过程呈现执行人员数量较大、临时性较强、流动性较高的特点，业内企业也多采取对外采购外包服务的形式执行。

综上，为避免人员冗余、合理控制管理成本，并将主要精力集中在数据设计、技术研发和质量控制等核心技术环节，发行人采取向人力资源服务公司等采购原

料数据采集与标注服务的方式。前述采购行为符合行业惯例，具备合理性。

(2) 发行人与服务公司、小外包人员、终端人员之间的合作情况

①合作模式

发行人与服务公司、小外包人员和终端人员共同合作完成原料数据的采集和标注工作，具体职责划分如下：

A. 发行人职责

- a. 设计训练数据集结构，确定训练数据集开发标准，制定原料数据采集方案；
- b. 确定数据采集、标注标准和规程并提供数据采集、标注过程中所需的技术、平台和工具；
- c. 监督采集、标注工作的具体开展；
- d. 对采集、标注形成的数据质量进行检测、验证。

B. 服务公司职责

a. 按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员。在部分具体训练数据生产项目中，涉及的终端采集、标注人员数量较多，且覆盖的语种、口音区、年龄特征等要求多种多样，服务公司也采取与不同类型小外包人员对接的方式，来组织终端人员为发行人提供相应服务；

b. 组织上述人员使用发行人研发的技术、平台、工具等，并对服务进度与交付进行管理，保证该等人员提供服务的质量；

c. 对小外包及直接对接的终端人进行人员管理，包括劳务合同签署、进度安排、支付款项、代扣代缴税款等。

C. 小外包人员职责

a. 按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员；

b. 组织终端人员为发行人提供相应服务。

D. 终端人员职责

- a.为发行人提供合乎标准的数据采集、标注服务；
- b.授权发行人使用其提供的数据用于业务经营。

②合同签署过程、实际业务执行过程

A.发行人与服务公司之间的合同签署及实际业务执行情况

a.发行人与服务公司等长期签订框架性的数据服务采购协议，约定发行人根据业务需要委托服务公司完成原料数据采集、标注工作，双方根据验收结算单确认的服务量和收费标准结算数据服务费；

b.发行人根据服务公司的地区覆盖能力、资源匹配能力和管理能力，对于不同的业务类型选择不同服务公司进行合作，例如智能语音业务涉及人员规模大、覆盖地区多，因此优先选择体量规模相对较大的善世系服务公司（包括善世（广东）企业服务外包有限公司、芜湖善慧人力资源服务有限公司、共青城宜邦人力资源服务有限公司等同一控制主体），服务公司根据下属主体划分的覆盖区域确定具体服务主体；

c.发行人在实际业务执行时通过会议、邮件等方式向服务公司沟通具体采购需求，确认服务工期要求，原料数据采集、标注服务量，被采集人的年龄、籍贯、性别、口音特点、采集场地等，以及标注人员的语言背景、专业技能和学历要求等；

d.服务公司根据寻找的人力资源情况，将价格信息反馈给发行人，发行人参考业务采购价格体系和过往执行的同类项目的价格，并结合当前项目可能存在的特殊要素要求，与数据服务提供商协商确认价格信息，双方达成一致后确定采集和标注单价；

e.发行人在数据服务采购过程中，对采集、标注人员或其组织人员（服务公司人员或小外包人员）进行培训，部署所需的平台和工具，并下达采集、标注的具体要求；

f.发行人把控项目执行的时间进度，敦促服务公司及其相关人员按时交付数据，并持续监督、关注交付的数据质量。

B.服务公司与小外包人员以及直接对接的终端人员之间的合同签署及实际

## 业务执行情况

a.服务公司与小外包人员以及直接对接的终端人员签署劳务服务合同；

b.服务公司在实际业务执行时对其进行监督、管理，包括进度时间安排、人员信息登记、直接对接的终端人员授权签署、人员纠纷解决等，确保其提供发行人所需的数据服务，以及提供数据的完整性；

c.服务公司在数据服务采购过程中会持续与发行人进行沟通，获取发行人对采标服务及数据质量的反馈情况，及时反馈给小外包人员和直接对接的终端人员，确保其为发行人提供服务的质量。

C.小外包人员与其管理的终端人员之间的实际业务执行情况

a.小外包人员根据发行人数据服务采购需求寻找符合条件的终端人员；

b.小外包人员组织并管理终端人员提供劳务，包括人员信息登记、终端人授权签署、人员工作时间安排及现场引导、对终端人员的服务进度和数据完整性进行把控等。

### ③资金支付情况

发行人按项目进度向服务公司结算采购费用，统一采用银行转账的形式。服务公司收到资金后，向小外包人员和其直接对接的终端人员支付劳务费用，统一采用银行转账的形式。小外包人员收到资金后向终端人员支付劳务费用。

### ④数据服务采购执行过程示例

以 2019 年某唤醒词命令词语音识别采集项目为例，该项目采集人数超过 200 人，采集对象要求多地区、可进行标准美国英语发音、年龄结构多样，采集工期为 15 天，发行人与服务公司、小外包人员和终端人员之间的数据服务采购过程如下：

A.发行人与各数据服务提供商签订采购框架协议，包括芜湖善慧人力资源服务有限公司（以下简称“芜湖善慧”），双方约定服务费率为 10%；

B.发行人综合考虑该项目所需要的语音识别采集人员数量较多、覆盖地区较广、工期较紧等实际情况，参考各服务公司的资源匹配能力和管理能力，并结合服务公司对于下属主体覆盖区域的划分，确定选择芜湖善慧作为该项目的供应商；

C.发行人自测标产比，并与芜湖善慧沟通后达成一致，按照项目所需标注的数据量与标产比的乘积计算所需的标注工时，该项目中所需标注的数据量为 200 小时，约定的标产比是 2.625；

D.发行人以电子邮件形式向芜湖善慧下达采购通知，写明项目简介、服务工期要求，并明确项目采购量为采集服务 210 人，标注服务 525 小时，其中标注服务 525 小时=所需标注的数据量 200 小时\*标产比 2.625；

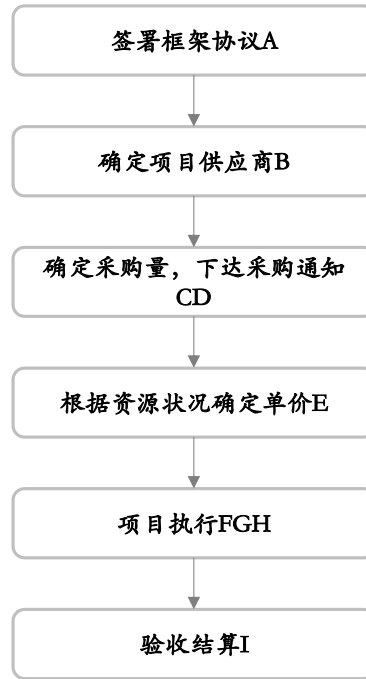
E.芜湖善慧通过小外包人员在上海、深圳、成都等地寻找到足够数量的终端采集人员，同时直接寻找了部分终端标注人员，将报价信息反馈给发行人，经发行人判断符合其业务采购价格体系及同类项目价格区间，双方达成一致确定采集单价和标注单价；

F.服务公司与小外包人员和直接对接的终端人员签订劳务合同，双方正式确定劳务关系；

G.发行人对小外包人员、终端人员进行培训，部署所需的平台和工具，小外包人员组织并管理终端人员开始进行原料数据的采集和标注，服务公司进行全程的监督、管理；

H.服务公司根据原料数据的采集和标注情况，保持与发行人、小外包人员、终端人员等各方的及时沟通，确保采购服务工期进度及数据质量，并向发行人交付采集标注后的数据；

I.发行人对数据进行质量验收后，与服务公司进行结算并出具验收结算单。验收结算单由发行人与服务公司双方盖章确认，记载经发行人验收合格的采集、标注工作量，以及经双方约定一致的采集、标注单价信息。发行人根据验收结算单计算的数据服务费金额(数据服务费=验收工作量\*单价)向服务公司进行支付，并按照框架协议中约定的服务费率按比例支付给服务公司。服务公司按照约定将发行人支付的数据服务费部分支付给小外包人员和直接对接的终端人员，小外包人员收到资金后向其寻找的终端人员支付劳务费用。



### (3) 数据服务采购质量控制措施情况介绍

发行人有效执行了一系列质量控制措施，以确保供应商提供的数据满足其采集需求。具体如下：

①与供应商充分沟通传达采集、标注需求：发行人在根据项目情况确定采购需求后，即通过会议、邮件等方式与供应商进行沟通，明确被采集人的年龄、籍贯、性别、口音特点、采集场地等，以及标注人员的语言背景、专业技能和学历要求等。对于重要项目，发行人会与供应商就上述要求进行书面确认，确保最终的采集、标注人员符合发行人需求。

②采集、标注人员属性的事前验证：在原料数据的采集、标注工作开始前，发行人会直接或通过供应商、小外包人员对终端人员进行身份验证，包括身份证明、学历证明、资质证明等，验证被采集人的年龄、籍贯、性别、口音等身份特征和标注人员的专业性。标注人员需要进行试标注，以检查其专业技能，并可能通过多个标注人员进行交叉验证。同时，发行人会对服务公司人员、小外包人员或终端人员进行规则培训，培训合格后正式开始原料数据的采集、标注。

③数据质量的事中控制：在原料数据的采集、标注过程中，发行人提供技术人员、管理人员，全程监督指导，确保交付的数据质量符合要求。数据采集、标注工作主要在发行人的平台、工具上执行，发行人通过其系统对数据进行自动检

验，通过检验的数据才能保留在系统中，进行下一步的生产环节、交付及验收。

④数据质量的事后验收：在原料数据的采集、标注工作结束后，发行人按照制定的验收标准对采集或标注形成的数据进行验收，包括工具质检、人工复核、抽检等形式，确保其符合发行人需求。对于可在采集、标注后数据中体现的人员属性（如性别、口音等），发行人通过事后对数据进行质检进一步复核是否符合前期提出的需求。

上述质量控制措施涉及的发行人内部控制手段为采购前下达采购通知和采购执行过程中的质量检查等。发行人采取的相关内部控制措施及其有效性的具体情况请参见下述“（4）数据服务采购相关的内部控制情况”。

综上，发行人采取的质量控制措施能够保证供应商提供的数据满足其采集需求，发行人采取了相关内部控制措施并有效执行。

#### （4）数据服务采购相关的内部控制情况

报告期内，发行人建立了《海天瑞声采购管理制度》、《海天瑞声项目管理办法》、《海天瑞声供应商管理制度》、《海天瑞声合同管理制度》和《海天瑞声业务采购实施细则》等规章制度作为对外采购数据服务的制度依据。

针对数据服务采购业务环节涉及的控制点（年度总体预算、供应商选择和持续管理、采购协议签署、采购价格管理、立项及预算管理、采购申请与审批、下达采购需求、供应商提供服务、验收结算及费用报销、供应商付款、项目决算、业财核对），发行人均制定了、采取了有效的控制措施/活动，并确保数据服务采购管内控制度有效执行。

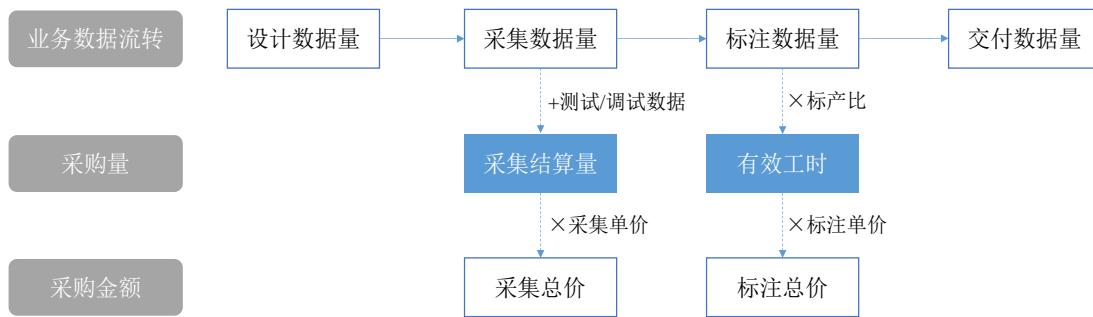
#### （5）与数据服务采购相关的内部控制的设计合理，符合发行人业务特征

发行人的上述内部控制制度符合《企业内部控制基本规范》相关规定，并结合了发行人数据服务采购业务流程的实际特征，可针对数据服务采购过程中的各个环节有效进行复核、控制，相关内部控制设计合理。具体说明如下：

①数据服务采购过程呈现“以销定采”、“同类业务标产比可比”的业务逻辑，可通过前述业务逻辑对数据服务采购量进行复核、确认

发行人训练数据定制服务和训练数据产品业务采购各环节的数据量和有效

工时之间的关系，以及与数据采购支出的关系如下图所示：



注：采集结算量、有效工时指发行人与供应商结算的采集环节、标注环节采购量；交付数据量指经过标注、质检环节，可作为成品交付客户的数据的数量。

发行人数据服务（采集服务、标注服务）采购量与交付给客户的数据量紧密挂钩。对于采集服务：发行人采集损耗率保持较低水平，保证了采集服务的结算量可通过对比设计数据量、采集数据量和交付数据量得到有效控制。对于标注服务：发行人以标注数据量为基数，用标产比（通过实际自测或与客户约定）折算有效工时的方式来核算标注工作量并结算，因此可通过对比标注数据量和交付数据量、复核标产比的合理性对标注服务的结算量（有效工时）进行有效控制。

例如，某“150人中文纯净人声录音棚录音项目”，客户订单要求录音人数为150人（交付数据量）实际采集人数为152人（采集数据量、采集结算量）。成品训练数据集需要交付多少数据量，则相应采集对应的数据量并留出少量冗余即可，不存在额外采集更多数据量的经济必要性。经试标注测速后，标产比确定为2.7，假设某标注人员完成了10小时的音频数据标注工作，则其有效工时核定为 $10 \times 2.7 = 27$ 小时，同类型项目（即数据的特征与领域基本相当）的标产比应基本一致。

在复核、确认采购量的同时，发行人也通过内部控制设计定期维护更新业务采购价格体系，从而充分适应行业下游训练数据需求的变化，在数据类别、数据形式等方面不断拓展细化，以确保该价格体系与发行人实际采购情况、市场价格相符。

## ②发行人的数据服务采购内部控制设计与上述业务逻辑相符，设计有效

发行人设计了完善的内部控制制度，对数据服务采购的各业务环节均施加有效控制。特别地，与上述业务逻辑相适应，发行人在立项及项目预算管理、验收



结算及费用报销、项目决算、业财核对各个关键业务点设置了针对数据服务采购量、采购价格、采购金额、业务数据与财务数据一致等的复核控制，确保数据服务采购核算真实、准确、完整。发行人的数据服务采购内部控制设计符合业务逻辑，设计合理、有效。

(6) 与数据服务采购相关的内部控制执行有效，可确保采购信息完整、准确记载，数据服务采购量可复核，财务信息与实际采购业务一致

① 发行人搭建了有效的内部信息系统，可记载完整、准确的采购信息等业务数据

发行人的内部信息系统包含三个组成部分，分别为一体化数据处理平台（前台生产系统）、业务管理平台（中台业务信息流转与管理系统）与用友财务系统（后台记账与报表系统）。各组成部分情况具体说明如下：

A. 一体化数据处理平台。该平台是发行人训练数据生产所需的各类工具、软件的集合，于 2014 年 9 月初始上线并持续优化，2018 年初开始保留与生产过程相关的基本日志信息，如每个录音人的录音时长信息、录音文件大小信息等，在 2019 年 9 月上线搭载了终端人管理系统及日志管理系统，能够更为全面记录并保留生产过程中的详细日志信息，包括终端劳务人员注册信息、生产过程中终端人员行为日志信息等。

B. 业务管理平台。该平台于 2017 年 1 月开始上线使用，主要功能包含合同和项目创建、项目执行过程和结算、验收管理等，可覆盖项目管理全生命周期，并可根据系统记录的业务数据自动生成收入、成本、存货、产品研发费用等相关的财务数据。发行人报告期内销售、采购业务涉及的量、价和金额信息均完整记载于业务管理平台。

C. 用友财务系统。该系统于 2015 年开始上线使用，自上线至 2019 年 4 月为用友 U8 系统，2019 年 4 月至今为用友 NC 系统，系统自上线以来未发生重大变更。

根据上述说明可知，公司内部信息系统均在报告期初上线运行并覆盖完整报告期间。发行人业务管理平台具有项目核算模块，能够按项目维度完整记录项目预算及支出相关信息，包括但不限于每个项目所对应采购的劳务服务数量、结算价

格及金额等，组成发行人业务采购的完整信息。

②发行人业务管理平台记录了完整、准确的采购信息，使得可以通过“以销定采”、“同类业务标产比可比”的业务逻辑对数据服务采购量进行复核、确认

发行人通过业务管理平台对项目进行管理控制，包括项目预算申请和复核，支出的填报和复核，对采集数据量、采集结算量、交付数据量进行交叉比对，对标注产出比、有效工时进行合理性复核，生成供应商验收结算确认单，进行验收和决算信息复核等。前述信息的准确记录，使得发行人可通过“以销定采”、“同类业务标产比可比”的业务逻辑对数据服务采购量进行复核、确认。

同时，发行人在项目预算申请环节参考业务采购价格体系对预算采购价格进行复核、确认，在项目支出报销环节参考预算采购价格对实际采购价格进行复核、确认，从而对采购价格进行有效的内部控制。

③发行人财务系统中数据服务采购信息均基于业务管理平台产生，可保证实际采购业务与财务信息一致

项目采购支出信息通过业务管理平台自动汇总并生成支出汇总表，人工录入财务系统并经专人复核<sup>9</sup>，进而生成存货、研发费用等信息，可保证采购相关财务信息与业务信息的一致性。

综上，报告期内，发行人依据《企业内部控制基本规范》要求，结合数据服务采购的具体流程、特征设置了内部控制体系，相关内部控制体系符合数据服务采购的业务逻辑，设计有效；发行人搭建了有效的内部信息系统和业务数据流转、复核机制，可确保采购信息完整、准确记载，数据服务采购量、采购单价可复核，财务信息和实际采购业务一致，数据服务采购相关内部控制制度有效执行。

## 5、数据服务采购的合规性说明

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。发行人主要向数据服务供应商采购原料数据采集、标注服务。发行人的数据服务采购过程中涉及的劳务用工形式以及原料数据获取及使用均符合相关法律法规规定。

---

<sup>9</sup> 报告期内，发行人持续改进、升级业务管理平台功能，于2019年4月后可实现自动将相关信息传输至财务系统，进一步确保采购相关财务信息与业务信息的一致性。

## (1) 发行人不存在以外包名义按劳务派遣用工形式使用劳动者的情形

根据《中华人民共和国劳动合同法》、《中华人民共和国劳动合同法实施条例》、《劳务派遣暂行规定》等的规定，劳务派遣是指用工单位与劳务派遣单位订立劳务派遣协议，劳务派遣单位与被派遣劳动者订立劳动合同，被派遣劳动者在约定的用工单位相应岗位上工作的制度。劳务外包是指企业将部分业务或工作发包给外包服务单位，由外包服务单位安排人员完成相关的业务或工作，企业按照业务或工作的完成情况与外包服务单位进行结算，劳务外包关系适用《中华人民共和国合同法》。

发行人与数据服务提供商之间的业务合作属于劳务外包关系，与劳务派遣的主要区别如下：

区别	劳务派遣	发行人与数据服务提供商之间的业务关系
合同关系	实际用工单位与劳务派遣单位签订劳动派遣协议，劳务派遣单位与劳务派遣人员签订劳动合同。	发行人和数据服务提供商签署《业务服务协议》，对服务提供商采购服务，服务提供商向发行人提供合同约定的服务。服务提供商根据项目需要与劳务人员签署劳务协议
费用结算	实际用工单位向劳务派遣单位支付劳务派遣费用，具体根据被派遣员工的人数、派遣用工的工作时间和约定的薪酬标准等支付对应费用。	发行人对数据服务提供商提供服务的成果进行验收，验收合格后根据数据量或服务量支付对应的采购费用。
服务人员的管理	被派遣人员在实际用工单位处工作，接受实际用工单位管理，需要遵守用工单位的相关规章制度，在用工单位的工作时间内专为用工单位提供劳务。	数据服务提供商提供服务对应的终端人员不在发行人处工作，也不接受发行人管理，并非专门为发行人提供劳务。发行人仅按照采购合同约定要求服务提供商提供服务。

综上，发行人与数据服务提供商签订合同，由数据服务提供商提供原料数据的采集、标注等服务，发行人要求数据服务提供商按照约定交付工作成果，并按照服务项目的验收成果进行结算，不对劳务人员具体进行管理，不要求劳务人员遵守发行人员工规章制度。因此，发行人委托数据服务提供商提供原料数据采集和标注服务构成劳务外包，发行人不存在以外包名义按劳务派遣用工形式使用劳动者的情形。

## (2) 数据信息获取和使用的合规性

## ①数据获取方式合法合规，取得相关数据主体授权许可

发行人生产训练数据过程中所需要的原料数据获取方式分为两种，分别为发行人主动收集数据方式以及客户自行提供数据方式。

对于主动收集数据方式，发行人主要通过数据服务供应商采集原料数据，数据服务供应商按照发行人的采集要求寻找匹配的终端人员，组织其使用发行人研发的技术、平台、工具等，为发行人提供数据采集服务。在个别业务中，由于采购量较小，作为补充性采购，发行人也存在少量直接采集数据的情形。在该模式下，被采集对象为个人且涉及采集个人语音、人像等个人信息的，发行人和原料数据采集供应商在数据采集前已明确告知被采集人或其监护人采集内容、目的、方式等规则，并经其授权同意，不存在非法获取的情形，符合相关法律规定。在采集上述个人语音、人像等业务数据之外，特定项目会对被采集人的年龄、口音区等设定具体要求，发行人会结合被采集人的身份证明文件等信息进行核验，收集身份信息涉及到个人敏感信息会进行去标识化处理。发行人主动收集非个人信息数据的，不涉及侵犯其他第三方的合法权益，不存在非法获取的情形。

对于客户提供数据的方式，发行人向客户提供数据转写、标注等纯加工定制服务时，是由客户向发行人提供业务数据并委托发行人进行加工。客户仅向发行人提供语音、图像、视频等业务数据，不提供涉及的个人身份信息。客户向发行人提供数据，主要通过加密传输等方式提供。该模式下发行人取得客户对于所提供数据进行加工处理的授权许可，符合相关法律规定。

综上，发行人取得数据的获取方式合法合规，获得了相关数据主体的授权许可。

## ②使用数据不存在超出授权许可范围的情形

发行人主动收集数据方式下，被采集对象为个人的，已取得被采集人或其监护人的授权许可。根据发行人制定、签署的授权声明或协议，相关授权许可存在使用范围、主体等方面的限制。发行人使用数据的用途包含在上述授权许可范围内。发行人不存在向第三方非法提供或公开披露业务数据、个人身份信息的情形，未超出授权许可的限制使用数据。

对于客户提供数据方式下，发行人将不同客户提供数据分开存储，严格按照协议约定对数据进行加工后再返还给客户。发行人未将客户提供数据用于其他用

途，或向第三方非法提供或公开披露，未超出客户授权许可限制使用数据。

发行人对数据使用了严格的内部控制措施保护数据安全，避免发生超出授权限制使用数据的情形：A.数据脱敏和加密传输； B.设置保存期限及收集数据超期删除； C.权限、人员设置及内部审批； D.人员保密协议及培训.发行人主要原料数据采集供应商也通过建立保密制度、签署保密协议、对人员进行数据保密相关培训、对数据信息储存设置密码和人员权限等措施，避免发生超出限制使用数据的情形。

### ③不存在数据内容涉及侵犯个人隐私或其他合法权益的情形

发行人数据信息收集及使用过程经过被采集对象的授权许可及明确同意，符合法律法规关于个人隐私权、肖像权、个人信息权益等方面的相关规定，不存在数据内容侵犯个人隐私或其他合法权益的情形。

## 6、销售模式

### (1) 目标客户

公司主营业务的目标客户为 AI 产业链上的各类机构，可分为三大类：大型科技公司，如阿里巴巴、腾讯、百度、微软、三星、亚马逊等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学等。

### (2) 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产品和服务吸引客户，并在持续服务客户的过程中提升服务价值和客户黏度。公司通过口碑传播、参与学术会议、官方网站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。

公司下游客户所处的人工智能行业具有较强的科技属性，行业内前沿技术的交流通常在学术会议上进行，无论是学术界的学者还是产业界的技术负责人、开发人员都会将学术会议作为学习和交流的重要场所。因此，公司也将这些会议视为树立品牌形象、建立和巩固客户关系、获取商业信息的重要机会。报告期内，公司参加的主要学术会议情况如下：

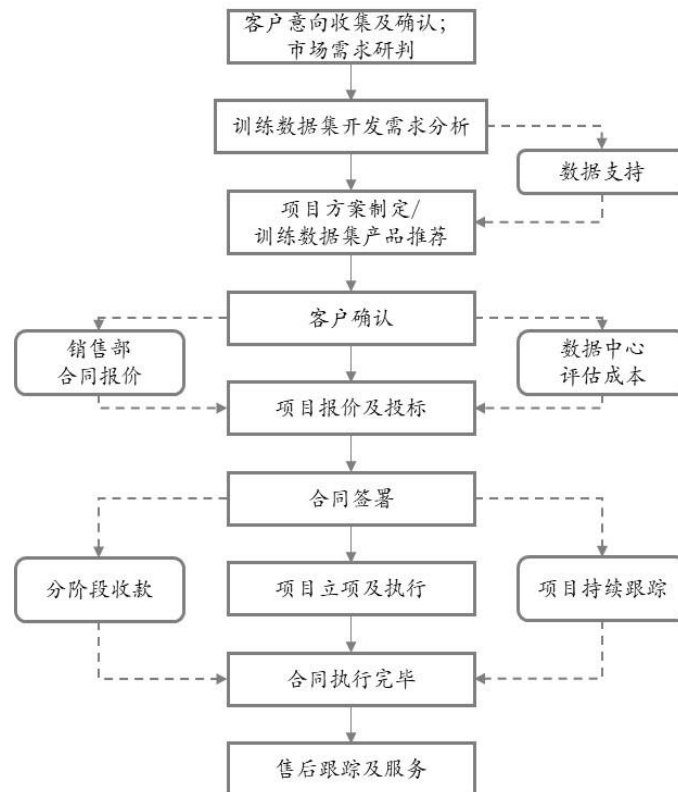
学术会议名称	接触客户业务方向	参展频次
INTERSPEECH	智能语音	一年一次
ICASSP	智能语音、计算机视觉	一年一次
APSIPA	智能语音	一年一次
ICCV	计算机视觉	一年一次
CVPR	计算机视觉	一年一次
LREC	智能语音	一年一次
O-COCOSDA	智能语音	一到两年一次
NCMMSC	智能语音	两年一次
SpeechTek	智能语音	一年一次
NLPCC	自然语言	一年一次
PRCV	计算机视觉	一年一次

报告期内，按主要产品及服务类型，公司与主要客户合同洽谈至最终签订的具体过程、合同履行过程中公司与客户沟通协商的具体情况如下：

序号	主要产品及服务类型	合同洽谈至最终签订的具体过程	合同履行过程中发行人与客户沟通协商的具体情况
1	训练数据定制服务	<p>(1) 合同洽谈开始、获得客户需求后，公司营销部门会对客户需求进行初步分析，并基于已有项目经验帮助客户进一步细化需求；</p> <p>(2) 客户需求基本确定后会引入技术人员进行技术支持，就客户需求制定基本解决方案，同时会基于需求进行成本和项目执行周期的评估，该过程使得客户的需求从直接的任务目标转变为可落地实施的项目方案；</p> <p>(3) 双方对合同细节进行沟通和优化，对项目工期、项目预算和项目报价等内容进行协商，形成可实施的最终解决方案；</p> <p>(4) 客户完成采购审批流程，双方最终签订合同。</p>	<p>合同履行过程中，公司会根据项目进展对客户进行常规项目汇报（如日报或周报等），也会通过不定期的专项会议与客户讨论过程中出现的问题和解决方案。</p> <p>对于训练数据定制服务，由于是根据客户定制化需求，所以在合同履行过程中，有时会先按初始计划生产一部分数据，客户使用这部分数据进行一定训练，然后反馈方案是否满足实际情况。</p> <p>如果发现跟实际有出入，客户会发出需求变更，公司会根据变更后的需求重新进行项目评估，并确定新的解决方案。如需求变化大，还需跟客户进行执行周期或合同金额变更的协商，待协商一致后继续执行项目，并签署相应的合同补充协议。</p>
2	训练数据产品	<p>(1) 合同洽谈开始、获得客户需求后，营销部门会对客户需求进行初步分析，并匹配公司现有的训练数据产品；</p> <p>(2) 营销部门对客户需求进行解答，阐述公司训练数据产品的优势，并结合现有模型训练结</p>	<p>此类合同执行即在合同要求的期间内，交付训练数据产品，通常也不会发生需求变更的情况。</p>

序号	主要产品及服务类型	合同洽谈至最终签订的具体过程	合同履行过程中发行人与客户沟通协商的具体情况
		果，给予客户相关量化指标参考，做出最优的训练数据产品推荐； (3) 给客户已有训练数据产品样例，说明文档及报价，让客户综合评估； (4) 客户确定训练数据产品后，启动采购审批流程，双方最终签订合同。	
3	训练数据相关的应用服务	(1) 合同洽谈开始、获得客户需求后，营销部门对客户进行初步分析，并基于已有项目经验帮助客户细化需求； (2) 客户需求基本确定后会引入技术人员进行技术支持，就客户提出的技术指标制定解决方案，同时会基于需求进行项目成本和项目执行周期等方面的评估； (3) 客户完成采购审批流程，双方最终签订合同。	公司在初期就开发进度和项目执行中可能出现的问题与客户及时沟通，确保最终服务成果达到或超过合同设定的指标。

### (3) 销售流程



图：销售流程图

#### (4) 签约模式

报告期内，发行人均采用直销方式直接对接客户并销售，不存在采用经销模式通过经销商对外销售产品或服务的情况。

其中，仅在训练数据产品业务中，存在少量与最终客户非直接签约的情形。在该种情形下，最终客户出于其自身经营管理需要，指定其认可的第三方主体与发行人签约，发行人根据销售合同约定，直接向最终客户指定的签约主体交付训练数据产品，并由该指定签约主体对交付物进行验收、确认及按照合同约定支付款项。报告期内，仅4家客户涉及该种情形，且该情形收入占比很低。

上述客户指定第三方签约模式主要是客户自身经营管理需要所致，例如：部分最终客户的内部流程较为繁琐，出于预算计划安排、缩短项目周期、便捷采购等角度考虑，其指定认可的第三方签约主体向发行人执行采购。该模式为发行人应最终客户要求形成，具备商业合理性，符合行业惯例。

报告期内，不同签约模式的收入及占比情况如下：

金额单位：万元

签约形式	2020年		2019年		2018年	
	销售收入	收入占比	销售收入	收入占比	销售收入	收入占比
直接签约	22,264.71	95.40%	22,932.74	96.54%	19,265.77	100.00%
客户指定第三方签约	1,072.68	4.60%	823.07	3.46%	-	-
合计	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>

#### (5) 定价模式

公司主要产品或服务的定价模式具体如下：

##### ①训练数据定制服务

定价策略和方法：一般采用成本加成定价法。公司根据客户的具体服务需求预估项目成本，在预估成本的基础上，参考公司制定的指导毛利率水平，结合项目技术难度、复杂程度、时限要求等进行报价，并根据市场环境与客户协商，最终确定价格。例如，采集环境和场景、采集对象年龄结构、语种等影响项目具体执行难度的因素，都会在定价中予以考虑。

定价过程：确定需求→评估成本→评估技术难度、复杂程度、时限要求等因



素→确定报价→与客户协商价格→确定价格、签署合同。

## ②训练数据产品

定价策略和方法：一般采用需求导向定价法。公司综合考虑训练数据集的开发支出、市场需求程度、预计未来重复销售的频率等因素，制定产品标准价格及价格区间，在销售过程中，根据客户的实际需求情况，以价格区间为基础向客户报价，经双方协商确定最终销售价格。训练数据产品通常以单个数据集为单位进行定价，定价比较灵活。

定价过程：确定需求→参考产品开发成本→评估市场需求程度、未来估计重复销售频率等因素→确定价格区间→与客户协商价格→确定价格、签署合同。

## ③训练数据相关的应用服务

该类业务通常为客户对算法模型的具体功能提出要求，发行人运用其训练数据为客户提供具体应用层面的开发、调试、优化、安装、使用培训等技术支持和实施服务工作，客户相应配合并支付项目费用。最终服务交付形式通常为软件授权或软硬件一体化销售，依据客户具体需求、项目实施难度、项目周期和预计成本、客户协商情况确定价格。

### （四）设立以来主营业务、主要产品或服务、主要经营模式的演变情况

自设立至今，发行人均主要从事训练数据的研发设计、生产及销售业务，主营业务、主要产品或服务、主要经营模式均未发生重大变化。

### （五）主要产品的工艺流程图或服务的流程图

发行人主要产品的工艺流程图或服务流程图及相关说明请参见本节“一、发行人主营业务、主要产品或服务的情况/（三）主要经营模式/2、生产或服务模式”。

### （六）生产经营中涉及的主要环境污染物、主要处理设施及处理能力

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务，生产经营过程主要为数据的采集和加工过程，不产生环境污染物，不涉及环境污染物的处理事宜。

## 二、发行人所处行业的基本情况和竞争状况

### （一）所属行业及确定所属行业的依据

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务，所属行业为软件和信息技术服务业。

根据国家统计局《战略性新兴产业分类（2018）》，发行人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《科创板企业发行上市申报及推荐暂行规定》中“新一代信息技术领域”的战略性新兴产业，符合科创板行业领域要求。因此，发行人通过设计训练数据集结构、执行数据加工处理过程，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，发行人所属行业为软件和信息技术服务业。

根据中国证监会颁布的《上市公司行业分类指引》（2012年修订），公司所属行业为“软件和信息技术服务业”，行业代码为“I65”。

### （二）所属行业的行业主管部门、行业监管体制、行业主要法律法规政策及对发行人经营发展的影响

#### 1、行业主管部门及监管体制

发行人所属行业为软件和信息技术服务业，所在细分领域为人工智能基础数据服务领域。

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、行业主要法律法规政策

序号	实施时间	颁布主体	主要法律法规及行业政策	相关内容
1	2010年10月	国务院	《关于加快培育和发展战略性新兴产业的决定》	将新一代信息技术作为七大重点支持发展的领域之一，着重提出了“加快建设宽带、泛在、融合、安全的信息网络基础设施”的要求
2	2015年5月	国务院	《中国制造2025》	瞄准新一代信息技术、高端装备、新材料、生物医药等战略重点，引导社会各类资源集聚，推动优势和战略产业快速发展
3	2015年7月	国务院	《国务院关于积极推进“互联网+”行动的指导意见》	明确提出人工智能作为11个重点布局的领域之一，促进其在智能家居、智能终端、智能汽车、机器人等领域的推广应用
4	2016年3月	国务院	《中华人民共和国国民经济和社会发展第十三个五年规划纲要》	支持新一代信息技术、新能源汽车、生物技术、绿色低碳、高端装备与材料、数字创意等领域的产业发展壮大。加强前瞻布局，在空天海洋、信息网络、生命科学、核技术等领域，培育一批战略性新兴产业。同时人工智能概念进入“十三五”重大工程
5	2016年5月	中共中央、国务院	《国家创新驱动发展战略纲要》	加快工业化和信息化深度融合，把数字化、网络化、智能化、绿色化作为提升产业竞争力的技术基点。提出到2020年进入创新型国家行列、2030年跻身创新型国家前列、到2050年建成世界科技创新强国“三步走”目标
6	2016年5月	国家发展改革委、科技部、工业和信息化部、中央网信办	《“互联网+”人工智能三年行动实施方案》	到2018年，打造人工智能基础资源与创新平台，人工智能产业体系、创新服务体系、标准化体系基本建立，基础核心技术有所突破，总体技术和产业发展与国际同步，应用及系统级技术局部领先。在重点领域培育若干全球领先的骨干企业，初步建成基础坚实、创新活跃、开放协作、绿色安全的产业生态，形成千亿级的市场应用规模
7	2016年8月	国务院	《“十三五”国家科技创新规划》	规定了“十三五”期间科技创新的总体思路、发展目标、主要任务和重大举措，致力于发展构建包括新一代信息技术在内的具有国际竞争力的现代产业技术体系
8	2017年2月	国家发改委	《战略性新兴产业重点产品和服务指导目录》	人工智能首次进入指导目录名单，包括面向社会开放的文本、语音、图像、视频、地图及行业应用数据等多类型海量训练资源库和标准测

序号	实施时间	颁布主体	主要法律法规及行业政策	相关内容
				试数据集在内的公共数据平台成为人工智能方向的重要子方向之一
9	2017年7月	国务院	《新一代人工智能发展规划》	到2020年,总体技术和应用与世界先进水平同步;到2025年,基础理论实现重大突破,部分技术与应用达到世界领先水平;到2030年,理论、技术与应用总体达到世界领先水平,成为世界主要人工智能创新中心
10	2017年10月	十八届中央委员会	《十九大报告》	人工智能写入十九大报告,将推动互联网、大数据、人工智能和实体经济深度融合
11	2017年12月	工业和信息化部	《促进新一代人工智能产业发展三年行动计划(2018-2020年)》	《行动计划》从推动产业发展角度出发,结合“中国制造2025”,对《新一代人工智能发展规划》相关任务进行了细化和落实,以信息技术与制造技术深度融合为主线,以新一代人工智能技术的产业化和集成应用为重点,推动人工智能和实体经济深度融合
12	2018年1月	国家电子技术标准化研究院	《人工智能标准化白皮书》	全面推进人工智能标准化工作,促进产业发展
13	2018年3月	国务院	《2018年国务院政府工作报告》	人工智能再次被列入政府工作报告:加强新一代人工智能研发应用;在医疗、养老、教育、文化、体育等多领域推进“互联网+”;发展智能产业,拓展智能生活
14	2018年4月	教育部	《高等学校人工智能创新行动计划》	到2020年,基本完成适应新一代人工智能发展的高校科技创新体系和学科体系的优化布局;到2025年,高校在新一代人工智能领域科技创新能力和人才培养质量显著提升;到2030年,高校成为建设世界主要人工智能创新中心的核心力量和引领新一代人工智能发展的人才高地
15	2018年10月	中共中央政治局	人工智能发展现状和趋势第九次集体学习	人工智能是引领新一轮科技革命和产业变革的战略性技术,具有溢出带动性很强的“头雁”效应。人工智能正在对经济发展、社会进步、国际政治经济格局等方面产生重大而深远的影响。加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手,是推动我国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源
16	2018年11月	工业和信息化部	《新一代人工智能产业创新重点任务揭榜工作方案》	基于《促进新一代人工智能产业发展三年行动计划(2018-2020年)》中部署的重点任务和目标,人工智能揭榜工作将在17个方向及细分领域,征集并遴选一批掌握关键核心技术、具备较强创新能力的创新主体,以突破产业发展的短板和瓶颈,加快我国人工智能产业与实体经济深度融合
17	2019年3月	中央全面深化改革委	《关于促进人工智能和实体经济深度融合的指导	促进人工智能和实体经济深度融合,要把握新一代人工智能发展的特点,坚持以市场需求为导向,以产业应用为目标,深化改革创新,优

序号	实施时间	颁布主体	主要法律法规及行业政策	相关内容
		员会	意见》	化制度环境，激发企业创新活力和内生动力，结合不同行业、不同区域特点，探索创新成果应用转化的路径和方法，构建数据驱动、人机协同、跨界融合、共创分享的智能经济形态
18	2019年8月	科学技术部	《国家新一代人工智能开放创新平台建设工作指引》	明确指出“开放、共享”是推动我国人工智能技术创新和产业发展的重要理念，鼓励开放创新平台面向细分领域建设标准测试数据集，促进数据开放和共享，形成标准化、模块化的模型、中间件及应用软件，以开放接口、模型库、算法包等方式向社会提供软硬件开放共享服务
19	2019年8月	科学技术部	《国家新一代人工智能创新发展试验区建设工作指引》	明确指出试验区建设以促进人工智能与经济社会发展深度融合为主线，重点任务包括加强网络基础设施、大数据基础设施、计算基础设施建设，提升传统基础设施的智能化水平，形成支撑新一代人工智能广泛应用的基础设施体系
20	2020年3-4月	中共中央政治局常委会、国家发改委	中央政治局常委会会议重要讲话	加快新型基础设施建设（“新基建”）进度，新基建范围包括以人工智能、云计算、区块链等为代表的新技术基础设施，以数据中心、智能计算中心为代表的算力基础设施等。
21	2020年4月	中共中央、国务院	《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》	首次将“数据”作为市场化要素写入国家顶层设计级别文件，提出要加快培育数据要素市场，发挥数据在市场化配置中的作用。
22	2020年8月	国家发改委、科技部、工业和信息化部	《国家新一代人工智能标准体系建设指南》	为加强人工智能领域标准化顶层设计，推动人工智能产业技术研发和标准制定，促进产业健康可持续发展
23	2021年3月	十三届全国人大四次会议	《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》	《十四五规划》指出要加快数字化发展，建设数字中国，同时打造数字经济新优势，充分发挥海量数据和丰富应用场景优势，促进数字技术与实体经济深度融合，赋能传统产业转型升级，催生新产业新业态新模式，壮大经济发展新引擎。同时指出要加强关键数字技术创新应用：聚焦高端芯片、操作系统、人工智能关键算法、传感器等关键领域；建设重点行业人工智能数据集，发展算法推理训练场景

### 3、行业主要法律法规政策对发行人经营发展的影响

训练数据是人工智能行业的基础支撑。2017年工业和信息化部颁布《促进新一代人工智能产业发展三年行动计划（2018-2020年）》（以下简称“《三年行动计划》”），明确提出到2020年人工智能产业支撑体系基本建立，具备一定规模的

高质量训练资源库、标准测试数据集建成并开放。面向语音识别、视觉识别、自然语言处理等基础领域及工业、医疗、金融、交通等行业领域，支持建设高质量训练资源库、标准测试数据集。到 2020 年，基础语音、视频图像、文本对话等公共训练数据量大幅提升，在工业、医疗、金融、交通等领域汇集一定规模的行业应用数据，用于支持创新创业。2018 年工业和信息化部启动了人工智能产业创新重点任务揭榜工作，明确指出揭榜工作将围绕《三年行动计划》确定的重点任务方向，在支撑体系方面选择高质量的行业训练资源库、标准测试、智能化网络基础设施、安全保障体系等作为揭榜攻关任务，从而加快完善基础环境、保障平台，形成我国人工智能创新发展的支撑能力。2019 年科学技术部出台《国家新一代人工智能开放创新平台建设指引》，鼓励开放创新平台面向细分领域建设标准测试数据集，促进数据开放和共享，形成标准化、模块化的模型、中间件及应用软件，以开放接口、模型库、算法包等方式向社会提供软硬件开放共享服务。

在国家产业政策的支持下，人工智能基础数据服务行业稳步发展，行业训练资源库等细分领域的产业价值逐步凸显。2017 年至 2020 年，发行人收入从 11,907.09 万元增长至 23,337.40 万元，年均复合增长率达 25.15%。未来，随着下游行业商业成熟度不断提高、覆盖领域广泛拓展，客户对多领域、多场景的训练数据的需求将呈现快速增长态势，公司的盈利能力有望持续增强。

### （三）所属行业发展情况和未来发展趋势

发行人所在细分领域为人工智能基础数据服务行业（以下简称“基础数据服务行业”），行业发展情况与下游人工智能技术与应用行业发展情况息息相关。

#### 1、下游人工智能行业整体发展情况及未来发展趋势

##### （1）训练数据是当前 AI 算法发展和演进的“燃料”

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。在当前人工智能行业发展进程中，有监督的深度学习算法是推动人工智能技术取得突破性发展的关键技术理论，而大量训练数据的支撑则是有监督的深度学习算法实现的基础，训练数据已成为算法模型发展和演进的“燃料”。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，2012-2016 年期间，人工智

能行业不断优化算法增加深度神经网络层级，利用大量的数据集训练提高算法精准性，ImageNet 数据集的超过 1,400 万张训练图片和 1,000 余种分类便在其中起到重要作用。然而，从自然数据源简单收集取得的原料数据并不能直接用于有监督的深度学习算法训练，必须经过专业化的采集、加工，形成相应的工程化训练数据集后才能供深度学习算法等训练使用。目前，应用有监督学习的算法对于训练数据的需求远大于现有的标注效率和投入预算，基础数据服务将持续释放其对于算法模型的基础支撑价值。

## （2）AI 产业对训练数据服务的需求持续产生

AI 产业对训练数据的需求主要来源于成熟算法模型的拓展性需求和新生算法模型的前瞻性需求。在成熟的拓展性需求方面，Mckinsey Global Institute 的研究报告表明：

深度学习模型对训练数据的数据量、多样性和更新速度方面提出较高要求。为充分发挥技术潜能，深度学习模型需要海量且涵盖图像、视频及语音在内等多种类型的训练数据进行模型训练。此外，人工智能技术要求算法模型根据潜在的应用场景变化而持续更新，因此，算法模型所使用的训练数据亦需要定期更新。具体而言，约 1/3 的算法模型每月至少更新一次，约 1/4 的算法模型每日至少更新一次，算法模型持续更新的特点将进一步拓展各领域训练数据的需求空间。

而在新生的前瞻性需求方面，随着人工智能商业化进程的演进，新兴应用场景如物联网 AIoT、AI PaaS、产业互联网等将展现出巨大的发展潜力，并逐步促进 AI 技术和算法模型的优化和创新。因此，在创新应用场景和新型算法的带动下，具有前瞻性的训练数据产品和高定制化的训练数据服务需求将逐步成为主流。

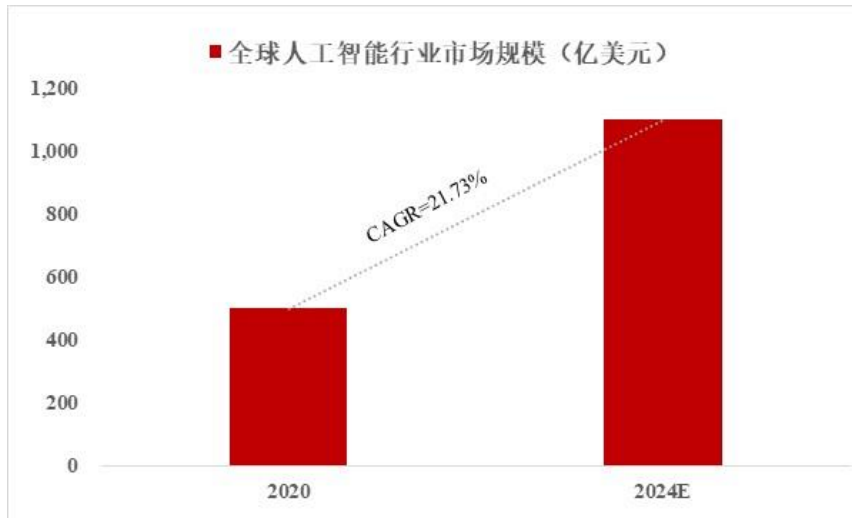
## （3）全球和中国的 AI 产业市场规模

### ①全球 AI 产业市场规模

经过多年的发展，人工智能技术已在金融、医疗、安防等多个领域实现技术落地，且应用场景愈来愈丰富，AI 产业已进入全方位商业化的发展阶段。

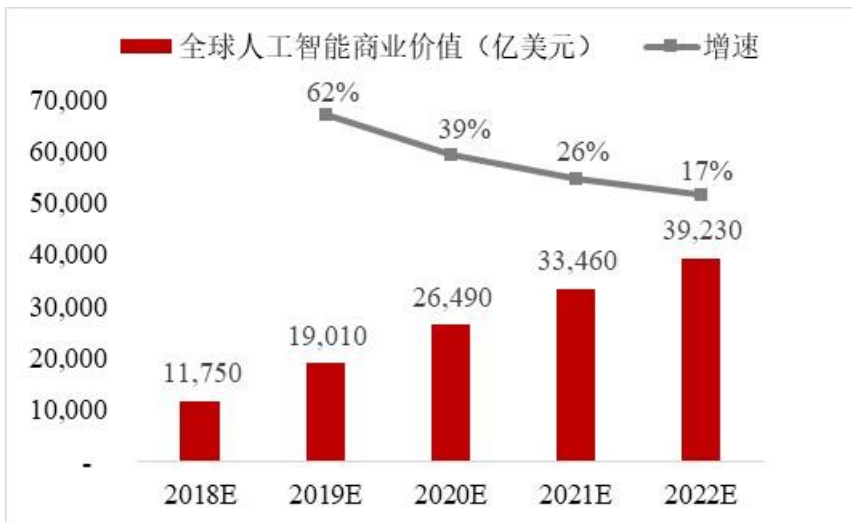
截至 2019 年年末，全球共有 5,386 家活跃 AI 企业，27,400 名高级研究人员，20 座重点发展城市。2019 年，全球人工智能行业共产生约 374 亿美元融资，其中，自动驾驶、药物医疗、人脸识别、视频内容和金融反欺诈是获得融资最多的

领域。根据国际数据公司（IDC）的数据，2020年，全球人工智能产业规模达到501亿美元，预计2024年将达到1,100亿美元，年复合增长率为21.73%。



数据来源：国际数据公司（IDC）

根据高德纳咨询（Gartner）2018年的预测数据，2019年，全球人工智能产业的商业价值将达到19,010亿美元，预计2022年将达到39,230亿美元，年复合增长率为27.31%。



数据来源：高德纳咨询（Gartner）

## ②中国 AI 产业市场规模

当前我国人工智能产业加速发展，从基础支撑、核心技术到行业应用的产业链条基本形成，一批创新活跃、特色鲜明的创新企业加速成长，新模式、新业态不断涌现，整体呈现蓬勃发展态势。政策支持、投资引导和巨头布局将推动中国 AI 产业的结构调整，进一步扩大市场规模。



2019 年中国与全球人工智能行业发展情况对比

项目	全球	中国	中国排名
论文数	AI 领域论文发布数量占全部论文的 3%	中国 AI 领域论文的发布数量占该领域全球论文发布总量的 28%	1
人才数/高级人才数	博士 22,000 名/专家 5,400 名	博士 413 名/专家 120 名	7
活跃企业数	5,386 家	1,189 家	2
融资总额	约 374 亿美元	约 166 亿美元	1
2008-2019 年 AI 相关专利申请	448,684 项	66,508 项	2

数据来源：艾瑞咨询

截至 2019 年年末，中国（不含港澳台）活跃 AI 企业达到 1,189 家，占全球总数的 22.08%，位居全球第二；中国高级研究人员共有 533 名，AI 人才储备方面综合排名全球第七，仍有较大的发展空间。2019 年期间，中国人工智能领域论文的发布数量占该领域全球论文发布总量的 28%，位列全球第一；2008-2019 年人工智能领域的专利申请共发生 448,684 项，中国拥有 66,508 项，位列全球第二，中国的 AI 技术水平处于全球前列。2019 年，中国人工智能行业的融资总额约为 166 亿美元，占全球的 44.39%，位居全球首位。根据国际数据公司（IDC）的数据，2020 年，中国人工智能产业规模将达到 62.70 亿美元，预计 2024 年将达 172.16 亿美元，年复合增长率达到 30.40%。



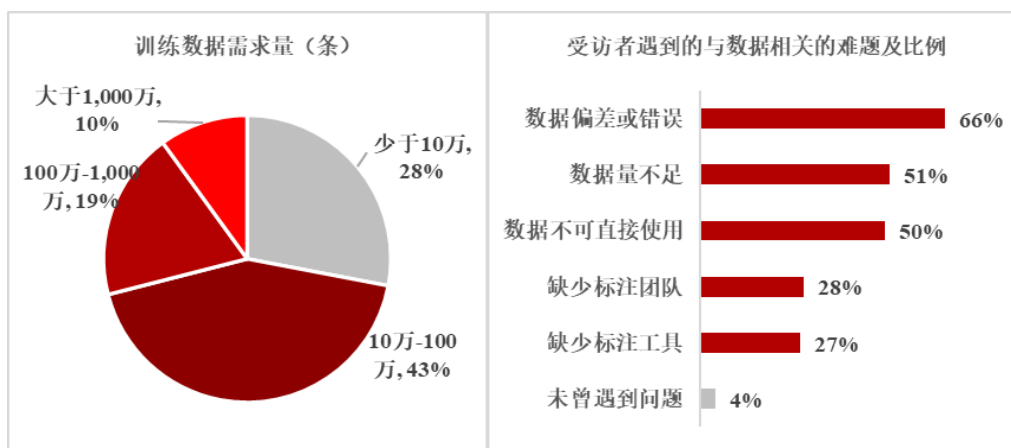
数据来源：国际数据公司（IDC）

## 2、人工智能基础数据服务行业发展情况及未来发展趋势

### （1）全球 AI 基础数据服务行业的发展情况

全球基础数据服务行业处于快速成长期，市场规模具有较大的增长空间。应

用场景的创新和机器学习算法的流行直接带动了训练数据需求的大幅增长，这种趋势导致训练数据难以获取和数据科学家、数据工程师等人力资源稀缺成为制约 AI 产业发展的两大挑战。根据 Dimensional Research 的全球调研报告，72% 的受访者认为至少使用超过 10 万条训练数据进行模型训练，才能保证模型有效性和可靠性，96% 的受访者在训练模型的过程中遇到训练数据质量不佳、数量不足、数据标注人员不足等难题。为应对训练数据所带来的多方面挑战，AI 企业开始从第三方购买原料数据收集、训练数据生产和数据专家咨询等服务，调研结果指出，外包服务能够有效加快算法模型落地应用的速度。因此，得益于训练数据需求增长和对外采购意识的形成，全球基础数据服务行业进入快速成长期，市场规模具有较大的增长潜力。



数据来源：Dimensional Research

## (2) 中国 AI 基础数据服务行业市场规模与需求类型

### ① 中国基础数据服务行业的市场规模

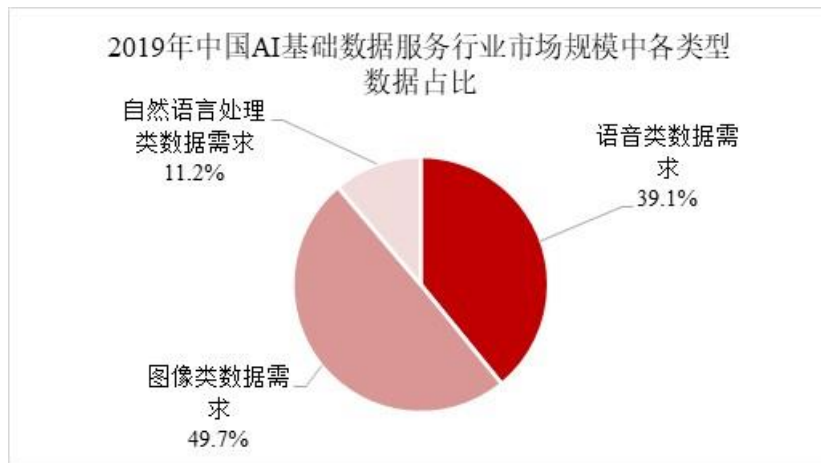
从 AI 产业链的发展情况和未来发展趋势来看，中国基础数据服务行业的市场规模将不断扩大。一方面，随着算法模型、技术理论和应用场景的优化和创新，AI 产业对训练数据的拓展性需求和前瞻性需求均快速增长；另一方面，随着行业内对训练数据需求类型的增加以及对服务标准要求的提高，产业链的专业化分工将愈加清晰，专业化的训练数据服务提供商将扮演更加重要的角色。根据艾瑞咨询《2020 年中国 AI 基础数据服务行业研究报告》，2019 年中国基础数据服务行业的市场规模达到 30.9 亿元，预计 2025 年将突破 100 亿元，复合年增长率达到 21.8%。



数据来源：艾瑞咨询

## ②中国 AI 基础数据服务行业需求类型

按数据类型划分，中国基础数据服务行业的市场需求可以分为图像类数据需求、语音类数据需求和自然语言处理类数据需求。2019 年，图像类、语音类和自然语言处理类数据需求规模占比分别为 49.7%、39.1%和 11.2%。



数据来源：艾瑞咨询

## （3）中国 AI 基础数据服务行业市场规模增长的驱动因素

随着基础数据服务行业的市场规模增速平稳向上，增量市场将替代存量市场成为主要拉力。从需求方的角度看，基础数据服务市场可以分为存量市场和增量市场：存量市场是指训练数据服务行业已有并且稳定的业务范畴，如人脸识别、车辆识别、图片识别、语音识别等相关业务，增量市场是指海外业务、新增需求方、新增业务场景等。在存量市场中，巨头互联网科技公司和 AI 公司为主要需求方，项目落地所需的训练数据逐渐成为需求核心，目前存量市场仍是基础数据

服务市场的需求主体。增量市场是相对于存量市场而存在的，以海外市场、国内新需求方市场、国内新兴业务拓展和国内新成立的 AI 创业公司的需求为主，目前增量市场对于整体市场规模的贡献率较低，但随着中国 AI 技术的不断深入与国际化，增量市场将在未来成为主要的拉动力量。

基础数据服务行业市场规模增长的驱动因素可以分为外部因素和内部因素。在外部因素方面，AI 产业支撑体系方面的政策支持、投资机构的资金投入、科技巨头企业的 AI 生态链布局将推动基础数据服务行业加速发展。在内部因素方面，现阶段有监督的深度学习算法的广泛应用为基础数据服务行业带来稳定的市场需求，未来 AI 新场景、新模式、新业态的不断涌现将对训练数据的数量和类型提出更高的要求，为基础数据服务行业带来源源不断的新生市场需求。

#### (4) 中国基础数据服务行业主要市场参与者

##### ①中国基础数据服务行业的产业链情况

中国基础数据服务行业的上游包括数据生产者和数据生产组织者，主要提供原料数据的采集服务；中游包括基础数据服务商，主要通过数据处理能力和项目管理能力完成训练数据集结构设计、数据加工和质量检测等工作，为下游客户提供训练数据产品和服务；下游包括科技公司、行业企业、AI 公司和科研单位，主要负责算法研发。发行人位于基础数据服务行业的中游，属于基础数据服务商。

中国人工智能基础数据服务行业的产业链图谱



资料来源：艾瑞咨询

##### ②中国基础数据服务行业的市场参与主体

当前，中国基础数据服务行业的市场参与主体主要包括下述几类：

一是学术机构，为开展相关研究工作，自行采集、标注，并建设学术训练资源库。这类训练数据主要用于算法的创新性验证、学术竞赛等，但通常其迭代速度较慢，难用于实际应用场景。

二是政府等中立机构，他们以公益形式开放的公共数据，主要包括政府、银行机构等行业数据及经济运行数据等，数据标注一般由使用数据的机构完成。

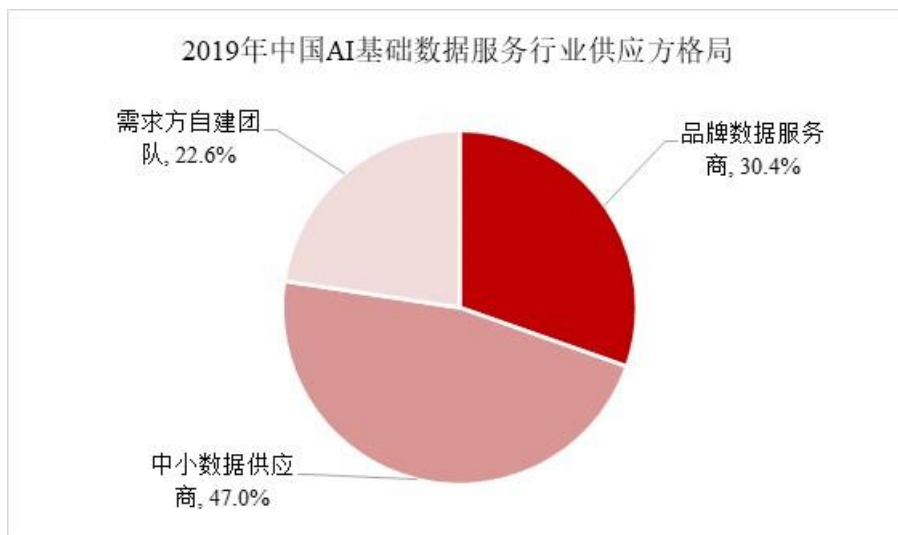
三是需求方自建基础数据团队，需求方科技公司或 AI 公司等为开展业务而自行建设训练资源库，一般自行采集、标注形成自用训练数据，或采购专业数据公司提供的数据外包服务。

四是基础数据服务商，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。这类公司业务包括出售现成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。

发行人主要属于上述第四类公司，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。

### ③中国基础数据服务行业的竞争格局

在上述的参与主体中，品牌数据服务商、中小数据供应商和需求方自建基础数据团队构成市场竞争关系，为基础数据服务市场的主要供应方，在 2019 年市场规模中的份额占比分别为 30.4%、47.0%和 22.6%，目前中小数据供应商是市场中的主要供应力量。



资料来源：艾瑞咨询

从供应方的发展来看，行业内部处于“洗牌”阶段，未来品牌数据服务商阵营将替代中小型供应商阵营，占据市场的主要份额。

目前，中小型数据供应商的整体体量仍然可观，但随着业务门槛提升、客户需求多样化、价格战中利润被压缩等情况成为常态，越来越多的中小型数据供应商面临经营困境，所占市场份额将呈现持续缩小的趋势。

艾瑞咨询的数据显示，2019年中小型数据供应商份额比预期值缩小了20.8%，而这部分份额按7:3的比例向品牌数据服务商和需求方自建基础数据团队释放，因此，品牌数据服务商将在行业内部调整阶段获益最多。

作为行业的头部阵营，海天瑞声等品牌数据服务商在品牌效益、团队建设、资质、服务意识、业务能力等方面均有优势，这些企业的盈利能力和市场份额将在行业内部调整的过程中逐渐提升，并将在未来增量市场成为主要拉动力的竞争阶段占有更大的主动性。

## （5）行业发展态势

### ①需求方对训练数据的要求向精细化转型

在行业发展初期，基础数据服务行业的门槛较低，玩家鱼龙混杂，行业标准模糊，服务质量参差不齐。随着AI产业落地成为主旋律、行业整体竞争愈发激烈，需求方对训练数据质量的要求不断提高，垂直场景的定制化训练数据需求成为主流，需求方市场对训练数据的要求逐渐向精细化转型。

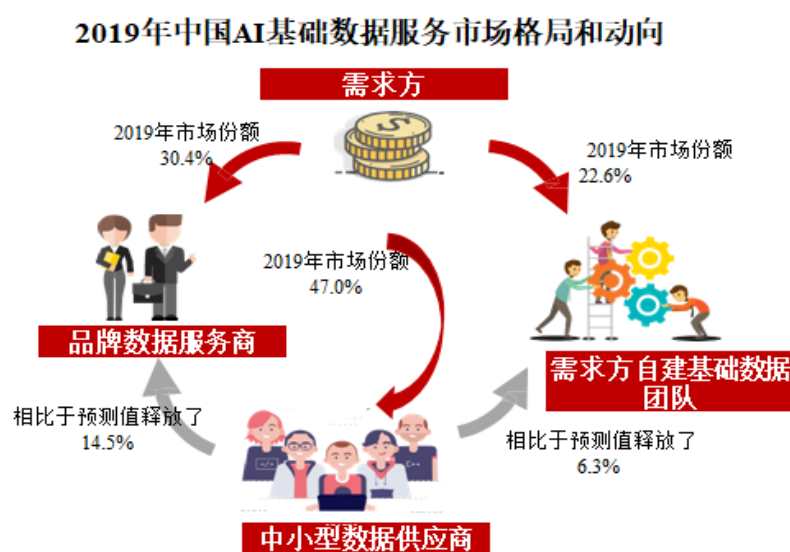
需求方对训练数据的精细化需求主要体现在两方面。一方面，人工智能算法应用要经历研发、训练和落地三个阶段，需求方根据算法应用的不同阶段对训练数据提出差异化需求：研发需求是对新拓展领域或新建算法的训练，对数据数量的要求较高，但数据标注内容倾向于标准化；训练需求一般是对算法的准确性和健壮性进行优化，对数据标注的内容需求较为丰富，对数据准确性要求较高；落地需求一般为算法较成熟的核心场景，对训练数据的内容有特定指向，采标难度较大，同时对基础数据服务商的技术能力、服务意识、稳定性和效率有较高要求。随着人工智能技术的发展及与应用场景的深入结合，训练需求和落地需求逐渐成为主流。另一方面，需求方对基础数据服务商的数据安全、采标能力、数据质量、管理能力、服务能力等核心能力提出了更高的要求：在数据安全方面，需求方对数据授权、数据传输、存储及结项后的数据销毁等环节更加重视；在采标能力方面，需求方更关注服务商在某些特定领域或垂直场景的采集能力、定制研发标注工具的能力以及对项目需求的理解程度；在数据质量方面，需求方对首次交付的准确率提出更高要求；在管理能力方面，需求方注重服务商的项目管理能力、服务效率以及执行团队的素养与信誉；在服务能力方面，需求方希望基础数据服务商具备快速响应能力和主动服务意识。

## ②中小型数据供应商市场份额缩小，品牌数据服务商价值凸显

中国基础数据服务行业的上游包括数据生产者和数据生产组织者，主要提供原料数据的采集服务；中游包括基础数据服务商，主要通过数据处理能力和项目管理能力完成训练数据集结构设计、数据加工和质量检测等工作，为下游客户提供训练数据产品和服务；下游包括科技公司、行业企业、AI公司和科研单位，主要负责算法研发。发行人位于基础数据服务行业的中游，属于基础数据服务商。

基础数据服务行业的竞争格局正发生剧烈变化。由于目前基础数据服务行业对劳动力存在密集型需求，人力输送和项目转包等服务形式仍然存在，因此中小型数据供应商仍占有一定的市场份额。但随着业务门槛提升、客户需求向精细化转型，众多中小型数据供应商在数据质量和采标能力方面达不到相关要求，将逐渐远离“利润中心”，面临被淘汰或被整合的局面。根据《行业研究报告》，2019年中小型数据服务商市场份额比预期值缩小20.8%，相应市场份额按7:3的比例

向品牌数据服务商和需求方自建基础数据团队转移，因此品牌数据服务商将获得大部分“红利份额”，商业价值逐步凸显。



数据来源：艾瑞咨询

### 3、发行人科技成果与产业深度融合情况

海天瑞声目前可以覆盖智能语音、计算机视觉、自然语言处理三大人工智能技术领域，产品和服务范围更是覆盖语音识别、语音合成、发音词典、图像与视频识别、手写体识别、光学字符识别、语义理解、机器翻译等多个技术方向。在以上领域和技术方向上，海天瑞声依靠专业的管理团队和成熟的技术架构，可以为下游客户提供所需的端到端、一站式数据研发及相关服务，服务链条贯穿上下游人工智能技术开发全周期。

经过多年的积累，截至2020年12月31日，海天瑞声已有760余个自主知识产权的可授权使用训练数据产品，并向下游客户提供了累计超3,600次/个训练数据定制服务或训练数据产品。这些产品和服务可支持超过160余个语种/方言，可覆盖生活交流、客服、家居、办公、行车、普通环境、噪声等多种特定场景中的语言现象和视觉呈现，构建成独具特色的训练数据集合，已应用于智能助理、智慧交通、智能搜索、智能家居、自动驾驶等16类应用领域，赋能人工智能技术与实体经济的深度融合。

### 4、发行人下游行业尚未接近饱和

结合下游人工智能整体行业发展情况、未来发展趋势及对训练数据的需求情况，发行人下游市场空间尚未接近饱和。具体说明如下：



### （1）全球和中国的 AI 产业规模持续增长，下游 AI 行业发展方兴未艾

全球、我国 AI 产业均持续快速发展，规模持续增长，具体参见本招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（三）所属行业发展情况和未来发展趋势/1、下游人工智能行业整体发展情况及未来发展趋势/（3）全球和中国的 AI 产业市场规模”。

（2）训练数据是当前 AI 算法发展和演进的“燃料”，对训练数据的需求持续增长

有监督的深度学习算法是实现 AI 技术的主要路径，大量训练数据是有监督的深度学习算法实现的重要基础，AI 产业对训练数据服务的需求持续产生。具体参见本招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（三）所属行业发展情况和未来发展趋势/1、下游人工智能行业整体发展情况及未来发展趋势/（1）训练数据是当前 AI 算法发展和演进的‘燃料’”。

### （3）小结

发行人服务的下游行业发展前景广阔，训练数据是算法发展演进的燃料，下游市场对训练数据的需求正在持续增长。发行人的下游市场空间未接近饱和。

## （四）发行人产品或服务的市场地位

### 1、发行人是国内领先的训练数据提供商

（1）发行人在基础数据服务行业深耕多年，拥有丰富的技术积累和行业经验，具备竞争优势

海天瑞声是我国最早进入基础数据服务行业的企业之一。自 2005 年成立以来，海天瑞声一直积极钻研核心技术、积攒行业经验，始终致力于为产业链上的各类机构提供算法模型开发训练所需的数据集，已成为我国领先的训练数据专业提供商。

通过多年的技术研发和业务实践，公司积累的核心技术横跨基础研究、平台工具、训练数据生产等三大领域，贯穿训练数据设计、原料数据采集和加工以及质量检测等应用环节，已成为公司为下游客户提供高质量训练数据产品、高效率训练数据定制服务以及高水准训练数据相关的应用服务的重要支撑。例如，在平

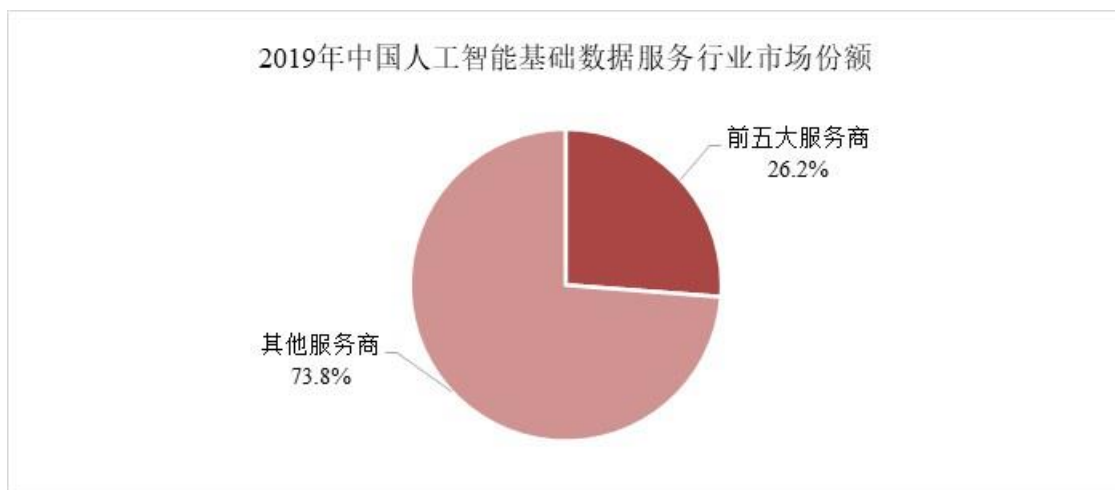
台工具方面，公司自主研发的“一体化数据处理平台”整合贯通了训练数据生产的设计、采集、加工、质检及数据安全管理等各个环节，充分挖掘了算法与技术所能创造的作用与效能，形成了多维度、多创新点的工具群，从而有效提高产能及效率、降低训练数据生产成本以及缩短开发处理时间。

凭借丰富的技术积累和行业经验，公司提供的训练数据产品及相关服务已覆盖三大技术领域、多个语种/方言和多项应用场景，赢得了阿里巴巴、腾讯、百度、微软、三星、亚马逊、科大讯飞、商汤科技、中国科学院、清华大学等众多大型优质客户的认可，核心技术、产品资源、优质客户等竞争优势逐步凸显。

## (2) 发行人是中国 AI 基础数据服务行业的头部企业

### ① 公司在中国基础数据服务行业的市场份额排名前五

海天瑞声是中国基础数据服务行业的头部企业，市场份额位居行业前五。根据艾瑞咨询发布的《2020 年中国 AI 基础数据服务行业研究报告》(以下简称“《行业研究报告》”)，2019 年，以海天瑞声、百度众包为代表的国内基础数据服务行业前五大企业的市场份额为 26.2%。行业整体仍处于低集中竞争阶段，行业活力充足，发展空间较大，作为中国领先的基础数据服务商，公司在品牌效益、团队建设、资质、服务意识、业务能力等方面均有优势，市场份额仍有较大的提升空间。



数据来源：艾瑞咨询

其中，2019 年基础数据服务行业前五大企业的市场份额情况如下：

企业	对应市场份额占比 (%)	备注
百度众包	11	-
海天瑞声	8	-
企业 3	2-3	对应的企业包括标贝科技、云测信息和数据堂, 未获得上述三家企业具体排名情况
企业 4	2-3	
企业 5	1-2	

注: 上述市场份额情况是艾瑞咨询在对市场规模调研推算的基础上调查得出的, 艾瑞咨询于 2019 年 12 月通过主要供应商(品牌服务商、中小数据供应商、自建数据团队的需求方)、及主要数据需求方(如互联网公司、智能硬件公司、AI 公司、汽车公司)的背对背调研获取, 样本的配比与单体样本的选择会对结果产生一定影响。

上表中, 百度众包为百度数据众包平台, 云测信息为北京云测信息技术有限公司, 数据堂为数据堂(北京)科技股份有限公司。

## ②公司在中国语音类基础数据服务领域排名第一

海天瑞声是中国语音类基础数据服务领域的头部企业。根据艾瑞咨询出具的《2020 年中国 AI 基础数据服务行业研究报告》, 海天瑞声在中国语音类基础数据服务领域企业中营收份额占比排名第一。按数据类型划分, 中国基础数据服务行业的市场需求可以分为语音类数据需求、图像类数据需求和自然语言处理类数据需求。现阶段国内基础数据服务商以提供图像类数据和语音类数据服务的公司为主。图像类数据业务内容涉及人像数据、OCR 数据、自动驾驶数据等, 百度众包在图像类基础数据服务领域的营收份额占比最高。而提供语音类数据产品和服务的供应门槛高于图像类数据, 业务内容包含语音识别数据、语音合成数据等, 海天瑞声在语音类基础数据服务领域的营收份额占比最高。

(3) 发行人荣膺多项资质荣誉、获评工信部国家专精特新“小巨人”企业、工信部“新一代人工智能产业创新重点任务揭榜优胜单位”, 技术实力得到行业、主管机构的高度认可。

凭借丰富的行业经验、先进的核心技术和优质的产品服务, 公司先后获得国家重点软件企业、国家高新技术企业、中关村高新技术企业等资质, 并成为中国人工智能产业发展联盟理事单位、中国语音产业联盟理事会员单位、中关村高新技术企业协会理事单位。近年来, 公司还荣膺多项国内外人工智能领域大奖: 2017 年公司被中国信息协会、信息化观察网评为“年度人工智能数据服务领域领军企业”; 同年作为亚太唯一一家人工智能基础数据供应商入选美国 IT 企业联盟杂志

APAC CIO Outlook 评选的“年度 25 家最具推荐价值人工智能企业”；2018 年公司获得中国语音产业联盟颁发的“智能语音创新大赛行业先锋奖”；同年获中国信息协会、信息化和软件服务网颁发的“2018 年人工智能行业最佳产品”奖项。

2019 年，经过多轮多方评议，海天瑞声入围“新一代人工智能产业创新重点任务入围揭榜单位”，开展“人工智能训练与测试数据资源库及服务平台”项目技术攻关工作。为贯彻落实国务院《新一代人工智能发展规划》和工业和信息化部《三年行动计划》，工业和信息化部于 2018 年 11 月启动“新一代人工智能产业创新重点任务揭榜工作”。围绕《三年行动计划》确定的重点任务方向，人工智能揭榜工作将在 17 个方向及细分领域，征集并遴选一批掌握关键核心技术、具备较强创新能力的创新主体，在人工智能主要细分领域，选拔领头羊、先锋队，以突破产业发展的短板和瓶颈，加快我国人工智能产业与实体经济深度融合，力图“重点突破一批技术先进、性能优秀、应用效果好的人工智能标志性产品、平台和服务，为产业界创新发展树立标杆和方向，培育我国人工智能产业发展的主力军。”揭榜工作主要涵盖智能产品、核心基础、智能制造关键技术装备和支撑体系四大方向，四大方向下囊括多个细分领域，海天瑞声所属的细分领域为支撑体系下的行业训练资源库。在其中，将“选择高质量的行业训练资源库等作为揭榜攻关任务。这些资源体系是影响人工智能健康发展的重要要素，需要加快完善基础环境、保障平台，加快形成我国人工智能产业发展的支撑能力。”

揭榜工作将在每个细分领域择优遴选不超过 5 家揭榜单位，择优公布揭榜成功单位不超过 3 家，旨在树立人工智能领域标杆，激发竞争活力、培育行业龙头，推动形成一批具有国际竞争力的人工智能企业。首次评选历时一年，全国参与评选的企业与机构共 1,248 家，最终 137 家单位榜上有名，入选率仅为 10.98%。除海天瑞声外，同时入选的还有阿里巴巴、腾讯、百度、科大讯飞等知名的人工智能企业与机构。2021 年 4 月，海天瑞声以优异表现获评工信部“新一代人工智能产业创新重点任务揭榜优胜单位”，充分体现了行业、主管部门对公司技术实力的高度认可，彰显了公司在基础数据服务行业的影响力。

2020 年 11 月，贯彻中办、国办《关于促进中小企业健康发展的指导意见》，根据《工业和信息化部办公厅关于开展第二批专精特新“小巨人”企业培育工作的通知》（工信厅企业函〔2020〕159 号）要求，经各省级中小企业主管部门初

核和推荐、行业协会限定性条件论证及专家审核等流程，发行人获工业和信息化部审核通过，评选为国家专精特新“小巨人”企业，其中，充分体现了行业主管部门对发行人创新性及技术实力的高度认可。

## 2、技术水平及特点

在业务经营中，发行人自主开发了一体化数据处理平台，在基础研究、平台工具、训练数据生产三个维度均积累下核心技术，有机搭建起发行人的训练数据开发一站式解决方案，提升开发效率、服务质量及数据安全性，为下游客户提供高质量的训练数据产品、高效率的训练数据定制服务及高水准的训练数据相关应用服务。

发行人拥有对 AI 核心算法的理解能力、前瞻性的专业训练数据集设计能力、丰富的语言覆盖能力及场景采集能力，以及 AI 辅助标注及精细标注能力，前述核心创新能力构建了发行人的技术储备，发行人产品及服务的技术水平及特点主要体现在下述三个方面：

### (1) 算法与数据处理技术并用

在算法、技术方面，公司掌握了 12 项核心技术，覆盖人工智能算法及训练数据设计生产过程所需，通过算法与数据处理技术并用，实现高效的数据加工及质检。

发行人将算法及数据处理技术应用于训练数据生产的各个环节，运用算法（语音识别算法、语音合成算法、计算机视觉算法等）执行自动化数据加工、质检环节，通过程序化操作结合人工检查提升训练数据生产效率，并用算法检验训练数据的训练效果、反哺指导训练数据生产；配合多种训练数据设计处理技术，如多语种多模态训练数据设计技术、数据同步技术、大数据驱动的高效数据处理技术、分布式高性能自动校验技术等，提升训练数据的生产效率和准确性，在提升自身产出效率的同时确保对客户算法模型的训练效果。

### (2) 工具和平台共建，实现训练数据生产过程的体系化管理

在工具、平台方面，公司自主开发了一体化数据处理平台，整合贯通了训练数据生产过程相关的设计、采集、加工、质检等业务环节，将项目管理、质量控制、数据安全的相关需求模块化、工具化、流程化、规范化、体系化并嵌入至一

体化数据处理平台中，充分提高了训练数据的生产效率及质量控制水平。

具体而言，一体化数据处理平台中集成了发行人自主开发积累的，训练数据生产各环节所需的工具：如采集环节使用的录音工具、视频采集工具，加工环节使用的校对工具、标注工具、声音切分工具等，质检工具等。发行人将多项技术研究成果整合至具体工具、平台中，通过技术研究成果的落地实践和应用提升训练数据生产效率。

### (3) 在语音语言学领域有深厚积累，语言覆盖能力行业领先

在语音语言学基础研究领域，发行人掌握了丰富的合作语言学家团队资源，成员遍布世界各地的学校及研究机构，在语音语言学领域具备丰富经验和技術储备；此外，发行人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，产品/服务覆盖 160 余个语种/方言，已积累下超 100 个语种/方言的发音词典，覆盖希伯来语、乌尔都语、缅甸语等小语种，累计词条数超 1000 万条，可构建高质量的智能语音训练数据，是发行人在智能语音训练数据生产领域的主要壁垒及核心技术之一。

截至目前，公司已取得 26 项专利授权（其中 24 项为发明专利，1 项为实用新型专利、1 项为外观设计专利）、拥有 134 项计算机软件著作权，另有 2 项专利已提交申请且已经受理，处于行业领先地位，具有技术优势。

## 3、行业内的主要企业

### (1) Appen

Appen Limited，1996 年成立于澳大利亚，是面向机器学习和人工智能开发的高质量人工标注数据集的提供商，已于 2015 年 1 月 7 日于澳大利亚证券交易所上市。Appen 在采集并丰富语音、文字、图像和视频等各种数据类型上积累多年经验，与全球技术、汽车和电子商务公司，以及政府部门建立了合作，帮助他们开发、利用和增强依赖自然语言和机器学习的产品。

### (2) 慧听科技

北京慧听科技有限公司成立于 2011 年，公司业务包括语音识别、语音合成、语音评测、语言文本类、多媒体类等多领域数据制作，以及语音合成、语音识别、

输入法系统的研发等。公司提供服务涵盖语音训练数据制作、音乐数据制作标注、语音质量评测等，经营模式包括定制开发和自有训练数据产品销售。

### （3）标贝科技

标贝（北京）科技有限公司成立于 2016 年，总部位于北京。主要提供智能语音交互相关服务，包括语音合成整体解决方案，以及语音合成、语音识别、图像识别等数据服务，经营模式包括定制开发和自有训练数据产品销售。

## 4、竞争优势与劣势

### （1）竞争优势

#### ①技术优势——多年积累的多项核心技术及数据处理使用的工具、平台

发行人在基础数据服务行业深耕多年，始终秉承基础研究与实际应用紧密结合的原则，持续开展基础研发等研究创新工作。

在算法、技术方面，发行人的专业研发团队结合多年训练数据生产经验需求，积累下 12 项核心技术（详见招股说明书“第六节 业务和技术/六、发行人的研发和技术/（四）各项核心技术具体情况”），尤其在多语种的语音语言学基础研究和高效数据处理技术方面积累下 5 项具备较强专业性、较高技术壁垒、较难为同行业公司或上下游行业突破的核心技术——语音语言学基础研究、多语种多模态训练数据设计技术、数据同步技术、大数据驱动的高效数据处理技术及分布式高性能自动校验技术。在主营业务领域，发行人已取得 24 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，另有 2 项专利已提交申请，较同行业公司处于领先水平。

在工具、平台方面，为使训练数据生产相关的技术落地，发行人自主开发了一体化数据处理平台，嵌入训练数据生产各环节所需的工具、软件模块，持续将训练数据生产相关算法、技术的基础研究成果运用至具体工具/平台之中，并结合市场及内部数据开发需求的变动持续调整技术应用、打磨技术细节、优化工具/平台，提升开发效率、服务质量及数据安全性，从而为客户提供高质量的训练数据产品和高效率的训练数据定制服务。

在语音语言学基础研究方面，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，产品/服务覆盖 160 余个语种/方言，已积累下超 100 个语种/方言的发音词典，覆盖希伯来语、乌尔都语、缅甸语等小语种，累计词条数超 1000 万条，可构建高质量的智能语音训练数据，是发行人在智能语音训练数据生产领域的主要壁垒及核心技术之一。

②资源优势——数百个自有知识产权的产品，多场景、多语言覆盖能力，丰富的多语种语言学家团队资源积累及稳定合作的供应商、客户等上下游资源

受益于前述专业技术积累以及多年专业客户服务经验，截至 2020 年 12 月 31 日，发行人已积累下 760 余个自有知识产权的训练数据产品，覆盖智能语音、计算机视觉及自然语言三大领域；产品/服务可覆盖 160 余个语种/方言，涉及生活交流、客服、家居、办公、行车、普通环境、噪声等多种特定场景中的语言现象和视觉呈现，可应用于智能助理、智慧交通、智能搜索、智能家居、自动驾驶等 16 类终端应用场景。

同时，发行人拥有丰富的多语种语言学家团队资源，成员遍布世界各地的学校和研究机构，在语音语言学领域具备丰富经验和技術储备。且多年来深耕基础数据服务行业，发行人与供应商、客户等上下游合作稳定，获得众多下游客户的认可。

上述资源优势根植于发行人超过 15 年的训练数据生产经验和资源积累，需要长时间的行业深耕和持续投入才能形成，存在明显优势。

③人才优势——公司高管及核心技术人员有着深厚的技术背景和丰富的行业经验。

公司高管及核心人员大多毕业于清华、北大、中科大、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。这些核心人员将其在业内优秀企业积累的技术经验和管理经验应用于公司实践，并通过传帮带培养了一批专业而精干的中层技术团队，有效提升了企业的技术水平和规范化运作水平，实现了对客户的快速响应、高品质交付。高管、核心技术人员及其培养出的专业而精干的技术团队所塑造的人才优势帮助公司创立了现在的优势市场地位，并为未来公司的长远发展提供了重要保证。



## （2）竞争劣势

①相比国际大型公司，发行人资金实力较弱、规模体量较小、覆盖的语种/方言数量相对较少

虽然发行人在国内基础数据服务行业拥有领先的行业地位，但与国际大型训练数据公司，如 Appen 相比，发行人在营收规模、净利润规模、员工数量等指标上均小于 Appen，在体量上具备一定劣势。具体来看，Appen 2020 年度取得营业收入 5.99 亿澳元，净利润 5,051 万澳元，发行人同期取得营业收入 2.33 亿人民币，净利润 8,208.10 万人民币。截至 2021 年 4 月，Appen 员工总数为 1,100 余人，同期发行人员工规模相对较小。

此外，Appen 已在基础数据服务行业深耕 20 余年，相较于发行人而言开展业务的时间更久，语种或方言的覆盖能力更强。截至 2021 年 4 月，根据公开信息可查到的 Appen 覆盖的语种或方言数量已达 230 余个，高于发行人覆盖的 160 余个。

②海外市场的业务拓展、及海外品牌影响力相对较弱

发行人通过优质的产品与服务以及与国内主要下游客户的长期合作，在国内市场获得了较大的品牌知名度，但在海外市场，发行人的品牌影响力相较于国际大型训练数据公司而言仍相对较弱。例如，Appen 已经在包括其所在国在内的全球主要国家和地区开展业务，覆盖的终端人范围已经超过 170 个国家和地区，相比之下，发行人的海外业务仍有进一步的拓展空间。同时，发行人员工、自营场地均主要分布在中国大陆地区，在海外的营销、业务人员、经营网点布局均相对有限，尚未形成较强的海外品牌影响力，在海外市场的竞争力及对海外客户的吸引力具有一定劣势，仍有较大提升空间。

③资金实力较弱，融资渠道单一

公司虽已是基础数据服务行业的领先企业，近年来发展迅速，但目前公司资金实力较弱、融资渠道单一，难以满足未来高速发展所需的大规模资金需求，亟需积极拓展多种融资渠道，以满足公司业务发展的资金需求，推动公司进一步发展壮大。

## 5、市场地位、竞争格局在报告期内的变化情况及未来变化趋势

### (1) 整体情况

发行人是国内领先的训练数据提供商。报告期内，受益于下游行业整体的高速发展以及公司自身技术和经营管理水平的提升，公司呈现出较高的成长性，营业收入稳步增长，市场地位更加稳固。

未来数年，预计国际和国内人工智能行业产业规模仍将快速增长，对训练数据产品和服务的市场需求还将不断扩大。一方面，随着下游行业的迅速发展和规模增长，需求方对训练数据的要求向精细化转型。另一方面随着行业内对训练数据标准的提高和需求类型的丰富，产业链的专业化分工将愈加清晰，专业化的训练数据提供商将扮演更加重要的角色，中小型数据供应商市场份额缩小，品牌数据服务商价值凸显。

发行人未来将继续受行业环境和自身发展的影响。在行业方面，国际和国内人工智能行业产业规模快速增长，需求方对训练数据的要求更加精细化，品牌数据服务商份额持续扩大，上述态势为公司发展提供了良好的外部环境。在自身方面，随着公司竞争优势的巩固、技术水平的提升及募投项目的实施，公司市场地位将进一步提升。

### (2) 发行人境外销售的主要进口国同类产品的竞争格局

发行人的产品和服务的主要出口地包括美国、日韩等国家或地区。在美国市场，发行人面对的主要竞争对手有 Appen、Globalme、DefinedCrowd、Scale AI 等；在日韩市场，发行人面对的主要竞争对手依然以 Appen、Globalme、DefinedCrowd 为主，同时还有一些本土的数据服务商如：Baobab（日本）和 AIMMO（韩国）等。

在美国市场上，由于美国 AI 技术发展比中国早，AI 数据服务商的整体成熟度高于中国。语言文字同源，文化同源等因素的影响，使得澳洲公司、欧美公司在美国市场有着先天的信任优势和沟通优势，发行人在这方面有一定竞争劣势；但发行人也存在以下竞争优势：1、截至 2020 年 12 月 31 日，公司已有 760 余个自有知识产权的训练数据产品，相对竞争对手有一定优势，可以帮助客户更快更好地进行算法优化；2、公司已经拥有 160 余个语种/方言的覆盖能力，不仅包括

含英法德日意等常见语种，还包括东南亚、一带一路等国家地区的罕见小语种，尤其在亚洲小语种的服务上，具备竞争优势。

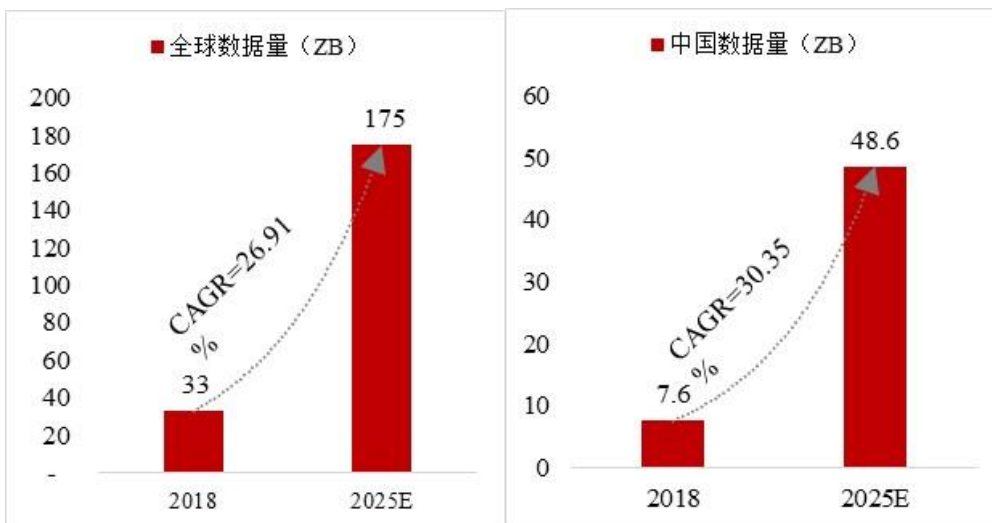
在日韩市场上，日韩科技公司或 AI 公司也会使用 Appen 以及部分美国数据供应商。同时，本土的数据服务供应商也会参与竞争。在日韩市场，与欧美竞争对手相比，公司在训练数据产品库和覆盖语言能力，尤其是亚洲语种覆盖方面具备竞争优势；与日韩本土服务商相比，公司在此领域深耕多年，具备多语种服务能力，数据资源来源广泛、储备丰富，具备竞争优势。

## 6、机遇与挑战

### (1) 机遇

#### ①数据时代来临，数据总量与处理需求量快速增长

大数据、云、物联网等信息技术的发展和互联网设备的普及产生了前所未有的海量数据，宣告了数据时代的来临。物联网的发展更使线下业务产生的大量数据被采集起来，数据量呈指数式增长，据国际数据公司（IDC）预测，全球每年的数据量将从 2018 年的 33ZB 快速增加到 2025 年的 175ZB，年复合增长率达 26.91%。其中，中国的数据量增速最快，平均每年的增长速度比全球快 3%。2018 年，中国的数据量为 7.6ZB，占全球总量的 23.4%，预计到 2025 年将增至 48.6ZB，占全球总量的 27.8%，年复合增长率达 30.35%。



数据来源：国际数据公司（IDC）

全球生产的数据中绝大部分为非结构化数据，需要经过清洗与标注才能符合

技术需求。据艾瑞咨询统计，在我国每年需要进行标注的语音数据超过 200 万小时，视觉类图片则有数亿张。数据时代的来临，产生了海量数据，也为基础数据服务行业提供了源源不断的清洗与标注数据业务需求。

### ②人工智能产业化落地，为基础数据服务行业提供发展红利

当前人工智能技术开始广泛应用于不同产业，展现出可观的商业价值和巨大的发展潜力。人工智能的产业化落地为基础数据服务行业提供巨大的发展红利。产业化应用新产品、新应用、新场景层出不穷，产生了大量新兴垂直领域的基础数据需求，作为算法训练支撑的大量落地场景内训练数据正成为行业需求痛点。下游的人工智能产业应用快速增长，为基础数据服务的发展提供了长期向好的基本面。

### ③良好的国家政策环境

我国高度重视人工智能产业的发展，出台了多个国家层面的政策规划推动行业发展。按照相关政策要求，我国将继续构建人工智能支撑体系，支持建设高质量训练资源库、标准测试数据集，进一步提升基础语音、视频图像、文本对话等训练数据量，在行业训练资源库领域取得突破。国家产业政策的支持，为基础数据服务行业发展提供了良好的政策环境。

### ④地方积极推动基础数据服务行业发展，促进人工智能创新融合

全球各地的主要城市在人工智能技术的发展中发挥了差异化作用，构建了各自的生态体系，并在赋能产业应用、助力区域经济发展方面实现初步效果。

在我国，发展基础数据服务是各地方推进人工智能建设的重要方向之一。贵州、山西、重庆等地相继出台指导意见，引入基础数据服务公司，共建数据基地、数据交易中心，打造具有地方特色的人工智能产业园。贵州省打造的惠水百鸟河数字小镇，园区自营超过 1,500 个席位的数据工场，提供了大量数据采集和标注服务，同时计划在全省范围内积极拓展新一代信息技术能力；成都市计划到 2020 年推进政府数据开放数据集 1,000 个以上，实现大数据从业人员规模和核心产业产值快速增长；山西省计划到 2022 年初步形成集数据采集、清洗、标注、交易、应用为一体的基础数据服务产业体系。

地方积极推动基础数据服务产业发展，有利于形成人工智能产业集聚，促进

创新融合，为基础数据服务行业发展创造有利局面。

## （2）挑战

### ①下游行业技术不断迭代

不断进步和发展的行业技术，对发行人研发能力和持续创新能力构成挑战。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，新的需求不断出现，技术不断迭代，发行人面临的下游需求随着行业发展不断发生变化。同时，下游客户对于基础数据服务商服务参与度以及产品质量的需求不断提高。发行人进行数据开发所需的各项技术也需要不断迭代。为紧跟下游需求变化，作为基础数据服务商需要对下游技术，如计算机视觉、智能语音、自然语言处理等算法有更深刻的理解，对行业发展趋势拥有洞察能力，同时加强对研发人才的有效组织和研发经费的经济投入，才能在下游行业技术快速迭代过程中持续保持技术先进性和技术优势。

### ②数据采集、标注业务门槛提高

随着基础数据服务行业发展，相关业务门槛不断提高。需求方对基础数据服务商的采集、标注能力提出了更高的要求，对发行人核心服务能力提出挑战。在数据采集能力方面，需求方希望基础数据服务商能覆盖更多的垂直场景，有更丰富的方言/小语种资源、全球采集渠道、场景搭建能力和特殊场景数据采集能力等；在数据标注能力方面，拥有如语音合成训练数据、3D 点云数据等高门槛数据标注能力的服务商将具有更强竞争力。

### ③业务流程各环节竞争壁垒提高

基础数据服务行业的产品形式主要为数据集产品和训练数据定制服务，二者在业务流程方面都按照训练数据设计、数据采集（或需求方提供）、数据加工、质检的步骤执行。基础数据服务商如何在各环节中建立壁垒，是决定行业地位的重要因素。

在训练数据设计环节，拥有对计算机视觉、智能语音、自然语言处理等算法更深刻理解，拥有更专业的训练数据设计能力，拥有更具前瞻性的数据集产品设计能力，以及参与过更多复杂、高难度项目的公司在获取新客户和新任务时具有明显优势。在数据采集环节，拥有更强采集能力，以及拥有稳定采集供应链的公

公司业务更具有长期发展实力。在数据加工环节，拥有更复杂数据加工能力，拥有算法辅助、人机协作能力，以及拥有全流程安全保障的公司业务更加稳定。在质检环节，拥有实时量化的可视化管理系统，拥有多重追责性的全查抽查机制，以及拥有批量检测能力与生物识别监控能力的公司在精细化管理、利润把控及客户合作方面更具优势。

#### ④发行人的管理与执行能力将面临考验

基础数据服务商的管理能力和执行能力是指在规定的周期内，高效地管理项目、高质量地服务客户的能力。随着训练数据需求多样化，以及复杂程度的提升，以往项目经理制的管理方式和使用单一工具应对单一需求的执行方式在能力和效率上都显得捉襟见肘。对于品牌数据服务商而言，客户类型丰富、数据需求多样、并发项目众多，仍使用传统方式，将会因产能天花板的压力，而限制发展规模。

基础数据服务商需要建立自主研发的贯通训练数据设计、数据采集、数据加工、质量检测、质量控制和数据安全管理等各环节于一体，并且能对图像、文本、语音、视频数据做到一站式加工处理的管理和执行一体化平台。通过建立管理和执行一体化平台，数据服务商可以提升人机协作效率、扩大产能，灵活可变地增加标注能力，准确的把控每一环节的数据安全和质量问题。

#### ⑤需求方市场向精细化转型，行业竞争加剧，行业格局将引来“洗牌”

需求方对训练数据需求的精细化主要体现在两方面。一方面，算法应用要经历研发、训练和落地三个阶段，需求方根据算法应用的不同阶段对训练数据提出了差异化需求；另一方面，需求方对基础数据服务商的数据安全、采标能力、数据质量、管理能力、服务能力等核心能力提出了更高的要求。

随着需求方市场向精细化转型，基础数据服务行业将面临项目要求提高、利润压缩、管理成本上升等问题。在未来一段时间内基础数据服务行业内竞争将加剧，行业格局将迎来“洗牌”。行业竞争格局的变化对品牌数据服务商的生产力、精细化管理能力、利润把控能力、营销能力和品牌影响力都带来了巨大的考验。

(五) 发行人与同行业可比公司在经营情况、市场地位、技术实力、衡量核心竞争力的关键业务数据、指标等方面的比较情况

1、整体情况

发行人的同行业可比公司、竞争对手主要包括 Appen、慧听科技以及标贝科技。结合前述竞争对手的公开披露信息，发行人与主要竞争对手在经营情况、技术指标、产品侧重点、研发投入、市场占有率、客户结构和客户数量等方面的比较情况如下：

项目	海天瑞声	Appen	慧听科技	标贝科技
<b>经营情况</b>				
成立年份	2005 年	1996 年	2011 年	2016 年
市场地位概述	我国领先的训练数据产品服务专业提供商，是我国最早从事训练数据产品服务研发销售的企业之一	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	-	-
行业机构评价	“北京市级企业科技研究开发机构”、“2019 中国信息技术人工智能行业优秀产品”、“2019 中国信息技术人工智能数据服务领军企业”、国家专精特新“小巨人”企业、“新一代人工智能产业创新重点任务揭榜优胜单位”等	2017-2019 年德勤亚太区“高科技高增长 500TM”、2012-2019 德勤澳洲“高科技高增长 50TM”、2019 美国常识顾问 (CSA Research) 第八大语言服务供应商	“慧听中文重口音语音数据库”曾获中国语音产业联盟“2015 中国语音创新产品”	中国企业联合会“2019 全球人工智能 TOP 50”、融资中国“新经济领域最具成长性企业 TOP30”、中国语音产业联盟会员单位、“2020 中国科创产业新锐 TOP50”、入选“2020 人工智能语义识别创新排行”等
员工数量	151 人 (截至 2020 年 12 月 31 日)	1,100 余人	未公开披露	未公开披露
市场占有率	中国 AI 基础数据服务行业第二名，行业前五大企业市场份额合计为 26.2%，其中海天瑞声的市场占有率 8%。	无公开信息披露	低于 1%-2% <sup>10</sup> (中国 AI 基础数据服务行业)	不高于 2%-3% <sup>11</sup> (中国 AI 基础数据服务行业)
<b>客户结构及客户数量</b>				

<sup>10</sup> 数据来源：艾瑞咨询。

<sup>11</sup> 数据来源：艾瑞咨询。

项目	海天瑞声	Appen	慧听科技	标贝科技
主要客户/合作伙伴情况	大型科技公司，如阿里巴巴、腾讯、百度、微软、三星等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学等	微软、谷歌等大型科技公司、汽车厂商及政府	未公开披露	大型科技公司，如阿里巴巴、腾讯、百度、网易、京东等；人工智能企业，如出门问问、旷视等；科研机构，如北京大学等。
客户数量	500 余家	未公开披露	数十家	100 余家
<b>技术指标</b>				
技术实力概述	海天瑞声自主开发了一体化数据处理支撑平台，在基础研究、平台工具、训练数据生产三个维度下均积累核心技术，将多项具体核心技术整合为发行人特有的核心技术体系。	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。	采用全程质量监控流程，执行完善的标注流程，配合保密管理手段，提供质量上乘的数据服务。	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。
语种/方言覆盖能力	160 余个	230 余个	20 个	10 个
应用领域覆盖	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	智能语音、计算机视觉、自然语言、音乐
已取得专利授权	26 项（24 项发明专利、1 项实用新型专利及 1 项外观设计专利）	2 项	无	1 项
正在申请专利数量	2 项	未公开披露	2 项	17 项
计算机软件著作权数量	134 项	未公开披露	13 项	29 项
<b>产品/服务及其侧重点</b>				
拥有的成品训练数据集数量	762 个 （截至 2020 年 12 月 31 日）	291 个	23 个	96 个
<b>其他经营指标</b>				
<b>主要财务指标（2020 年度/2020 年末）</b>				
营业收入	23,337.40 万元	59,938 万澳元	未公开披露	未公开披露



项目	海天瑞声	Appen	慧听科技	标贝科技
净利润	8,208.10 万元	5,051 万澳元	未公开披露	未公开披露

注：艾瑞咨询出具的《2020 年中国 AI 基础数据服务行业研究报告》中 AI 基础数据服务行业 CR5 涉及的百度众包、云测信息、数据堂所实际从事业务与发行人并不相同，故未列入比较。

数据来源及说明：

1、Appen、慧听科技、标贝科技数据：截至 2021 年 4 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（<http://cpquery.sipo.gov.cn/>）、中国版权保护中心 CPCC 微平台等公开信息查询渠道及第三方机构查询信息。

2、发行人数据：除特别标注外，均为截至 2021 年 5 月 7 日数据。考虑到可比性、可参考性，主要财务指标选取 2020 年度/2020 年末指标进行比较。

与 Appen 相比，发行人规模较小，营收及净利润规模、员工数量等均低于 Appen，在语种/方言覆盖能力方面也具备一定劣势；训练数据的应用领域覆盖能力相当；发行人在成品训练数据集储备情况以及我国客户的覆盖程度方面具备一定优势——发行人在训练数据服务领域深耕多年，目前与人工智能产业链上的各类机构都建立了长期的战略合作伙伴关系，国内主要合作伙伴包括阿里巴巴、腾讯、百度、科大讯飞、商汤科技、云知声、海康威视、中国科学院、清华大学等知名机构。

与慧听科技、标贝科技相比，根据艾瑞咨询数据，发行人为中国 AI 基础数据服务行业前五大企业之一，市场份额领先于慧听科技、标贝科技。发行人在训练数据覆盖的语种/方言覆盖能力、成品训练数据集储备数量、计算机软件著作权数量方面均高于慧听科技、标贝科技（限于公开信息可查询范围），优势明显；在应用领域覆盖方面，除智能语音、计算机视觉、自然语言外，标贝可覆盖音乐类训练数据；发行人未单独列示音乐类训练数据，但其语音合成类业务可提供歌曲合成类服务，也已覆盖音乐类训练数据，发行人在应用领域覆盖方面不存在明显劣势。

在市场地位方面，Appen 是较早从事训练数据开发的训练数据提供商，经营历史较长，规模、体量等相比发行人均具备优势；而发行人是我国最早从事训练数据研发、生产、销售的企业之一，在我国市场具备领先地位。多年来，发行人深耕训练数据服务领域，伴随了众多国内客户在人工智能领域特别是智能语音领域的开拓、成长，为其持续提供了全球语种语音训练数据的高质量的本土服务，降低了对国外同类训练数据的依赖。

在主要财务指标方面，慧听科技、标贝科技未公开披露其财务数据信息。根据艾瑞咨询数据，发行人为中国基础数据服务行业前五大企业之一，市场份额领先于慧听科技、标贝科技，发行人收入规模高于慧听科技、标贝科技；发行人在营业收入、净利润规模及营业收入增长率方面相比 Appen 存在一定劣势，综合毛利率则高于 Appen。

在专利储备方面，通过公开信息渠道可获悉的 Appen、慧听科技、标贝科技的专利储备数量较少，发行人在专利技术储备方面具备明显优势。在计算机软件著作权方面，慧听科技及其子公司共拥有计算机软件著作权 13 项，标贝科技及其子公司共拥有计算机软件著作权 29 项；发行人及其子公司共拥有计算机软件著作权 134 项，远高于公开信息可查询的慧听科技、标贝科技的计算机软件著作权数量，具备一定优势。

在语种/方言覆盖能力方面，根据 Appen、慧听科技、标贝科技官方网站的信息，发行人的产品和服务可以覆盖超过 160 余个语种/方言，覆盖的语种/方言数量少于 Appen，但发行人在自有知识产权训练数据产品数量上具备一定优势。与慧听科技、标贝科技相比，发行人在产品和服务覆盖的语种/方言个数、自有知识产权训练数据产品数量方面均高于公开信息可查询的慧听科技、标贝科技的相关数量，具有明显的优势。

## 2、技术竞争优劣势

### （1）竞争优势

结合上述比较可以看出，相比主要竞争对手，发行人的竞争优势主要体现在：经过多年深耕，以长期的业务实践和创新积累为基础，整合了多项核心技术，构建起围绕基础研究、平台工具和训练数据生产三个层次的完整核心技术体系。与同行业竞争对手相比，发行人的语种/方言覆盖能力处于领先水平，积累的自主知识产权的训练数据产品数量和规模远大于竞争对手。在专利储备方面具备显著优势，截至本招股说明书签署日，发行人已取得 26 项专利授权（包含 24 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），显著优于同行业竞争对手，充分反映了发行人的技术竞争优势。

### （2）竞争劣势

相比同行业竞争对手，发行人的技术竞争劣势主要体现在：与国内竞争对手相比，发行人无明显劣势；在语种/方言覆盖能力方面，发行人的覆盖能力略低于 Appen。

### 3、发行人设计的训练数据集结构与同行业竞争对手或下游客户自身数据集结构之间的差异<sup>12</sup>情况，及发行人数据集结构设计的优势说明

#### (1) 与同行业竞争对手的对比情况及优势体现

项目	海天瑞声	Appen	慧听科技	标贝科技
<b>1、智能语音类训练数据产品</b>				
数量	579	155	19	86
覆盖语种/方言数量	108	92	19	10
录音场景类别	9	7	4	2
采集设备种类	10	4	5	2
采集通道覆盖数	6	4	4	1
<b>2、自然语言类训练数据产品</b>				
数量	101	130	2	10
覆盖语种/方言数量	52	119	2	5
<b>3、计算机视觉训练数据产品</b>				
数量	82	6	2	0

针对各应用领域的训练数据产品，对其结构具体差异进一步对比如下：

#### ①智能语音训练数据产品结构对比

对于智能语音类训练数据而言，其在内容、分布方面的结构属性体现在文本内容、噪声环境、录音通道等方面。文本内容指智能语音训练数据集对应的录音文本覆盖的具体内容或类别，覆盖种类越多则该训练数据集可针对更广泛的应用领域和场景实现更好的算法模型训练效果；噪声环境指覆盖的噪声环境类型数量及分布情况，训练数据集覆盖的噪声环境越多，则语音识别算法模型经训练后可实现更多环境下的可靠识别率；录音通道指训练数据集中的语音数据来源的录音设备数量，通道数越多则该训练数据集可适配的录音设备种类越多。

<sup>12</sup> 注：本部分对比中，Appen、慧听科技、标贝科技的训练数据产品数量等相关信息来源为其官方网站或公开信息披露所列举的信息，数据截至2021年4月；发行人的训练数据产品信息为截至2020年12月31日之情况。

选取较有代表性的几类智能语音训练数据产品，与同行业主要竞争对手在训练数据结构方面比较情况如下：

#### A.通用语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
单一训练数据产品覆盖的文本类型数量	1-18 类不等	1~10 类不等	1~3 类不等	1 类
文本类型	拼读词、命令词、数字串、自然数、时间、日期、人名、地名、歌曲名、机构名、长句等	数字串、自然数、字符串、人名地名企业名、命令词、申请词、拼音词、时间和日期、长句等	命令短语、普通短语、日常短句	拼读词、命令词、数字串、自然数、时间、日期、人名、地名、歌曲名等
单一训练数据产品覆盖的噪声环境类别数量（个）	1~6	1~6	1~4	1~3
噪声环境类型	车内、办公室、家庭、医院、地铁、游戏场、餐厅/咖啡厅、商场、街道	办公室、家庭、街边、车内、公共场所、录音棚	办公室/宿舍/家、餐厅/咖啡厅、街道、车载	安静室内、车载、录音棚
单一训练数据产品覆盖的录音通道数量	1~6	1~5	1~6	1

注：Appen、慧听科技、标贝科技的智能语音训练数据产品情况据其公开披露信息整理。

#### B.车载语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
文本内容类型总数	55 类	8 类	未披露具体信息	36 类
文本内容类型示例	数字串、街道名称、控制词、地名、命令词、英语、数字串、时间、日期、度量衡、长句、自然口语等	数字、自然数、街道名称、通用指令和特定指令、字母串、地点、企业名称、自然语句等	未披露具体信息	导航找音乐，城市导航，机车系统控制，查火车，查酒店，查餐厅，查地铁，路况查询等
各训练数据产品覆盖的噪声环境类别数量	7	2	5	7
覆盖车载噪声环境具体类别	怠速、低速、低速噪音、高速、高速噪音、城市道路、城市道路噪音	怠速、高速	高速路况、低速路况、停车待机、车窗打开、车窗关闭	怠速、低速、低速噪音、高速、高速噪音、城市道路、城市道路噪音
各训练数据产品覆盖的录音通道数量	4	5	4	1

注：Appen、慧听科技、标贝科技的智能语音训练数据产品情况据其公开披露的信息整理。

### C. 特色语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
稀有语言覆盖数量	33	41	0	0

注：Appen、慧听科技、标贝科技的智能语音训练数据产品情况据其公开披露信息整理。

与同行业主要竞争对手公开披露、可在训练数据产品结构维度加以比较的智能语音类训练数据产品相对比，发行人的训练数据产品在结构方面的差异主要体现在：覆盖的噪声类型、录音通道数量、录音文本内容类型等方面。以通用语音识别训练数据产品、车载语音识别训练数据产品、特色语音识别训练数据产品这几类典型的智能语音类训练数据产品为例，发行人的训练数据产品结构整体而言可覆盖更丰富的录音文本内容类型、噪音环境类型、录音通道数量，相比境内竞争对手可覆盖更多稀有语言，具备广泛性、强适用性。

#### ② 计算机视觉训练数据产品结构对比

计算机视觉训练数据产品主要是对现实生活中的人体（包含人脸和身体姿态等）、物体（包含车辆等）、生活场景中的文字图片等图像信息等进行采集和标注所形成的训练数据产品。以人脸训练数据产品为例，计算机视觉训练数据产品的设计结构差异主要体现在人脸图像角度、光线、背景，被采集人脸对应的肤色、年龄、性别，人脸图像的标注点数等方面。针对发行人及同行业竞争对手公开披露的计算机视觉数据产品信息，将其在结构方面的差异梳理对比如下：

训练数据产品结构属性	海天瑞声	Appen	慧听科技	标贝科技
人像识别检测相关	包含了约3万不同人脸数据，覆盖黑、白、黄、棕等不同肤色。覆盖闸机口安检库、手机拍照库、人物表情库等应用场景。	包含1万多张黑人在不同角度和光线下的人脸图片；以及1,000名中国人身体动作视频等	包含2万张人物轮廓图片，覆盖不同年龄段男女老少、各类着装打扮，各类行动姿势（站立、行走、蹲坐等）	未公开披露
OCR（光学字符识别）	覆盖10余个语种，覆盖菜单、路牌、收据、书籍等17类不同场景	包含中文、泰语、芬兰语3个语种，覆盖书籍、报刊、门票、路牌、菜单等16个不同场景	未公开披露	未公开披露

注：Appen、慧听科技、标贝科技的智能语音训练数据产品情况据其公开披露信息整理。

与同行业主要竞争对手公开披露、可在训练数据产品结构维度加以比较的计算机视觉类训练数据产品相对比，发行人的训练数据产品在结构方面的差异主要体现在：在人像识别检测相关数据产品领域，发行人与同行业竞争对手均可覆盖不同类型的人脸、姿势、场景等，因计算机视觉数据产品个性化特征较强，各类训练数据产品在具体的人脸类型、姿势范围、场景等方面具备一定差异，发行人与同行业竞争对手相比不存在显著劣势。在 OCR 训练数据产品方面，慧听科技、标贝科技未披露其在该领域的产品储备情况；与 Appen 相比，发行人的 OCR 训练数据产品可覆盖更多语种的光学字符识别需要，具备广泛性、强适用性特征。

### ③自然语言训练数据产品结构对比

自然语言训练数据产品主要是对现实生活中的文本类数据进行采集标注所形成的训练数据产品。自然语言训练数据产品类型较多，例如文本来源可为新闻、短信、聊天记录、小说、诗歌、翻译句对等；标注点种类也较为繁多，如分词标注、命名实体标注、词性标注、句法结构标注、情感标注、极性标注、领域标注、意图标注等。不同自然语言训练数据产品的结构差异主要体现在文本来源、内容、标注属性等方面。同行业主要竞争对手公开信息披露的自然语言训练数据产品数量、信息均较少，未披露训练数据产品结构相关属性信息，因此无法直接对具体训练数据产品的细节结构信息进行比较。

### ④发行人数据集结构设计的优势说明

结合上述对比情况，发行人与竞争对手在公开信息渠道列示披露的训练数据产品相比，发行人设计的训练数据产品的差异之处主要体现在：

**训练数据产品数量更多：**发行人在智能语音、自然语言及计算机视觉各领域均具备较为丰富的训练数据产品储备。与国内竞争对手相比，发行人在训练数据产品数量上具备显著优势；与境外竞争对手 Appen 相比，发行人在智能语音、计算机视觉领域具备产品数量优势；在自然语言训练数据产品数量方面略有劣势。

**覆盖语种/方言数量更多：**发行人的智能语音训练数据产品覆盖的语种/方言数量相比境内竞争对手有显著优势，与境外竞争对手 Appen 基本相当；自然语言类训练数据产品覆盖的语种方言数量低于 Appen，并显著优于慧听科技、标贝科技。

采集方案复杂度更高：发行人智能语音训练数据产品的录音场景类别丰富于慧听科技、标贝科技，类别数量略低于 Appen；采集设备种类显著高于竞争对手；采集通道覆盖数与 Appen、慧听科技相当，相比标贝科技存在一定优势。综合而言，发行人的智能语音数据产品在采集方案复杂度相关指标上更为突出，可提供更为复杂的采集方案。

上述训练数据产品结构差异表明，与竞争对手相比发行人有能力设计覆盖多语种/方言、多场景、多领域的，采集方案更为复杂的训练数据产品，体现了发行人在训练数据产品设计领域的全面性、专业性优势。

(2) 与下游客户自身的训练数据产品之间的差异及体现的数据集结构设计优势

发行人的下游客户主要为人工智能应用层、技术层的各类机构，下游客户自身的训练数据产品通常服务于与其算法模型或产品应用相关的特定语种/方言及应用场景，而发行人在多场景、多领域、多语种/方言数据资源的全面性、综合性设计能力方面更有优势。此外，下游客户也常提供其获取的各类原料数据给予发行人进行专业化处理，形成可供算法模型训练使用的数据资源，这也反映了发行人在数据处理方面的全面性、专业性优势。

### 三、销售情况和主要客户

#### (一) 报告期内销售情况

##### 1、整体情况

报告期内，发行人销售收入按主要产品及服务分类具体如下：

金额单位：万元

类别	2020年		2019年		2018年	
	金额	占营业收入比重	金额	占营业收入比重	金额	占营业收入比重
训练数据定制服务	11,729.91	50.26%	14,232.91	59.91%	12,369.55	64.20%
训练数据产品	10,613.71	45.48%	9,176.47	38.63%	6,601.67	34.27%
训练数据相关的应用服务	993.77	4.26%	346.44	1.46%	294.55	1.53%
<b>主营业务收入</b>	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>

类别	2020年		2019年		2018年	
	金额	占营业收入比重	金额	占营业收入比重	金额	占营业收入比重
营业收入	23,337.40	100.00%	23,755.81	100.00%	19,265.77	100.00%

报告期内，发行人主要产品及服务均采用直接销售模式销售，不存在经销商销售情形。训练数据定制服务和训练数据产品为发行人最主要的收入来源，报告期内各年收入占比合计均超过 95%；训练数据相关的应用服务收入尚处于起步拓展阶段，收入占比较低。

对训练数据定制服务及训练数据产品销售情况具体分析如下：

## 2、训练数据定制服务销售情况

报告期内，发行人训练数据定制服务收入规模先增长后有所回落，2018 年、2019 年及 2020 年收入规模分别达到 12,369.55 万元、14,232.91 万元及 11,729.91 万元；2018 年至 2020 年占营业收入的比重分别达到 64.20%、59.91% 及 50.26%。

### (1) 按领域分类情况

按照下游应用领域划分，训练数据定制服务收入情况如下：

金额单位：万元

训练数据定制服务	2020年		2019年		2018年	
	金额	占比	金额	占比	金额	占比
智能语音	8,459.50	72.12%	10,950.46	76.94%	9,889.68	79.95%
计算机视觉	2,182.31	18.60%	2,078.46	14.60%	1,668.72	13.49%
自然语言	1,088.10	9.28%	1,203.98	8.46%	811.15	6.56%
合计	11,729.91	100.00%	14,232.91	100.00%	12,369.55	100.00%

在智能语音、计算机视觉及自然语言三大领域中，智能语音领域训练数据为发行人最早开始生产、销售的训练数据类型，为发行人的优势领域和主要收入来源，2018 年、2019 年及 2020 年占训练数据定制服务整体收入比重分别达到 79.95%、76.94% 及 72.12%；随着发行人业务的发展壮大，与下游行业发展路径相适应，发行人也逐步将业务领域拓展至计算机视觉、自然语言领域。

### (2) 产量、销量、销售价格变动情况

训练数据定制服务为发行人根据客户需求提供的定制服务，属于“以销定产”



模式，因而该类服务产量即销量。报告期内各年，训练数据定制服务的主要收入由智能语音领域贡献，各年占比均超过 70%。各领域训练数据定制服务的产/销量及平均销售价格情况如下：

### ①智能语音

业务细分类别	2020 年度		2019 年度		2018 年度	
	产/销量 (小时)	价格 (元/小时)	产/销量 (小时)	价格 (元/小时)	产/销量 (小时)	价格 (元/小时)
语音识别采集	42,638	294	33,417	616	58,341	482
语音识别转写	107,041	319	78,709	326	90,923	326
语音合成	3,580	10,590	2,733	23,153	1,500	27,458
合计	153,259	552	114,859	953	150,765	656

### ②计算机视觉、自然语言

细分领域	2020 年度		2019 年度		2018 年度	
	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)
计算机视觉	172	12.69	213	9.76	185	9.02
自然语言	247	4.41	225	5.35	250	3.24

上述产/销量及平均售价变动的具体分析请参见本招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/4、主要产品和服务的销量和价格分析”。

### 3、训练数据产品销售情况

报告期内，发行人训练数据产品收入稳步上升，2018 年、2019 年及 2020 年收入规模分别达到 6,601.67 万元、9,176.47 万元及 10,613.71 万元，占营业收入比重分别为 34.27%、38.63% 及 45.48%。

报告期内，发行人各期间训练数据产品的库存量、新增量和销售量如下：

单位：个数/次数

类别	2020 年	2019 年	2018 年
期末 / 年末库存量 (个数)	762	619	516
新增量 (个数)	143	103	51
销售量 (个数)	395	293	183
销售量 (授权次数)	576	414	265

报告期内，发行人不存在将定制服务产生的数据直接或修改后用于训练数据产品的情形。训练数据定制服务为发行人根据客户的个性化要求，为客户量体裁衣地提供定制化训练数据集的设计及开发服务，训练数据定制服务所形成的训练数据所有权在销售时已经转移至客户，公司不再拥有该等训练数据的所有权，并且根据公司内部相关业务制度，由专人在有效期截止日清除训练数据定制服务已交付的业务数据，有效期需要根据合同约定确定，合同中无明确约定的，有效期为6个月。发行人严格执行上述规定，不存在将定制服务形成的训练数据直接或修改后用于训练数据产品的情形。

#### (1) 按领域分类整体情况

按照下游应用领域划分，训练数据产品收入情况如下：

金额单位：万元

训练数据产品	2020年		2019年		2018年	
	金额	占比	金额	占比	金额	占比
智能语音	9,648.81	90.91%	8,170.50	89.04%	6,017.52	91.15%
计算机视觉	408.29	3.85%	629.25	6.86%	186.34	2.82%
自然语言	556.61	5.24%	376.72	4.11%	397.82	6.03%
<b>合计</b>	<b>10,613.71</b>	<b>100.00%</b>	<b>9,176.47</b>	<b>100.00%</b>	<b>6,601.67</b>	<b>100.00%</b>

对于训练数据产品而言，智能语音领域同样为发行人的优势领域和主要收入来源，报告期内占比均超过89%。

#### (2) 产量、销量、销售价格变动整体情况

发行人拥有训练数据产品的知识产权，可通过销售训练数据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，因此各年训练数据产品的销量主要反映训练数据产品在当年的销售情况，而产量主要反映发行人当年新增的训练数据产品储备情况。

发行人的训练数据产品中，智能语音为收入贡献最多的业务领域，占各年训练数据产品收入的比重约为90%。报告期内各年，发行人智能语音领域训练数据产品的产量、销量及平均销售价格变动情况如下：

业务细分类别	2020 年度			2019 年度			2018 年度		
	产量 (小时)	销量 (小时)	价格 (元/小时)	产量 (小时)	销量 (小时)	价格 (元/小时)	产量 (小时)	销量 (小时)	价格 (元/小时)
语音识别采集	33,018	198,951	413	27,288	136,411	549	19,902	115,636	504
语音合成	428	1,029	13,832	15	274	24,946	15	96	20,124
合计	<b>33,445</b>	<b>199,980</b>	<b>482</b>	<b>27,303</b>	<b>136,685</b>	<b>598</b>	<b>19,917</b>	<b>115,732</b>	<b>520</b>

对产量变动情况分析如下：2018 年-2019 年发行人训练数据产品呈现较为突出的产量上升趋势，主要原因是：2018 年，发行人训练数据定制服务业务需求、订单较多，发行人将主要的研发资源及人员配备向前述业务倾斜，使得当年训练数据产品储备增加速度略有放缓；2019 年发行人发力推进训练数据产品储备，当年产量有较大上升。2020 年，训练数据产品产量相比 2019 年有所增长，主要原因是：新冠肺炎疫情爆发以来，成品数据集发挥了较强的销售带动作用，发行人结合成品训练数据集在当期销售量较大的市场形势，除对已产成的训练数据产品进行升级完善以更好地满足当期市场变化外，也加大力度研发新训练数据产品。

对销量及平均售价变动的具体分析请参见本招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/4、主要产品和服务的销量和价格分析”。

### （3）训练数据产品销售具体情况

报告期内，发行人训练数据产品的具体销售情况（主要内容、资源类型、数量、授权使用次数、平均售价）如下：

注：发行人的训练数据产品开发完成后可授权予客户使用，一次生产完成后可重复多次销售使用权。下文列示的“授权使用次数合计”指标为各期实现销售的训练数据产品在当期销售使用权的次数合计值。

#### ①2020 年训练数据产品的具体销售情况

##### A. 智能语音领域

序号	资源类型		主要内容	2020 年销售情况				发行人积累的训练数据产品总数（个，截至 2020 年 12 月 31 日）
				当期实现销售的训练数据产品数量（个）	授权使用次数合计（次）	销量（小时）	平均售价（元/小时）	
1	智能语音	中文	唤醒词、单句、对话等	101	166	91,455	255	180

序号	资源类型	主要内容	2020 年销售情况				发行人积累的 训练数据产品 总数（个，截 至 2020 年 12 月 31 日）
			当期实现销 售的训练数 据产品数量 （个）	授权使用次 数合计（次）	销量 （小时）	平均售价 （元/小时）	
	外语种	唤醒词、单 句、对话等	233	341	108,525	674	399

## B. 计算机视觉、自然语言领域

序号	资源类型	主要内容	2020 年销售情况			发行人积累的训 练数据产品总数 （个，截至 2020 年 12 月 31 日）
			当期实现销售的训 练数据产品数量 （个）	授权使用次数 合计（次）	平均售价 （万元）	
1	自然语言	词典	33	36	15.46	101
2	计算机视觉	人脸、人像 目标等	28	33	12.37	82

## ②2019 年训练数据产品的具体销售情况

## A. 智能语音领域

序号	资源类型	主要内容	2019 年销售情况				发行人积累的 训练数据产品 总数（个，截 至 2019 年 12 月 31 日）
			当期实现销 售的训练数 据产品数量 （个）	授权使用次 数合计（次）	销量 （小时）	平均售价 （元/小时）	
1	智能 语音	中文	79	114	55,957	317	166
		外语种	171	245	80,728	793	317

## B. 计算机视觉、自然语言领域

序号	资源类型	主要内容	2019 年销售情况			发行人积累的训 练数据产品总数 （个，截至 2019 年 12 月 31 日）
			当期实现销售的 训练数据产品数 量（个）	授权使用次 数合计（次）	平均售价 （万元）	
1	自然语言	词典、语料等	26	30	12.56	82
2	计算机视觉	人脸、手势、活体对抗、 自然场景、人车 2D、交 通标志等	17	25	25.17	54

## ③2018 年训练数据产品的具体销售情况

## A. 智能语音领域

序号	资源类型		主要内容	2018年销售情况				发行人积累的 训练数据产品总数 (个,截至2018 年12月31日)
				当期实现销售 的训练数据产 品数量(个)	授权使用次 数合计(次)	销量(小 时)	平均售 价(元/ 小时)	
1	智能 语音	中文	唤醒词、单句、 对话等	57	88	68,587	364	151
		外语种	唤醒词、单句、 对话等	90	123	47,145	747	256

## B. 计算机视觉、自然语言领域

序号	资源类型	主要内容	2018年销售情况			发行人积累的训 练数据产品总数 (个截至2018年 12月31日)
			当期实现销售的训 练数据产品数量 (个)	授权使用次数 合计(次)	平均售价 (万元)	
1	自然语言	词典、语料等	17	29	13.72	72
2	计算机视觉	人脸、人车2D、行 车、前/后置人像等	19	25	7.45	37

## (二) 主要客户情况

### 1、前五大客户情况

报告期内，发行人向前五名客户（受同一控制的客户合并计算销售额）销售的情况如下：

单位：万元

2020年				
序号	客户品牌/名称	与发行人是否存在关联关系	营业收入	占比
1	微软	否	4,332.18	18.56%
2	阿里巴巴	否	3,030.38	12.99%
3	腾讯	否	1,830.75	7.84%
4	三星	否	1,383.66	5.93%
5	字节跳动	否	1,149.06	4.92%
合计			<b>11,726.03</b>	<b>50.25%</b>
营业收入			<b>23,337.40</b>	<b>100.00%</b>
2019年				
序号	客户品牌/名称	与发行人是否存在关联关系	营业收入	占比
1	三星	否	4,762.06	20.05%
2	微软	否	2,876.20	12.11%
3	阿里巴巴	否	2,208.26	9.30%
4	亚马逊	否	1,492.98	6.28%

5	腾讯	否	1,459.31	6.14%
合计			<b>12,798.80</b>	<b>53.88%</b>
营业收入			<b>23,755.81</b>	<b>100.00%</b>
<b>2018年</b>				
序号	客户品牌/名称	与发行人是否存在关联关系	营业收入	占比
1	阿里巴巴	否	5,179.07	26.88%
2	三星	否	2,436.81	12.65%
3	腾讯	否	1,574.16	8.17%
4	微软	否	1,299.73	6.75%
5	百度	否	888.29	4.61%
合计			<b>11,378.06</b>	<b>59.06%</b>
营业收入			<b>19,265.77</b>	<b>100.00%</b>

报告期内各年，发行人前五名客户主要为人工智能产业链上的各类知名机构，如阿里巴巴、腾讯、百度、微软、三星、亚马逊、字节跳动等，各年前五大客户与发行人均不存在关联关系。

2018年、2019年及2020年，发行人向前五名客户合计销售额占当期营业收入比重分别为59.06%、53.88%及50.25%，向单个客户的销售比例均不超过当期营业收入总额的50%，不存在对单一客户的依赖。

## 2、前五大客户销售内容及营业收入变动分析

报告期内各期，发行人前五大客户的直接签约方或接受服务方名称、销售内容及其对应的营业收入和占比明细具体如下：

单位：万元

2020年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	微软	微软（中国）有限公司、Microsoft Corporation、Microsoft India R&D Private Limited	4,332.18	18.56%	智能语音-训练数据定制服务、自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务、训练数据相关的应用服务
2	阿里巴巴	淘宝（中国）软件有限公司、阿里云计算有限公司、支付宝（中国）网络技术有限公司、浙江天猫技术有限公司、斑马网络技术有限公司、支付宝（杭州）信息技术有限公司、阿里巴巴丝路有限公司、阿里巴巴科技（北京）	3,030.38	12.99%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务及训练数据产品、训练数据相关的应用服务

		有限公司、北京高德云图科技有限公司、浙江菜鸟供应链管理有限公司、阿里巴巴（中国）网络技术有限公司			
3	腾讯	腾讯科技（成都）有限公司、Tencent Mobile International Limited、腾讯科技（深圳）有限公司、财付通支付科技有限公司	1,830.75	7.84%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务及训练数据产品
4	三星	Samsung Electronics Co., Ltd.、Samsung Research America, Inc.、Samsung R&D Institute India-Bangalore Pvt. Ltd.、北京三星通信技术研究有限公司	1,383.66	5.93%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务、训练数据相关的应用服务
5	字节跳动	北京有竹居网络技术有限公司、北京吉云互动科技有限公司、北京字节跳动网络技术有限公司	1,149.06	4.92%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据产品
合计			<b>11,726.03</b>	<b>50.25%</b>	
营业收入总额			<b>23,337.40</b>	<b>100.00%</b>	
<b>2019年</b>					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	三星	Samsung Electronics Co., Ltd.、北京三星通信技术研究有限公司、Samsung R&D Institute India-Bangalore Pvt. Ltd.	4,762.06	20.05%	智能语音-训练数据定制服务及训练数据产品；自然语言-训练数据定制服务
2	微软	微软（中国）有限公司、Microsoft Corporation、Microsoft India R&D Private Limited	2,876.20	12.11%	智能语音-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务、自然语言-训练数据定制服务及训练数据产品
3	阿里巴巴	淘宝（中国）软件有限公司、浙江天猫技术有限公司、广州神马移动信息科技有限公司、广州阿里巴巴文学信息技术有限公司、阿里云计算有限公司、支付宝（中国）网络技术有限公司、浙江菜鸟供应链管理有限公司、阿里巴巴（中国）网络技术有限公司、上海盒马网络科技有限公司、阿里巴巴丽人购（上海）电子商务有限公司、阿里巴巴科技（北京）有限公司	2,208.26	9.30%	智能语音-训练数据定制服务及训练数据产品；计算机视觉-训练数据定制服务；自然语言-训练数据定制服务及训练数据产品
4	亚马逊	Amazon Web Services, Inc.	1,492.98	6.28%	智能语音-训练数据产品、智能语音-训练数据定制服务
5	腾讯	腾讯科技（成都）有限公司、腾讯科技（深圳）有限公司、深圳市腾讯计算机系统有限公司、财付通支付科技有限公司	1,459.31	6.14%	智能语音-训练数据定制服务及训练数据产品；自然语言-训练数据定制服务及训练数据产品；计算机视觉-训练数据定制服务及训练数据产品
合计			<b>12,798.80</b>	<b>53.88%</b>	
营业收入总额			<b>23,755.81</b>	<b>100.00%</b>	

2018年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	阿里巴巴	淘宝(中国)软件有限公司、阿里巴巴(中国)网络技术有限公司、浙江天猫技术有限公司、广州阿里巴巴文学信息技术有限公司、支付宝(中国)网络技术有限公司、阿里云计算有限公司、广州阿里巴巴文学信息技术有限公司北京分公司	5,179.07	26.88%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务
2	三星	Samsung Electronics Co., Ltd.、北京三星通信技术研究有限公司、Samsung R&D Institute India-Bangalore Pvt. Ltd.	2,436.81	12.65%	智能语音-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务;自然语言-训练数据定制服务及训练数据产品
3	腾讯	腾讯科技(深圳)有限公司、深圳市腾讯计算机系统有限公司、腾讯科技(成都)有限公司、腾讯科技(上海)有限公司、腾讯数码(天津)有限公司	1,574.16	8.17%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务
4	微软	微软(中国)有限公司、Microsoft Corporation	1,299.73	6.75%	计算机视觉-训练数据定制服务;智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务
5	百度	北京百度网讯科技有限公司、百度国际科技(深圳)有限公司、百度时代网络技术(北京)有限公司、百度在线网络技术(北京)有限公司	888.29	4.61%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务、及训练数据产品
合计			11,378.06	59.06%	
营业收入总额			19,265.77	100.00%	
2017年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	腾讯	腾讯科技(深圳)有限公司、深圳市腾讯计算机系统有限公司、腾讯科技(上海)有限公司	1,770.05	14.87%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务及训练数据产品
2	三星	北京三星通信技术研究有限公司、Samsung Electronics Co., Ltd.	1,553.08	13.04%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务;训练数据相关的应用服务
3	阿里巴巴	阿里云计算有限公司、淘宝(中国)软件有限公司、浙江天猫技术有限公司、广州神马移动信息科技有限公司、阿里巴巴(北京)软件服务有限公司、支付宝(中国)网络技术有限公司、浙江阿里巴巴机器人有限公	1,132.01	9.51%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务及训练数据产品;训练数据相关的应用服务



		司、阿里巴巴（中国）网络技术有限公司			
4	微软	微软（中国）有限公司、Microsoft Corporation	1,058.87	8.89%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务及训练数据产品
5	百度	北京百度网讯科技有限公司	585.29	4.92%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务
合计			<b>6,099.29</b>	<b>51.22%</b>	
营业收入总额			<b>11,907.09</b>	<b>100.00%</b>	

对发行人在报告期内向上述客户销售收入变化原因具体分析如下：

报告期内，发行人各期前五大客户整体较为稳定，主要包括三星、微软、阿里巴巴、亚马逊、腾讯、百度及字节跳动共计 7 个人工智能产业链上的知名机构。

对于前述客户在报告期内贡献的营业收入变动情况分析如下：

前述主要客户在报告期内的营业收入变动情况：

金额单位：万元

项目	营业收入		
	2020 年	2019 年	2018 年
主要客户合计	<b>12,412.96</b>	<b>14,391.30</b>	<b>12,219.40</b>

前述 7 个主要客户营业收入在 2018-2019 年间有所增长，2020 年略有回落，整体保持在较高水平。2020 年，一方面由于部分客户前期在一些领域向发行人采购了较大规模的训练数据，可以支撑其在一段时间内的产品应用拓展和性能迭代等需求，或因部分客户的研发节奏安排，其在 2020 年针对特定领域的训练数据需求量有所减少，导致当期采购量有所下降；另一方面，新冠肺炎疫情也对部分境外客户在 2020 年下半年采购训练数据规模造成了一定影响，因此 2020 年主要客户贡献的收入规模相比 2019 年有所回落。

### 3、各类型目标客户情况

发行人的目标客户可分为大型科技公司、人工智能企业、科研机构<sup>13</sup>三类。截至 2020 年 12 月 31 日，发行人不同目标客户类型的客户数量情况如下：

<sup>13</sup> 大型科技公司主要为三星、阿里巴巴、微软、腾讯、亚马逊等业务范围较广的知名综合性科技企业，人工智能企业主要为海康威视、科大讯飞等主要专长于人工智能具体技术领域（如智能语音、计算机视觉等）的企业；科研院所主要指中国科学院、清华大学等开展人工智能相关科学研究的研究院、研究所、大专院校等机构。

目标客户类型	客户数量	客户示例
大型科技公司	24 个品牌、80 个客户主体	三星、阿里巴巴、微软、腾讯、亚马逊等
人工智能企业	301 个品牌、394 个客户主体	海康威视、科大讯飞、Naver、Verint 等
科研机构	66 个品牌、76 个客户主体	中国科学院、清华大学等

注：上表统计客户均为历史上曾为发行人贡献业务收入或已与发行人签署业务合作协议/合同并实际开展了业务往来的客户群体。

不同目标客户类型的产品销售或服务的具体情况如下：

金额单位：万元

销售收入	2020 年		2019 年		2018 年	
	营业收入	占比	营业收入	占比	营业收入	占比
<b>大型科技公司</b>	<b>16,187.06</b>	<b>69.36%</b>	<b>17,003.58</b>	<b>71.58%</b>	<b>14,188.80</b>	<b>73.65%</b>
其中：训练数据定制服务	6,455.04	27.66%	11,626.58	48.94%	9,788.06	50.81%
训练数据产品	9,239.07	39.59%	5,376.49	22.63%	4,400.74	22.84%
训练数据相关的应用服务	492.95	2.11%	0.51	0.00%	-	-
<b>人工智能企业</b>	<b>6,507.15</b>	<b>27.88%</b>	<b>6,049.60</b>	<b>25.47%</b>	<b>4,641.46</b>	<b>24.09%</b>
其中：训练数据定制服务	3,725.63	15.96%	2,139.52	9.01%	2,187.33	11.35%
训练数据产品	2,405.92	10.31%	3,583.13	15.08%	2,160.50	11.21%
训练数据相关的应用服务	375.59	1.61%	326.95	1.38%	293.64	1.52%
<b>科研机构</b>	<b>643.19</b>	<b>2.76%</b>	<b>702.63</b>	<b>2.96%</b>	<b>435.50</b>	<b>2.26%</b>
其中：训练数据定制服务	433.05	1.86%	466.81	1.97%	394.17	2.05%
训练数据产品	84.92	0.36%	216.85	0.91%	40.43	0.21%
训练数据相关的应用服务	125.22	0.54%	18.98	0.08%	0.91	0.00%
<b>营业收入合计</b>	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100.00%</b>	<b>19,265.77</b>	<b>100.00%</b>

报告期内，大型科技公司类客户贡献的收入金额最大、占比最高，发行人主要向其销售训练数据定制服务和训练数据产品，其中 2018 年、2019 年销售以训练数据定制服务为主，2020 年因疫情对现场采集、标注工作产生了一定影响，销售以训练数据产品为主。大型科技公司类客户贡献的收入金额较高的原因是：一方面是由于阿里巴巴、腾讯、百度、微软、三星、亚马逊等大型科技公司近些年加大了在人工智能领域的投入，相应提升了对训练数据产品和服务的采购规模；另一方面是由于发行人通过长期经营树立了品牌形象和市场地位，对大型客户保持了较高的吸引力和黏性。

人工智能企业类别客户贡献收入仅次于大型科技公司，发行人向该类客户销售呈现训练数据定制服务和训练数据产品并重的现象，2018年、2019年训练数据产品整体销售收入略高于训练数据定制服务，2020年训练数据定制服务占比较高，主要是因为随部分人工智能企业业务和研发进程发展，其采购的训练数据从成品训练数据集向更个性化、更符合其算法模型研发需要的定制型训练数据转移。

科研机构类客户贡献收入占比较低，发行人向其销售的也主要是训练数据定制服务和训练数据产品。

#### 4、按产品类别分类的前五大客户情况

##### (1) 训练数据定制服务

训练数据定制服务各期前五大客户及其收入情况如下：

金额单位：万元

2020年				
序号	客户名称（品牌）	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	微软	3,071.03	26.18%	13.16%
2	阿里巴巴	1,995.34	17.01%	8.55%
3	腾讯	1,473.93	12.57%	6.32%
4	某国际消费电子产品厂商	928.33	7.91%	3.98%
5	中国某大型科技公司	508.99	4.34%	2.18%
合计		<b>7,977.63</b>	<b>68.01%</b>	<b>34.18%</b>
训练数据定制服务收入总额		<b>11,729.91</b>	<b>100.00%</b>	<b>50.26%</b>
营业收入总额		<b>23,337.40</b>		<b>100.00%</b>
2019年				
序号	客户名称（品牌）	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	三星	4,189.52	29.44%	17.64%
2	阿里巴巴	2,150.34	15.11%	9.05%
3	微软	1,385.40	9.73%	5.83%
4	腾讯	1,074.74	7.55%	4.52%
5	某国际消费电子产品厂商	920.39	6.47%	3.87%

合计		<b>9,720.39</b>	<b>68.30%</b>	<b>40.92%</b>
训练数据定制服务收入总额		<b>14,232.91</b>	<b>100.00%</b>	<b>59.91%</b>
营业收入总额		<b>23,755.81</b>	-	<b>100.00%</b>
<b>2018 年</b>				
序号	客户名称 (品牌)	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	阿里巴巴	4,452.41	35.99%	23.11%
2	三星	2,269.89	18.35%	11.78%
3	腾讯	1,193.79	9.65%	6.20%
4	微软	972.26	7.86%	5.05%
5	百度	391.19	3.16%	2.03%
合计		<b>9,279.52</b>	<b>75.02%</b>	<b>48.17%</b>
训练数据定制服务收入总额		<b>12,369.55</b>	<b>100.00%</b>	<b>64.20%</b>
营业收入总额		<b>19,265.77</b>	-	<b>100.00%</b>

## (2) 训练数据产品

训练数据产品各期前五大客户及其收入情况如下：

金额单位：万元

<b>2020 年</b>				
序号	客户名称 (品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	三星	1,257.80	11.85%	5.39%
2	微软	1,250.52	11.78%	5.36%
3	字节跳动	1,020.22	9.61%	4.37%
4	Google	669.74	6.31%	2.87%
5	阿里巴巴	552.22	5.20%	2.37%
合计		<b>4,750.50</b>	<b>44.76%</b>	<b>20.36%</b>
训练数据产品收入总额		<b>10,613.71</b>	<b>100.00%</b>	<b>45.48%</b>
营业收入总额		<b>23,337.40</b>		<b>100.00%</b>
<b>2019 年</b>				
序号	客户名称 (品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	微软	1,490.80	16.25%	6.28%
2	亚马逊	901.45	9.82%	3.79%
3	客户 2	666.66	7.26%	2.81%
4	三星	572.03	6.23%	2.41%

5	中国某大型科技公司	525.53	5.73%	2.21%
合计		<b>4,156.47</b>	<b>45.29%</b>	<b>17.50%</b>
训练数据产品收入总额		<b>9,176.47</b>	<b>100.00%</b>	<b>38.63%</b>
营业收入总额		<b>23,755.81</b>	-	<b>100.00%</b>
<b>2018年</b>				
序号	客户名称(品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	Google	804.89	12.19%	4.18%
2	阿里巴巴	726.67	11.01%	3.77%
3	中国某大型科技公司	622.51	9.43%	3.23%
4	百度	497.10	7.53%	2.58%
5	亚马逊	422.42	6.40%	2.19%
合计		<b>3,073.60</b>	<b>46.56%</b>	<b>15.95%</b>
训练数据产品收入总额		<b>6,601.67</b>	<b>100.00%</b>	<b>34.27%</b>
营业收入总额		<b>19,265.77</b>	-	<b>100.00%</b>
营业收入总额		<b>11,907.09</b>	-	<b>100.00%</b>

### (3) 主要客户的获取方式、合作历史

发行人获取客户的方式主要包括：

情形一：于公司成立初期即开始合作，早期客户开拓主要来自公司创始人贺琳多年积累的人脉基础；

情形二：通过学术会议接触，之后经过客户拜访、意向接触、商务洽谈等步骤建立合作关系；

情形三：因公司在行业内有良好的口碑，合作过的客户向业内其他客户推荐发行人从而建立联系，进而开展合作；

情形四：通过网络挖掘，获得客户的公共联系方式，然后有针对性地向客户发送邮件主动进行宣传推广，进而建立联系、开展合作；

情形五：通过电话销售，获得客户联系人信息，进而通过邮件沟通、客户拜访等方式获取客户需求，进行合作；

情形六：客户通过公司官网主动向发行人留言咨询，后续通过意向接触、商务洽谈等步骤建立合作。

上述主要客户的获取方式、合作历史情况如下：

主要客户名称（品牌）	获取方式	合作起始时间
阿里巴巴	情形二	2010
百度	情形三	2011
三星	情形二	2010
腾讯	情形二、情形三	2012
微软	情形一	2005
中国某大型科技公司	情形二、情形四	2012
亚马逊	情形二	2012
某国际消费电子产品厂商	情形四	2010
Google	情形二、情形五	2017
客户 2	情形三、情形六	2015
字节跳动	情形三	2017

报告期内，发行人的销售模式与行业普遍模式相同，即主要采用直接对接并服务客户的直销模式进行销售。基于多年的行业经验和知识储备，公司以高品质的训练数据产品及服务吸引客户，并在持续服务的过程中提升服务价值和客户黏度。

公司获取客户的具体方式主要是：通过口碑传播（客户介绍）、参与学术会议、官方网站展示等方式建立品牌知名度以及与客户之间的沟通联系，后续按照客户的采购要求以商务谈判、招投标等形式获取业务机会。

初次获取客户后，发行人会通过常规方式维持与客户的稳定交流和持续沟通，不断挖掘新的业务机会、开展合作，具体举措包括：及时响应客户主动下达的订单需求，与客户及时沟通并相应提供产品服务；不定期联系、拜访客户，了解客户业务最新发展动向和潜在的训练数据需求，并向客户推介发行人新增的训练数据产品、服务等。

## 5、境内外前五大客户及其销售情况

### （1）境内前五大客户及其情况

金额单位：万元

2020年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	微软	3,252.99	21.73%	13.94%	智能语音-训练数据定制服务、智能语音-训练数据产品、计算机视觉-训练数据定制服务、自然语言-训练数据定制服务、训练数据相关的应用服务
2	阿里巴巴	3,030.38	20.24%	12.99%	智能语音-训练数据定制服务、智能语音-训练数据产品、计算机视觉-训练数据定制服务、计算机视觉-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品、训练数据相关的应用服务
3	腾讯	1,763.40	11.78%	7.56%	智能语音-训练数据定制服务、智能语音-训练数据产品、计算机视觉-训练数据定制服务、计算机视觉-训练数据产品、自然语言-训练数据定制服务
4	字节跳动	1,149.06	7.68%	4.92%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品
5	中国某大型科技公司	822.41	5.49%	3.52%	智能语音-训练数据定制服务、智能语音-训练数据产品、计算机视觉-训练数据定制服务、计算机视觉-训练数据产品、自然语言-训练数据定制服务
合计		<b>10,018.23</b>	<b>66.93%</b>	<b>42.93%</b>	
境内收入总额		<b>14,968.83</b>	<b>100.00%</b>	<b>64.14%</b>	
营业收入总额		<b>23,337.40</b>		<b>100.00%</b>	
2019年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	阿里巴巴	2,208.26	18.44%	9.30%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务
2	腾讯	1,459.31	12.19%	6.14%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品、计算机视觉-训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务
3	微软	1,415.71	11.82%	5.96%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、自然语言-训练数据定制服务、智能语音-训练数据产品
4	中国某大型科技公司	1,019.94	8.52%	4.29%	计算机视觉-训练数据产品、智能语音-训练数据定制服务、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品
5	百度	544.90	4.55%	2.29%	智能语音-训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据定制服务
合计		<b>6,648.11</b>	<b>55.52%</b>	<b>27.99%</b>	
境内收入总额		<b>11,973.75</b>	<b>100.00%</b>	<b>50.40%</b>	
营业收入总额		<b>23,755.81</b>		<b>100.00%</b>	
2018年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	阿里巴巴	5,179.07	38.96%	26.88%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品、计算机视觉-训练数据定制服务

2	腾讯	1,574.16	11.84%	8.17%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务
3	微软	957.75	7.20%	4.97%	计算机视觉-训练数据定制服务、智能语音-训练数据定制服务、自然语言-训练数据定制服务
4	百度	888.29	6.68%	4.61%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品
5	中国某大型科技公司	650.97	4.90%	3.38%	智能语音-训练数据产品、自然语言-训练数据产品、计算机视觉-训练数据产品、自然语言-训练数据定制服务、智能语音-训练数据定制服务
合计		<b>9,250.24</b>	<b>69.58%</b>	<b>48.01%</b>	
境内收入总额		<b>13,293.77</b>	<b>100.00%</b>	<b>69.00%</b>	
营业收入总额		<b>19,265.77</b>		<b>100.00%</b>	

## (2) 境外前五大客户及其情况

金额单位：万元

2020年					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	三星	1,336.20	15.97%	5.73%	智能语音-训练数据定制服务、智能语音-训练数据产品、训练数据相关的应用服务
2	微软	1,079.20	12.90%	4.62%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品、训练数据相关的应用服务
3	某国际消费电子产品厂商	928.33	11.09%	3.98%	计算机视觉-训练数据定制服务、自然语言-训练数据定制服务
4	Google	669.74	8.00%	2.87%	智能语音-训练数据产品
5	客户 1	490.48	5.86%	2.10%	智能语音-训练数据产品
合计		<b>4,503.94</b>	<b>53.82%</b>	<b>19.30%</b>	
境外收入总额		<b>8,368.56</b>	<b>100.00%</b>	<b>35.86%</b>	
营业收入总额		<b>23,337.40</b>		<b>100.00%</b>	
2019年					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	三星	4,269.02	36.23%	17.97%	智能语音-训练数据定制服务、智能语音-训练数据产品
2	亚马逊	1,492.98	12.67%	6.28%	智能语音-训练数据产品、智能语音-训练数据定制服务
3	微软	1,460.49	12.40%	6.15%	智能语音-训练数据产品、自然语言-训练数据产品
4	某国际消费电子产品厂商	920.39	7.81%	3.87%	计算机视觉-训练数据定制服务、自然语言-训练数据定制服务
5	客户 2	666.66	5.66%	2.81%	智能语音-训练数据产品、自然语言-训练数据产品
合计		<b>8,809.54</b>	<b>74.77%</b>	<b>37.08%</b>	



<b>境外收入总额</b>		<b>11,782.07</b>	<b>100.00%</b>	<b>49.60%</b>	
<b>营业收入总额</b>		<b>23,755.81</b>		<b>100.00%</b>	
<b>2018年</b>					
序号	客户品牌	营业收入	占境外收入 比重	占营业收入 比重	销售内容
1	三星	1,833.67	30.70%	9.52%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品
2	Google	804.89	13.48%	4.18%	智能语音-训练数据产品
3	亚马逊	465.00	7.79%	2.41%	智能语音-训练数据产品、智能语音-训练数据定制服务
4	纽康 <sup>14</sup>	371.15	6.21%	1.93%	智能语音-训练数据定制服务、训练数据相关的应用服务
5	微软	341.98	5.73%	1.78%	智能语音-训练数据产品、自然语言-训练数据产品
<b>合计</b>		<b>3,816.69</b>	<b>63.91%</b>	<b>19.81%</b>	
<b>境外收入总额</b>		<b>5,972.00</b>	<b>100.00%</b>	<b>31.00%</b>	
<b>营业收入总额</b>		<b>19,265.77</b>		<b>100.00%</b>	

<sup>14</sup> 纽康：即纽康品牌旗下企业，报告期内与发行人合作的签约主体包括日本纽康株式会社、海南纽康信息系统有限公司。

## 6、境外销售具体情况

### (1) 境外销售主要国家地区、产品服务种类、各国家地区主要客户

金额单位：万元

序号	国家地区	产品服务类别	2020年		2019年		2018年		主要客户
			营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	
1	日韩	总收入	2,993.85	12.83%	5,688.34	23.95%	2,531.45	13.14%	三星（韩国）、纽康、Naver、客户2、客户1
		其中：训练数据定制服务	316.57	1.36%	3,973.37	16.73%	2,473.59	12.84%	
		训练数据产品	2,677.28	11.47%	1,714.97	7.22%	48.96	0.25%	
		训练数据相关的应用服务	-	-	-	-	8.90	0.05%	
2	美国	总收入	4,651.52	19.93%	4,628.51	19.48%	2,496.28	12.96%	微软、亚马逊、某国际消费电子产品厂商、Google
		其中：训练数据定制服务	1,088.86	4.67%	1,521.43	6.40%	342.36	1.78%	
		训练数据产品	3,552.97	15.22%	3,107.08	13.08%	2,153.92	11.18%	
		训练数据相关的应用服务	9.69	0.04%	-	-	-	0.00%	
3	其他	总收入	723.19	3.10%	1,465.21	6.17%	944.27	4.90%	Verint <sup>15</sup> 、三星（印度）、Nuance <sup>16</sup>
		其中：训练数据定制服务	205.34	0.88%	742.31	3.12%	624.16	3.24%	
		训练数据产品	517.85	2.22%	713.23	3.00%	319.64	1.66%	

<sup>15</sup> Verint：即 Verint 品牌，报告期内发行人与该品牌合作的签约主体包括 Verint Systems Inc.（已于纳斯达克主板上市，股票代码 VRNT）与 Verint CES Ltd。

<sup>16</sup> Nuance：即 Nuance 品牌，报告期内发行人与该品牌合作的签约主体包括 Nuance Communications International BVBA、Nuance Communications Ireland Limited、Nuance Communications Services Ireland Limited 及纽昂司软件技术（北京）有限公司，前述主体与纳斯达克上市公司 Nuance Communications Inc.（NUAN）隶属同一控制。

序号	国家地区	产品服务类别	2020年		2019年		2018年		主要客户
			营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	
		训练数据相关的应用服务	-	-	9.67	0.04%	0.47	0.00%	
境外收入合计			8,368.56	35.86%	11,782.07	49.60%	5,972.00	31.00%	
营业收入总额			23,337.40	100.00%	23,755.81	100.00%	19,265.77	100.00%	

注：训练数据相关的应用服务各年境外收入占比较低，下文销量、单价情况主要列示训练数据定制服务、训练数据产品相关情况。

## (2) 境外销售主要国家地区的销售量、销售单价情况

### ①训练数据定制服务

类别	收入（万元）				销量					单价				
	2020年	2019年	2018年	2017年	销量单位	2020年	2019年	2018年	2017年	单价单位	2020年	2019年	2018年	2017年
智能语音	587.01	5,032.28	3,002.51	915.06	小时	4,676	11,302	17,219	6,884	元/小时	1,255.50	4,452.74	1,743.67	1,329.26
计算机视觉	812.14	807.02	266.86	531.45	个	19	6	15	8	元/个	427,443.85	1,345,028.75	177,907.56	664,313.80
自然语言	211.61	397.81	170.74	147.93	个	61	55	63	37	元/个	34,690.39	72,329.31	27,102.24	39,979.88

### ②训练数据产品

类别	收入（万元）				销量					单价				
	2020年	2019年	2018年	2017年	销量单位	2020年	2019年	2018年	2017年	单价单位	2020年	2019年	2018年	2017年
智能语音	6,310.80	5,292.45	2,477.03	1,705.12	小时	65,427	57,366	46,493	15,991	元/小时	964.56	922.58	532.77	1,066.28

类别	收入（万元）				销量				单价					
	2020年	2019年	2018年	2017年	销量单位	2020年	2019年	2018年	2017年	单价单位	2020年	2019年	2018年	2017年
计算机视觉	34.88	-	2.28	-	个	2	-	1	-	元/个	174,415.44	-	22,822.21	-
自然语言	402.43	242.84	43.20	92.89	个	25	16	5	9	元/个	160,970.17	151,775.42	86,398.93	103,210.58

### ③说明

在境外销售主要国家/地区，发行人同类业务产品/服务单价基本稳定，但受客户议价情况、具体项目需求、项目复杂程度，以及涉及的语种/方言等多方面因素影响，也存在一定波动。各年计算机视觉、自然语言按项目（销售的训练数据集）个数计算单价，因境外客户上述类别项目规模通常较大，促使销售单价均较高。

## 7、涉及公开招投标程序的情况说明

### (1) 应履行公开招投标程序的情形说明

根据《中华人民共和国政府采购法》及其实施条例，政府采购是指各级国家机关、事业单位和团体组织，使用财政性资金采购依法制定的集中采购目录以内的或者采购限额标准以上的货物、工程和服务的行为，政府采购方式包括公开招标、邀请招标、竞争性谈判、单一来源采购、询价等；政府采购限额标准，属于中央预算的政府采购项目，由国务院确定并公布；属于地方预算的政府采购项目，由省、自治区、直辖市人民政府或者其授权的机构确定并公布。

报告期内，发行人客户主要为人工智能产业链上的各类知名机构，如阿里巴巴、腾讯、百度、微软、三星、亚马逊等提供服务，仅少量业务涉及根据上述规定应执行政府采购制度的国家机关、事业单位等客户（包含中央预算单位、地方预算单位），该等情形极少且销售额较小，具体参见下述（2）论述。

### (2) 发行人不存在应履行公开招投标程序而未履行的情形

#### ①向中央预算单位客户销售的情形

根据《国务院办公厅关于印发中央预算单位 2017-2018 年政府集中采购目录及标准的通知》（国办发〔2016〕96 号）、《财政部关于中央预算单位政府集中采购目录及标准有关问题的通知》（财库〔2019〕2 号）、《国务院办公厅关于印发中央预算单位政府集中采购目录及标准（2020 年版）的通知》（国办发〔2019〕55 号），报告期内，中央预算单位进行政府采购，单项采购金额达到 200 万元以上的，必须采用公开招标方式。

报告期内，发行人涉及向科研院所等中央预算单位销售的销售合同共 83 份，合同金额合计 1,167.15 万元，单份合同金额最高为 140 万元，未达到依据上述规定需采用公开招标方式的 200 万元单项采购金额标准，不存在应履行公开招投标程序而未履行的情形。

#### ②向地方预算单位客户销售的情形

报告期内，发行人涉及向地方预算单位销售的销售合同共 2 份，合同金额合计 20.56 万元，单份合同金额最高为 20 万元。经查询相应各地招投标相关规定，

上述合同金额均未达到执行政府采购制度的限额标准，不存在应履行公开招标程序而未履行的情形。

综上，报告期内发行人不存在应履行公开招标程序而未履行的情形。

#### 四、采购情况和主要供应商

##### （一）报告期内采购情况

##### 1、报告期内的分类采购明细数据

报告期内，发行人生产经营直接相关的采购支出主要包含：数据服务费，即训练数据生产过程所需的、非核心技术环节的原料数据采集、标注服务的采购支出；另有少量办公房租支出，以及设备及运输工具等固定资产及办公软件等无形资产的采购支出；项目租赁费，即训练数据生产过程中所需的场地、设备租赁费用；少量岗位服务费用支出。

报告期内，发行人分类采购明细具体如下：

单位：万元

采购细分类别	2020年		2019年		2018年	
	金额	占比	金额	占比	金额	占比
数据服务费	7,644.40	84.69%	6,099.95	79.84%	7,352.66	81.22%
办公房租	776.49	8.60%	694.66	9.09%	545.17	6.02%
资产采购	164.94	1.83%	287.12	3.76%	552.90	6.11%
项目租赁费	128.30	1.42%	274.06	3.59%	345.08	3.81%
岗位服务费	311.71	3.45%	284.89	3.73%	256.62	2.83%
<b>采购总额</b>	<b>9,025.84</b>	<b>100.00%</b>	<b>7,640.69</b>	<b>100.00%</b>	<b>9,052.42</b>	<b>100.00%</b>

2018年、2019年及2020年，发行人采购总额分别为9,052.42万元、7,640.69万元及9,025.84万元，采购额呈现先下降后上升的趋势。2019年采购总额相比2018年有所下降，主要是数据服务采购支出下降所致，详见下文具体分析。

##### 2、数据服务采购情况

2018年、2019年及2020年，发行人采购原料数据采集、标注服务所支付的数据服务费占采购总额的比重分别达到81.22%、79.84%及84.69%，各年占比基本稳定。

根据业务特点，发行人向数据服务提供商的采购集中分布在发行人业务的三大领域（智能语音、自然语言和计算机视觉）中的两大业务流程（采集、标注）中，其中原料数据采集环节主要以按字/词、人、工作小时为单位进行结算，标注环节主要以工作小时为单位进行结算，以下数据中金额单位为：万元，单价为：元/（字/词、人、工作小时）。

报告期内数据服务采购情况具体如下：

（1）2020 年

领域	环节	结算单位	2020 年采购情况		
			服务数量	采购单价 (元)	采购金额 (万元)
智能语音	采集	字/词	30,578,865	0.44	1,337.92
		人	49,654	216.25	1,073.76
	标注	工作小时	1,590,448	20.85	3,315.33
自然语言	采集	工作小时	483	73.38	3.54
	标注	工作小时	156,240	26.13	408.28
计算机视觉	采集	工作小时	88,426	50.48	446.40
	标注	工作小时	282,639	18.67	527.60
<b>合计</b>					<b>7,112.82</b>
<b>其他<sup>(注)</sup></b>					<b>531.58</b>
<b>数据服务费合计</b>					<b>7,644.40</b>

（2）2019 年

领域	环节	结算单位	2019 年采购情况		
			采购数量	采购单价 (元)	采购金额 (万元)
智能语音	采集	字/词	15,591,810	0.46	717.58
		人	36,831	263.90	971.98
	标注	工作小时	1,211,907	22.40	2,714.89
自然语言	采集	工作小时	3,045	234.67	71.45
	标注	工作小时	138,624	23.56	326.61
计算机视觉	采集	工作小时	29,771	76.09	226.53
	标注	工作小时	481,344	16.43	790.91
<b>合计</b>					<b>5,819.94</b>

领域	环节	结算单位	2019 年采购情况		
			采购数量	采购单价 (元)	采购金额 (万元)
其他 <sup>(注)</sup>					280.01
数据服务费合计					6,099.95

## (3) 2018 年

领域	环节	结算单位	2018 年采购情况		
			采购数量	采购单价 (元)	采购金额 (万元)
智能语音	采集	字/词	19,213,402	0.52	1,003.14
		人	53,311	202.28	1,078.38
	标注	工作小时	1,570,977	20.47	3,216.41
自然语言	采集	工作小时	5,529	104.39	57.71
	标注	工作小时	170,559	20.38	347.52
计算机视觉	采集	工作小时	96,872	37.69	365.08
	标注	工作小时	397,722	16.84	669.61
合计					6,737.85
其他 <sup>(注)</sup>					614.81
数据服务费合计					7,352.66

注：报告期内，发行人采购的数据服务中，智能语音领域支出较大，三年合计金额占总数据服务费比例超过 70%，采购情况（服务数量、采购单价）已完整列示；自然语言和计算机视觉两大领域采购支出相对较少，但因结算单位种类较多，故选择列示主要结算单位的采购明细（服务数量、采购单价），其余采购情况在“其他”中列示。已披露服务数量、采购单价的数据服务采购情况合计占当年数据服务费采购金额比例均超过 90%。

2018 年-2020 年，发行人数据服务采购规模整体呈现先上升后下降的趋势，其中智能语音领域支出较大，报告期合计占比超过 70%，与智能语音领域收入占发行人收入总体比例相当；自然语言和计算机视觉两大领域采购支出相对较少，单个项目需求差异化特征较强，采购量和采购单价变动较大。

对智能语音领域数据服务采购情况进一步分析如下：

## (1) 采集服务

## ① 采购数量方面：

2019 年采集数量相比 2018 年有所下降，主要原因是：按字/词为结算单位列



示的主要是语音合成相关的采集服务采购情况，按人为单位列示的主要是语音识别相关的采集服务采购情况。在语音合成领域，项目生产周期较长，部分大额项目的外语种原料数据采集工作主要在 2018 年完成，因而 2018 年语音合成相关的采集数量增长、2019 年相应下降。在语音识别领域，中文语音识别训练数据市场竞争加剧，使得 2019 年当年发行人获得的中文语音识别训练数据领域在手项目减少，相应采购的中文原料数据采集服务也随之下降。

2020 年，语音识别、语音合成领域的采集服务的采购量均有所增长，其中语音合成领域采集服务的采购量相比 2019 年有大幅增长，主要原因是：在语音识别领域，发行人调整价格策略，对中文业务采取较低报价应对市场竞争的加剧，促使当期语音识别采集业务量有所增长，采集服务采购量的增长与该业务增长趋势相匹配。而在语音合成领域，当期新增执行了多个大规模素人采集、有声读物项目，前述项目对原料数据的采集需求较大，提升了当期采集服务的采购规模；同时为了丰富自有训练数据产品及储备发音人资源，发行人在当年加大了平均音色类训练数据产品开发，也提升了当期采集服务的采购规模。

②采购价格方面：报告期内采集单价的变动主要是高单价的外语种项目的占比、数量变动的结果。

对于以字/词为结算单位列示的语音合成领域原料数据采集服务：报告期内单价整体较为稳定，逐年略有变动主要是因为各期执行项目对发音人和领域多样化的要求有所不同造成的。

对于以人为结算单位列示的语音识别领域的原料数据采集服务：2018-2020 年单价整体较为平稳。2019 年单价相比 2018 年略有增长，主要原因是发行人业务逐步向采集单价较高的外语种、自由对话类型训练数据拓展，推高了当年采集的平均单价。2020 年采集单价相比 2019 年有所下降，一方面是由于当期完成了部分较为简单的中文项目，例如中文唤醒词项目等，使得采集单价较低；另一方面是因为当期执行的较多英语采集项目引入了中国发音人录制，成本较英语母语地区人员相比较低。

## （2）标注服务

### ①采购数量方面：

2019 年标注服务采购量相比 2018 年有所下降，主要原因是：部分客户需求受其研发计划进展节奏影响有所波动，使得 2019 年当年发行人执行的中文语音识别转写项目减少，当年相应采购的中文原料数据标注服务数量也随之下降；此外，发行人在数据标注过程应用了更多自主开发的辅助标注工具，例如增加语音识别算法在标注及质检环节的应用等，提升了标注过程的效率和准确性、减少了需要人工校对的数据量，也使得当年采购的标注服务数量有所下降。

2020 年标注服务采购量相比 2019 年有所回升，与 2018 年采购量基本相当，主要是智能语音内部细分业务类别采购量变动的整体结果。①语音识别采集：2020 年语音识别采集业务采购的标注服务数量与 2019 年基本相当，主要是因为部分当期执行项目（如唤醒词项目等）应客户需求所需进行的标注工作较少甚至无需标注，同时部分项目涉及手机单通道录音设备，在标注环节无需在通道间进行数据对齐，标注工作量相对较少，造成标注服务量相较采集量的增长变动幅度较小。②语音合成：语音合成类业务采购的标注服务数量则有较大幅度增加，主要原因是：部分主要客户在语音合成业务领域加大投入，新增的歌曲类、平均音色类项目对标注服务也有较大需求，使得整体标注服务采购量有所上升。③语音识别转写：语音识别转写业务的标注服务采购量相比 2019 年有大幅增长，主要是因为：随市场竞争加剧，发行人为开拓市场、维持客户关系，调整了该类业务的报价策略，该类业务体量规模有所增长，相应需采购的标注服务数量也随之增长。

②采购单价方面：报告期内标注服务单价较为稳定，受各年中文、外语种标注单价影响略有变动。

### 3、发行人向自然人、服务公司采购数据服务情况

发行人向自然人和服务公司（包括服务公司采取小外包和直接对接两种形式）的采购内容均为原料数据采集、标注服务，其中自然人为直接向发行人提供服务，服务公司为按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员，组织其使用发行人研发的技术、平台、工具等，为发行人提供合乎标准的数据采集、标注服务。

报告期内，发行人向自然人和服务公司（包括服务公司采取小外包和直接对接两种形式）的采购金额情况如下：

单位：万元

采购方式	2020 年度		2019 年度		2018 年度	
	采购金额	占比	采购金额	占比	采购金额	占比
通过服务公司采购	7,340.92	96.03%	5,912.58	96.93%	7,231.42	98.35%
其中：通过小外包采购	5,115.67	66.92%	4,281.03	70.18%	4,567.55	62.12%
直接对接采购	2,225.24	29.11%	1,631.55	26.75%	2,663.87	36.23%
直接向自然人采购	303.48	3.97%	187.37	3.07%	121.24	1.65%
<b>数据服务费采购额</b>	<b>7,644.40</b>	<b>100.00%</b>	<b>6,099.95</b>	<b>100.00%</b>	<b>7,352.66</b>	<b>100.00%</b>

## （二）主要供应商情况

### 1、主要供应商情况

报告期内发行人向前五名供应商（受同一控制的供应商合并计算采购额）采购情况如下表所示：

金额单位：万元

2020 年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	善世（广东）企业服务外包有限公司 <sup>17</sup>	否	4,403.16	48.78%	原料数据采集、标注服务
	共青城宜邦人力资源服务有限公司	否			原料数据标注服务
	才星（广州）科技服务有限公司	否			原料数据采集、标注服务
	芜湖善慧人力资源服务有限公司	否			原料数据采集、标注服务
2	北京博睿驰信人力资源有限公司	否	1,592.40	17.64%	原料数据采集、标注服务；岗位服务
	恩泽尔商贸有限公司	否			原料数据采集、标注服务
3	李宏	否	494.44	5.48%	租赁办公用房
4	北京冠华英才国际经济技术有限公司	否	252.39	2.80%	原料数据标注服务
5	Voices.com Inc	否	229.83	2.55%	原料数据采集服务

<sup>17</sup> 曾用名“广州善世人力资源服务有限公司”。

合计			6,972.23	77.25%	
<b>2019年</b>					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	芜湖善慧人力资源服务有限公司	否	3,153.41	41.27%	原料数据采集、标注服务
	善世（广东）企业服务有限公司 <sup>18</sup>	否			原料数据采集、标注服务
	共青城宜邦人力资源服务有限公司	否			原料数据标注服务
	才星（广州）科技服务有限公司 <sup>19</sup>	否			原料数据标注服务
2	北京博睿驰信人力资源服务有限公司	否	1,573.46	20.59%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否			原料数据采集、标注服务
3	李宏	否	428.89	5.61%	租赁办公用房
4	北京冠华英才国际经济技术有限公司	否	388.90	5.09%	原料数据采集、标注服务
5	赵峻	否	168.70	2.21%	租赁办公用房
合计			5,713.36	74.78%	
<b>2018年</b>					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	广州善世人力资源服务有限公司	否	3,324.46	36.72%	原料数据标注服务
	芜湖善慧人力资源服务有限公司	否			原料数据采集、标注服务
	广州金社宝网络技术有限公司	否			原料数据标注服务
2	北京博睿驰信人力资源服务有限公司	否	2,078.46	22.96%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否			原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	905.61	10.00%	原料数据采集、标注服务
4	李宏	否	410.29	4.53%	租赁办公用房
5	TWIN STUDIOS	否	229.97	2.54%	原料数据采集服务
合计			6,948.80	76.76%	

<sup>18</sup> 曾用名“广州善世人力资源服务有限公司”。

<sup>19</sup> 曾用名“广州金社宝网络技术有限公司”。

报告期内，发行人各年前五名供应商主要为人力资源服务公司等，发行人向其采购的内容是训练数据生产过程中的、非核心技术环节的原料数据采集、标注服务；此外发行人也有部分办公房租采购支出等。发行人向单一供应商的采购比例未超过 50%，不存在对单一供应商的严重依赖情形。发行人与各年前五大供应商均不存在关联关系。

## 2、主要数据服务供应商情况

### (1) 报告期内前五大数据服务供应商情况

报告期内发行人向前五名数据服务供应商（受同一控制的供应商合并计算采购额）采购情况如下表所示：

金额单位：万元

2020 年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	善世（广东）企业服务外包有限公司 <sup>20</sup>	否	1,918.58	21.26%	原料数据采集、标注服务
	共青城宜邦人力资源服务有限公司	否	1,518.82	16.83%	原料数据标注服务
	才星（广州）科技服务有限公司	否	965.59	10.70%	原料数据采集、标注服务
	芜湖善慧人力资源服务有限公司	否	0.16	0.002%	原料数据采集、标注服务
2	北京博睿驰信人力资源有限公司	否	1,321.62	14.64%	原料数据采集、标注服务；岗位服务
	恩泽尔商贸有限公司	否	270.78	3.00%	原料数据采集、标注服务
4	北京冠华英才国际经济技术有限公司	否	252.39	2.80%	原料数据标注服务
3	Voices.com Inc.	否	229.83	2.55%	原料数据采集服务
5	IVO EMEA Ltd.	否	88.16	0.98%	原料数据采集服务
合计			<b>6,565.95</b>	<b>72.75%</b>	
2019 年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容

<sup>20</sup> 曾用名“广州善世人力资源服务有限公司”。

1	芜湖善慧人力资源服务有限公司	否	1,512.55	19.80%	原料数据采集、标注服务
	善世（广东）企业服务外包有限公司 <sup>21</sup>	否	1,374.11	17.98%	原料数据采集、标注服务
	共青城宜邦人力资源服务有限公司	否	261.18	3.42%	原料数据标注服务
	才星（广州）科技服务有限公司 <sup>22</sup>	否	5.57	0.07%	原料数据标注服务
2	北京博睿驰信人力资源有限公司	否	1,432.77	18.75%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否	140.69	1.84%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	388.90	5.09%	原料数据采集、标注服务
4	Atrium Studio Medienproduktion GmbH	否	138.71	1.82%	原料数据采集服务
5	SAKURA JAPAN LLC	否	114.55	1.50%	原料数据采集服务
合计			<b>5,369.03</b>	<b>70.27%</b>	
<b>2018年</b>					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	广州善世人力资源服务有限公司	否	1,755.67	19.39%	原料数据标注服务
	芜湖善慧人力资源服务有限公司	否	1,541.91	17.03%	原料数据采集、标注服务
	广州金社宝网络技术有限公司	否	26.89	0.30%	原料数据标注服务
2	北京博睿驰信人力资源有限公司	否	1,829.37	20.21%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否	249.09	2.75%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	905.61	10.00%	原料数据采集、标注服务
4	TWIN STUDIOS	否	229.97	2.54%	原料数据采集服务
5	THINKWILD STUDIOS, S.L.	否	129.85	1.43%	原料数据采集服务
合计			<b>6,668.36</b>	<b>73.66%</b>	

注：（1）善世（广东）企业服务外包有限公司、芜湖善慧人力资源服务有限公司、才星（广州）科技服务有限公司、共青城宜邦人力资源服务有限公司因共同关联股东持股及共同人员管理而被认定为同一控制主体；（2）北京博睿驰信人力资源有限公司、恩泽尔商贸有限公司均受同一股东控制，且该股东拥有北京零点起航营销策划有限公司的委托经营权，上述三个主体被认定为同一控制主体。

<sup>21</sup> 曾用名“广州善世人力资源服务有限公司”。

<sup>22</sup> 曾用名“广州金社宝网络技术有限公司”。

## (2) 主要数据服务提供商业资质情况说明

由于相关法律法规并未就原料数据的采集、标注服务规定专门的业务资质许可，主要原料数据采集、标注服务提供商向发行人提供原料数据采集、标注服务不需要取得专门的业务资质。根据中国境内主要原料数据采集、标注服务提供商的营业执照，其为发行人提供数据采集、标注服务未超出其《营业执照》上核准的经营范围。主要原料数据采集、标注服务提供商与发行人合作，具备提供相关服务的能力，能够按照与发行人的协议约定提供原料数据采集、标注服务。

## 3、主要数据服务供应商服务能力、服务提供情况说明

## (1) 主要数据服务供应商具有足够能力为发行人提供相关服务

## ①主要原料数据采集、标注服务提供商的规模、员工数量情况

序号	供应商名称	成立日期	注册资本	股权结构	注册地	实际经营地	业务经营情况	合作期间员工数量(人)	合作期间营业收入(万元/年)
1	善世(广东)企业服务外包有限公司	2014年2月	2,100万元	陈玦霏持股 40%	广州市白云区齐富路自编5号718房		人力资源外包、劳务承揽、劳务派遣服务等	90-150	13,000-28,000
				白云持股 40%					
				高洁持股 20%					
	共青城宜邦人力资源服务有限公司	2019年3月	200万元	夏建平持股 70%	江西省九江市共青城市科技一大道66号3楼308室				
				陈礼秋持股 30%					
	芜湖善慧人力资源服务有限公司	2016年11月	200万元	陈玦霏持股 32%	安徽省芜湖市镜湖区世纪城绿地新都会办公座15层077号	安徽省芜湖市镜湖区世纪城绿地新都会D座16层			
				莫天勤持股 32%					
				徐洪超持股 30%					
				欧阳刚持股 3%					
	才星(广州)科技服务有限公司	2016年5月	680万元	王宏志持股 40%	广州市海珠区阅江中路832号3605、3606房				
莫天勤持股 30%									
陈玦霏持股 30%									
2	北京博睿驰信人力资源有限公司	2013年5月	500万元	王驰持股 100%	北京市大兴区滨河街27号14层1402	劳务服务、劳务派遣、企业	45-65	2,000-5,200	
	恩泽尔商贸有限公司	2016年9月	5万美元	王驰持股 100%	马绍尔群岛	北京市大			

序号	供应商名称	成立日期	注册资本	股权结构	注册地	实际经营地	业务经营情况	合作期间员工数量(人)	合作期间营业收入(万元/年)
	北京零点起航营销策划有限公司	2013年12月	200万元	张立文持股 50% 刘凡茂持股 50%	北京市大兴区滨河街27号14层1402	兴区滨河街27号14层1402	管理咨询等		
3	北京冠华英才国际经济技术有限公司	1999年11月	11,000万元	齐宝春持股 50% 孙进荣持股 50%	北京市朝阳区北三环贞东古冠华商会馆2层	北京市朝阳区北三环贞东古冠华商会馆2层	专业承包、劳务分包、劳务派遣等	约 100	19,000-26,000
4	信阳市联智网络科技有限公司	2016年1月	1,000万元	洪琪持股 100%	信阳市羊山新区新八街信阳电子商务产业园大厦二层205		网络技术开发与服务、计算机领域的技术服务等	约 100	180-470
5	SAKURA JAPAN LLC	2016年12月	50万日元	高見一廣持股 60% 劉曉傑持股 40%	栃木県小山市横倉新田95-208		人力资源服务、技术服务等	约 50	600-1,000
6	THINKWILD STUDIOS, S.L.	2009年5月	3,010欧元	Rossana Giacomelli 和 Carlos Gomez-Mira 共同持股	C/ Castillo de Fuensaldaña Nº 4 Las Rozas, 28232 Madrid Spain		录音制作等	--	--
7	TWIN STUDIOS	1996年1月	48,000欧元	--	2 AVENUE DU MARECHAL MAUNOURY 75016 Paris 16th Arrondissement		录音制作等	--	--
8	Atrium Studio Medienproduktion GmbH	1998年	--	--	Zielstattstr. 33 81379 München		录音制作等	--	--
9	Voices.com Inc	2003年12月	--	David Ciccarella 和 Stephanie	100 Dundas St., Suite 700 London, ON		录音制作	约 100	--



序号	供应商名称	成立日期	注册资本	股权结构	注册地	实际经营地	业务经营情况	合作期间员工数量(人)	合作期间营业收入(万元/年)
				Cicarelli 共同持股	N6A 5B6 Canada		等		
10	IVO EMEA LTD.	1998 年	-	-	32 Merrion Street Upper, D02 KW80, Ireland		录音制作	15-20	100 万美元 -500 万美元 之间

备注：(1) 北京博睿驰信人力资源有限公司、恩泽尔商贸有限公司、北京零点起航营销策划有限公司的合并员工数量及营业收入计算包含与其受同一控制的北京天创共盈科技发展有限公司、北京中惠商客科技有限公司；(2) 信阳市联智网络科技有限公司的员工数量及营业收入计算包含与其受同一控制的信阳市晨曦电子科技有限公司；(3) 上述部分境外供应商未提供相关资料信息。

### ② 发行人对原料数据采集、标注服务提供商的主要需求

发行人从事数据采集、标注所需劳务人员数量较大、临时性较强、对人员响应速度要求较快，且该类工作的执行为非核心技术环节，无需较高技术门槛，而服务提供商专业从事大量人员招募和管理的工作，能够满足发行人相关服务需求，因此发行人主要通过向服务提供商采购的方式完成数据采集、标注的业务环节。

发行人对服务提供商的主要需求如下：

A. 接收并传达采集、标注需求：发行人与服务提供商沟通确定数据采集、标注的数量、具体要求、时间进度等，并对服务提供商进行培训，服务提供商接受发行人的服务需求，并将服务要求传达给所联系的劳务人员（包括可以直接提供服务的终端人或有相应能力的小外包人员）；

B. 人员寻找和管理：服务提供商按发行人指定要求寻找上述劳务人员，并对该等劳务人员进行管理，与其签署劳务协议，按服务量向其支付劳务报酬，并为其代扣代缴个税等；

C. 工作进度与数据交付管理：服务提供商或小外包人员对终端人进行数据采集、标注的进度等进行监督管理；服务提供商持续与发行人进行沟通，获取发行人对数据服务质量的反馈情况，及时反馈给小外包、直接对接的终端劳务人员，保证其提供服务的质量。

### ③ 主要原料数据采集、标注服务提供商有足够能力为发行人提供服务

#### A. 主要原料数据采集、标注服务提供商具有相关人员资源

数据采集、标注服务不涉及专门的业务资质。上述表格所列 1-5 项主要原料

数据采集、标注服务提供商具体从事对外提供劳务/技术服务的业务经营，积累了相关人员资源；上述表格所列 6-10 项主要原料数据采集、标注服务提供商为专业录音机构，具有专业发音人资源。服务提供商均能够及时响应发行人对数据采集、标注人员的需求。

#### B. 主要原料数据采集、标注服务提供商具有人员管理等业务经验

上述表格所列 1-5 项主要原料数据采集、标注服务提供商具有人员管理、项目管理或技术服务的业务经验，经发行人培训后能够掌握数据采集、标注的规范和流程，能够按发行人要求对数据采集、标注的进度、质量进行管控。上述表格所列 6-10 项主要原料数据采集、标注服务提供商为专业录音机构，具有专业的录音场地和人员及项目管理经验，能够为发行人提供对录音指标、录音环境要求较高的数据采集服务。

上述主要原料数据采集、标注服务提供商在与发行人合作过程中，均按照协议约定提供数据采集、标注服务，未发生因其自身能力问题与发行人停止合作的情形。

综上，主要原料数据采集、标注服务提供商有足够能力为发行人提供相关服务。

#### (2) 相关服务提供商不存在以发行人名义招聘实习生或兼职人员的情形

根据相关服务提供商的确认及网络公开检索，相关服务提供商报告期内不存在以发行人名义招聘实习生或者兼职人员的情形。相关服务提供商通过其自身招聘渠道寻找劳务人员，并直接与劳务人员签署劳务协议。根据实际业务需要，发行人存在少量以自身名义临时招聘实习生或兼职人员的情形。

### (三) 中介机构核查情况

保荐机构、申报会计师对发行人数据服务采购进行的核查情况如下：

序号	核查程序	核查目的及核查结论
1	与数据服务采购支出相关的内部控制测试	确认与数据服务采购支出相关的关键内部控制存在且执行有效。 经核查，报告期内与采购支出相关的关键内部控制存在且得到有效执行。
2	数据服务采购的平均单价分析	通过检查数据服务采购单价的合理性，确认数据服务采购支出的真实性与完整性。 经核查，报告期内的不同业务领域内全部原料数据采集与标

序号	核查程序		核查目的及核查结论	
			注的平均采购单价整体合理。	
3	数据服务采购的采集损耗率和标注产出比分析		通过检查数据服务采购量与销售量是否配比，确认数据服务采购支出的真实性与完整性。 经核查，抽取项目的采集损耗率和标注产出比整体合理，采集量、标注量与销售量相匹配，数据服务采购交易相关的采购数量真实、准确。	
4	核查主要数据服务供应商的采购支出的主要支持性依据		检查与确认数据服务采购支出的真实性与准确性。 经核查，抽查的交易金额与各项支持性依据的金额核对一致。	
5	采购支出在资产负债表日前后的截止性测试		检查与确认数据服务采购支出的完整性。 经核查，未发现在资产负债表日前后数据服务费采购支出存在重大截止性差异。	
6	信息系统相关测试		确认与数据服务采购支出相关的信息系统运行有效，并通过数据分析程序确认系统记录数据的真实性和准确性。 经核查，信息系统的一般性控制测试和应用控制测试的结果显示发行人信息系统总体有效，系统记录数据真实、准确。	
7	主要数据服务供应商的函证程序		确认服务公司采购支出的真实性、完整性和期末余额的准确性。 经核查，发行人与服务公司的采购交易真实、完整、准确。	
8	主要数据服务供应商的访谈走访程序		核实与服务公司采购交易的真实性，识别是否存在未披露的关联交易。 经核查，发行人与服务公司的采购交易真实、完整。	
9	对服务公司、小外包人员和终端人员执行的进一步的穿透核查程序	核查主要服务公司的银行流水	核实与服务公司、小外包人员和终端人员的采购交易的真实性，识别是否存在未披露的关联交易，识别是否存在发行人关联主体体外承担成本的情形	识别是否存在未披露的关联交易，是否存在关联主体体外承担成本的情形。 经核查，发行人不存在未披露的关联交易，不存在发行人关联主体体外承担成本费用的情形。
		核查主要服务公司对小外包和终端人员的款项支付情况		核实与小外包和终端人员的采购交易的真实性。 经核查，与小外包和终端人员的采购交易真实、准确。
		抽样对小外包人员进行电话访谈和深度访谈		核实与小外包采购交易的真实性。 经核查，与小外包的采购交易真实、准确。
		抽样对完成项目的终端劳务人员进行电话访谈		核实与终端人员的采购交易的真实性。 经核查，与终端人员的采购交易真实发生。
		抽样对终端人员提供劳务的签字文件进行核查		核实与终端人员的采购交易的真实性。 经核查，与终端人员的采购交易真实发生。

经核查，保荐机构、申报会计师认为：发行人与服务公司、小外包和终端人员之间的采购交易真实存在，发行人对数据服务费采购的核算真实、准确、完整。发行人不存在未披露的关联交易，不存在发行人关联主体为发行人体外承担成本费用的情形。

## 五、发行人的主要固定资产和无形资产

### （一）主要固定资产

截至 2020 年 12 月 31 日，发行人的主要固定资产情况如下：

单位：万元

项目	原值	累计折旧	账面价值
房屋及建筑物	2,102.20	99.85	2,002.34
电子设备	805.80	487.10	318.69
运输工具	158.48	134.45	24.03
办公家具	58.86	46.60	12.26
<b>合计</b>	<b>3,125.33</b>	<b>768.01</b>	<b>2,357.32</b>

报告期内，发行人的固定资产主要为自购的房屋及建筑物，日常业务经营使用的录音设备、电脑等电子设备及运输工具、办公家具等。

发行人的房屋及建筑物、租赁房产情况具体如下：

### 1、房屋及建筑物

序号	所有权人	不动产权证书编号	房屋坐落	使用用途	建筑面积
1	海天瑞声	京（2018）海不动产权第0066374号	海淀区知春路1号1号楼15层1501	办公	165.55 m <sup>2</sup>
2	海天瑞声	京（2018）海不动产权第0065368号	海淀区知春路1号1号楼15层1513	数据采集	178.22 m <sup>2</sup>

### 2、主要租赁房产

#### （1）主要租赁房产情况

截至本招股说明书签署日，发行人主要租赁房产的具体情况如下表所示：

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积（m <sup>2</sup> ）	是否已办理房屋租赁备案
1	发行人	李宏	北京市海淀区成府路28号4座8层801、02、03、06、07、08、09、11	2020年5月17日至2022年5月16日	1,255.63	是
2	发行人	赵峻	北京市海淀区成府路28号3-801、3-803、3-807、3-809	2021年5月16日至2022年5月15日	500.31	是
3	发行人	荣成市科学技术局	荣成经济开发区海湾南路86号科技企业孵化器内	—	约400	否
4	发行人	三河市城美房地产开发有限公司	维多利亚D座1801、1803	2019年12月25日至2021年12月24日	152.7	否

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积 (m <sup>2</sup> )	是否已办理房屋租赁备案
5	发行人	三河市城美房地产开发有限公司	维多利亚 D 座 1805	2019 年 6 月 25 日至 2022 年 6 月 24 日	78.54	否

截至本招股说明书签署日，发行人承租的 3 处房屋未办理租赁备案登记。上述未办理房屋租赁备案登记的租赁物业存在被所在地主管机构责令限期改正及如未能按要求改正而受到罚款的风险，但其潜在处罚的金额较小。截至本招股说明书签署日，发行人未曾收到其租赁房产所在地有关主管部门责令限期改正的通知，亦未受到房屋租赁管理部门的行政处罚。

截至本招股说明书签署日，除上述第 3 项租赁房产外，其余出租方均已办理房屋所有权证，双方签订的租赁合同合法有效。上述第 3 项租赁房产的出租方荣成市科学技术局未提供房屋所有权证，根据其出具的《说明》，该房产为荣成市人民政府孵化器场地，所有权人系荣成盛泉科技园发展有限公司，荣成市科学技术局具体负责孵化器运营场地并作为出租方与入孵企业签署租赁协议；发行人租赁上述房屋系用于荣成分公司日常办公，对租赁房屋并无特殊要求，且易于搬迁。

发行人控股股东、实际控制人贺琳对上述情形出具了说明和承诺：“关于北京海天瑞声科技股份有限公司（以下简称“公司”）和/或其子公司在公司首次公开发行股票并上市前租赁使用的房屋，如因租赁房屋的出租方未就出租房屋取得房屋所有权证、房屋租赁事宜未办理房屋租赁备案或存在其他不符合相关法律、法规的情形，公司和/或其子公司被要求搬迁、或被处以任何形式的处罚、或被要求承担任何形式的法律责任，本人将全额承担该部分搬迁费用、被处罚或被追索的支出及费用，且在承担后不向公司和/或其子公司追偿，保证公司和/或其子公司不会因此遭受任何损失。”

(2) 发行人向个人租赁房产的实际用途与其法定用途相符

发行人目前向个人租赁的主要房产的用途等情况如下：

序号	出租方	租赁房屋坐落	租赁期限	实际用途	不动产权证记载用途
1	李宏	北京市海淀区成府路 28 号 4 座 8 层 801、02、03、06、	2020 年 5 月 17 日至 2022 年 5 月 16 日	办公	办公

序号	出租方	租赁房屋坐落	租赁期限	实际用途	不动产权证记载用途
		07、08、09、11			
2	赵峻	北京市海淀区成府路 28 号 3-801、3-803、3-807、3-809	2021 年 5 月 16 日至 2022 年 5 月 15 日	办公	办公

上述出租方提供了租赁房屋的不动产权证，发行人实际使用用途与其产权证上记载房屋用途均为办公，因此，发行人向个人租赁房产的实际用途与其法定用途相符。

## （二）主要无形资产

### 1、专利

截至 2021 年 5 月 7 日，发行人及其子公司共拥有 26 项已授权专利，包括 24 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，均已取得国家知识产权局出具的《专利证书》。具体情况如下：

序号	专利名称	专利权人	专利类型	专利号	专利申请日	授权公告日	期限	取得方式
1	语音对齐方法及装置	发行人	发明专利	ZL 2018 1 0449585.3	2018 年 5 月 11 日	2020 年 6 月 23 日	自申请日起 20 年	原始取得
2	音素误标注的检测方法和装置	发行人	发明专利	ZL 2018 1 0827500.0	2018 年 7 月 25 日	2020 年 1 月 31 日	自申请日起 20 年	原始取得
3	语音检测方法和装置	发行人	发明专利	ZL 2018 1 0883930.4	2018 年 8 月 6 日	2020 年 6 月 23 日	自申请日起 20 年	原始取得
4	韵律标注方法、装置和设备	发行人	发明专利	ZL 2018 1 0988973.9	2018 年 8 月 28 日	2020 年 1 月 7 日	自申请日起 20 年	原始取得
5	中英混合语料的生成方法、装置、设备及存储介质	发行人	发明专利	ZL 2019 1 0022453.7	2019 年 1 月 10 日	2020 年 3 月 24 日	自申请日起 20 年	原始取得
6	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	发明专利	ZL 2019 1 0072150.6	2019 年 1 月 25 日	2020 年 1 月 14 日	自申请日起 20 年	原始取得
7	音频质量评估方法、装置、电子设备及存储介质	发行人	发明专利	ZL 2019 1 0073390.8	2019 年 1 月 25 日	2019 年 10 月 29 日	自申请日起 20 年	原始取得
8	视频中音频聚类的处理方法和装置	发行人	发明专利	ZL 2019 1 1289077.4	2019 年 12 月 16 日	2020 年 5 月 5 日	自申请日起 20 年	原始取得
9	语音采集设备	发行人	实用新型专利	ZL 2019 2 2276799.8	2019 年 12 月 18 日	2020 年 3 月 31 日	自申请日起 10 年	原始取得
10	个性语料获取方法及个性语料获取装置	发行人	发明专利	ZL 2019 1 1314319.0	2019 年 12 月 19 日	2020 年 5 月 5 日	自申请日起 20 年	原始取得
11	语音库的语音获取方法、装置、电子设备及存储介质	发行人	发明专利	ZL 2019 1 1314504.X	2019 年 12 月 19 日	2020 年 5 月 5 日	自申请日起 20 年	原始取得

序号	专利名称	专利权人	专利类型	专利号	专利申请日	授权公告日	期限	取得方式
12	短语语料获取方法及短语语料获取装置	发行人	发明专利	ZL 2019 1 1352915.8	2019年12月25日	2020年5月5日	自申请日起20年	原始取得
13	端到端的语音合成方法、装置及存储介质	发行人	发明专利	ZL 2020 1 0007974.8	2020年1月6日	2020年5月19日	自申请日起20年	原始取得
14	视频筛选方法、装置及存储介质	发行人	发明专利	ZL 2020 1 0096575.3	2020年2月18日	2020年6月26日	自申请日起20年	原始取得
15	图像标注方法、图像标注装置及计算机存储介质	发行人	发明专利	ZL 2020 1 0096586.1	2020年2月18日	2020年8月4日	自申请日起20年	原始取得
16	视频抽选方法、装置及存储介质	发行人	发明专利	ZL 2020 1 0100697.5	2020年2月19日	2020年6月2日	自申请日起20年	原始取得
17	跨语言非标准词识别方法及装置	发行人	发明专利	ZL 2020 1 0122519.2	2020年2月27日	2020年7月14日	自申请日起20年	原始取得
18	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	发明专利	ZL 2020 1 0164556.X	2020年3月11日	2020年6月19日	自申请日起20年	原始取得
19	语音转写方法、语音转写装置及计算机存储介质	发行人	发明专利	ZL 2020 1 0283135.9	2020年4月13日	2020年7月28日	自申请日起20年	原始取得
20	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	发明专利	ZL 2020 1 0309303.7	2020年4月20日	2020年7月28日	自申请日起20年	原始取得
21	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	发明专利	ZL 2020 1 0659644.7	2020年7月10日	2020年10月16日	自申请日起20年	原始取得
22	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	发明专利	ZL 2020 1 0659647.0	2020年7月10日	2020年11月10日	自申请日起20年	原始取得
23	点云数据标注方法、点云数据标注装置及存储介质	发行人	发明专利	ZL 2021 1 0018470.0	2021年1月7日	2021年4月20日	自申请日起20年	原始取得
24	一种基于三元组的语音识别准确率计算方法	发行人	发明专利	ZL 2021 1 0026253.6	2021年1月8日	2021年4月20日	自申请日起20年	原始取得
25	基于麦克风的音频处理方法和装置	中瑞智	发明专利	ZL 2017 1 0539306.8	2017年7月4日	2020年3月31日	自申请日起20年	原始取得
26	智能语音采集设备	中瑞智	外观设计专利	ZL 2020 3 0062563.X	2020年2月27日	2020年8月18日	自申请日起10年	原始取得

除上述已授权专利外，发行人另已提交 2 项专利申请。

截至 2021 年 5 月 7 日，上述已授权专利中的 9 项发明专利和 1 项实用新型专利被提出宣告无效请求。发行人已经收到上述专利中 7 项发明专利和 1 项实用新型专利的无效宣告请求审查决定书，其中 7 项发明专利维持专利权有效，1 项实用新型专利专利权维持部分有效，其余 2 项专利的无效宣告请求审查事项尚处于正常审理阶段，国家知识产权局专利局尚未作出裁定。涉及无效宣告请求的专利具体情况如下：

序号	专利名称	所有人	专利类型	专利号	专利申请日	授权公告日	发行人收到《无效宣告请求受理通知书》的时间	发行人收到《无效宣告请求审查决定书》的时间	《无效宣告请求审查决定书》的结果
1	短语语料获取方法及短语语料获取装置	发行人	发明	ZL 2019 1352915.8	2019年12月25日	2020年5月5日	2020年10月9日	2021年2月3日	维持有效
2	语音采集设备	发行人	实用新型	ZL 2019 2276799.8	2019年12月18日	2020年3月31日	2020年10月9日	2021年2月24日	维持部分有效
3	端到端的语音合成方法、装置及存储介质	发行人	发明	ZL 2020 0007974.8	2020年1月6日	2020年5月19日	2020年10月9日	2021年2月9日	维持有效
4	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	ZL 2019 0022453.7	2019年1月10日	2020年3月24日	2020年10月27日	2021年2月20日	维持有效
5	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	ZL 2019 1314504.X	2019年12月19日	2020年5月5日	2020年10月27日	2021年3月8日	维持有效
6	个性语料获取方法及个性语料获取装置	发行人	发明	ZL 2019 1314319.0	2019年12月19日	2020年5月5日	2020年10月27日	-	-
7	视频中音频聚类的处理方法和装置	发行人	发明	ZL 2019 1289077.4	2019年12月16日	2020年5月5日	2020年10月27日	2021年2月25日	维持有效
8	音素误标注的检测方法和装置	发行人	发明	ZL 2018 0827500.0	2018年7月25日	2020年1月31日	2020年10月27日	2021年2月20日	维持有效
9	韵律标注方法、装置和设备	发行人	发明	ZL 2018 0988973.9	2018年8月28日	2020年1月7日	2020年10月29日	2021年3月1日	维持有效
10	基于麦克风的音频处理方法和装置	中瑞智	发明	ZL 2017 0539306.8	2017年7月4日	2020年3月31日	2020年10月27日	-	-

## 2、计算机软件著作权

截至2021年5月7日,发行人及其子公司共拥有134项计算机软件著作权,均已取得著作权证书,具体情况如下:



序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
1	发行人	2009SRBJ2960	语音自动切分系统 V1.0	2009年3月31日	2009年5月12日	原始取得	全部权利
2	发行人	2009SRBJ3067	外文词典信息提取软件 V1.0	2009年4月13日	2009年5月13日	原始取得	全部权利
3	发行人	2009SRBJ3063	递增式中文语料库抽选系统 V1.0	2009年4月13日	2009年5月13日	原始取得	全部权利
4	发行人	2009SRBJ3231	音频文件及文本校对软件 V1.0	2009年4月15日	2009年5月31日	原始取得	全部权利
5	发行人	2009SRBJ3239	多通道录音软件 V1.0	2009年4月15日	2009年5月31日	原始取得	全部权利
6	发行人	2009SRBJ3344	文本声音处理软件 V1.0	2009年4月17日	2009年6月1日	原始取得	全部权利
7	发行人	2010SRBJ2391	KDCRecorderWinMo6.1手机录音系统 V1.0	2010年4月20日	2010年5月28日	原始取得	全部权利
8	发行人	2010SRBJ2361	工时录入系统 V1.0	2010年4月20日	2010年5月28日	原始取得	全部权利
9	发行人	2010SRBJ2360	智能语音自动调幅软件 V1.0	2010年4月20日	2010年5月28日	原始取得	全部权利
10	发行人	2011SR095548	ASIORecord 多通道录音软件 V3.4	2011年11月1日	2011年12月15日	原始取得	全部权利
11	发行人	2011SR095630	Android 智能平台录音软件 V1.0	2011年11月1日	2011年12月15日	原始取得	全部权利
12	发行人	2011SR096789	Symbian 平台语音采集软件 V1.0	2011年11月1日	2011年12月17日	原始取得	全部权利
13	发行人	2011SRBJ4898	yTrans 语音转写标注软件 V1.0	2011年11月1日	2011年12月28日	原始取得	全部权利
14	发行人	2012SR115506	合成语音评测软件 V1.0	2012年9月11日	2012年11月28日	原始取得	全部权利
15	发行人	2012SR115664	多通道同步视频采集器软件 V2.0	2012年9月11日	2012年11月28日	原始取得	全部权利
16	发行人	2012SR115704	ASIORecord 多通道录音软件 V3.6	2012年9月11日	2012年11月28日	原始取得	全部权利
17	发行人	2012SR115707	IVR 录音平台软件 V1.0	2012年9月11日	2012年11月28日	原始取得	全部权利
18	发行人	2012SR116312	语料管理平台软件 V1.0	2012年9月11日	2012年11月29日	原始取得	全部权利
19	发行人	2012SR119967	语料提取软件 V1.0	2012年9月11日	2012年12月6日	原始取得	全部权利
20	发行人	2013SR048176	WindowsPhone 平台 AudioRecord 录音工具软件 V1.0	2013年3月12日	2013年5月22日	原始取得	全部权利
21	发行人	2013SR048355	视频搜索与定位 GFFGUI 软件 V1.0	2013年3月12日	2013年5月22日	原始取得	全部权利
22	发行人	2013SR048586	WindowsMobile 平台 KDCRecorder 录音工具软件 V1.0	2013年3月12日	2013年5月22日	原始取得	全部权利
23	发行人	2013SR048590	Windows8RT 平台 rtAudioRecord 录音工具软件 V1.0	2013年3月12日	2013年5月22日	原始取得	全部权利
24	发行人	2013SR048687	IOS 平台 iPAudio 录音软件 V1.0	2013年3月12日	2013年5月22日	原始取得	全部权利
25	发行人	2013SR051739	Android 智能平台录音软件 V2.0	2013年3月12日	2013年5月29日	原始取得	全部权利
26	发行人	2014SR068565	WindowsPhone 平台 AudioRecord 录音工具	2014年3月21日	2014年5月29日	原始取得	全部权利

序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
			软件 V2.0				
27	发行人	2014SR068567	语音识别（ASR）数据后期处理系统 V1.0	2014年3月12日	2014年5月29日	原始取得	全部权利
28	发行人	2014SR068789	BingImgSpider 快速下载软件 V1.0	2014年3月12日	2014年5月29日	原始取得	全部权利
29	发行人	2014SR068893	uTrans 在线语音转写标注质检平台系统 V1.0	2014年3月24日	2014年5月29日	原始取得	全部权利
30	发行人	2014SR068894	SpeechOcean 语料管理系统 V1.0	2014年3月21日	2014年5月29日	原始取得	全部权利
31	发行人	2014SR069088	SMSDataFilter 软件 V1.0	2014年3月21日	2014年5月29日	原始取得	全部权利
32	发行人	2014SR069160	Android 平台 AudioRec 录音软件 V2.0	2014年3月12日	2014年5月29日	原始取得	全部权利
33	发行人	2015SR107314	web 平台标注与切音软件 V1.1.2	2015年4月22日	2015年6月16日	原始取得	全部权利
34	发行人	2015SR108035	Android 平台众包录音 Sayit 软件 V1.0.0	2015年4月25日	2015年6月16日	原始取得	全部权利
35	发行人	2015SR108036	对话语音库软件 V1.0.0.0	2015年4月22日	2015年6月16日	原始取得	全部权利
36	发行人	2015SR108037	识别语音库软件 V1.0.0.0	2015年4月22日	2015年6月16日	原始取得	全部权利
37	发行人	2015SR129568	文本数据库软件 V1.0.0.0	2015年5月15日	2015年7月10日	原始取得	全部权利
38	发行人	2015SR130013	词典数据库软件 V1.0.0.0	2015年5月15日	2015年7月10日	原始取得	全部权利
39	发行人	2015SR130378	语音合成数据库软件 V1.0.0.0	2015年5月15日	2015年7月10日	原始取得	全部权利
40	发行人	2016SR221475	图形图像数据库软件 V1.0.0.0	2016年6月6日	2016年8月16日	原始取得	全部权利
41	发行人	2016SR232131	语音识别（ASR）数据后期处理软件 V2.3	2016年5月3日	2016年8月24日	原始取得	全部权利
42	发行人	2016SR232133	web 平台转写与校对软件 V1.0	2016年4月25日	2016年8月24日	原始取得	全部权利
43	发行人	2016SR261243	《天天有赚》众包软件 V1.0.0	2016年5月5日	2016年9月14日	原始取得	全部权利
44	发行人	2016SR369165	设备管理软件 V1.0	2016年10月5日	2016年12月13日	原始取得	全部权利
45	发行人	2016SR373382	在线文字属性标注软件 V1.0	2016年10月5日	2016年12月15日	原始取得	全部权利
46	发行人	2016SR373656	在线图片文字标注软件 V1.0	2016年10月5日	2016年12月15日	原始取得	全部权利
47	发行人	2016SR374245	人脸图片标注软件 V1.0	2016年10月7日	2016年12月15日	原始取得	全部权利
48	发行人	2016SR375819	汽车图片局部特征标注软件 V1.0	2016年10月7日	2016年12月15日	原始取得	全部权利
49	发行人	2017SR315167	在线语句标注软件 V1.0.1	2017年5月5日	2017年6月27日	原始取得	全部权利
50	发行人	2017SR315182	行车视频标注软件 V1.0.1	2017年5月1日	2017年6月27日	原始取得	全部权利
51	发行人	2017SR315190	人脸五官信息标注软件 V1.0.1	2017年5月3日	2017年6月27日	原始取得	全部权利
52	发行人	2017SR315196	图片中物体部分上色软件 V1.0.1	2017年5月5日	2017年6月27日	原始取得	全部权利

序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
53	发行人	2017SR315205	在线语句扩展软件 V1.0.1	2017年5月1日	2017年6月27日	原始取得	全部权利
54	发行人	2017SR494426	韩国语文本正则化软件 V1.5.0	2017年7月10日	2017年9月7日	原始取得	全部权利
55	发行人	2017SR499808	视频标注软件 V1.0.1	2017年7月22日	2017年9月8日	原始取得	全部权利
56	发行人	2017SR642743	视频数据库软件 V1.0.1	2017年11月6日	2017年11月22日	原始取得	全部权利
57	发行人	2017SR607198	天气图片标注软件 V1.0.1	2017年9月13日	2017年11月6日	原始取得	全部权利
58	发行人	2017SR609165	在线语句意图与属性标注软件 V1.0.1	2017年9月12日	2017年11月7日	原始取得	全部权利
59	发行人	2017SR609168	动漫标注软件 V1.0.1	2017年9月13日	2017年11月7日	原始取得	全部权利
60	发行人	2017SR609174	词典编辑软件 V1.0.1	2017年9月8日	2017年11月7日	原始取得	全部权利
61	发行人	2018SR280360	语音混音检查采样转换软件 V1.0.0	2017年11月6日	2018年4月25日	原始取得	全部权利
62	发行人	2018SR280367	分布式图片采集和分析软件 V1.0.0	2017年11月20日	2018年4月25日	原始取得	全部权利
63	发行人	2018SR280375	安卓图片采集软件 V1.0.3	2017年12月20日	2018年4月25日	原始取得	全部权利
64	发行人	2018SR280382	音频视频对齐分割软件 V1.0.0	2017年11月6日	2018年4月25日	原始取得	全部权利
65	发行人	2018SR383360	基于移动端手写体软件 V1.2.3	2017年11月6日	2018年5月25日	原始取得	全部权利
66	发行人	2018SR441160	多语言分布式文本采集和分析软件 V1.0.0	2017年11月20日	2018年6月12日	原始取得	全部权利
67	发行人	2018SR686811	桌面端 xml_json 通用软件 V1.0.0.0	2018年7月4日	2018年8月27日	原始取得	全部权利
68	发行人	2018SR686725	桌面端 OCR 标注软件 V1.1.1.2	2018年7月5日	2018年8月27日	原始取得	全部权利
69	发行人	2018SR766500	AudioBook 自动切分与文本对齐软件 V1.0.0	2018年7月14日	2018年9月20日	原始取得	全部权利
70	发行人	2018SR767437	Kaldi 多线程并行解码软件 V1.0.0	2018年7月19日	2018年9月20日	原始取得	全部权利
71	发行人	2018SR767419	多语言多种类语料设计软件 V1.0.0	2018年7月13日	2018年9月20日	原始取得	全部权利
72	发行人	2018SR766507	文本易读性打分软件 V1.0.0	2018年7月14日	2018年9月20日	原始取得	全部权利
73	发行人	2018SR751066	基于 ASR 的中文语音合成音字一致性校对软件 V1.0.0	2018年7月13日	2018年9月17日	原始取得	全部权利
74	发行人	2018SR746304	基于 hts 的语音合成软件 V1.0.0	2018年7月20日	2018年9月14日	原始取得	全部权利
75	发行人	2018SR751080	基于 Kaldi 的语音数据库质量评测样本选取软件 V1.0.0	2018年7月13日	2018年9月17日	原始取得	全部权利
76	发行人	2018SR748025	说话人自适应合成软件 V1.0.0	2018年7月25日	2018年9月14日	原始取得	全部权利
77	发行人	2019SR0351410	海天瑞声工作平台 V1.9.6.1	2019年3月6日	2019年4月19日	原始取得	全部权利
78	发行人	2019SR0515275	基于 PC 端视频标注软件 V1.0.1.3	2019年3月28日	2019年5月24日	原始取得	全部权利

序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
79	发行人	2019SR0515678	基于 PC 端 2D 标注软件 V1.1.1.8	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
80	发行人	2019SR0515233	在线文本推理标注软件 V1.0.1	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
81	发行人	2019SR0515243	海天瑞声问卷调查系统 V1.0.0.1	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
82	发行人	2019SR0515289	在线滑音输入任务分配软件 V1.0.1	2019 年 3 月 28 日	2019 年 5 月 24 日	原始取得	全部权利
83	发行人	2019SR0754830	文本语音质检平台 V1.0.0.1	2019 年 6 月 3 日	2019 年 7 月 22 日	原始取得	全部权利
84	发行人	2019SR0754102	基于 PC 端工具插件平台 V1.0.0.5	2019 年 6 月 3 日	2019 年 7 月 22 日	原始取得	全部权利
85	发行人	2019SR0980463	西班牙语音素自动切分软件 V1.0.0	2019 年 7 月 30 日	2019 年 9 月 23 日	原始取得	全部权利
86	发行人	2019SR0990577	基于 ios 多通道录音软件 V3.0.0.2	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
87	发行人	2019SR0990559	德语逆正则化软件 V1.0.0	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
88	发行人	2019SR0987596	语音数据库基频一致性评估软件 V1.0.0	2019 年 7 月 30 日	2019 年 9 月 24 日	原始取得	全部权利
89	发行人	2019SR0991560	中英数字复杂混合文本注音及韵律标注软件 V1.0.0	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
90	发行人	2019SR0991574	基于 window 的重音标注软件 V1.0.0.0	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
91	发行人	2019SR0994723	基于 Android 多通道录音软件 V3.0.0.2	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
92	发行人	2019SR0994729	基于 window 的新闻语料标注软件 V1.0.0.0	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
93	发行人	2019SR0994774	海天瑞声录音棚预约平台 V1.0.0	2019 年 7 月 30 日	2019 年 9 月 25 日	原始取得	全部权利
94	发行人	2019SR0987594	跨平台 3D 点云标注软件 V1.0.0.0	2019 年 7 月 30 日	2019 年 9 月 24 日	原始取得	全部权利
95	发行人	2020SR1132497	海天瑞声语音合成在线校对平台 V4.0.0.0	2020 年 4 月 30 日	2020 年 9 月 21 日	原始取得	全部权利
96	发行人	2020SR1135478	海天瑞声手机录音实时检测后端 ASR 服务软件 V1.0.0	2020 年 4 月 30 日	2020 年 9 月 22 日	原始取得	全部权利
97	发行人	2020SR1135486	海天瑞声双语对齐检测软件 V1.0.0	2020 年 4 月 30 日	2020 年 9 月 22 日	原始取得	全部权利
98	发行人	2020SR1219259	海天瑞声线上 MOS 评测软件 V1.3.1	2020 年 4 月 30 日	2020 年 10 月 15 日	原始取得	全部权利
99	发行人	2020SR1219262	海天瑞声安卓版 ASR 检测版在线录音软件 V3.0.0.3	2020 年 4 月 30 日	2020 年 10 月 15 日	原始取得	全部权利
100	发行人	2020SR1219265	海天瑞声移动版调查问卷软件 V1.0.0	2020 年 4 月 30 日	2020 年 10 月 15 日	原始取得	全部权利
101	发行人	2020SR1219268	中文文本易读性打分软件 V1.0.0	2020 年 4 月 30 日	2020 年 10 月 15 日	原始取得	全部权利
102	发行人	2020SR1210313	海天瑞声图形图像语义分割校对软件 V0.1.0.0	2020 年 4 月 30 日	2020 年 10 月 13 日	原始取得	全部权利
103	发行人	2020SR1210166	海天瑞声安卓版 OCR 图片采集软件 V1.0.9	2020 年 4 月 30 日	2020 年 10 月 13 日	原始取得	全部权利
104	发行人	2020SR1210161	海天瑞声 IOS 版 ASR 检测版在线录音软件	2020 年 4 月 30 日	2020 年 10 月 13 日	原始取得	全部权利

序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
			V1.0.5				
105	发行人	2020SR1244943	海天瑞声中考作文打分标注软件 V1.0.0	2020年4月30日	2020年10月27日	原始取得	全部权利
106	发行人	2020SR1244920	海天瑞声日语手写体采集标注软件 V1.0.0	2020年4月30日	2020年10月27日	原始取得	全部权利
107	发行人	2020SR1244944	海天瑞声多音字标注软件 V1.0.0	2020年4月30日	2020年10月27日	原始取得	全部权利
108	发行人	2021SR0358467	海天瑞声安卓版天天AI录软件 V1.5.6	2021年3月3日	2021年3月9日	原始取得	全部权利
109	中瑞智	2018SR152885	智能音频采集系统软件 V1.1.4	2017年12月25日	2018年3月8日	原始取得	全部权利
110	中瑞智	2018SR152956	电子书记员系统服务软件 V1.0.0	2018年3月2日	2018年3月8日	原始取得	全部权利
111	中瑞智	2018SR289027	电子书记员系统服务软件 V2.0.0	2018年1月29日	2018年4月27日	原始取得	全部权利
112	中瑞智	2018SR288025	法官助手软件 V1.0	2017年12月22日	2018年4月27日	原始取得	全部权利
113	中瑞智	2018SR288404	实时字幕系统软件 V1.0.0	2018年1月8日	2018年4月27日	原始取得	全部权利
114	中瑞智	2018SR288393	智能语音输入法软件 V1.0.0	2018年1月8日	2018年4月27日	原始取得	全部权利
115	中瑞智	2018SR381241	智能会议系统软件 V1.0.1	2018年3月21日	2018年5月25日	原始取得	全部权利
116	中瑞智	2018SR595610	电子书记员 Word 插件软件 V1.0.1	2018年4月21日	2018年7月30日	原始取得	全部权利
117	中瑞智	2018SR689924	定制模型训练工具软件 V1.0	2018年6月12日	2018年8月28日	原始取得	全部权利
118	中瑞智	2018SR689931	会议记录软件 V1.0	2018年6月13日	2018年8月28日	原始取得	全部权利
119	中瑞智	2018SR689917	潘多拉魔盒软件 V1.0	2018年6月12日	2018年8月28日	原始取得	全部权利
120	中瑞智	2018SR689936	音频转写助手软件 V1.0.1	2018年6月12日	2018年8月28日	原始取得	全部权利
121	中瑞智	2018SR689912	智能语音问答记录软件 V1.0.1	2018年6月13日	2018年8月28日	原始取得	全部权利
122	中瑞智	2018SR714798	智能语音输入法软件 V2.0.0	2018年6月19日	2018年9月5日	原始取得	全部权利
123	中瑞智	2018SR840024	智控软件 V1.0	2018年10月12日	2018年10月22日	原始取得	全部权利
124	中瑞智	2018SR875343	语音合成系统软件 V1.0	2018年10月22日	2018年11月1日	原始取得	全部权利
125	中瑞智	2019SR0835377	会议管理软件 V1.0	2018年12月12日	2019年8月12日	原始取得	全部权利
126	中瑞智	2019SR0835481	钉钉会议预约软件 V1.0	2019年1月2日	2019年8月12日	原始取得	全部权利
127	中瑞智	2019SR1422105	庭审智能语音助手软件 V1.0	2019年10月25日	2019年12月24日	原始取得	全部权利
128	中瑞智	2019SR1399297	音频检索软件 V1.0	2019年10月25日	2019年12月19日	原始取得	全部权利
129	中瑞智	2019SR1399303	语音标注转写软件 V1.0	2019年10月25日	2019年12月19日	原始取得	全部权利
130	中瑞智	2020SR0669044	智能监听管理软件 V1.0	2020年3月29日	2020年6月23日	原始取得	全部权利

序号	著作权人	登记号	软件名称	首次发表日期	登记日期	权利取得方式	权利范围
131	中瑞智	2020SR0669052	产品授权软件 V1.0	2020年3月9日	2020年6月23日	原始取得	全部权利
132	中瑞智	2020SR1554929	自学习管理软件 V1.0	2020年9月16日	2020年11月9日	原始取得	全部权利
133	中瑞智	2020SR1831515	智能庭审记录系统软件 V1.0.0	未发表	2020年12月16日	原始取得	全部权利
134	中瑞智	2020SR1831516	锐听管理客户端软件 V1.0	未发表	2020年12月16日	原始取得	全部权利

### 3、商标

截至2021年5月7日，发行人及其子公司拥有的主要注册商标共30项，均已取得商标注册证书，具体情况如下：

序号	商标	注册人	注册号	注册有效期限	类别	取得方式
1	<b>海天瑞声</b>	发行人	37713178	2019年12月14日至2029年12月13日	9	自行申请
2	<b>海天瑞声</b>	发行人	37718092	2019年12月14日至2029年12月13日	35	自行申请
3	<b>海天瑞声</b>	发行人	37711077	2019年12月14日至2029年12月13日	38	自行申请
4	<b>海天瑞声</b>	发行人	37722765	2019年12月14日至2029年12月13日	41	自行申请
5	<b>海天瑞声</b>	发行人	37711715	2019年12月14日至2029年12月13日	45	自行申请
6	<b>海天瑞声</b>	发行人	8563299	2011年08月14日至2031年08月13日	42	自行申请
7	<b>海天瑞声</b>	发行人	48514307	2021年3月14日至2031年3月13日	9	自行申请
8	<b>海天瑞声</b>	发行人	48534118	2021年3月14日至2031年3月13日	35	自行申请
9	<b>海天瑞声</b>	发行人	48541063	2021年3月14日至2031年3月13日	38	自行申请
10	<b>海天瑞声</b>	发行人	48534178	2021年3月14日至2031年3月13日	41	自行申请

序号	商标	注册人	注册号	注册有效期限	类别	取得方式
11	海天瑞声	发行人	48546346	2021年3月14日至2031年3月13日	42	自行申请
12	speechacean	发行人	37724912	2019年12月14日至2029年12月13日	9	自行申请
13	speechacean	发行人	37733773	2019年12月14日至2029年12月13日	35	自行申请
14	speechacean	发行人	37711082	2019年12月14日至2029年12月13日	38	自行申请
15	speechacean	发行人	37711106	2019年12月14日至2029年12月13日	41	自行申请
16	speechacean	发行人	8563300	2011年08月14日至2031年08月13日	42	自行申请
17	speechacean	发行人	37720041	2019年12月21日至2029年12月20日	45	自行申请
18	speechacean	发行人	48514304	2021年3月14日至2031年3月13日	9	自行申请
19	speechacean	发行人	48546256	2021年3月14日至2031年3月13日	35	自行申请
20	speechacean	发行人	48533036	2021年3月14日至2031年3月13日	38	自行申请
21	speechacean	发行人	48523324	2021年3月14日至2031年3月13日	41	自行申请
22	speechacean	发行人	48530058	2021年3月14日至2031年3月13日	42	自行申请
23	speechacean 海天瑞声	发行人	48534096	2021年3月14日至2031年3月13日	9	自行申请
24	speechacean 海天瑞声	发行人	48533021	2021年3月14日至2031年3月13日	35	自行申请
25	speechacean 海天瑞声	发行人	48534157	2021年3月14日至2031年3月13日	38	自行申请
26	speechacean 海天瑞声	发行人	48534180	2021年3月14日至2031年3月13日	41	自行申请

序号	商标	注册人	注册号	注册有效期限	类别	取得方式
27		发行人	48546350	2021年3月14日至2031年3月13日	42	自行申请
28		发行人	018325007	2020年10月22日至2030年10月22日	9、35、42	自行申请
29		发行人	1567635	2020年10月22日至2030年10月22日	9、35、42	自行申请
30		中瑞智	36771818A	2019年11月21日至2029年11月20日	9、35、37、38、41、42	自行申请

## 六、发行人的研发和技术

### （一）核心技术整体概述



图：海天瑞声科技创新实力示意

海天瑞声拥有对人工智能核心算法的理解能力、前瞻性的专业数据集设计能力、丰富的语言覆盖能力及场景采集能力，以及 AI 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：公司积累了 12 项核心技术，其中在多语种的语音语言学基础研究和高效数据处理两大领域积累了 5 项关键核心技术。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 2020 年 12 月 31 日，公司积累有 760 余个自有知识产权的训练数据产品，可提供超过 160 余个语种/方言的训练数据。

2、工具和平台共建：公司自主开发了一体化数据处理平台，融入项目流程管理、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。

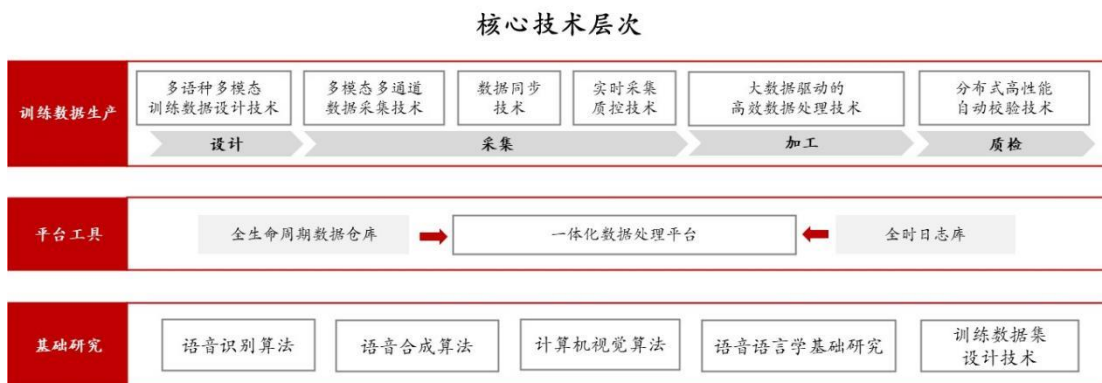
3、在语音语言学基础研究方面有深厚积累：公司建立了成熟的发音词典构建流程、积累了深厚的语音语言学基础研究成果。截至本招股说明书签署日，公司的产品/服务已覆盖 160 余个语种/方言，公司已积累下超过 100 个语种/方言的发音词典，累计词条数超过 1,000 万条，可构建高质量的智能语音训练数据。

截至目前，公司已经取得 24 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，134 项计算机软件著作权，另有 2 项专利已提交申请。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”及“中关村高新技术企业”称号，并于 2018 年被评为国家规划布局内重点软件企业。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。2017 年，发行人被中国信息协会、信息化观察网评选为“年度人工智能数据服务领域领军企业”；2018 年，中国语音产业联盟向发行人颁发了“智能语音创新大赛行业先锋奖”，中国信息协会、信息化和软件服务网向发行人颁发了“2018 年人工智能行业最佳产品”奖项；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”，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，技术实力得到行业、主管部门的高度认可。

发行人的 12 项核心技术已持续应用于训练数据生产的各个环节（设计、采集、加工、质检）。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，按照从底层自下而上的顺序对上述三个层次的核心技术情况介绍如下：



图：各层次核心技术示意图

## 1、基础研究

该层次技术主要包含训练数据生产相关的人工智能算法技术、语音语言学研究、以及训练数据集设计技术等基础技术，是发行人多项核心技术的基础。其中：

（1）人工智能算法技术（语音识别算法、语音合成算法、计算机视觉算法等）：可用于检查训练数据集对算法模型的训练效果，进而反哺指导训练数据集的设计；也可应用于训练数据生产的各个环节，构建训练数据设计、加工相关的核心技术。

（2）语音语言学基础研究：主要涉及音素集、发音规则、发音词典等语音语言学领域的基础研究成果和技术。语音语言学领域的研究成果与发行人的其他项核心技术结合，构建了发行人的多语种训练数据生产能力。

（3）训练数据集设计技术：该层次技术主要为训练数据集设计相关的技术，主要用于提升发行人的训练数据设计及加工能力，实现以既定的数据量提供更好的算法模型训练效果。

发行人生产、销售的训练数据，服务于下游 AI 算法技术行业，该行业技术发展日新月异。基础研究层次的技术积累不仅构建了发行人训练数据生产环节的

技术基础，也使得发行人能够紧跟 AI 领域的前沿算法知识和应用动态，从而更好地把握下游技术的发展方向、深入理解下游客户对训练数据的运用逻辑和需求痛点，为客户提供更加优质的产品服务。

## 2、平台工具

该层次核心技术构建了发行人的一体化数据处理平台，整合贯通了训练数据生产过程相关的设计、采集、加工、质检等业务环节，将项目管理、质量控制、数据安全的相关需求模块化、工具化、流程化、规范化、体系化并嵌入至一体化数据处理平台中，充分提高了训练数据的生产效率及质量控制水平。

具体而言，一体化数据处理平台中集成了发行人自主开发积累的，训练数据生产各环节所需的工具：如采集环节使用的录音工具、视频采集工具，加工环节使用的校对工具、标注工具、声音切分工具等，以及质检工具等。发行人将多项技术研究成果整合至具体工具、平台中，通过技术研究成果的落地实践和应用提升训练数据生产效率。

围绕一体化数据处理平台，发行人还建立了全时日志库，构建了终端人管理系统，对管理人员、研发人员、供应商管理人员、被采集的终端人、标注人员及其组织人员等进行管理，确保人员信息在系统内完整保存、可供追溯；构建了日志管理系统，覆盖训练数据生产全过程，对生产过程数据进行自动化记录和保存，以日志数据形式在信息系统中保留经营业务相关信息，实现关键业务数据、日志数据和用户数据的妥善记录和保存。

此外，发行人正在计划建设全生命周期数据仓库，将进一步加强平台数据安全，保障平台自身及基于平台相关工作的优化和提升。

## 3、训练数据生产

该层次核心技术为发行人训练数据生产中使用的直接技术群，包括设计、采集、加工、质检等环节的多项具体核心技术，例如多语种多模态训练数据设计技术、多模态多通道数据采集技术、数据同步技术、实时采集质控技术、大数据驱动的高效数据处理技术、分布式高性能自动校验技术等。

发行人将上述技术直接运用于训练数据的生产过程，提升了训练数据的生产效率和准确性，在提升自身产出效率的同时确保对客户算法模型的训练效果。

## （二）报告期内主要依靠核心技术开展生产经营的情况

如上所述，在经营过程中，发行人高度重视并坚持科技创新，通过持续的研发投入积累形成核心技术——发行人积累了 12 项核心技术，覆盖基础研究、平台工具、训练数据生产三个层次，应用于训练数据生产的设计、采集、加工、质检全流程之中；上述核心技术构成了发行人生产经营的基础，发行人通过上述核心技术的成果转化，形成了基于核心技术的训练数据相关产品和服务。

报告期内，发行人的产品/服务均通过其核心技术应用开发，核心技术产品/服务的生产、销售数量请参见本招股说明书“第六节 业务和技术/三、销售情况和主要客户”；细分市场占有率情况请参见本招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（四）发行人产品或服务的市场地位”。

发行人的核心技术广泛应用于产品/服务的各开发环节中，报告期实现的营业收入基本均为发行人依靠其核心技术开展经营所产生的收入。报告期内，发行人核心技术产品收入占营业收入的比例如下：

单位：万元

项目	2020年度	2019年度	2018年度
核心技术产品收入	23,323.82	19,246.54	23,733.16
营业收入	23,337.40	19,265.77	23,755.81
核心技术产品收入占营业收入的比例	99.94%	99.90%	99.90%

营业收入具体构成、占比、变动情况及原因等请参见本招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析”。

## （三）核心技术与训练数据定制服务、训练数据产品之间的关系

发行人从事的训练数据定制服务、训练数据产品相关业务，实质均在于为客户提供可供 AI 算法模型训练使用的专业训练数据集。发行人训练数据定制服务的业务过程、训练数据产品的生产过程，就是发行人生产训练数据并交付给客户的过程。

截至目前，发行人共积累了 12 项核心技术，覆盖基础研究、平台工具及训练数据生产三个层次，均应用于训练数据生产全流程（设计、采集、加工、质检）

之中；上述核心技术构成了发行人生产经营的基础，发行人通过应用上述核心技术、实现成果转化，形成了基于核心技术体系的训练数据相关产品及服务。

发行人的各层次核心技术在训练数据生产环节中的应用情况请参见下文具体论述：

#### （四）各项核心技术具体情况

发行人共积累了 3 个层次、12 项核心技术，广泛运用在训练数据生产的各个环节之中，具体如下表所示。其中，5 项核心技术具备较高技术壁垒（下表中以★标示），反映了发行人的技术先进性，将在下文介绍中具体介绍。

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				√
	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、行人检测技术、运动轨迹跟踪技术			√	√
	★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	N 元语言模型训练与优化技术、文本正则化技术、基于语言模型的文本易读性评测技术	√			
平台工具	一体化数据处理平台	基于 C/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 架构的文本词性标注、文本属性标注技术，基于 C/S 架构和 B/S 架构的图像标注技术，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、AudioBook 自动切分与文本对齐技术	√			
	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			√	
	实时采集质控技术	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			√	

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
		HMM/CRF 的词性预测技术、基于 SVM 的文本分类技术、基于 CRF 的命名实体识别技术、基于前后向最大匹配的中文分词技术				
	★分布式高性能自动校验技术	音素标注正确率校验技术、相似说话人自动筛查技术、音字一致校对技术				√

各项核心技术的应用情况及 5 项具备壁垒的核心技术的先进性情况具体如下：

### 1、语音识别算法

发行人拥有基于 DNN-HMM 等模型的多语言语音识别技术，并掌握当前语音识别中流行的多种深度学习技术。在语音识别算法领域，发行人已掌握细分技术语音数据库质量预估技术，已取得 4 项发明专利授权（“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”、“语音数据库的质量评估方法、装置及计算机存储介质”、“基于线性预测残差负熵的语音音质度量评价方法及装置”及“一种基于三元组的语音识别准确率计算方法”）。

发行人将上述人工智能算法技术运用在训练数据生产过程中，提升了训练数据生产效率，具体如下：

#### （1）检验语音识别训练数据的训练效果、反哺指导训练数据生产

发行人用其生产的训练数据，对发行人自有的语音识别算法、模型进行训练，从而测试、检验语音识别领域训练数据的算法模型训练效果，反哺指导训练数据的生产过程，促进发行人优化训练数据的结构开发、内容设计，进而提供具备更佳训练效果的训练数据，提升发行人产品服务的附加值和竞争力。

#### （2）用于执行自动化数据加工、质检环节，程序化操作结合人工检查提升训练数据生产效率

发行人将语音识别算法集成在其开发的数据采集、加工工具之中，一方面可在采集过程向被采集对象提供实时反馈，进而提高采集的效率和准确率；另一方面可运用于语音数据生产的各环节中，通过语音识别算法对智能语音数据进行预标注，与人工检查环节结合，提升单位时间加工语音数据的能力，提高智能语音训练数据的生产效率和质量；也可通过语音识别算法对智能语音训练数据进行质

检，反向检验人工数据加工结果。



图：语音识别算法在采集环节中的实时质检反馈示例——海天瑞声自主开发的手机端采集软件中，集成了运用语音识别技术校验采集数据情况的功能，并实时向被采集对象提供反馈。

(3) 直观展示语音识别训练数据的训练效果，为客户提供评判质量优劣的直观标准

发行人提供的语音识别算法模型可用于其训练数据的模拟训练和测试，为客户直观展示发行人训练数据的模型训练、测试效果，使得客户能够直观衡量发行人语音识别训练数据的质量水平，从而选择与其预期的算法模型训练、测试效果更为匹配的语音识别训练数据。

例如，发行人运用不同语种的训练数据产品对同一语音识别算法模型进行了训练，通过算法模型训练结果直观展示发行人的训练数据产品具备较优的测试效果，各语种训练数据训练后的语音算法模型的字符错误率/词错误率均低于 5%：

产品编号	产品名称	训练数据产品中的音频数据总时长(小时)	测试集比例	字符错误率/词错误率
King-ASR-137	韩语语音识别库(手机)	342.00	10%	0.95%
King-ASR-124	俄语语音识别库(手机)	341.30	10%	0.94%
King-ASR-151	德语语音识别库(手机)	444.90	10%	1.04%
King-ASR-054	英式英语语音识别库(手机)	428.40	10%	2.83%

产品编号	产品名称	训练数据产品中的音频数据总时长(小时)	测试集比例	字符错误率/词错误率
King-ASR-142	西班牙裔西班牙语语音识别库(手机)	496.20	10%	1.83%
King-ASR-241	广州粤语语音识别库(手机)	3,220.80	10%	4.89%
King-ASR-329	加拿大法语语音识别库(手机)	405.00	10%	1.70%
King-ASR-072	法国法语语音识别库(手机)	496.50	10%	2.38%

注：1、“测试集比例为 10%”的含义为：从训练数据产品中随机选取 90%数据对语音识别算法模型进行训练，并使用剩余的 10%数据作为测试集，测试经过训练后的算法模型的识别正确率情况。

2、字符错误率、词错误率为语音识别算法模型识别正确率的反向衡量指标，错误率越高则模型识别准确率越低、识别效果越差。

#### (4) 运用语音识别算法模型技术为客户提供部分算法模型拓展服务

发行人基于其生产的训练数据，运用语音识别算法模型技术为客户提供部分算法模型测试、拓展和训练服务，如协助下游客户完成算法模型的语言拓展、特定算法模块拓展、垂直领域应用拓展等，可为客户定制针对特定行业和口音的专属算法模型。

## 2、语音合成算法

发行人拥有参数合成技术、基于神经网络的端到端语音合成技术，拥有语音合成的 AI 算法模型，已取得 2 项发明专利授权（“端到端的语音合成方法、装置及存储介质”及“基于线性预测残差负熵的语音音质度量评价方法及装置”），正在申请 1 项发明专利。

发行人将上述技术运用于其训练数据生产过程，提升了训练数据生产效率，具体包括：

#### (1) 检验语音合成训练数据的训练效果、反哺指导训练数据生产

发行人运用上述语音合成技术，基于其生产的语音合成训练数据做了大量的算法模型训练、语音合成效果验证和测试。经过持续的算法迭代和优化，逐步提升语音合成算法技术的研究水平，将算法模型用于检验语音合成训练数据的合成效果，反过来指导训练数据生产部门对训练数据的设计、采集、加工和质检等过程进行相应的调整和改进，提升训练数据质量。



(2) 运用语音合成算法模型展示训练数据的合成效果、帮助客户选择合意的训练数据

发行人使用上述语音合成算法模型，基于不同的训练数据集进行了语音合成效果测试，为客户直观展示训练数据的合成样音，使得客户可以在实际开展算法模型训练和测试工作之前初步了解特定训练数据集的语音合成效果，从而选择风格、音色等更符合需求的语音合成训练数据，最终取得更为理想的语音合成效果。

### 3、计算机视觉算法

计算机视觉领域是发行人训练数据生产销售的三大应用领域之一，发行人积极在该领域积累算法技术能力，从而更好地理解下游客户需求，相应生产、提供高质量的训练数据集。在计算机视觉算法领域，发行人已取得 4 项发明专利授权（“视频中音频聚类的处理方法和装置”、“视频抽选方法、装置及存储介质”、“视频筛选方法、装置及存储介质”及“图像标注方法、图像标注装置及计算机存储介质”），另有 1 项发明专利正在申请中。

发行人现已掌握计算机视觉领域相关的算法技术，并将其运用至训练数据生产过程的降本增效之中，具体如下：

发行人将计算机视觉算法技术嵌入至自主研发的分布式标注平台之中，全面提升了计算机视觉领域的原料数据加工效率：

以发行人的 1,000 人手机人像短视频训练数据集（King-AV-025）为例，该训练数据集是发行人针对手机短视频领域人像识别应用所开发的训练数据集，数据集中包括 1,000 个黄种人共计 1,000 段短视频数据，视频时间长度合计达到 25 小时。发行人采集完构建该训练数据集所需的短视频原料数据后，需要对上述短视频数据进行检查和进一步的加工。为保证训练数据集的质量，发行人需要检查、确认上述视频原料数据不存在重复。在上述“去重”的数据检验排查过程中，发行人借助其在计算机视觉算法领域积累的“视频筛选方法、装置及存储介质”技术，针对各个短视频批量自动截取关键帧并执行相似度计算，对于算法检查出的、相似度较高的短视频原料数据，再做人工排查。该技术降低了人工比对成本，提供了数据加工过程的准确性。

#### 4、★语音语言学基础研究

##### (1) 核心技术及其应用介绍

语音语言学领域的专业知识是构建高质量语音识别算法和语音合成算法的关键要素。以语音合成为例，在语音合成系统中，发音词典提供了从单词到音素之间的映射关系，将语言模型建模单位解构为声学模型的建模单元，为后续合成发音奠定基础。语音合成系统接收到文本信息后，首先运用发音词典对其进行语言处理、韵律处理，将文本（单词、字符等）转换并解构为一系列对应的发音符号（类似于国际音标）；随后，系统中的语音合成器接收到前述发音符号，运用语音库合成转换为语音对外输出，最终实现文本到语音的语音合成过程（参见下图）。可见，高质量的发音词典在语音合成系统中具备重要作用。



图：语音合成系统框架示意

由上述示例可以看出，要获得高准确率的语音合成算法模型，就要求智能语音训练数据结构中包含高质量的发音词典。要在大词汇量的连续语音交互中正确、合理运用智能语音相关的语言模型、语法及词法模型，则必须有效地运用计算语言学方面的基础知识和研究成果。语音语言学领域的基础研究成果和专业知识构建了发音规则、发音词典的形成基础，进而为构建高准确率的语音识别、合成训练数据提供了条件。

经过多年积累，发行人积累有深厚的语音语言学基础研究成果，现已积累下超过 100 个语种/方言的发音词典，累计词条数超 1,000 万条，并将其运用至构建

高质量的智能语音训练数据。

## （2）技术先进性情况

在语音语言学基础研究领域，发行人的领先性、技术先进性主要体现在：

### ①丰富的多语种语言学家团队资源积累、多语种发音词典积累

经过多年积累，发行人掌握有丰富的语言学家团队资源，成员遍布世界各地的学校和研究机构，在语音语言学领域具备丰富经验和技術储备。以印地语、他加禄语等外语种、小语种为例，这类语言并没有国际通用的发音系统，构建这类语言的发音词典就需要语言学家综合考虑该语言的发音特点、并从易于语音识别系统区分的角度出发，进行该语言的音素集设计，并总结提炼对该语言的发音规则，构建基本的字/词到音素的对应关系；然后对该语言的常见词语进行标音，最终形成发音词典。

### ②成熟的发音词典构建技术和流程积累

在词典构建领域，发行人建立了成熟的词典构建流程，并掌握了成熟的词典构建技术——如基于词典与模型的发音预测技术等。发行人对基于规则和统计的发音预测算法进行了多年研究，并利用长期经营获得的大量数据和经验积累，研发出了基于发音词典和模型的文本发音预测技术，可运用于发音词典设计制作、训练数据集的设计当中。

## 5、训练数据集设计技术

训练数据集设计是构建高质量智能语音、计算机视觉和自然语言处理等算法模型开发、训练、优化和拓展所需训练数据的关键环节。

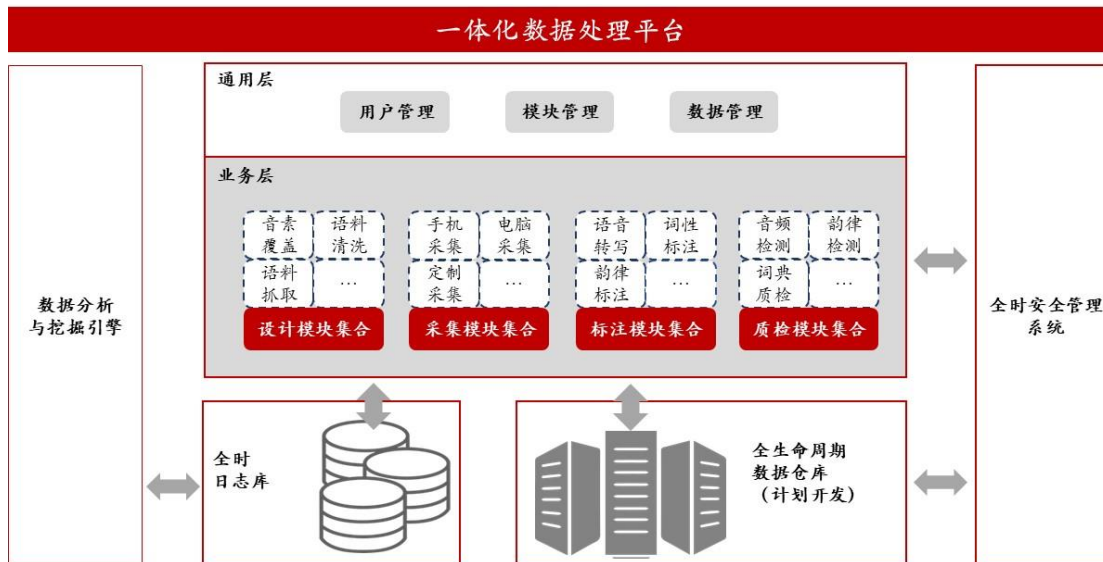
以语音识别、语音合成领域的训练数据集为例，在原料数据的采集环节，发音人（被采集对象）需要朗读发行人提供的基础语料，并用指定的录音设备录制形成原料音频数据。因此，在设计阶段，发行人就需要考虑如何设计基础语料，才能使得容量有限的训练数据集能够覆盖尽可能多的自然语言现象，如覆盖更多的语音、语气、语调情况，兼顾多种句子类型（陈述句、疑问句、感叹句等）和句长分布，考虑不同的发音习惯、语言特点，最终达到更好的语料覆盖效果，从而使得算法模型获得更好的训练结果——使得语音识别模型能够“听懂”更多实

际场景中的语音数据；使得语音合成模型能够“说出”更为自然的语音效果。

通过多年经营积累，发行人已具备 160 余个语种/方言的训练数据设计开发能力，并掌握了 N 元语言模型训练与优化技术、文本正则化技术、基于语言模型的文本易读性评测技术等算法技术，可结合音素均衡、语料主题、覆盖领域、发音人属性、设备特征等因素综合设计满足算法模型开发、训练、拓展需要的训练数据集。在训练数据集设计技术领域，发行人已取得 1 项发明专利授权（“跨语言非标准词识别方法及装置”）。

## 6、一体化数据处理平台

在多年的行业实践过程中，发行人始终注重充分挖掘和培养自身的核心技术优势，秉承基础研究与实际应用紧密结合的原则，持续建设和完善训练数据生产相关的平台和工具。发行人自主研发的一体化数据处理平台整合贯通了训练数据生产各环节（设计、采集、加工、质检）及数据安全管理的需求，整合多维度、多创新点的工具群，构建了流程化、规范化和体系化的训练数据生产体系，提高了产能及效率，降低训练数据生产成本、缩短了训练数据生产处理时间。



图：“一体化数据处理平台”示意图

发行人的一体化数据处理平台具备下述优势和先进性：

(1) 统一入口：为全体训练数据生产人员（项目经理，数据采集、加工人员，系统维护人员等）提供进入系统的统一入口，并根据不同人员职责差异分配不同的访问和使用权限，方便记忆和管理。

(2) 统一风格：提供一致的用户体验，提升项目管理人员及数据采集、加工人员等的操作效率，有利于为客户提供质量稳定、标准统一的训练数据集。

(3) 统一流程：结合多年行业深耕经验，简化并统一训练数据生产流程，融入发行人多年的项目管理、进度把控经验，并提供模块式的项目生成和管理方式，供训练数据生产人员根据实际项目需求有机结合、灵活调整。

(4) 统一框架：在一体化数据处理平台设计开发过程中，发行人注重总结并抽象出共有模块，提高代码的复用率，降低开发成本，也有利于快速新增模块；优化代码组织框架，降低程序错误率；充分提高了一体化数据处理平台的使用效率。

与一体化数据处理平台相配合，发行人已建立了全时日志库，构建了终端人管理系统，对管理人员、研发人员、供应商管理人员、被采集的终端人、标注人员等进行管理，确保人员信息在系统内完整保存、可供追溯；同时构建了完善的日志管理系统，覆盖训练数据生产全过程，对生产过程数据进行自动化记录和保存，以日志数据形式在信息系统中保留经营业务相关信息，如记录采集事项发生时间、单次采集数据量、累计采集数据量、测试标产比、各标注人员的有效工时、产出量、实际标产比等关键业务数据、日志数据和用户行为数据。

此外，发行人计划建设全生命周期数据仓库，将进一步加强数据安全管理及隐私保护，实现一体化数据处理平台功能的拓展及安全性的提升。

在与一体化数据处理平台相关的技术领域，发行人已取得 3 项发明专利授权（“语音转写方法、语音转写装置及计算机存储介质”、“图像标注方法、图像标注装置及计算机存储介质”及“点云数据标注方法、点云数据标注装置及存储介质”），正在申请 1 项发明专利。

## 7、★多语种多模态训练数据设计技术

### (1) 核心技术及其应用介绍

设计环节是整个训练数据生产流程中的第一个环节，是高质量训练数据的生产基础。要生产高质量的训练数据，就需要在生产之初即设计好训练数据集结构并相应制定合理的原料数据采集方案。这要求对发行人对客户算法和应用场景具备较高的理解能力，也要求发行人在自然语言理解、语音语言学等诸多基础领域

具备较高的技术水平。

发行人基于多年的行业经验和技術储备，针对不同类型的数据生产需求，自主研发了一系列算法、技术与工具，形成了多语种多模态<sup>23</sup>训练数据设计这一核心技术。

在多语种训练数据设计领域，发行人自主研发掌握了多语种音素均衡语料设计技术、混合语言语料设计技术等子技术，可解决智能语音训练数据集设计时，多种语言的音素覆盖、句长分布、领域内容分布问题，可高效生产混合语料训练数据。

在多模态训练数据设计领域，发行人可结合具体场景、应用领域特性，设计与之相匹配的、可用于算法模型开发、训练、拓展及优化的多模态训练数据。如通过多种终端设备同时获取人发出的对话语音信息、唇部动作信息、声音来源方向信息等多维度信息，随后综合运用计算机视觉技术、智能语音技术等算法技术实现视觉、听觉等融合的多维度交互。

在多语种多模态训练数据设计技术领域，发行人已取得 4 项发明专利授权（“语料选取处理方法、装置、设备及计算机可读存储介质”、“个性语料获取方法及个性语料获取装置”、“短语语料获取方法及短语语料获取装置”及“中英混合语料的生成方法、装置、设备及存储介质”）。

## （2）技术先进性情况

发行人通过持续进行自主研发、原始创新并有机整合多种训练数据设计技术、数据采集技术，掌握了上述技术。在多语种多模态训练数据设计技术领域，发行人的技术壁垒和竞争优势主要体现在其掌握的下述细分技术中：

多语种音素均衡语料设计技术：一个高质量的语音训练数据开发方案既需要考虑相关语言的音素覆盖问题，又需要考虑句长分布、领域内容分布等其他因素的均衡问题。发行人通过运用自主研发的多语种音素均衡语料设计技术，解决了

---

<sup>23</sup> 多模态数据：指同时来源于多元感知渠道信息的多维度数据集，以人机交互过程为例，在人机交互过程中，机器首先通过多种终端设备同时获取人发出的对话语音信息、唇部动作信息、声音来源方向信息等多维度信息，随后综合运用计算机视觉技术、智能语音技术等 AI 算法技术判断声音指令来源，并对同一时间、状态、环境下的多模态数据进行综合分析，进而相应做出反应和输出，实现视觉、听觉等融合的多维度交互。

各音素之间互相影响的问题，避免了训练数据集结构的缺失，保证了训练数据的质量。发行人已在该技术领域获得了 1 项发明专利授权“语料选取处理方法、装置、设备及计算机可读存储介质”，可根据原始句长分布需求，从原始语料中选取满足句数要求和句长要求、且与所述原始句长分布相匹配的语料，作为初始句长分布模型，得到与原始句长分布一致或者较接近的初始句长分布模型；可通过修正处理初始句长分布模型得到满足总字数要求、句数要求和句长要求的最终句长分布模型，使得最终句长分布模型中的句长分布接近原始句长分布，符合语料设计中对于句长分布的要求。

混合语言语料设计技术<sup>24</sup>：发行人的混合语言语料设计技术运用于训练数据生产的设计阶段，可基于原始中英混语料的特征提取，通过转化纯中文语料生成中文汉字、英文单词及标点符号的中英混合语料，并兼顾易读性和常见性，从而设计出具备良好中英文混合效果的训练数据。发行人已针对该项技术取得了 1 项发明专利授权“中英混合语料的生成方法、装置、设备及存储介质”，可通过对中文语料数据进行分词处理及词性标注，获得标注后的中文数据；根据标注后的中文数据，生成中英混合语料，可有效扩充中英混合语料库，为中英混合语音识别的模型训练提供充足的训练数据，从而提高模型的准确性。

## 8、多模态多通道数据采集技术

多模态多通道采集过程是指通过部署多台设备实现多通道多模态原料数据采集，具体工作方式举例如下：针对同一发音人，同时部署多台录音设备、多路视频采集设备，从而同时采集发音人的语音、视频、唇形等多维度不同形态的数据。

在多模态多通道数据采集过程中，各录音设备、视频采集设备等需接入同一服务器，可能出现多路设备冲突导致丢失数据包、视频帧、语音视频不一致的情形。发行人针对上述现象，自主研发了多模态多通道数据采集技术，并开发完善了相应的采集软件、软硬件一体系统，能够实现多维度原料数据的采集和有机整合。在多模态多通道数据采集技术领域，发行人已取得 2 项发明专利授权（“语音库的语音获取方法、装置、电子设备及存储介质”及“视频中音频聚类的处理

---

<sup>24</sup> 混合语言语料：指混合了多种语言的语料，如中文和英文混合：“快去帮我买杯 coffee”。

方法和装置”）及 1 项实用新型专利授权（“语音采集设备”）。

发行人将上述技术充分运用至高质量、复杂训练数据的生产过程之中。以发行人的 200 人中文普通话语音识别数据库（King-ASR-010）和 250 人唇语语音视频数据库（King-AV-018）生产过程为例：

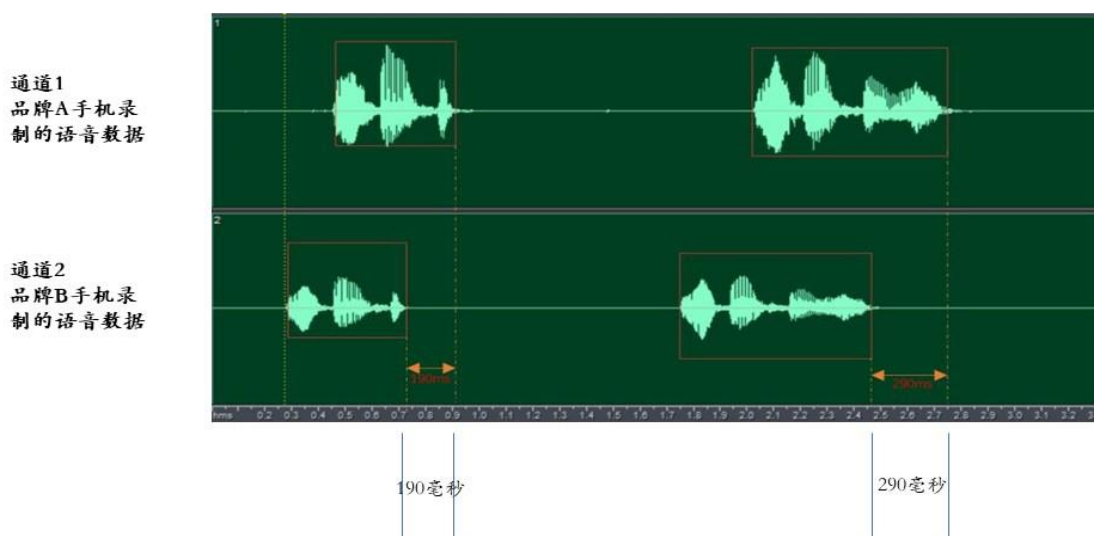
King-ASR-010 是发行人生产的中文普通话语音识别训练数据集，该数据集覆盖了 4 个通道（在同一桌面不同位置布设的 4 个录音设备）的音频数据，每个通道覆盖 200 人共 6,000 句语音数据。发行人自主研发的多通道采集工具很好地支持了 4 路语音数据同时录入、自动对齐整合的采集需求。

King-AV-018 是发行人针对虚拟主播需求开发的视频训练数据集，该数据集中含有 250 个中国人共计 610 句的唇语视频信息。发行人在常见的语音数据采集设备上叠加了针对唇语信息的视频采集设备，运用自主研发的多模态数据采集软件同时录入不同硬件设备的数据，并保持多维度数据的对齐。

## 9、★数据同步技术

### （1）核心技术及其应用介绍

数据同步性是训练数据集质量的重要评价标准。而在数据采集过程中，往往会出现多通道（多个设备）同时采集的需求，如多个收音设备同时记录声音、多个拍摄设备同时记录影像等。因设备的物理限制，受设备采集开始时间不同步、设备本身时钟不同步、结束时间不同步等因素影响，多个设备采集的原始数据很容易发生不同步的情况（参见下图示例），使得后续数据加工难度大大增加。





图：不同通道录制的语音数据不对齐情况示例——针对同一音源，不同品牌手机同时录制的语音数据的时间起点不一致，且不同语音片段的终点也不一致，且起点至终点之间并不是完全线性平移的关系，例如第一段语音的终点差距为 190 毫秒、第二段语音的终点差距则达到了 290 毫秒

针对上述问题，发行人自主研发了数据同步技术，通过获取不同录音设备采集的同一语音内容对应的多个语音数据，并从任一语音数据选取任一语音片段作为样本，通过确定所选语音样本的帧数提取其语音特征参数，在各其他语音数据中确定与上述样本相似度最高的目标语音片段，进而实现多个语音数据的时间轴对齐处理工作。该技术可实现多通道采集原料数据的自动对齐，提升了训练数据生产效率和产品质量。在该技术领域，发行人已取得 1 项发明专利授权（“语音对齐方法及装置”）。

## （2）技术先进性

在数据同步技术领域，发行人自主研发了音频对齐技术，将其运用至智能语音数据的生产过程中，通过计算多个语音特征向量的自相关系数，解决了多个音频文件起始点自动对齐的问题，对齐误差小于 1 毫秒，大大提升了智能语音训练数据集的质量。

以在发行人生产的训练数据集山东重口音普通话语音数据库（手机）（King-ASR-384-12）的生产过程为例：

King-ASR-384-12 是发行人针对山东普通话手机端语音识别应用开发的典型训练数据集。该数据集的原料音频数据共涉及三个通道，分别由三个不同的手机同时采集，每个通道采集约 542 个小时的语音，对应句子数达 500,866 句。

在 King-ASR-384-12 的制作过程中，需要对录制的 50 多万组、每组 3 句话的语音数据进行比对和对齐，工作量巨大，采用纯人工操作难以较快完成。发行人通过运用核心技术数据同步技术项下的音频对齐技术，对上述 50 多万组语音数据进行了并行处理及对齐，高效地完成了前述音频文件的同步需求。

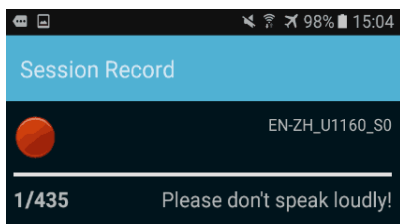
## 10、实时采集质控技术

在原料数据采集过程中，受采集设备的物理限制、被采集人的主客观情况或者采集环境变化等因素影响，无法完全保证采集到的每条原料数据都满足后期加工的质量要求。例如，在手机录音采集任务中，当发音人说话声音太响，在音频

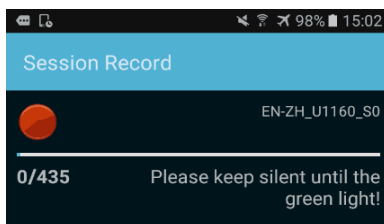
采集时会出现截幅现象；但说话声音太小时，又会造成所采集到的语音片段难以听清内容，不符合后期加工要求；此外，音频录制时，若周围噪音过大，也会导致音频数据的信噪比过小，不符合质量要求。

为解决上述采集过程中的原料数据质量问题，提升采集数据的有效率、降低无效数据占比，发行人总结多年行业经验自主研发了实时采集质控技术，将声音能量检测、音频截幅检测、静音检测、信噪比检测等数字信号处理技术整合到实际数据采集设备中，可在采集的同时进行质量检查，及时发现采集数据的质量问题并及时提供现场反馈（参加下图示例）。这项技术提高了采集数据的有效性，从而保证最终产品的质量。

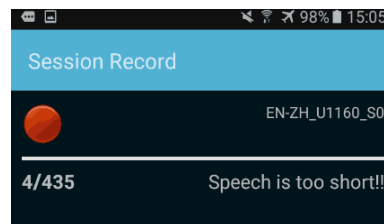
在实时采集质控技术领域，发行人已取得 1 项发明专利授权（“语音库的语音获取方法、装置、电子设备及存储介质”）及 1 项实用新型专利授权（“语音采集设备”）。



图一：音量过大



图二：未保留足够静音



图三：语音段太短

图：原料数据采集过程中实时质检反馈示意

以发行人生产的西班牙语手机对话类语音数据库（King-ASR-321）的生产过程为例：

King-ASR-321 是发行人针对西班牙语自由对话语音识别应用开发的训练数据集。该训练数据集涉及普通环境下的呼叫中心+手机双通道，每个通道采集了约 213 个小时的语音数据，对话内容覆盖 21 个主题。

在 King-ASR-321 原料对话数据的采集过程中，应用了发行人实时采集质控技术的录音设备能够实时检测信噪比、语音能量和静音时长等信息，一旦出现录制声音过大或过小、信噪比过低、首尾静音时长等不符合要求的语句，会即时提示要求发音人重录；同时，该设备将依据长音频中切分的各段语音长度和对应文本字数计算发音人的语速分布，后续可执行统计分析、找到异常情况并校正。发行人有效地提高了该训练数据产品的整体质量。

## 11、★大数据驱动的高效数据处理技术

### (1) 核心技术及其应用介绍

发行人在训练数据研发、生产领域深耕多年，积累了深厚的开发经验，并在训练数据的设计、采集和处理环节积累了大量数据，具体包括：发行人自行采集的大量原料数据、发行人拥有自主知识产权的训练数据产品；以及训练数据生产过程中产生的大量日志数据等。

大数据驱动的高效数据处理技术是指发行人充分利用前述数据，通过分析训练数据生产过程中的数据分布规律和行为数据模型设计开发出了高效率的数据处理平台与工具集。在该技术领域，发行人已取得 4 项发明专利授权（“韵律标注方法、装置和设备”、“视频中音频聚类的处理方法和装置”、“视频抽选方法、装置及存储介质”及“视频筛选方法、装置及存储介质”）。

发行人运用上述技术充分提高了数据加工工作效率和数据产品服务质量。具体如下：

发行人对原料数据及加工后数据的异同点进行分析，运用统计模型归纳出一般性规律，将之运用于开发和优化采集、标注过程所运用的辅助工具与算法，也运用于优化数据加工任务类型的划分和流程管控，从而提升数据加工的工作效率。例如：发行人基于多年积累的训练数据生产经验开发了词性预测工具，针对不同语种/方言反复调整参数和算法，不断提升预测的准确率，提升数据加工效率、降低人力需求；此外，发行人分析了标注人员的专业领域、属性特征，与其从事数据加工的效率情况和工作质量情况，从而指导自身针对具体数据加工需求更科学地进行数据处理任务分配、提升数据加工效率；发行人也对数据加工过程中的人工行为的模式及原因加以分析，用以指导数据加工流程的优化和工具界面调整设计，最终达到提升数据开发效率的目的。

发行人现已积累了 25 个运用于标注过程的预测算法或工具，利用上述工具提升了数据加工过程的效率及准确率。以词性预测算法工具为例，其将词性预测算法的准确率由 86% 提升至 96%<sup>25</sup>。

---

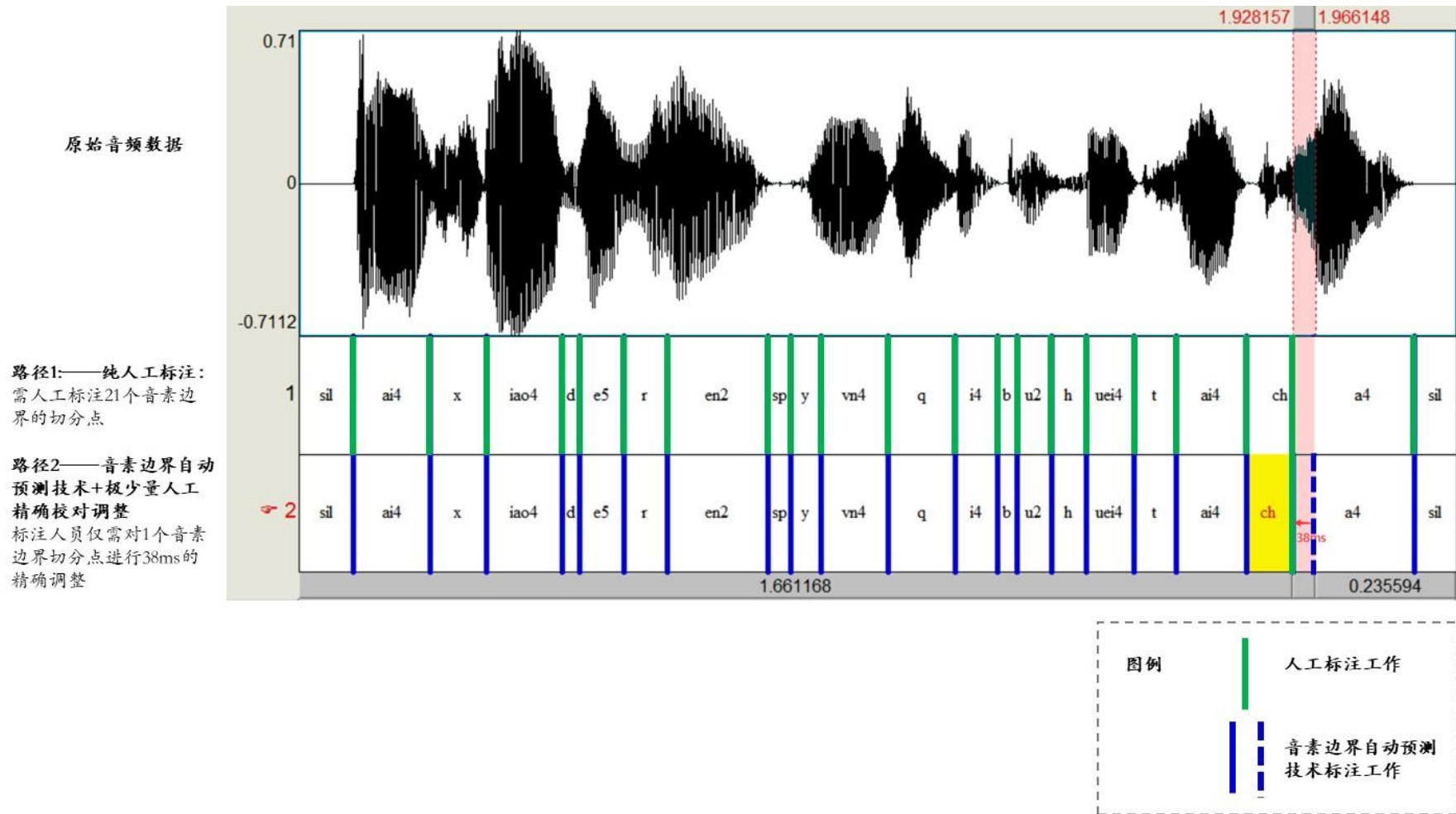
<sup>25</sup> 上述数据来自发行人实验室模拟运算分析。

## （2）技术先进性情况

发行人在大数据驱动的高效数据处理技术领域构建的技术壁垒和竞争优势主要体现在音素边界自动预测技术、基于语音特征的韵律预测技术等子技术中。前述技术的成熟和普遍运用提升了发行人的数据加工效率，有效减少了数据加工过程中的人工工作量，举例来看：

### ① 音素边界自动预测技术

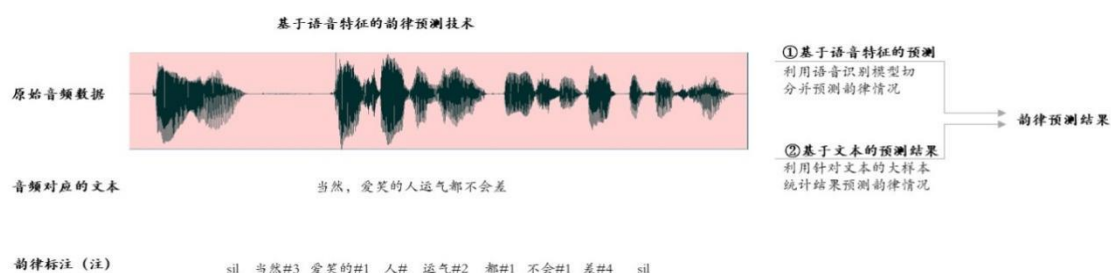
音素是根据语音的自然属性划分出来的最小语音单位，依据音节里的发音动作来分析，一个动作构成一个音素。音素分为元音与辅音两大类。如汉语音节啊（ā）只有一个音素，爱（ài）有两个音素，代（dài）有三个音素等。音素及音素边界的正确性是衡量语音合成训练数据质量的重要指标。发行人通过大量数据训练声学模型自主研发掌握了音素边界自动预测技术，通过语音识别算法的运用得到音素边界的预测值以辅助标注人员的工作，有效提升了音素标注效率。



图：音素边界自动预测技术对标注工作效率提升示例

以上图为例，对于该原始音频片段，在未引入音素边界自动预测技术的情况下，标注工作人员需要通过查看原始音频片段对应的波形图、语谱图，并结合人工听辨，手动标注 21 个音素边界点；在引入了音素边界自动预测技术后，标记人员可基于音素边界自动预测技术的标注结果执行人工复核，仅需要对 1 个音素边界点进行精确人工调整。音素边界自动预测技术的引入极大地提升了音素标注的效率和准确性。

## ②基于语音特征的韵律预测技术



图：基于语音特征的韵律预测技术示例

韵律反映的是语音语句中各字词之间的停顿时长情况，韵律标注即需标出句子中不同字词的具体停顿长度，如以“#1”代表语法词的停顿长度、“#2”代表 $\leq 90$  毫秒的停顿长度等。由于语音韵律具有较强的差异性和个人特色，因此韵律标注通常由标注人员通过语音听辨人工完成，标注人员根据韵律标注规范，判断每个语法词的韵律边界类型。

常见的韵律预测技术主要是从文本出发，根据大量的文本统计，得出通常情况下该语句中各字词的停顿时长预测结果（参见图示中的路径②）；发行人所掌握的基于语音特征的韵律预测技术，可以利用语音识别模型，对音素进行自动切分并预测韵律边界（参见路径①），利用语音识别模型预测结果，修正路径①中基于文本得到韵律预测结果；后续再由标注人员结合语音听辨，在预测结果上核对修改。该韵律预测技术通过引入语音识别模型，结合语音特征预测结果和文本特征预测结果，提升了韵律预测与实际音频数据的贴合度，提高了韵律预标的准确率，有效减少了标注过程中约 25% 的人工标注工作量。

## 12、★分布式高性能自动校验技术

### (1) 核心技术及其应用介绍

质检环节是训练数据生产过程的关键步骤，是发行人保证训练数据集质量的重要手段。发行人生产的训练数据集将直接运用于客户算法模型的训练过程中，对训练数据集的质量、准确率均有较高的要求。

以语音合成训练数据的音素信息为例，对音素信息标注的准确率往往需要达到 99% 以上。对大规模数据进行质检并避免重复人工检查、提升最终的训练数据准确率是质检环节面临的挑战。基于传统的信号处理技术已无法很好地满足要求，发行人自主研发了分布式高性能自动校验技术，运用成熟的人工智能算法来辅助完成训练数据质检。

在分布式高性能自动校验技术领域，发行人已取得 4 项发明专利授权（“音素误标注的检测方法和装置”、“音频质量评估方法、装置、电子设备及存储介质”、“语音检测方法和装置”及“语音转写方法、语音转写装置及计算机存储介质”）。

根据多年行业经验，发行人制订了一整套完整有效的质检流程，以保证给客户id提供高质量的训练数据集。其中多个质检环节都需要对海量的数据文件做复杂处理，比如：通过比对语音文件与文本文件的内容，检测具有质量缺陷的数据并予以更正。如果在单一服务器上完成运行上述比对、处理等复杂计算工作，则由于数据量巨大（单个数据库就可能有上百万个语音文件）和计算过程复杂（很多信号处理要求）将耗用数以周计的运行时间，无法满足快速交付实际需求。

发行人部署了分布式集群系统，并在前述分布式高性能集群系统上运行自主研发的平台架构、工具及全自动校验技术相关算法。发行人自主研发的质检平台实现了各计算节点间的数据共享、安全传输、任务动态分配及消息队列自动管理等功能，同时支持自动添加、删除和更新节点服务器，弹性地支撑实际业务需求，可实现质检、比对、处理等计算任务的高效运行，可在约 24 小时内完成约 660G 的数据质检工作。分布式高性能集群系统为发行人的分布式高性能集群自动校验技术提供了基础算力支撑，与发行人自主研发的平台、架构工具以及全自动校验技术算法一起，构成了发行人在质检环节使用的分布式高性能集群自动校验技术体系。

## （2）技术先进性

在分布式高性能自动校验技术领域，发行人的技术优势主要体现在其自主研

发的相似说话人自动筛查技术及音素标注正确率校验技术这两个细分领域。上述细分技术情况具体如下：

①相似说话人自动筛查技术：为覆盖不同说话人的声学特性，一个大规模语音识别训练数据集往往包含上千个说话人录制的语音数据。为避免同一说话人录制多份数据影响训练数据的质量，发行人自主研发了相似说话人自动筛查技术。该技术首先建立说话人的 GMM 模型，然后对说话人进行相似度打分，筛选出疑似相同说话人，指导后续人工校验过程。该技术将校验工作的人工工作量减少了 98%以上，同时提高了工作的准确度，在该技术领域，发行人已取得 1 项发明专利授权（“语音检测方法和装置”）。

②音素标注正确率校验技术：音素及音素边界的正确性是衡量语音合成训练数据质量的重要指标。发行人利用基于三音子模型的音素边界自动预测技术，通过训练 GMM-HMM 声学模型，对音素进行识别，基于一定数学分布的假设先行筛选出错误可能性较高的音素，交由人工检查进一步修正，有效提高了音素标注质量和效率。在该技术领域，发行人已取得 1 项发明专利授权（“音素误标注的检测方法和装置”）。

## （五）大数据和算法在训练数据生产过程中的运用情况

### 1、大数据和算法在训练数据生产过程中的运用情况

#### （1）大数据在训练生产中的主要运用情况

发行人在训练数据研发、生产领域深耕多年，积累了深厚的开发经验，在训练数据的设计、采集和处理环节涉及大量数据，具体包括采集的原料数据、拥有自主知识产权的训练数据产品以及训练数据生产过程中产生的大量日志数据等。

发行人对前述大数据的运用主要体现在：积累了大数据驱动的高效数据处理技术，通过分析训练数据生产过程中的数据分布规律和行为数据模型，设计开发出高效率的数据处理平台与工具集，并充分运用上述技术、平台、工具集提高数据加工效率和训练数据质量。具体包括：

①对原料数据及加工后数据的异同点进行分析，运用统计模型归纳出一般性规律，将之运用于开发和优化采集、标注过程所运用的辅助工具与算法，也运用于优化数据加工任务类型的划分和流程管控，从而提升数据加工的工作效率。例



如：发行人基于多年积累的训练数据生产经验开发了词性预测工具，针对不同语种/方言反复调整参数和算法，不断提升预测的准确率，提升数据加工效率、降低人力需求；

②对标注人员的专业领域、属性特征，与其从事数据加工的效率情况和工作质量情况等大数据进行分析，从而指导自身针对具体数据加工需求更科学地进行数据处理任务分配、提升数据加工效率；发行人也对数据加工过程中的人工行为的模式及原因加以分析，用以指导数据加工流程的优化和工具界面调整设计，最终达到提升数据开发效率的目的。

通过大数据积累和分析运用，发行人现已积累了 25 个运用于标注过程的预测算法或工具，并利用上述工具提升了数据加工过程的效率及准确率。以词性预测算法工具为例，其将词性预测算法的准确率由 86% 提升至 96%<sup>26</sup>。

## (2) 人工智能算法在训练生产中的主要运用情况

发行人通过运用语音识别算法、语音合成算法、计算机视觉算法，对训练数据集的算法模型训练效果加以检查，从而反哺指导训练数据集的设计；同时将前述算法运用于训练数据生产环节，构建了训练数据设计、加工相关的核心技术。

对于各人工智能算法的运用情况具体说明如下：

①语音识别算法：发行人将语音识别算法运用于 A. 检验语音识别训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别算法模型进行训练，测试训练效果；B. 执行自动化数据加工、质检环节，程序化操作结合人工检查提升训练数据生产效率——通过语音识别算法对智能语音数据进行预标注，结合人工检查，降低人工工作量、提高数据处理能力；C. 直观展示语音识别训练数据的训练效果，为客户提供评判质量优劣的直观标准——运用不同训练数据集对同一语音识别算法模型进行训练，直观展现训练数据质量水平，方便客户选择更符合其预期的语音识别训练数据。

②语音合成算法：发行人将语音合成算法运用于 A. 检验语音合成训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别

<sup>26</sup> 上述数据来自发行人实验室模拟运算分析。

算法模型进行训练，测试训练效果；B. 运用语音合成算法模型展示训练数据的合成效果、帮助客户选择合意的训练数据——运用不同训练数据集对自有语音合成进行训练和效果测试，使得客户可以直观听取合成样音，进而选择风格、音色更符合要求的语音合成训练数据。

③计算机视觉算法：发行人将计算机视觉算法运用于训练数据生产过程的降本增效之中，通过将计算机视觉算法技术嵌入至自主研发的分布式标注平台之中，全面提升了计算机视觉领域的原料数据加工效率。例如运用计算机视觉算法对采集的原料数据短视频进行批量识别检查，确认是否存在采集内容重复的情况；以此为基础再进行人工排查。

## (3) 大数据和算法在训练数据生产过程情况的具体说明

类别	大数据、算法在训练数据生产各环节的具体应用情况				运用举例及对数据处理能力、数据处理效率提升的量化指标说明 <sup>27</sup>
	设计	采集	加工	质检	
大数据	<p>大数据的运用主要表现在，发行人基于生产经营积累的大数据，总结出多项技术诀窍（know-how），助力发行人设计训练数据集结构、为生产优质的训练数据集奠定基础：</p> <p>1、发音预测技术：发行人基于大数据积累了核心技术“语音语言学基础研究”项下的词典与模型的发音预测技术，为后续实现智能语音训练数据集的音素均衡奠定基础。</p> <p>2、语料清洗相关技术：发行人基于大数据积累了拼写检查、敏感词过滤、语种预测等多项技术并持续优化，运用于设计环节中的语料清洗步骤。</p> <p>3、文本易读性评判相关技术：基于长期积累的录音采集过程中的行为日志，积累上下文易读性打分技术，可对设计的语料的易读性进行评判，提升语料易读性、降低发音人朗读错误概率，提升训练数据集质量。结合发行人的“个性语料获取方法及个性语料获取装置”这一专利技术，可围绕特定发音人的基础属性、行为数据，进一步提升语料适配发音人的能力、提升训练数据集录制质量。</p>	无	<p>大数据的运用主要体现在：发行人大数据积累，开发训练数据生产工具、优化生产过程，实现产品质量的提升和生产成本降低：</p> <p>1、数据加工工具开发：以自有训练数据产品积累对初始算法模型进行训练，将前述算法模型运用至预标注过程汇总，结合后续人工标注过程中的修改痕迹对算法模型进行迭代优化，逐步得到一整套完整的数据加工工具群。</p> <p>2、生产过程中的任务分配优化：基于生产过程中的日志数据，结合日常生产管理的反馈信息，辅助优化数据加工任务类型的划分、数据任务包的分配和标注人员的管理，提升生产效率。</p> <p>3、行为模式分析驱动工具、平台优化：基于生产过程中的日志数据，分析一体化平台上的操作行为模式，辅助优化工具界面和交互，打造更优的工具集，进一步提升生产效率。</p>	<p>大数据在质检环节的运用主要体现在“分布式高性能自动校验技术”的运用之中，包括以下方面：</p> <p>1、多质检点高效质检：发行人的质检平台可实现多个质检点的自动高效质检，前述功能的实现即来源于：发行人通过分析大数据中的内在规律，基于统计学习和分布预测技术，找到离群点，从而对数据瑕疵进行定位。</p> <p>2、标注人员反常行为数据检测：质检平台上的部分质检点会对数据加工过程中记录的标注人员行为日志大数据进行分析，从而发现“反常”的数据处理行为，以判断其关联的数据是否存在质量问题。</p>	<p>发行人通过大数据分析积累，开发了 25 个可运用于标注过程之中的预测算法或工具，提升了数据加工过程的效率及准确率。例如：将中文词性预测工具的准确率逐步从 86% 提升至 96%；运用音素边界预测工具简化标注人员的人力成本；基于语音特征的韵律预测技术相比行业通用的基于文本的韵律预测技术降低了 25% 的人工标注工作量。</p>

<sup>27</sup> 该部分量化指标均来自发行人测试结果。

类别	大数据、算法在训练数据生产各环节的具体应用情况				运用举例及对数据处理能力、数据处理效率提升的量化指标说明 <sup>27</sup>
	设计	采集	加工	质检	
语音识别算法	训练效果模拟、指导设计： 检验语音识别训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别算法模型进行训练，测试训练效果。	实时识别、检测采集情况： 将语音识别算法运用于实时采集质检工具中，实时比对录制的语音数据与原始设计的文本语料是否相符。	执行预标注、提升标注效率： 执行自动化数据加工、质检环节，程序化操作结合人工检查提升训练数据生产效率——通过语音识别算法对智能语音数据进行预标注，结合人工检查，降低人工工作量、提高数据处理能力，例如： 1、智能语音训练数据集的音字一致性检查：音字一致性即训练数据集中语音文件和文本文件的一致性，是衡量训练数据集质量的重要指标。因此需要在加工环节对音字一致性进行多轮校对。发行人将语音识别算法模型引入上述校对环节，运用语音识别算法将语音文件转换为对应的文字，并与文本文件核对，提高了校对工作的准确率和效率。 2、嵌入运用至加工环节的几个核心技术中，进行场景适配和修改创新，如发行人的专利技术：基于语音特征的韵律预测技术等。	自动化质检运用： 将语音识别相关算法中的部分技术嵌入运用至质检技术中，以适配多个质检点的质检任务。例如： 质检环节中运用了相似说话人技术，从而筛选数据集中同一个说话人可能录制的多份文本，该技术将语音识别算法的方法论运用其中。	在运用了实时语音识别算法的采集项目中，原始语音数据的准确性（语音文件与原始设计文本的相同率）相比不包含实时语音识别的项目的准确率提高了 15.9%。
语音合成算法	无	无	无	自动检测数据质量： 发行人发现语音合成数据库的音色一致性和基频稳定性对于最终算法训练效果的影响很大，故定向研发了相关质检工具，进一步提高了语音合成训练数据的质量	无
计算机视觉算法	无	无	计算机视觉算法在发行人加工环节的应用主要为： 1、部分标注工作的自动处理：运用	自动检测数据质量： 发行人运用人脸识别相关计算机视觉算法技术研发了短	运用于人脸数据的重复性检查，相比人工抽检的方式，运用了计算机视觉

类别	大数据、算法在训练数据生产各环节的具体应用情况				运用举例及对数据处理能力、数据处理效率提升的量化指标说明 <sup>27</sup>
	设计	采集	加工	质检	
			<p>于原料数据的脱敏处理之中：当所采集的图片或者视频数据的需求和特定人群无关时，发行人会利用计算机视觉算法自动对原始数据中采集到的人脸进行快速定位，并执行脱敏（图片模糊化）处理。</p> <p>2、数据预测、辅助人工标注：发行人深度结合通用计算机视觉算法和具体数据加工场景，研发了数据预测工具，运用至加工环节中，对人工标注提供辅助，提升数据质量。</p>	<p>视频数据中相似人脸检查技术，可用于检查、保证计算机视觉训练数据集中人脸数据的独立性、提升训练数据质量。</p>	<p>算法的处理效率提升约96%。</p>

## 2、大数据、算法对训练数据集训练效果检查方式说明

发行人运用 AI 算法检查训练数据集的训练效果的具体方式包括：

(1) 语音识别算法：基于开源或自有的语音识别算法，以发行人生产的语音识别训练数据集中 90% 的数据作为训练数据，对该算法模型进行训练，再以同一训练数据集中剩余的 10% 数据作为测试数据集，检验前述经过训练的算法模型对该 10% 数据进行语音识别的正确率情况，从而对语音识别训练数据集的质量进行检查和验证。

例如，发行人运用不同语种的训练数据产品对同一语音识别算法模型进行了训练，通过算法模型训练结果直观展示发行人的训练数据产品具备较优的测试效果，各语种训练数据训练后的语音算法模型的字符错误率/词错误率均低于 5%，反映了发行人训练数据产品较好的训练效果。具体训练效果情况可参见招股说明书“第六节 业务和技术/六、发行人的研发和技术/（四）各项核心技术具体情况/1、语音识别算法”中相关介绍。

(2) 语音合成算法：与语音识别算法类似，发行人运用语音合成算法模型，基于其生产的语音合成训练数据做了大量的算法模型训练、语音合成效果验证和测试。以发行人生产的语音合成训练数据集对语音合成算法模型进行训练，通过样音合成效果检验数据集对语音合成算法模型的训练效果。

## 3、发行人的数据处理能力、数据处理效率情况

经过多年经营，发行人已形成了覆盖多应用领域、多应用场景、多语种的产品和服务体系，具备可覆盖智能语音、计算机视觉和自然语言 3 个应用领域，160 余个语种/方言（包括多个小语种）的数据处理能力。发行人数据处理能力、数据处理效率情况如下：

整体数据处理能力指标	定量描述	说明
应用领域覆盖能力	3 种	智能语音领域、计算机视觉领域、自然语言领域
语言覆盖能力	160 余个语种/方言	-
小语种覆盖能力	多种	南非荷兰语、拉脱维亚语、斯洛文尼亚语等

通过大数据、算法并用和驱动，发行人开发了多项与训练数据生产相关的核心技术，可运用于数据采集、标注过程中，提升了数据处理能力和效率。报告期内发行人数据处理能力、效率量化指标如下：

应用领域分类	数据处理能力/数据处理效率定量描述情况
语音识别	报告期内平均每月采集近 4,000 人语音数据 每月最高可处理 2 万小时语音数据
语音合成	报告期内平均每月采集近 150 万字/词语音数据 每月最高可处理 18 万句语音数据
计算机视觉	每月最高可在图片数据中标注超过 500 万个目标对象
自然语言	该类业务占比较小、非标准化、类型多样 发行人每月标注自然语言数据可达 2 万工作小时

#### （六）发行人核心技术与通用技术的差异情况说明

##### 1、发行人的核心技术以通用技术为理论基础，经持续的原始创新、自主研发和实践积累形成，是发行人的特有技术

通用技术无法直接满足训练数据生产研发的生产经营需求，发行人以其为理论基础，进行了持续的原始创新和自主研发，形成自己的特有技术。

发行人生产经营所涉及的技术主要包括人工智能算法类技术、语音语言学基础研究、训练数据集设计等基础技术，训练数据生产过程中使用的直接技术群（设计、采集、加工、质检等环节的多项具体核心技术），以及训练数据研发生产过程所需的工具、平台等。

前述领域的通用技术为一般性的数据采集、处理环节提供了基础，但在训练数据领域，不同语种/方言、领域、场景的训练数据在设计、采集、加工、质检各个环节均具备特异性要求，需要运用在相应领域积累的技术诀窍(know-how)，例如在设计阶段，发行人就需要运用训练数据集设计技术，在设计中充分考虑基础语料的结构、内容比例，才能使得容量有限的训练数据集能够覆盖尽可能多的自然语言现象，如覆盖更多的语音、语气、语调情况，兼顾多种句子类型（陈述句、疑问句、感叹句等）和句长分布，考虑不同的发音习惯、语言特点，最终实现更好的语料覆盖效果，从而生产出的训练数据集能够帮助算法模型获得更好的训练结果。

## 2、发行人的特有技术与通用技术的差异点

发行人将多领域的技术创新性、有机性地整合在一起，并与研发生产中积累的实践经验相结合，构建了适应自身生产经营所需的多项特有技术，同时还在特定领域、场景对前述特有技术进行了综合性地创新运用，提升了训练数据开发过程各环节的生产效率。截至本招股说明书签署日，发行人已经取得 24 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，充分说明了发行人的技术实力和技术独特性。

发行人特有技术与通用技术的差异点具体如下：



发行人核心技术层次	发行人的核心技术	行业通用技术简述	发行人核心技术与行业通用技术的具体差异点	在各核心技术领域取得的发明专利情况
基础研究	语音识别算法	语音识别算法指的是 Automatic Speech Recognition, 是让机器能够“听懂”人类语音的技术, 通过机器自动将语音信号转换为文本, 是实现智能人机交互的关键技术之一。	与行业通用技术相比, 发行人核心技术的差异点主要体现在: 基于对训练数据各生产环节细分场景的理解, 发行人对通用的语音识别算法进行了自主深度调优, 充分适配并嵌入运用至训练数据生产环节的各个工具中, 提升数据质量和生产效率。在语音识别算法领域, 发行人通过原始创新取得了包括语音数据库质量预估技术在内的专利技术。	已取得 4 项发明专利授权
	语音合成算法	语音合成算法指的是 Text to Speech 或者 Speech Synthesis, 是让机器能够“说出”人类语音的技术, 能将任意文字信息转化为流畅的语音朗读出来, 相当于给机器装上了人工嘴巴。	发行人对通用的参数合成、基于神经网络的端到端等语音合成技术进行了自主深度调优和落地化开发, 形成了特有的核心技术, 运用至训练数据生产过程中。在语音合成算法领域, 发行人通过原始创新取得积累的专利技术可有效提升端到端算法对语音合成训练数据集的运用效果, 实现更好的训练成果、提升发行人训练数据集的质量。	已取得 2 项发明专利授权, 正在申请 1 项发明专利
	计算机视觉算法	计算机视觉指的是 Computer Vision, 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	发行人基于通用的计算机视觉算法进行了原始创新和落地研发, 积累了适用于训练数据生产过程的技术诀窍 (know-how)。发行人在该领域进行了大量原始创新, 掌握了可应用于训练数据生产过程中的核心技术, 具体如下: 1、发行人运用人脸识别相关计算机视觉算法技术研发了短视频中相似人脸检查技术, 在该细分领域取得了 3 项发明专利授权, 可用于检查、保证计算机视觉训练数据集中人脸数据的独立性、提升训练数据质量。 2、发行人创新性结合人脸识别相关的计算机视觉算法和语音处理算法, 研发了短视频音频聚类技术, 并在该技术领域已取得 1 项发明专利授权, 可对多模态训练数据集执行数据关联性、重复性检测, 为提升训练数据质量提供了特有技术手段。 3、在物体识别和场景分割等技术领域研发了多项可用于图像分割、图像标注等计算机视觉训练数据生产的核心技术。	已取得 4 项发明专利授权, 正在申请 1 项发明专利
	★语音语言学基础研究	语音语言学基础研究主要指各个语种的高质量发音词典的构建能力, 包括音素集设计、发音规则制定、常见词标音等细分技术。	发行人通过积累语音语言学基础研究成果, 积累了 160 余个语种/方言的发音词典构建能力, 特有核心技术包括: 成熟的发音词典构建技术和流程积累, 可针对不同语言的特点, 从语音识别训练数据集、语音合成训练数据的模型训练角度出发, 优化细分技术和算法模型。	无
	训练数据集设计技术	训练数据集设计技术主要涉及的通用技术包括语言模型相关技术和文本正则化技术。语言模型是对语句概率分布的建模, 用来衡量一句话是否合理存在的方式, 对语言	在训练数据的设计环节, 与通用技术相比, 发行人核心技术的差异点主要在于: 1、发行人的设计流程和自有算法融入了发行人在多年训练数据生产过程中积累的实践经验和技术诀窍 (know-how, 如关于音素均衡、语料主题、覆盖领域、发音人属性、设备特征、自然环境等因素分布、设计的经验化技术积累); 2、发行人对语言模型、文本正则化等通用技术自主进行了系统性实现与落地优化,	已取得 1 项发明专利授权

发行人核心技术层次	发行人的核心技术	行业通用技术简述	发行人核心技术与行业通用技术的具体差异点	在各核心技术领域取得的发明专利情况
		模型的研究和应用能帮助人工智能算法找到最可能出现的语句。文本正则化技术在语音合成领域的应用是将文本中的数字、符号、缩写等转换成语言文字的技术,为了消除文本在读音层面的歧义,以确保后续合成的语音读音准确。	将其嵌入自主开发的一体化数据处理平台中,使公司具备高质量训练数据集的设计能力。例如,在文本正则化技术领域,发行人可处理7种主流语言、26种不同种类非标准词;具备超过2,000万句文本的累计处理经验,积累了大量技术诀窍(know-how)和经验。	
平台工具	一体化数据处理平台	为发行人自主开发的训练数据生产过程使用的数据处理平台,无直接对应的通用技术。	该数据处理平台由发行人自主开发,整合了发行人多年的训练数据开发生产实践经验,完整覆盖训练数据生产环节,拥有统一的人员管理、统一的数据管理、统一的风格管理、统一的流程管理等,整合贯通了发行人各类核心算法、技术,高度适用发行人训练数据开发的实际业务场景,是技术与生产实践的高度结合。可支持大规模并发数据处理的能力、支持严格的数据安全和隐私保护(如统一的加密数据存储、数据传输,对所有数据操作人员试行统一、严格的账号和密码管理等),可运用生产过程中的实时数据和行为日志等大数据对算法进行即时反馈和优化,最大化算法对生产效率和数据质量的提升作用。	已取得3项发明专利授权,正在申请1项发明专利
训练数据生产	★多语种多模态训练数据设计技术	此技术点包括的行业通用技术一般包括:拼写检查、敏感词过滤、语种检测等基础的文本处理技术,和发音预测、音素均衡等基础的语音处理技术。	与通用技术相比,发行人核心技术的差异点主要体现在: 1、将拼写检查等十多种文本相关技术整合在自主研发的语料清洗工具中,并加入自主研发的易读性检测技术等重要模块,提高清洗后语料数据的质量;同时,发行人将前述语料清洗技术与发行人的语种/方言覆盖能力相结合,可支持约70种不同语种/方言的数据清洗。同时发行人将语音、视觉、文本等领域积累的丰富的训练数据集设计能力整合起来,融会贯通后形成了高质量多模态训练数据的设计能力。 2、训练数据集的设计是一个研究与工程相结合的综合问题,存在诸多实践所需的技术诀窍(know-how),发行人已在该领域积累的专利技术分别创新性地解决了:中英混合训练数据集的设计问题、最优句长分布与音素覆盖的问题、中文稀缺短句的设计问题、个性化语料设计以优化朗读效果的问题。	已取得4项发明专利授权
	多模态多通道数据采集技术	此技术点包括的行业通用技术一般包括:手机录音APP的开发、计算机连接麦克风进行录音等基本的语音采集技术,手机拍照及照片管理技术、专业摄像头拍摄采集技术、手写体与光学字符采集技术等图片视频类采集技术。	与行业通用技术相比,发行人核心技术的差异点主要体现在:有机整合了基础的单通道数据采集技术,解决了多设备数据冲突、多来源数据协调一致等关键技术难点并积累了2项专利技术,实现多通道(多录音设备、多视频设备等)同时采集的能力。	已取得2项发明专利授权及1项实用新型专利授权

发行人核心技术层次	发行人的核心技术	行业通用技术简述	发行人核心技术与行业通用技术的具体差异点	在各核心技术领域取得的发明专利情况
	★数据同步技术	此技术点仅适用于特定应用场景，行业多以人工标注的方式处理，一般不认为有行业通用技术。	数据同步性是衡量训练数据集质量的重要指标。在多通道数据采集过程中，发行人自主研发了1项专利技术，解决语音数据采集过程中多通道语音数据的一致性问题。发行人的专利音频对齐技术，可将数据对齐误差控制在1毫秒以内，并支持每小时处理100小时以上的双通道音频文件对齐需求。	已取得1项发明专利授权
	实时采集质控技术	此技术点包括的行业通用技术一般都是语音信号处理相关的技术，包括：声音能量检测、音频截幅检测、静音检测、信噪比检测等技术。	与行业通用技术相比，发行人自主研发的实时采集质控技术综合运用了声音能量检测、音频截幅检测、静音检测、信噪比检测等技术，解决了数据采集的实时反馈难点，可支持多个采集工具几乎零延时地对数据瑕疵给出即时反馈，保证原料数据的质量。	已取得1项发明专利授权及1项实用新型专利授权
	★大数据驱动的高效数据处理技术	此技术点涵盖了训练数据加工环节的众多算法、技术和工具，和这些相关的行业通用技术一般包括：统计学习、语音信号处理、SVM等基础技术，以及各语种拼写检查、词性预测、分词处理等常见应用技术。	<p>发行人积累的大数据主要是指：大量自有知识产权的训练数据产品积累、大量生产过程中的行为日志数据积累等。发行人基于前述自有大数据积累，对生产经营中的生产模式、流程进行针对性地优化，技术的差异点主要在于运用前述大数据训练自有算法模型、积累自有技术，实现数据生产过程的优化和产品质量、生产效率的提升，具体如下：</p> <ol style="list-style-type: none"> <li>1、以自有训练数据产品对初始模型进行训练，将前述算法模型运用于预标注过程中，结合后续人工标注过程的修改痕迹对前述模型持续迭代优化，逐步得到一系列实用的特定算法模型和特有工具。发行人自主开发的中文词性预测工具准确率经逐步优化已从86%提升至96%；音素边界预测工具简化了标注人员的人力成本；基于语音特征的韵律预测技术将人工标注工作量降低了25%。</li> <li>2、发行人对生产过程积累的日志数据和日常生产管理的反馈信息等大数据进行分析，从而辅助优化数据加工任务类型的划分、数据任务包的分配和标注人员的管理，提升生产效率。</li> <li>3、发行人基于生产过程中的自有日志数据积累，分析一体化数据处理平台上的操作行为模式，辅助优化工具界面和交互，打造更优的工具集，进一步提升生产效率。</li> </ol>	已取得4项发明专利授权
	★分布式高性能自动校验技术	此技术点相关的行业通用技术包括两类：一类是高性能分布式集群计算技术；一类是语音信号处理等基础技术。	<p>与行业通用技术相比，发行人的核心技术差异点主要在于自主研发了多种关键质检技术，并将其运行于分布式高性能质检平台上，具体包括：</p> <ol style="list-style-type: none"> <li>1、针对训练数据具体质检需求研发多种实用的算法技术，例如相似说话人筛查技术将人工处理工作量降低了98%以上；音素标注正确率校验技术缩减了音素错误标注的检测时间，可满足音素标注99%的准确率要求等。发行人在前述技术领域均已获得发明专利授权。</li> <li>2、发行人在通用的分布式集群系统上针对数据质检的特定场景和需求进行了特有的整体架构优化和升级；并将前述核心算法的计算功能以及大量质检工具部署到该集群上，已达到日质检660G训练数据的能力。</li> </ol>	已取得4项发明专利授权

### （七）发行人核心技术申请专利权情况及其他技术保护措施

发行人的核心技术可按基础研究、平台工具、训练数据生产三个层次划分，各项核心技术均运用于其训练数据生产过程中。按照前述三个层次，区分训练数据生产的各环节，发行人核心技术、技术名称和技术保护措施的具体对应关系如下：

核心技术层次	核心技术	核心技术项下的子技术名称	技术保护措施
基础研究	语音识别算法	语音数据库质量预估技术	已取得4项发明专利授权、2项计算机软件著作权
	语音合成算法	语音合成数据库评估技术	已取得2项发明专利授权、2项计算机软件著作权，正在申请1项发明专利
		说话人自适应语音合成技术	已取得1项计算机软件著作权
		语音合成系统评测技术	已取得1项计算机软件著作权
	计算机视觉算法	人脸检测和识别技术	已取得3项发明专利授权
		物体识别技术	已取得1项发明专利授权
		光学字符识别技术	已取得1项计算机软件著作权
		场景分割技术	已取得1项计算机软件著作权、正在申请1项发明专利
		行人检测技术	-
	运动轨迹跟踪技术	-	
	语音语言学基础研究	基于词典与模型的发音预测技术	已取得1项计算机软件著作权
	训练数据集设计技术	N元语言模型训练与优化技术	已取得1项计算机软件著作权
		文本正则化技术	已取得1项发明专利授权、2项计算机软件著作权
		基于语言模型的文本易读性评测技术	已取得2项计算机软件著作权
平台工具	一体化数据处理平台	基于C/S架构的大规模语音处理平台	已取得1项发明专利授权、5项计算机软件著作权
		基于C/S架构的音素标注技术	已取得2项计算机软件著作权
		多语言分布式文本处理技术	已取得5项计算机软件著作权
		基于C/S架构的文本词性标注、文本属性标注技术	已取得2项计算机软件著作权
		基于C/S架构和B/S架构的图像标注技术	已取得2项发明专利授权、4项计算机软件著作权，正在申请1项发明专利
		大规模数据采集及标注平台	已取得12项计算机软件著作权
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	已取得2项发明专利授权、1项计算机软件著作权
		多语种音素均衡语料设计技术	已取得1项发明专利授权、1项计算机软件著作权
		混合语言语料设计技术	已取得1项发明专利授权、3项计算机软件著作权

核心技术层次	核心技术	核心技术项下的子技术名称	技术保护措施
		AudioBook 自动切分与文本对齐技术	已取得 1 项计算机软件著作权
	多模态多通道数据采集技术	多通道录音技术	已取得 1 项实用新型专利授权、13 项计算机软件著作权
		移动设备上的语音采集技术	已取得 1 项发明专利授权、11 项计算机软件著作权
		分布式图片内容分析技术	已取得 1 项计算机软件著作权
		移动设备上的图片采集技术	已取得 3 项计算机软件著作权
		多语言手写体数据采集技术	已取得 3 项计算机软件著作权
		多模态数据采集技术	已取得 1 项发明专利授权、3 项计算机软件著作权
	数据同步技术	音频对齐技术	已取得 1 项发明专利授权、2 项计算机软件著作权
	实时采集质控技术	语音信号质量检测技术	已取得 1 项实用新型专利授权
	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	已取得 3 项发明专利授权
		音素边界自动预测技术	已取得 2 项计算机软件著作权
		基于语音特征的韵律预测技术	已取得 1 项发明专利授权
		基于 CRF 的韵律预测技术	已取得 1 项计算机软件著作权
		基于 HMM/CRF 的词性预测技术	-
		基于 SVM 的文本分类技术	-
		基于 CRF 的命名实体识别技术	-
		基于前后向最大匹配的中文分词技术	-
	分布式高性能自动校验技术	音素标注正确率校验技术	已取得 1 项发明专利授权
		相似说话人自动筛查技术	已取得 1 项发明专利授权
音字一致校对技术		已取得 2 项发明专利授权、4 项计算机软件著作权	

注：部分计算机软件著作权、专利同时涉及多项子技术，在上面各子技术下均进行了列示。

#### （八）公司技术是否成熟或存在快速迭代风险的情况说明

发行人已将其核心技术体系运用于设计、采集、加工、质检等训练数据生产相关的各个环节，构成了发行人业务持续经营的技术基础，并在报告期内保障了发行人业务规模和经营业绩的成长性。总体而言发行人的核心技术体系整体已较为成熟。

为了保持并加强公司的技术实力、提升核心技术竞争力，发行人采取了一系列战略举措、构建起技术创新机制，促进公司技术的持续升级和创新。发行人积

极通过文化建设营造鼓励创新的工作氛围，结合一系列机制保障开拓创新的工作环境，充分利用薪酬、奖励、表彰等多种方式激发员工的主观能动性，结合公司的主营业务，形成勇于开拓、积极创新的创新环境。报告期内，发行人每年平均完成 10 余次平台升级，涉及百余个功能新增拓展。

但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，发行人进行训练数据生产所需的各项技术也面临着快速迭代更新的风险。

## （九）研发人员和核心技术人员情况

### 1、核心技术人员、研发人员占员工总数的比例

截至 2020 年 12 月 31 日，公司现有研发人员 44 名，占公司员工总数的比例达 29.14%；技术人员 59 名，占比 39.07%。其中核心技术人员 6 名，占比 3.97%。

### 2、核心技术人员的认定情况及认定依据

发行人制定了《海天瑞声岗位管理办法》，将核心技术人员定义如下：核心技术人员为“对公司战略目标和核心技术竞争优势有关键影响力的管理人员、公司核心技术研发的主要负责人、研发成果及知识产权的主要发明人”。

发行人根据上述认定标准认定了 6 名核心技术人员，分别为唐涤飞（董事、总经理）、李科（董事、副总经理）、黄宇凯（董事、副总经理、技术总监）、郝玉峰（副总经理、技术总监、首席科学家）、曹琼（计算机视觉业务部总监）及邵志明（高级研发工程师）。

前述核心技术人员中，唐涤飞全面领导公司的研发工作；李科、黄宇凯主要负责领导及实施一体化数据处理平台相关的研发工作以及全时日志库和全生命周期数据仓库重要模块的研发；黄宇凯、郝玉峰负责领导公司技术研发工作，郝玉峰同时负责领导公司语音语言学基础研究工作；曹琼负责领导公司计算机视觉领域训练数据产品和服务的开发及相关技术研发；邵志明主要负责智能语音数据相关研发工作以及训练数据生产过程中相关软件、工具的开发。

前述核心技术人员均为计算机、信息工程、自动化等领域的专业人士，具备多年从业经验。核心技术人员的专业背景，及其对公司研发的具体贡献（在公司

的产品研发、技术研究、专利和软件著作权积累等方面发挥的具体作用)请参见下文“3、核心技术人员基本情况及对公司研发的具体贡献”。

### 3、核心技术人员基本情况及对公司研发的具体贡献

姓名	职务	专业背景和对公司研发的具体贡献
唐涤飞	董事、总经理	拥有 20 余年人工智能和软件行业的研发和管理经验,领导跨国团队在语音语义、数据应用和移动平台等技术、产品、服务取得重要突破,是 6 项美国、中国专利的发明人。加入海天瑞声后,带领团队打造了以基础算法研究、数据处理平台,算法技术与工具平台相结合验证数据库效果为核心的公司技术战略布局,确立了公司训练数据产品在业内的竞争优势。
李科	董事、副总经理	主导研发了公司一体化数据处理平台。此平台拥有超过 300 多个技术工具或模块,实现了数据开发的智能化和高效率。在此平台上,发行人运用前述平台、工具为客户提供超过超 3,600 次/个训练数据定制服务或训练数据产品,涵盖智能语音、计算机视觉、自然语言等领域,支持 160 余个语种或方言,处理了超过 90 万小时语音时长,超过 5,000 多万张图片、2,500 多万行文本数据。组织、负责公司知识产权保护工作,带领公司研发、技术团队为公司取得 130 余项软件著作权。 是公司“语音库的语音获取方法、装置、电子设备及存储介质”、“端到端的语音合成方法、装置及存储介质”等共计 17 项已授权专利及 2 项在申请专利的共同发明人。
黄宇凯	董事、副总经理、技术总监	重构一体化数据处理平台的软件架构;组织开发并改进数据处理各环节所需工具以高效支撑智能语音、计算机视觉、自然语言处理等公司业务;遴选公司人工智能基础算法的研究成果并转化为软件系统使之应用于数据开发业务。 是公司“语音库的语音获取方法、装置、电子设备及存储介质”、“端到端的语音合成方法、装置及存储介质”等共计 17 项已授权专利、2 项在申请专利的共同发明人。
郝玉峰	副总经理、技术总监、首席科学家	负责通过语音识别、语音合成、自然语言处理、图像处理等 AI 技术手段,提高公司训练数据设计生产的工作效率,提升训练数据质量,主持或参与了 30 余项软件著作权的研发工作,任职期间设计研发多语种语料清洗和语料设计流程与相关算法;基于 VAD 技术和 SNR 计算技术的录音质量监控技术、基于 ASIO 的多通道语音采集软件、以及基于 ASR 的语音自动切分系技术等信号处理技术。以上算法和技术被广泛地应用于公司数据设计、开发和质量控制流程,构成了公司一体化数据处理平台的主要技术来源。是中国中文信息学会语音信息专业委员会委员、中国人民大学信息学院工程硕士导师。 是公司“音素误标注的检测方法和装置”、“韵律标注方法、装置和设备”等共计 24 项已授权专利、2 项在申请专利的共同发明人。
曹琼	计算机视觉业务部总监	设计并实现使用计算机视觉数据自动检测算法与人工标注结合的人机交互数据生产方案,主导开发了计算机视觉数据处理通用平台,并在实际项目中得到应用,提高了生产

姓名	职务	专业背景和对公司研发的具体贡献
		效率，保障了数据的隐私安全性。其成果被应用于公司计算机视觉业务中，成为公司竞争优势之一。是“目标检测方法及设备”、“图像标注方法与装置”等 14 项中国、美国及日本专利发明人。 是公司“视频中音频聚类的处理方法和装置”等共计 18 项已授权专利、2 项在申请专利的共同发明人。
邵志明	高级研发工程师	从事语音信号处理、语音合成、说话人识别、语种识别、自然语言处理等技术研究，其主导研发的语音多通道对齐技术和 ASR 唤醒词训练数据自动转写系统，在相应的业务上节省了三分之一的人工成本。 是公司“个性语料获取方法及个性语料获取装置”、“短语语料获取方法及短语语料获取装置”等共计 7 项已授权专利、1 项在申请专利的共同发明人。

#### 4、对核心技术人员的约束和激励措施

##### (1) 发行人对核心技术人员的约束措施

发行人制定了完善的保密管理制度，对包括技术机密在内的各类保密信息及材料制定了相应的保密要求，并制定了保密教育、泄密惩治机制等管理制度。同时发行人与包括核心技术人员在内的全体员工均签署了《保密协议》，并与核心技术人员均签署了《竞业禁止协议》，对核心技术人员在职期间和离职后一定年限内的保密、竞业和侵权事项进行了严格约定。

##### (2) 发行人对核心技术人员采取的激励措施

为了保障持续创新能力，发行人对核心技术人员采取了全方位、多元化的激励制度：

①创新环境建设：积极通过企业文化和制度建设，营造鼓励创新的工作氛围、保障开拓创新的工作环境，为核心技术人员提供鼓励创新、有利创新的工作氛围和文化环境。

②科研激励制度建设：制定以《海天瑞声科研激励计划》为核心的科研激励制度，激励核心技术人员及其他员工持续开展科研创新工作，鼓励他们在专利、技术、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。

③股权激励体系建设：发行人设立了中瑞安、中瑞立两个员工持股平台，通过授予员工持股平台的合伙份额的形式对核心技术人员实施股权激励。上述激励



制度使得发行人的核心技术人员得以持有公司股份，享受公司业绩增长带来的增值收益，有助于维持公司核心技术团队稳定、保持公司核心竞争力。

## 5、报告期内核心技术人员的变动情况

截至本招股说明书签署日，发行人的核心技术人员为唐涤飞、李科、黄宇凯、郝玉峰、曹琼和邵志明。除黄宇凯为 2018 年新增核心技术人员外，其他核心技术人员自报告期初即在公司任职且未发生变动。核心技术人员的持续加入，提高了公司研发创新能力，促使公司持续、健康发展。

### （十）科研实力和成果

经过多年积累，发行人先后积累了以下知识产权、荣誉资质及产学研协作成果：

#### 1、知识产权

截至 2021 年 5 月 7 日，发行人已取得 26 项专利授权（其中发明专利 24 项、实用新型专利 1 项、外观设计专利 1 项），另有 2 项专利正在申请之中。

#### 2、奖项、荣誉及认证、称号情况

截至本招股说明书签署日，发行人及其产品服务所获奖项、荣誉或认证、称号情况如下：

序号	名称	颁发单位	颁发时间
1	《人工智能训练数据资源全生命周期风险识别与管理应用案例》 人工智能企业典型应用案例	中共中央网络安全和信息化委员会办公室	2021 年
2	新一代人工智能产业创新重点任务揭榜优胜单位	工业和信息化部	2021 年
3	国家专精特新“小巨人”企业	工业和信息化部中小企业局	2020 年
4	北京市专精特新“小巨人”企业	北京市经济和信息化局	2020 年
5	北京市新技术新产品（服务） <sup>28</sup>	北京市科委、北京市发展改革委等六家单位联合颁发	2020 年
6	北京市级企业科技研究开发机构	北京市科学技术委员会	2019 年
7	新一代人工智能产业创新重点任务入围揭榜单位	工业和信息化部	2019 年

<sup>28</sup>发行人产品/服务：uTrans 在线语音转写标注质检平台服务、King-Lexicon-051 上海话发音词典、King-ASR-384 中文口音语音识别系列数据库获评北京市新技术新产品（服务）称号。

序号	名称	颁发单位	颁发时间
8	2019 中国信息技术人工智能行业优秀产品	中国信息协会、信息化观察网	2019 年
9	2019 中国信息技术人工智能数据服务领军企业	中国信息协会、信息化观察网	2019 年
10	2019 中关村高成长企业 TOP100	中关村高新技术企业协会	2019 年
11	腾讯 2019 年服务类供应商“杰出贡献奖”	深圳市腾讯计算机系统有限公司	2019 年
12	菜鸟网络“驼峰计划”突出贡献奖	菜鸟网络	2019 年
13	GB/T 22080-2016/ISO/IEC27001:2013《信息技术 安全技术 信息安全管理体系要求》标准认证	北京赛西认证有限责任公司	2019 年
14	中国人工智能产业发展联盟理事单位	中国人工智能产业发展联盟	2018 年
15	国家规划布局内重点软件企业	国家发展改革委、工业和信息化部、财政部、商务部、国家税务总局	2018 年
16	中关村高新技术企业	中关村科技园区管理委员会	2018 年
17	国家高新技术企业	北京科学技术委员会、北京市财政局、北京市国家税务局和北京市地方税务局联合颁布	2018 年
18	中关村高新技术企业协会理事单位	中关村高新技术企业协会	2018 年
19	信息化影响中国·2018 年人工智能行业最佳产品	中国信息协会、信息化和软件服务网	2018 年
20	2018“中国声谷”杯智能语音创新大赛行业先锋奖	中国语音产业联盟	2018 年
21	2018 中关村高成长企业 TOP100	中关村高新技术企业协会	2018 年
22	中国语音产业联盟理事会员单位	中国语音产业联盟	2017 年
23	2017 中国信息技术年度人工智能数据服务领域领军企业奖	中国信息协会、信息化观察网	2017 年
24	25 Most Recommended Artificial Intelligence Companies 2017(2017 年度 25 家最具推荐价值人工智能企业)	美国 APAC CIO Outlook 杂志	2017 年
25	Oriental COCOSDA 2016 ITN Best Paper Award	Oriental COCOSDA	2016 年

### 3、产学研（生产、教育、科研）协作

发行人积极参与产学研协作，促进科研、教育、生产的协同与集成化，推动科技创新和技术合作：发行人现任中国人工智能产业发展联盟、中国语音产业联盟理事，国家人工智能标准化总体组成员，中国电子工业标准化技术协会《信息技术-人工智能-面向机器学习的数据标注规程》标准参与起草单位；与北京语言

大学合作建立“海天瑞声-北京语言大学汉语学院联合实验室”；与清华大学就“语音数据库构建与验证平台”开展技术开发合作（详见下文“4、合作研发”）等。

#### 4、合作研发

报告期内，发行人的合作研发项目为与清华大学就“语音数据库构建与验证平台”的技术开发合作，主要情况如下：

##### （1）协议主要内容

项目目标：针对语音识别、语音合成训练数据领域，构建训练数据的验证规范及技术，建立验证平台，为语音训练数据提供质量验证。

期限：2019年11月12日至2021年12月31日。

##### （2）权利义务划分及技术成果归属约定

在合作形式上，由发行人提供开发工作所需经费、设备、相关数据及资料，清华大学提供人员、研究开发的技术与方法，双方发挥各自优势开展合作。清华大学向发行人交付的开发成果包括含有软件代码的载体和相关文档。发行人和清华大学将共同撰写论文或申请专利。利用研究开发经费，用于清华大学进行技术设计和实验而购置的设备、器材和资料归清华大学所有。

在本合作研发项目产生的全部研究开发成果及阶段性成果的知识产权均归双方共同所有。双方均有权对开发成果及阶段性成果进行修改及二次开发，由此产生的后续开发成果属于开发方所有。清华大学在向发行人交付研究开发成果之前，不得自行将研究开发成果转让予发行人以外的其他第三方；若一方欲转让开发成果均需首先通知对方，在同等条件下，另一方享有优先受让权。发行人有权使用项目下产生的开发成果及阶段性成果，后续继续用于研发、制作、销售、服务，无需向清华大学支付任何费用。

##### （3）采取的保密措施

发行人、清华大学在协议中约定了双方针对技术情报、资料等的保密义务及赔偿责任，主要包括：约定双方在讨论、订立及履行协议的过程中均承担保密义务，禁止向第三方传播保密内容；约定双方妥善保管所获信息，不得擅自复制，防止因管理不善造成信息泄露；除因合作产生的成果载体外，合作期间取得的对

方的技术秘密的其他载体都要归还对方或予以销毁。任何一方因违反保密条款而造成对方秘密信息泄露，都应承担由此引起的法律责任，并赔偿因此给对方造成的损失。保密条款不因合作协议的终止而失效。

### （十一）在研项目情况

在多年的行业实践及经营过程中，发行人高度重视自身的核心技术优势的挖掘和培养，秉承基础研究与实际应用紧密结合的原则，持续研究、开发、完善训练数据生产所需的基础工具和管理平台。截至招股说明书签署日，发行人正在进行的基础工具和平台研发情况如下：

金额单位：万元

在研项目名称	技术类型	所处阶段	进展情况	研发目标（拟达到的目标）	研发人员数量（人）	截至2021年3月31日已投入的研发支出	研发项目与行业技术水平的比较情况
海天瑞声OCR在线生产平台二期	计算机视觉	产品开发	升级维护中	优化对光学字符体识别（OCR）数据的采集、标注和质检功能	8	338	进一步优化OCR数据库的生产过程，提升工作效率15%
支持无人车场景的语义分割平台	计算机视觉	产品开发	产品开发过程中	支持无人车场景下的图形数据的基础标注、语义分割、数据质检及流程管理等功能	11	181	覆盖无人车场景的语义分割数据的生产过程，提升工作效率35%
支持语音数据库的语料标注在线平台	语音	产品开发	产品开发过程中	支持语音数据库制作过程中的语料编辑、处理、质检及流程管理等功能	9	145	提升语料生产标注效率15%

### （十二）研发支出情况

公司自成立以来，始终坚持以技术创新引领产品和服务创新，不断完善研发体系建设，丰富技术、产品和服务模式，持续打造公司核心竞争力。2018年至今，公司不断加大研发投入，2018年、2019年及2020年公司研发费用分别为2,734.53万元、4,169.61万元及4,349.75万元，呈逐年上升趋势。报告期内，公司的累计研发投入达到11,253.89万元，累计研发投入占收入比重为16.96%。

报告期内，发行人的研发费用主要由训练数据产品开发相关支出及基础研发支出两部分构成。训练数据产品开发相关支出主要为开发训练数据产品对应的数据服务费及部分职工薪酬等，该部分支出未计入营业成本，详见本招股说明书

“第八节 财务会计信息与管理层分析/九、经营成果分析/(四)期间费用分析/3、研发费用分析”。基础研发支出主要为发行人研究开发训练数据开发生产所需的算法、技术以及工具、平台等的支出。2018年、2019年及2020年，发行人基础研发支出分别为1,469.94万元、1,951.79万元及2,127.07万元，基本呈逐年上升趋势；报告期内累计基础研发投入合计达5,548.80万元，占报告期营业收入比例为8.36%。

报告期内，发行人整体研发费用、基础研发费用及其占营业收入比例如下：

项目	2020年	2019年	2018年
研发费用（万元）	4,349.75	4,169.61	2,734.53
研发费用/营业收入	18.64%	17.55%	14.19%
报告期平均占比	16.96%		
基础研发费用（万元）	2,127.07	1,951.79	1,469.94
基础研发费用/营业收入	9.11%	8.22%	7.63%
报告期平均占比	8.36%		

注：报告期内平均占比=报告期内研发费用合计/营业收入合计

报告期内，发行人整体研发费用构成情况如下：

单位：万元

项目	2020年度		2019年度		2018年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	1,989.22	45.73%	1,858.94	44.58%	1,374.24	50.25%
数据服务费	1,844.91	42.41%	1,844.97	44.25%	957.04	35.00%
语言研究	144.78	3.33%	182.04	4.37%	175.46	6.42%
折旧与摊销	139.19	3.20%	128.12	3.07%	69.86	2.55%
房租费	86.63	1.99%	74.34	1.78%	75.33	2.75%
其他	145.02	3.33%	81.19	1.95%	82.61	3.02%
<b>合计</b>	<b>4,349.75</b>	<b>100.00%</b>	<b>4,169.61</b>	<b>100.00%</b>	<b>2,734.53</b>	<b>100.00%</b>

报告期内，发行人的研发支出主要包括研发相关的职工薪酬支出，训练数据产品生产过程中的原料数据采集、标注相关的采购支出，以及少量折旧摊销、语言研究支出、房租支出等。研发费用详细分析请参见本招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（四）期间费用分析/3、研发费用分析”。

### （十三）技术创新机制及技术创新相关安排

为了保持并加强在训练数据领域的优势，发行人采取了一系列战略举措、构建起技术创新机制，促进公司技术和产品的持续创新，有机结合行业前沿技术革新、下游客户新产品和新技术的研发及公司自身训练数据业务的三大创新要素，实现创新目标，形成产业互动与协同。同时，发行人积极通过文化建设营造鼓励创新的工作氛围，结合一系列机制保障开拓创新的工作环境，充分利用薪酬、奖励、表彰等多种方式激发员工的主观能动性，结合公司的主营业务，形成勇于开拓、积极创新的创新环境。发行人技术创新机制及技术创新相关安排具体如下：

#### 1、保持科研人才的持续研发活力

##### （1）建立健全创新奖励机制

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的原创性工作成果。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。

##### （2）明确工作成果的评价机制

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；对于数据制作类项目，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。

##### （3）加强培训、促进交流合作

发行人创建了导师培养机制及核心员工培养计划，通过持续开展技术交流讲座，组织员工参与内部培训、外部培训等方式，不断拓展员工的眼界，提高员工创新性发现问题和解决问题的能力。

#### 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，数据中心、语言研究部、质量保证部、市场部全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软件服务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) 时刻关注行业技术的发展

发行人每年均派出核心技术人员和市场人员参加人工智能技术领域的国际知名会议,包括 Interspeech、ICASSP、APSIPA、ICCV、CVPR、LREC、O-COCOSDA、NCMMSC、SpeechTek 等,关注技术与算法发展,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。

### (十四) 发行人符合科创板定位情况说明

根据《科创属性评价指引(试行)》和《科创板企业发行上市申报及推荐暂行规定》的科创属性要求,发行人符合科创板定位。具体说明如下:

#### 1、发行人符合科创板行业领域要求

公司所属行业领域	<input checked="" type="checkbox"/> 新一代信息技术	根据国家统计局《战略性新兴产业分类(2018)》,发行人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
	<input type="checkbox"/> 高端装备	
	<input type="checkbox"/> 新材料	

	<input type="checkbox"/> 新能源	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。
	<input type="checkbox"/> 节能环保	
	<input type="checkbox"/> 生物医药	
	<input type="checkbox"/> 符合科创板定位的其他领域	

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。根据《上市公司分类与代码》（2012年修订），发行人所属行业为软件和信息技术服务业。根据国家统计局《战略性新兴产业分类（2018）》，发行人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。因此，发行人属于《科创板企业发行上市申报及推荐暂行规定》中“新一代信息技术领域”中的战略性新兴产业，符合科创板行业领域要求。

## 2、发行人符合科创属性要求

逐条对照科创属性评价指引，海天瑞声符合科创属性指标要求情况具体如下：

科创属性评价标准一	是否符合	发行人指标情况
最近三年累计研发投入占最近三年累计营业收入比例 $\geq 5\%$ ，或最近三年累计研发投入金额 $\geq 6,000$ 万元	<input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否	最近三年累计研发投入占最近三年累计营业收入比例： 16.96%
研发人员占当年员工总数的比例 $\geq 10\%$	<input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否	2020年研发人员数量占员工总数比例为29.14%。
形成主营业务收入的发明专利（含国防专利） $\geq 5$ 项	<input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否	形成主营业务收入的发明专利：24项
最近三年营业收入复合增长率 $\geq 20\%$ ，或最近一年营业收入金额 $\geq 3$ 亿元	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	最近三年营业收入复合增长率：10.06%
科创属性评价标准二	是否符合	发行人情况说明
拥有的核心技术经国家主管部门认定具有国际领先、引领作用或者对于国家战略具有重大意义	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	-
作为主要参与单位或者发行人的核心技术人员作为主要参与人员，获得国家科技进步奖、国家自然科学奖、国家技术发明奖，并将相关技术运用于公司主营业务	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	-
独立或者牵头承担与主营业务和核心技术相关的国家重大科技专项项目	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	-
依靠核心技术形成的主要产品（服务），属于国家鼓励、支持和推动的关键设备、关键产品、关键零部件、关键材料等，并实现了进口替代	<input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否	参见下文论述
形成核心技术和主营业务收入的发明专利（含国防专利）合计50项以上	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	-



发行人符合“科创属性评价标准二”第四条“依靠核心技术形成的主要产品（服务），属于国家鼓励、支持和推动的关键设备、关键产品、关键零部件、关键材料等，并实现了进口替代”的相关要求。对发行人符合科创属性评价标准二（四）相关要求的情况具体说明如下：

（1）发行人依靠核心技术形成的主要产品/服务，属于国家鼓励、支持和推动的关键设备、关键产品、关键零部件、关键材料等

①根据国家战略、产业政策，发行人依靠核心技术形成的主要产品/服务，属于国家鼓励、支持和推动的关键产品

人工智能是国家战略、产业政策大力支持的关键技术领域，多个国家部门在该领域频频出台政策，推进人工智能关键技术和人工智能训练数据资源建设。发行人提供的训练数据服务于人工智能关键技术的训练测试及领域拓展进程，是人工智能技术发展的关键产品。

工业和信息化部在关于《促进新一代人工智能产业发展三年行动计划（2018-2020年）》（以下简称“行动计划”）的解读中指出：目前我国人工智能发展的痛点问题之一就是缺少有效的行业资源训练库等公共服务支撑体系，业界普遍反映已经影响了人工智能技术发展及在行业中的应用；《行动计划》注意到了这一关键问题，加大对产业公共服务平台的支持，将形成有效引导，不断完善产业发展环境。工业和信息化部相应制定了《新一代人工智能产业创新重点任务揭榜工作方案》，提出“选择高质量的行业训练资源库……等作为揭榜攻关任务”，并指出“这些资源体系是影响人工智能健康发展的重要要素，需要加快完善基础环境、保障平台，加快形成我国人工智能产业创新发展的支撑能力”。国务院《关于积极推进“互联网+”行动的指导意见》提出建设海量训练数据资源库，进而推进各项人工智能关键技术的研发和产业化。国家发展改革委、科技部、工业和信息化部、中央网信办出台的《“互联网+”人工智能三年行动实施方案》（以下简称“实施方案”）则将加快建设海量训练数据资源库认定为“互联网+”人工智能三年行动实施方案的主要任务之一，整体服务于“突破人工智能关键技术”的总体思路。

结合上述产业政策，发行人提供的训练产品/服务从属于产业政策拟大力发

展的人工智能训练数据资源库范畴，是建设海量训练数据资源库、突破人工智能关键技术这一产业任务的主要构成部分。发行人的训练数据产品响应于国家发展人工智能关键技术的战略意图，是国家鼓励、支持和推动的关键产品。

依据上述政策，发行人于 2019 年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”名单并于 2021 年 4 月揭榜优胜，根据工业和信息化部《新一代人工智能产业创新重点任务揭榜工作方案》，揭榜优胜单位应是掌握关键核心技术、具备较强创新能力以突破产业发展的短板和瓶颈的创新主体。同时，公司于 2020 年入选工业和信息化部第二批专精特新“小巨人”企业名单，根据工业和信息化部《关于开展第二批专精特新“小巨人”企业培育工作的通知》，专精特新“小巨人”企业主导产品应符合该《工业“四基”发展目录》所述核心基础零部件、先进基础工艺、关键基础材料；或“重点领域”相关产品、产业链供应链关键环节及关键领域“补短板”产品或国家/省重点鼓励发展的支柱优势特色产业领域。以上由国家部委授予的荣誉、认证均验证了发行人掌握关键核心技术，具备较强的科创能力，发行人提供的产品/服务是突破产业发展、国家战略鼓励的关键产品。

②根据国家主管部门证明及行业专家评定，发行人依靠核心技术形成的主要产品/服务，属于国家鼓励、支持和推动的关键产品

为贯彻落实党的十九大精神，加快我国人工智能产业创新发展，按照国务院《新一代人工智能发展规划》（国发〔2017〕35号）和《促进新一代人工智能产业发展三年行动计划（2018-2020）》（工信部科〔2017〕315号）部署，工信部于 2018 年底启动了新一代人工智能产业创新重点任务揭榜工作，旨在选拔一批在我国人工智能主要细分领域掌握关键核心技术，聚焦并能突破人工智能产业发展短板瓶颈，具有国际竞争力、领先性的标杆企业。

根据工信部科技司出具的《关于对北京海天瑞声科技股份有限公司相关情况说明的复函》工科函〔2021〕277号认定：“为加快推动我国新一代人工智能产业创新发展，我部于 2018 年底印发《新一代人工智能产业创新重点任务揭榜工作方案》（工信厅科函〔2018〕80号），聚焦人工智能产业发展重点领域，遴选培育掌握关键核心技术、具备较强创新能力的企事业单位，壮大发展产业力量。北京海天瑞声科技股份有限公司揭榜承担了“行业训练资源库”方向任务，经专

家评议和专业机构测评，成果达到预期指标，完成了揭榜任务，被我部确认为新一代人工智能产业创新重点任务揭榜优胜单位。”

此外，经中国信息通信研究院组织，共计 5 名中国工程院院士、中国科学院等科研机构正高研究员及教授、行业资深人士等行业专家对发行人产品/服务情况及技术水平进行了评定。经前述行业专家确认：“发行人所提供的训练数据已经成为国家鼓励、支持和推动的人工智能产业发展所必需的关键产品，是形成我国人工智能产业创新发展的关键性基础能力，是攻克我国人工智能产业基础发展瓶颈的关键一环。”

综上主管部门说明及行业专家评定，发行人在训练数据领域掌握了关键核心技术，具备较强的创新能力，突破了人工智能产业发展短板瓶颈，是具有国际竞争力、领先性的标杆企业，所提供的训练数据产品或服务是国家鼓励、支持和推动的人工智能产业发展所必需的关键产品。

## （2）发行人主要产品/服务的进口替代依据和技术先进性情况说明

### ①实现进口替代的发行人产品/服务主要为智能语音类训练数据产品/服务

发行人实现进口替代的主要产品/服务为智能语音类训练数据。在该产品/服务领域，发行人在国内企业中起步较早，并持续陪伴我国人工智能产业发展，为下游人工智能行业实现了智能语音训练数据的国产、独立、自主供应，避免海外进口冲击、达到了进口替代的作用。

②发行人通过核心技术优势和经营成果，实现了基础数据服务行业的国产优势，有效地降低进口依赖和发展产业的独立自主，实现了进口替代的目的

进口替代的最终目的是通过国内市场主要由本国生产者供应，减少或者完全消除该种产品的进口，实现经济上的独立自主。发行人基于自身拥有的和积累的核心技术优势和经营成果，实现了基础数据服务行业的国产优势，从而有效地降低进口依赖并发展产业的独立自主，实现了进口替代的目的。具体说明如下：

A.中国人工智能产业蓬勃发展，Appen 作为从事训练数据服务的主要国外厂商，其竞争实力较强，拓展中国市场的战略方向明确

近年来，我国人工智能产业从基础支撑、核心技术到行业应用的产业链条基

本形成，一批创新活跃、特色鲜明的创新企业加速成长，新模式、新业态不断涌现，整体呈现蓬勃发展态势。我国基础数据服务行业日益增长的行业需求使得国内市场成为了包括国外厂商在内的训练数据服务厂商的必争之地。从事训练数据服务的主要国外厂商包括 Appen、Definedcrowd、Zen3Tech 等公司。其中 Appen 为上市公司，通过其信息披露可以获知，其整体竞争实力较强，且中国市场是其重点业务方向。Appen 已建立专门的运营和产品团队服务中国市场，中国市场对于 Appen 业绩的持续增长变得日趋重要。Appen 负责中国市场的大中国区总经理已出任全球高级副总裁，截至目前 Appen 已在中国设立 3 家全资子公司，中国现已成为 Appen 除美国外设立子公司数量最多的国家，人员和机构设置均体现了 Appen 对中国市场的高度重视。

B. 海天瑞声是国内最早从事人工智能基础数据服务的公司，并实现了基础数据服务的国产优势，达到了进口替代的目的

发行人是我国领先的训练数据专业提供商，自 2005 年成立以来始终致力于为 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。发行人的进口替代体现在两个方面：

第一，发行人成功替代 Appen 等公司成为全球科技巨头（如 IBM、微软、英特尔等）在中国设立的分支研究机构的供应商。

第二，发行人的成立和发展与我国 AI 产业发展和对训练数据的需求是同步的，始终保持了该领域的国产优势，实现了训练数据供应的独立、自主：

在智能语音这一优势领域，发行人的产品线已包含 160 余个主要语种及方言，发行人提供的多语种智能语音训练数据帮助众多国内客户实现了多功能、多地域等国际业务扩张，提高国内厂商在国际上的综合竞争力。尤其近年来持续响应国家“一带一路”整体规划，语言研究能力已覆盖规划内的 54 个国家的 42 个语种及方言，累计词条数近 350 万条。

截至目前，发行人智能语音训练数据相关的产品和服务获得了阿里巴巴、腾讯、百度、科大讯飞、中国科学院、清华大学等国内大型科技公司、人工智能企业和科研机构的广泛认可，累计覆盖国内客户 274 家。

除向上述国内的商业客户提供智能语音训练数据产品及服务外，也为中央网

信办下属单位国家计算机网络与信息安全管理中心等政府客户提供其产品及服务。随着人工智能技术在我国社会安全、互联网安全等方面的介入和应用，国家有关部门对训练数据的需求也越来越高，由于该类训练数据涉及的领域比较特殊，涉及国家公共安全特征，因此对国产化要求高。发行人所拥有的核心技术能力和丰富语种资源，能够充分满足现有国产化要求，使建设国家公共安全的训练数据研发与生产的核心技术掌握在中国人手中，对我国公共安全信息化建设具有重要支撑作用。

根据国家计算机网络与信息安全管理中心出具的说明函：“海天瑞声在为中中心提供训练数据产品、服务的过程中，符合国内企业身份，在企业实力、关键技术、产品/服务能力、质量控制、管理能力等条件上具备明显优势，有效地满足了中心的相关训练数据采购需求，协助中心实现了同类训练数据国产化采购的要求。”

③发行人通过自主研发，构建了人工智能算法技术和大数据处理技术结合的驱动机制，动态提升数据处理自动化程度，解决了智能语音训练数据规模化、场景化生产的技术难题，并保证数据质量、提高生产效率、降低数据成本。发行人基于核心技术所生产的智能语音训练数据在核心指标上与主要国外竞争对手相比具有一定的领先优势，具备技术先进性，能够有效地实现进口替代。

发行人通过持续的研发，积累了核心技术并充分运用于智能语音训练数据的生产工作中，实现降本增效，可高质高量地生产体量规模更大、场景覆盖更加全面、成本更为优化的智能语音训练数据，奠定了行业内的竞争地位。

发行人在智能语音训练数据研发核心技术发明专利数量、拥有自主知识产权的智能语音训练数据产品数量和规模、智能语音产品服务内容的全面性和专业性、智能语音主流语种成品训练数据集的核心指标、智能语音产品详细参数指标等方面普遍领先于 Appen，其智能语音训练数据产品及服务具备技术先进性，能够有效地实现进口替代。

④发行人已在业务实践中受到行业的高度认可。作为我国最早进入基础数据服务行业的企业之一，发行人提供的智能语音训练数据产品及服务已覆盖多个语种/方言和多项应用场景，赢得了阿里巴巴、腾讯、百度、微软、三星、亚马逊、

科大讯飞、字节跳动、中国科学院、清华大学等众多大型优质客户的认可。同时，公司先后获得国家重点软件企业、国家高新技术企业等资质，成为中国人工智能产业发展联盟理事单位、中国语音产业联盟理事会员单位、中关村高新技术企业协会理事单位，并荣膺多项国内外人工智能领域奖项和荣誉称号。

同时，根据由中国信息通信研究院组织、共计 5 名中国工程院院士、中国科学院等科研机构正高研究员及教授、行业资深人士等行业专家出具的专家意见：

“与国际主要竞争对手相比，海天瑞声在主营业务覆盖广度、核心技术发明专利数量、拥有自主知识产权的训练数据产品数量和规模、主流语种语音类成品训练数据集的核心指标对比、语音类产品服务内容的全面性和专业性等方面已经具备领先优势。同时，海天瑞声在整体技术实力、应用领域覆盖能力等方面已与国际主要竞争对手保持同步领先水平。”

在国际产业竞争日趋激烈的大环境下，海天瑞声依托其 100%自主研发的国产化技术体系、生产体系，实现了同类产品服务的进口替代，其产出的大规模、高质量的训练数据资源集群对带动我国人工智能训练数据建设与发展起到了引领作用，对加快构建我国人工智能产业发展体系、完成人工智能在国家战略层面的系统布局起到了积极的推动作用，具备广阔的市场应用前景。”

### （3）发行人主要产品/服务的进口替代效果及前景

发行人主要产品/服务的进口替代效果和前景主要体现在：①人工智能基础数据服务保持了较高国产化率，发行人市场份额排名领先；②人工智能基础数据服务在未来仍具备广阔的应用前景，发行人具有持续增长的能力，并将持续满足进口替代需求。

（4）结论：综上，发行人依靠核心技术形成的主要产品/服务，属于国家鼓励、支持和推动的关键产品，发行人通过核心技术优势和经营成果，实现了基础数据服务行业的国产优势，有效的降低进口依赖和发展产业的独立自主，实现了进口替代的目的，符合科创属性评价标准二（四）的要求；因此，发行人具备科创属性。

## 七、发行人的境外经营及境外资产情况

### （一）基本情况

发行人主要在中华人民共和国境内进行生产经营，其全资子公司香港海天瑞声在境外生产经营，承担发行人少量境外业务拓展和服务采购职能。香港海天瑞声简要基本情况及最近一年的主要财务数据请参见本招股说明书“第五节 发行人基本情况/六、发行人控股子公司及参股公司情况/（一）子公司/2、香港海天瑞声”。

### （二）发行人境外经营符合当地规定

发行人主要在中华人民共和国境内进行生产经营，其全资子公司香港海天瑞声在境外承担发行人的少量境外业务拓展和服务采购职能。除香港子公司外，发行人在境外无经营场所或经营主体；发行人及其子公司为境外客户提供产品/服务，涉及境外销售。

#### 1、香港子公司经营情况

根据君合律师事务所出具的法律意见书，截至2021年4月22日，香港海天瑞声有效登记、存续，其依据与发行人签署的业务协议内容开展业务不需要向香港监管机构或部门申请、取得任何执照或许可。香港子公司经营符合当地规定。

#### 2、境外销售情况

发行人的境外销售主要涉及美国、韩国、日本等地，上述国家对信息技术等货品进口的基本规定主要如下：

国家	法律法规	相关规定
美国	《特别指定国民和被封锁人员清单》	美国人被禁止与特别指定国民清单中的各方进行交易，并且必须封锁其在全球各地占有或控制的特别指定国民清单上的各方的任何资产。任何人寻求与被封锁方进行交易，必须首先向海外资产控制办公室申请特别许可，除非海外资产控制办公室一般豁免许可允许交易。
韩国	《对外贸易法》	1、在属于以下各款之一情况下，产业资源部长可根据总统令规定，采取特别措施，限制或禁止物品的进口或出口：（1）在韩国或韩国的贸易对象国发生战争、事变或自然灾害时；（2）在交易对象国否认一般公认的国际协定中所规定的韩国的权益时；（3）当交易对象国对同韩国的贸易增加不正当或不合理的负担或限制时；（4）为履行依据宪法缔结并公布的有关贸易条约和一般公认的国际法规中规定的维持国际和平与安全的义务必需时；（5）为保护人类健康和生命安全或动植物的健康、生命和环境安全以及保护国内资源所必需时；

国家	法律法规	相关规定
		<p>2、当国内产业处于以下各项之一的情况时，与该产业有利害关系者或掌管该产业的有关行政机关之长可向贸易委员会提出对该特定产品的进口、贸易及流通服务对国内产业造成的影响进行调查。</p> <p>（1）特定物品进口增加对生产同类物品或生产有直接竞争关系的物品的国内产业的发展造成严重影响或可能造成影响时；（2）外国人（包括依据大韩民国法律设立的法人企业和二分之一股份由外国人持有的法人企业）从事贸易、流通服务的增加对从事同类贸易、流通服务的国内产业或有直接竞争关系的国内产业造成严重影响或可能造成影响时；（3）进口商品损害到受国内法律保护的特许权、实用新案权、外观设计权、商标权、著作权、著作邻接权、软件开发权及半导体集成电路的配置设计权，使得生产该产品或与该产品有直接竞争关系的产品的国内产业受到影响或有可能受到影响时；</p> <p>3、贸易委员会通过对国内产业被侵害的调查认定国内产业遭受侵害后，自判定日起 45 日内可建议有关行政机关之长在一定时日内采取以下各项之一的措施。（1）限制进口物品的数量（2）调整进口关税税率（3）根据有关法律对工业农林水产业、矿业的中小企业提供技术支援（4）终止或禁止特定物品或特定贸易交易者的进口（5）为救济其他国内产业实行总统令规定的措施</p>
日本	《外汇与外国贸易法》	<p>1、为对外贸易和国民经济的健康发展而进口货物的业者，依据政令的规定，有义务取得进口许可。</p> <p>2、居住者与非居住者之间拟转让非居住者实施的工业产权及其他有关技术的权利，或订立有关这些权利的使用权，以及签订或更新关于事业经营的技术指导合同，或者变更该合同的条款，如果在这些当中有政令规定的属于需要审查的，依据政令的规定，事先必须把签订该引进技术合同的条款及其他政令规定的事项，向大藏大臣和行业主管大臣申报；在接到申报的情况下，大藏大臣及行业主管大臣，认为有必要审查该申报的引进技术合同，如存在会损害国家安全、妨碍维护公共秩序或妨碍保护人民群众的安全、给予日本经济的顺利发展带来严重的不利影响的引进技术合同等时，可以延长不得签订与该申报有关的引进技术合同的期限。期限为自受理之日起四个月。</p>

发行人境外销售主要涉及美国、韩国、日本等国家，前述国家主要针对可能严重损害数字经济安全、国家安全、国民健康、产业安全等的情形进行了限制，未直接限制、禁止从发行人进口数据产品或接受发行人提供数据服务。发行人严格遵守境外销售国家/地区的相关规定，为境外客户提供训练数据服务或产品，不存在被列入负面清单或违反当地规定的情形。

综上，发行人香港子公司境外经营符合当地规定；发行人境外销售主要涉及的国家并无禁止或限制数据进口的相关规定，发行人向境外客户销售训练数据产品及服务符合当地规定。



## 第七节 公司治理与独立性

### 一、公司治理

(一) 发行人股东大会、董事会、监事会、独立董事、董事会秘书制度的建立健全及运行情况

#### 1、发行人公司治理结构建立健全情况

自公司整体变更为股份公司以来，公司根据《公司法》、《证券法》、《上市公司章程指引》等相关法律法规的要求，逐步建立健全了由股东大会、董事会、独立董事、监事会和高级管理人员组成的治理结构。公司建立了符合上市公司治理规范性要求的《股东大会议事规则》、《董事会议事规则》、《监事会议事规则》、《独立董事工作制度》、《关联交易管理办法》、《规范与关联方资金往来的管理制度》、《对外担保管理办法》、《总经理工作细则》、《董事会秘书工作细则》、《对外投资管理办法》、《信息披露事务管理制度》、《投资者关系工作管理制度》、《募集资金管理制度》等制度，并建立了战略委员会、审计委员会、提名委员会、薪酬与考核委员会等董事会下设专门委员会及其相应的议事规则。

#### 2、报告期内发行人公司治理存在的缺陷及改进情况

公司改制成为股份有限公司后，公司股东大会、董事会、监事会及相关职能部门按照有关法律法规和公司内部制度规范运行，形成了职责明确、相互制衡、规范有效的公司治理机制，没有违法违规情况发生，报告期内发行人不存在公司治理缺陷。

#### 3、股东大会、董事会、监事会运行情况

根据《公司章程》的规定，发行人股东大会由全体股东组成，是发行人的权力机构，行使法律法规和《公司章程》规定的职权。发行人董事会由9名董事组成（含3名独立董事），对股东大会负责，行使法律法规和《公司章程》规定的以及股东大会授权的职责；发行人现任董事中4人兼任高级管理人员，少于董事总人数的二分之一。发行人监事会由3名监事组成，设监事会主席1名，职工代表监事不少于三分之一，监事会执行法律法规和《公司章程》赋予的监督职能。

发行人的董事和监事每届任期均为三年，高级管理人员采用合同聘任制，其

任职期限均符合《公司法》、《公司章程》及《公司章程（草案）》等的相关规定。

发行人自股份有限公司设立以来，截至 2021 年 5 月 7 日，共召开了 17 次股东大会会议（含创立大会）、31 次董事会会议、26 次监事会会议。上述会议在召集方式、议事程序、表决方式和决议内容等方面，均符合有关法律、法规和《公司章程》、《股东大会议事规则》、《董事会议事规则》、《监事会议事规则》的规定。

#### **4、独立董事制度的建立和运行情况**

公司按照《中国证监会关于在上市公司设立独立董事指导意见》等规定，设置了独立董事，并制定了《独立董事工作制度》。公司有独立董事 3 名，其中 1 名为会计专业人士，独立董事人数达到公司董事会总人数的三分之一。

公司独立董事自任职以来，依据《公司章程》、《独立董事工作制度》等要求积极参与公司决策，充分发挥在财务、行业、管理等方面的特长，就公司规范运作和经营情况提出意见，就报告期内的董事、监事、高级管理人员选聘和利润分配等事宜发表独立意见。同时，独立董事积极参与董事会下设专门委员会的相关运作。公司独立董事制度的建立和运行，有利于公司治理结构的改善，维护了全体股东的利益。

#### **5、董事会专门委员会的运行情况**

根据《公司章程》规定，公司董事会下设战略委员会、审计委员会、提名委员会、薪酬与考核委员会。公司战略委员会成员为唐涤飞、贺琳、仪晓辉，唐涤飞担任召集人；公司审计委员会成员为杨涛、志鹏、张东晖，杨涛为会计专业人士并担任召集人；公司提名委员会成员为仪晓辉、贺琳、杨涛，仪晓辉担任召集人；公司薪酬与考核委员会成员为张东晖、杨涛、吕思遥，张东晖担任召集人。

公司董事会各专门委员会自设立以来，按照法律法规、《公司章程》以及公司各专门委员会议事规则等相关规定履行相关职责。公司各专门委员会会议在召集方式、议事程序、表决方式和决议内容等方面，均符合有关法律、法规和《公司章程》等的规定。

#### **6、董事会秘书制度的建立和运行情况**

根据《公司章程》规定，公司制定了《董事会秘书工作细则》。公司设董事

会秘书 1 名,经董事长提名,由董事会聘任。董事会秘书为公司的高级管理人员,对公司和董事会负责,具体职责如下:

(1) 公司上市后负责公司信息披露事务,协调公司信息披露工作,组织制订公司信息披露事务管理制度,督促公司及相关信息披露义务人遵守信息披露相关规定;(2) 公司上市后负责公司投资者关系管理和股东资料管理工作,协调公司与证券监管机构、股东及实际控制人、证券服务机构、媒体等之间的信息沟通;(3) 组织筹备董事会会议和股东大会,参加股东大会、董事会会议、监事会会议及高级管理人员相关会议,负责董事会会议记录工作并签字确认;(4) 公司上市后负责公司信息披露的保密工作,在未公开重大信息出现泄露时,及时向公司上市的证券交易所报告并公告;(5) 公司上市后关注公共媒体报道并主动求证真实情况,督促董事会及时回复公司上市的证券交易所的问询;(6) 组织董事、监事和高级管理人员进行证券法律法规、相关上市规则及公司上市的证券交易所其他相关规定的培训,协助前述人员了解各自在信息披露中的权利和义务;(7) 公司上市后督促董事、监事和高级管理人员遵守法律、法规、规章、规范性文件、相关上市规则、公司上市的证券交易所其他相关规定及《公司章程》,切实履行其所作出的承诺;公司上市后,在知悉公司作出或者可能作出违反有关规定的决议时,应当予以提醒并立即如实地向公司上市的证券交易所报告;(8) 《公司法》、《证券法》以及证券监管机构要求履行的其他职责。

股份公司设立以来,公司董事会秘书依法筹备了历次董事会会议及股东大会会议,确保了公司董事会和股东大会的依法召开,在信息披露、公司治理、投资者关系管理等方面发挥了重要作用,提升了公司的规范运作水平。

(二) 公司管理层对内部控制完整性、合理性及有效性的自我评估意见以及注册会计师对公司内部控制的审核意见

### 1、管理层的内部控制自我评价意见

公司董事会认为,公司根据国家法律法规和公司经营现状,已建立了一套适合自身特点的内部控制制度,形成了健全、有效的内部控制体系,并在各项经营活动中得到有效的执行,起到了保护公司资产安全和完整,保障经营活动有效运行,保证会计记录和其他相关信息真实、准确、完整的控制目标。公司内部控制

制度的设计不存在重大缺陷，可以对风险进行有效控制，并符合国家相关法律、法规和规章制度的要求。

公司董事会认为，公司按照《企业内部控制基本规范》及相关规定，针对所有重大事项均建立了健全、合理的内部控制制度，并于 2020 年 12 月 31 日在所有重大方面保持了与财务报表相关的有效的内部控制。

## 2、注册会计师对公司内部控制的审核意见

毕马威华振事务所于 2021 年 4 月 30 日出具了《内部控制审核报告》（毕马威华振专字第 2103529 号），认为公司于 2020 年 12 月 31 日在所有重大方面保持了按照《企业内部控制基本规范》标准建立的与财务报表相关的有效的内部控制。

### （三）发行人报告期内的违法违规情况

报告期内，公司严格按照法律法规的规定开展经营活动，不存在重大违法违规行为以及受到相关主管部门行政处罚且情节严重的情况，不存在因涉嫌犯罪被司法机关立案侦查或其他严重损害投资者合法权益和社会公共利益的情况。

### （四）报告期内资金占用或及担保情况

报告期内发行人不存在资金被控股股东、实际控制人及其控制的其他企业以借款、代偿债务、代垫款项或者其他方式占用的情况，也不存在为控股股东、实际控制人及其控制的其他企业担保的情况。

## 二、独立性情况

公司在资产、人员、财务、机构、业务等方面均独立于公司控股股东、实际控制人及其控制的其他企业，具有独立完整的资产、业务体系，具备直接面向市场独立持续经营的能力。

### （一）资产完整方面

公司拥有独立、完整的与经营相关的业务体系和相关资产，包括独立的采购、研发、销售体系。公司资产权属清晰、完整，合法拥有与生产经营有关的主要专利申请、非专利技术的所有权或者使用权，不存在对控股股东、实际控制人及其控制的其他企业的依赖情况。

## （二）人员独立方面

公司总经理、副总经理、技术总监、财务负责人和董事会秘书等高级管理人员不在控股股东、实际控制人及其控制的其他企业中担任除董事、监事以外的其他职务，不在控股股东、实际控制人及其控制的其他企业领取薪酬。公司的财务人员不在控股股东、实际控制人及其控制的其他企业中兼职。

## （三）财务独立方面

公司已建立独立、完整的财务核算体系，独立作出财务决策，具有规范的财务会计制度和对分公司、子公司的财务管理制度。公司不存在与控股股东、实际控制人及其控制的其他企业共用银行账户的情形。报告期内，公司不存在资金被控股股东、实际控制人及其控制的其他企业以借款、代偿债务、代垫款项或者其他方式占用的情况，不存在为控股股东、实际控制人及其控制的其他企业担保的情况。

## （四）机构独立方面

公司建立了独立、健全的内部管理机构，独立行使管理职权。公司的生产经营和办公场所与控股股东、实际控制人及其控制的其他企业严格分开，不存在与控股股东、实际控制人及其控制的其他企业混合经营、合署办公的情形。

## （五）业务独立方面

公司具有独立完整的采购、研发、销售系统，具有独立完整的业务和面向市场自主经营的能力，不存在依赖控股股东、实际控制人及其控制的其他企业进行生产经营活动的情况，与控股股东、实际控制人及其控制的其他企业之间不存在同业竞争，不存在严重影响独立性或显失公平的关联交易。

## （六）最近两年主营业务、管理层情况及控制权情况

报告期内，发行人主营业务为训练数据的研发设计、生产及销售，控制权、管理团队和核心技术人员稳定，最近2年内公司主营业务和董事、高级管理人员及核心技术人员均未发生重大不利变化；控股股东和受控股股东、实际控制人支配的股东所持发行人的股份权属清晰，最近2年实际控制人未发生变更，不存在导致控制权可能变更的重大权属纠纷。

### （七）发行人不存在对持续经营具有重大影响的事项说明

截至本招股说明书签署日，公司核心技术均拥有清晰产权，不存在主要资产、核心技术、商标的重大权属纠纷，不存在重大偿债风险，不存在重大担保、诉讼、仲裁等或有事项。公司经营环境良好，不存在已经或将要发生的对持续经营有重大影响的经营环境变化事项。

## 三、同业竞争

### （一）公司与控股股东、实际控制人及其控制的其他企业不存在同业竞争

截至本招股说明书签署日，公司控股股东、实际控制人为贺琳。贺琳控制的除发行人及发行人子公司以外的其他企业包括中瑞安、创世联合和海讯科技。

中瑞安为海天瑞声员工持股平台，除持有发行人股权外未实际经营业务，创世联合为中瑞安的普通合伙人，除持有中瑞安出资外未实际经营业务，因此中瑞安、创世联合未以任何形式直接或间接从事与公司相同或相似的业务，与公司不存在同业竞争的情况。

贺琳及其配偶蔡惠智合计持有海讯科技 50%的股权，其中贺琳持股 40%，蔡惠智持股 10%。海讯科技为贺琳和蔡惠智共同控制的企业。截至本招股说明书签署日，海讯科技的基本情况如下：

名称	北京中科海讯科技有限公司	
统一社会信用代码	91110108749398794B	
法定代表人	何国建	
注册资本	500 万元	
实收资本	500 万元	
成立时间	2003 年 4 月 7 日	
营业期限	2003 年 4 月 7 日至 2023 年 4 月 6 日	
经营范围	能源、节能环保领域内的技术开发、技术咨询、技术服务；自然科学研究与试验发展。（企业依法自主选择经营项目，开展经营活动；依法须经批准的项目，经相关部门批准后依批准的内容开展经营活动；不得从事本市产业政策禁止和限制类项目的经营活动。）	
住所	北京市海淀区地锦路 33 号院 1 号楼 4 层 W407	
股东构成及持股比例	股东	持股比例
	贺琳	40%

	陈丹平	25%
	何国建	25%
	蔡惠智	10%
<b>控制情况</b>	贺琳和蔡惠智共同控制	
<b>主营业务</b>	报告期内，公司未实际开展与经营范围相关的业务，主营收入来自房屋出租。	

2018 年以来，海讯科技未以任何形式直接或间接从事与公司相同或相似的业务，与公司不存在同业竞争的情况。

## （二）控股股东、实际控制人为避免同业竞争而出具的承诺

为避免未来可能发生的同业竞争情况，公司控股股东、实际控制人贺琳于 2020 年 5 月 29 日出具《关于避免同业竞争的承诺》，承诺：

“本人及其控制的企业目前不存在且不从事与北京海天瑞声科技股份有限公司（下称“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；

同时，本人承诺：

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海天瑞声及其子公司相同、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海天瑞声及其子公司相同、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海天瑞声及其子公司相同、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、不向其他业务与海天瑞声及其子公司相同、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、客户信息等商业机密；

5、如果未来本人拟从事的业务可能与海天瑞声及其子公司存在同业竞争，本人将本着海天瑞声及其子公司优先的原则与海天瑞声协商解决。

本承诺函自出具之日起生效，并在本人作为海天瑞声关联方的整个期间持续有效。”

## 四、关联方及关联交易

### （一）关联方及关联关系

根据《公司法》、《企业会计准则第 36 号——关联方披露》和《上海证券交易所科创板股票上市规则》等有关法律法规关于关联方和关联关系的有关规定，公司的主要关联方和关联关系如下：

#### 1、发行人控股股东、实际控制人

序号	关联方名称	关联关系
1	贺琳	直接和间接控制发行人 42.44% 股份，为发行人控股股东、实际控制人

#### 2、其他直接或间接持有发行人 5% 以上股份的自然人

序号	关联方名称	关联关系
1	唐涤飞	直接持有发行人 11.15% 股份，其持股 50% 的创慧科瑞持有中瑞立（持有发行人 5.83% 股份）的 29.41% 出资

根据《公司法》、《企业会计准则第 36 号——关联方披露》和《上海证券交易所科创板股票上市规则》等有关法律法规，其他间接持有公司 5% 以上股份的自然人股东为公司的关联方。

#### 3、发行人董事、监事及高级管理人员

发行人董事、监事及高级管理人员请参见本招股说明书“第五节 发行人基本情况/九、董事、监事、高级管理人员与核心技术人员的简要情况”。

#### 4、上述 1-3 项所述关联自然人的关系密切的家庭成员，包括配偶、年满 18 周岁的子女及其配偶、父母及配偶的父母、兄弟姐妹及其配偶、配偶的兄弟姐妹、子女配偶的父母

根据《企业会计准则第 36 号——关联方披露》、《上海证券交易所科创板股票上市规则》等法律法规及规范性文件规定，发行人实际控制人、董事、监事、高级管理人员和直接或间接持有发行人 5% 以上股份的自然人关系密切的家庭成员，包括配偶、年满 18 周岁的子女及其配偶、父母及配偶的父母、兄弟姐妹及其配偶、配偶的兄弟姐妹、子女配偶的父母为发行人的关联方。



## 5、直接持有发行人 5%以上股份的法人或其他组织

序号	关联方名称	关联关系
1	中瑞安	直接持有发行人 15.43%股份
2	中移投资	直接持有发行人 12.01%股份
3	清德投资	直接持有发行人 7.93%股份
4	上海丰琬	直接持有发行人 5.86%股份
5	中瑞立	直接持有发行人 5.83%股份

## 6、上述 1-5 项所述关联法人或关联自然人直接或者间接控制的，或者由前述关联自然人（独立董事除外）担任董事、高级管理人员的法人或其他组织

根据《企业会计准则第 36 号—关联方披露》、《上海证券交易所科创板股票上市规则》等法律法规及规范性文件规定，上述 1-5 项所述关联法人或关联自然人直接或者间接控制的，或者由前述关联自然人（独立董事除外）担任董事、高级管理人员的法人或其他组织为公司的关联方。

序号	关联方名称	关联关系
1	创世联合	贺琳持有 100% 股权，并任执行董事、经理
2	北京中科海讯科技有限公司	贺琳持有 40% 股权；贺琳配偶蔡惠智持有 10% 股权
3	北京海讯瑞声管理咨询有限公司	贺琳配偶蔡惠智持有 100% 股权
4	宁波梅山保税港区中科海讯科技投资合伙企业（有限合伙）	贺琳配偶蔡惠智为有限合伙人，持有 90.74% 的出资；北京海讯瑞声管理咨询有限公司持股为普通合伙人、执行事务合伙人，持有 1.08% 的出资
5	北京中科海讯数字科技股份有限公司（股票代码：300810）	贺琳配偶蔡惠智持有 4.57% 股权并担任董事长，宁波梅山保税港区中科海讯科技投资合伙企业（有限合伙）持有 29.26% 股权
6	武汉海晟科讯科技有限公司	北京中科海讯数字科技股份有限公司持有 67% 股权
7	秦皇岛中科雷声信息技术有限公司	北京中科海讯数字科技股份有限公司持股 100%，蔡惠智担任董事
8	青岛海讯数字科技有限公司	北京中科海讯数字科技股份有限公司持股 100%，蔡惠智任执行董事
9	武汉范思合成数字科技有限责任公司	北京中科海讯数字科技股份有限公司持有 51% 股权
10	武汉中科海讯电子科技有限公司	北京中科海讯数字科技股份有限公司持股 100%，蔡惠智任执行董事
11	创慧科瑞	唐涤飞持有 50% 股权；志鹏持有 50% 股权，并任执行董事、经理
12	BING & DIFEI BELL-TOWER INVESTMENT LLC	唐涤飞及其配偶持有 100% 股权并任董事

序号	关联方名称	关联关系
13	BING & DIFEI YARROW POINT INVESTMENT LLC	唐涤飞及其配偶持有 100% 股权并任董事
14	北京贺兮科技有限公司	吕思遥持有 100% 股权并任执行董事, 其父亲担任经理
15	北京海淀科技金融资本控股集团股份有限公司	志鹏任董事
16	昆山青石投资管理有限公司	志鹏持有 25% 股权并任执行董事、总经理; 志鹏配偶钟山持有 50% 股权
17	北京清融投资管理有限公司	志鹏持有 40% 股权、其配偶钟山持有 60% 股权并任执行董事及经理
18	清源华动(北京)科技有限公司	志鹏配偶钟山任经理
19	西藏金钟达信息科技有限责任公司	志鹏配偶钟山持股 100% 并任执行董事、经理
20	北京耐飞科技有限公司	西藏金钟达信息科技有限责任公司持股 48.40%, 志鹏配偶钟山任董事、经理
21	西藏耐飞影视有限公司	北京耐飞科技有限公司持股 100%, 志鹏配偶钟山任董事、经理
22	北京耐飞影视有限公司	北京耐飞科技有限公司持股 100%, 志鹏配偶钟山任执行董事
23	北京联合网视文化传播有限公司	北京耐飞影视有限公司持股 100%, 志鹏配偶钟山任执行董事
24	耐飞影视有限公司	北京耐飞科技有限公司持股 100%
25	潍坊发生影业有限公司	北京耐飞科技有限公司持股 51%
26	北京凡喜影业有限公司	潍坊发生影业有限公司持股 100%
27	喀什青柠影业有限公司	北京耐飞科技有限公司持股 100%
28	天津兔子洞影视文化传媒有限公司	北京耐飞科技有限公司持股 100%
29	北京有格影视传媒有限公司	天津兔子洞影视文化传媒有限公司持股 100%
30	佛山耐飞影视有限公司	北京耐飞科技有限公司持股 100%
31	海宁双盈影视有限公司	北京耐飞科技有限公司持股 100%
32	苏州清桥电子科技有限公司	志鹏配偶钟山任董事长、经理
33	上海万彧企业管理中心(有限合伙)	黄宇凯为有限合伙人, 持有 54.35% 出资
34	上海缪阳企业管理咨询有限公司	黄宇凯姐姐的配偶持有 100% 的股权并任执行董事
35	上海缪阳医疗科技中心(有限合伙)	上海缪阳企业管理咨询有限公司为普通合伙人, 执行事务合伙人, 出资比例 1%; 黄宇凯姐姐的配偶为有限合伙人, 出资比例 99%
36	上海心瑞医疗科技有限公司	黄宇凯姐姐的配偶持有 44.93% 的股权并任董事长、总经理, 上海缪阳医疗科技中心(有限合伙)持有 4.41% 的股权
37	上海勤和互联网技术软件开发有限公司	张小龙任董事
38	上海文华财经资讯股份有限公司	张小龙担任董事

序号	关联方名称	关联关系
39	上海兴富平龙创业投资有限公司	张小龙担任经理
40	兴富投资管理有限公司	张小龙担任副总经理
41	深圳市博阅科技股份有限公司	张小龙担任董事
42	南通市久正人体工学股份有限公司	张小龙担任董事
43	杭州碧橙数字技术股份有限公司	张小龙担任董事
44	景宁行止企业管理合伙企业(有限合伙)	方彦彬出资比例为 99%，为有限合伙人
45	景宁仰止企业管理合伙企业(有限合伙)	方彦彬出资比例为 1%，并任执行事务合伙人
46	景宁观止企业管理合伙企业(有限合伙)	方彦彬出资比例为 1%，并任执行事务合伙人
47	云仰科技(杭州)有限公司	方彦彬任董事、总经理

## 7、发行人子公司、合营公司及联营公司

报告期内，发行人拥有 2 家纳入合并报表范围的全资子公司，无合营公司或联营公司，具体情况如下：

序号	关联方名称	关联关系
1	北京中瑞智科技有限公司	发行人全资子公司
2	海天瑞声(香港)科技有限公司	发行人全资子公司

## 8、间接持有发行人 5%以上股份的法人或其他组织

根据《上市规则》，间接持有发行人 5%以上股份的法人或者其他组织为发行人的关联方。

## 9、报告期内曾经存在的关联方

序号	关联方类别	关联方名称	关联关系	备注
1	实际控制人的关联方	北京中科海讯软件科技有限公司	北京中科海讯科技有限公司持股 100%	2018 年 11 月已注销
2		北京中科海讯数字设备有限公司	北京中科海讯科技有限公司持股 100%	2018 年 11 月已注销
3		北京中科海讯声学科技有限公司	贺琳配偶蔡惠智持有 24% 股权，并担任执行董事、经理	2019 年 1 月已注销
4	公司董事志鹏的关联方	深圳市云网拜特科技有限公司	志鹏配偶钟山曾任执行董事	2018 年 6 月已离任
5		南京智亚源电力科技有限公司	志鹏配偶钟山曾持股 0.98% 并任董事	2018 年 12 月已转让全部股权并离任
6		西藏微霜夜影视文	北京耐飞科技有限公司持	2019 年 9 月已注销

序号	关联方类别	关联方名称	关联关系	备注
		化有限公司	股 51%，志鹏配偶钟山任执行董事、经理	
7		天津卓腾文化传媒有限公司	北京耐飞科技有限公司持股 100%	2019 年 12 月已注销
8		霍尔果斯耐飞影视有限公司	北京耐飞科技有限公司持股 100%，志鹏配偶钟山任执行董事	2020 年 1 月已注销
9		西藏新七组影视文化有限公司	北京耐飞科技有限公司持股 80%，志鹏配偶钟山任董事长、经理	已于 2020 年 10 月注销
10		北京影拓星瀚网络科技有限公司	志鹏配偶钟山曾任董事长、经理	钟山已于 2020 年 8 月离任
11		西藏雄关峻卡影视有限公司	北京耐飞科技有限公司持股 100%，志鹏配偶钟山任董事长、经理	已于 2020 年 10 月注销
12		北京智能广宣科技有限公司	志鹏配偶钟山曾任董事	钟山已于 2020 年 11 月离任
13		北京影娱互动科技有限公司	西藏金钟达信息科技有限公司曾持股 60%	西藏金钟达信息科技有限公司已于 2020 年 11 月转出全部股权
14		上海烛龙影视文化有限公司	北京耐飞科技有限公司曾持股 40%，志鹏配偶钟山曾任董事长	2021 年 2 月，北京耐飞科技有限公司转出全部股权，钟山离任
15	公司董事黄宇凯的关联方	恩脉（上海）医疗科技有限公司	黄宇凯持股 54.35% 的上海万彧企业管理中心（有限合伙）曾持股 41.40%	已于 2020 年 9 月转出
16		杭州全垒投资管理有限公司	方彦彬曾担任董事	2019 年 6 月已离任
17		杭州云深处科技有限公司	方彦彬曾担任董事	2020 年 5 月已离任
18	公司监事方彦彬的关联方	杭州慧钉网络科技有限公司	方彦彬曾担任董事	2020 年 7 月已离任
19		飞米光年（北京）科技有限公司	方彦彬持股 60%，并任执行董事、经理，	已于 2019 年 11 月注销
20		杭州行止数据科技有限公司	飞米光年（北京）科技有限公司持股 100%，方彦彬任执行董事兼总经理	已于 2020 年 7 月注销
21	公司前独立董事	吕一凡	曾任发行人独立董事	2018 年 11 月已离任
22	公司前副总经理	程显峰	曾任发行人副总经理	2019 年 1 月已离任
23	发行人原 5% 以上股东及	上海兴富	发行人原 5% 以上股东	2019 年 12 月上海兴富持股比例降至 5% 以下
24	其相关关联方	宁波高新区兴佳创业投资中心（有限合伙）	发行人原 5% 以上股东上海兴富持有 59.99% 的出资，为有限合伙人	

序号	关联方类别	关联方名称	关联关系	备注
25		德清兴富睿宏投资管理合伙企业（有限合伙）	发行人原 5% 以上股东上海兴富持有 52.63% 的出资，为有限合伙人	

## （二）关联交易

### 1、经常性关联交易

报告期内，公司向董事、监事及高级管理人员支付报酬的金额如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
董事、监事及高级管理人员薪酬	720.95	898.19	811.78

### 2、偶发性关联交易

报告期内，公司不存在偶发性关联交易。

### 3、报告期内关联交易汇总情况

单位：万元

		经常性关联交易		
关联方	交易内容	2020 年	2019 年度	2018 年度
董事、监事及高级管理人员	发放薪酬	720.95	898.19	811.78

### 4、报告期内关联方往来余额汇总情况

报告期内各期末，公司与各关联方之间往来余额为 0。

## （三）规范关联交易的制度安排

发行人根据相关法律法规在《公司章程》、《股东大会议事规则》、《董事会议事规则》、《独立董事工作制度》、《关联交易管理办法》和《规范与关联方资金往来的管理制度》等内部规章制度中对关联方、关联关系、关联交易价格、关联交易的批准权限、关联交易的回避与决策程序、关联交易的信息披露等做了明确的规定，保证了公司与关联方之间订立的关联交易符合公平、公开、公正的原则。

## （四）报告期内关联交易履行规定程序的情况及独立董事意见

发行人报告期内各年度董事、监事及高级管理人员薪酬已经发行人董事会或股东大会审议通过。因此，报告期内发生的关联交易不存在损害发行人及其他股

东利益的情况。

#### （五）关于规范及减少关联交易的承诺

公司控股股东、实际控制人贺琳已经出具了《关于规范及减少关联交易的承诺》，主要承诺内容如下：

“1、本人及关联方将尽量避免和减少与海天瑞声及其下属子公司之间的关联交易，对于海天瑞声及其下属子公司能够通过市场与独立第三方之间发生的交易，将由海天瑞声及其下属子公司与独立第三方进行。本人及关联方将严格避免向海天瑞声及其下属子公司拆借、占用海天瑞声及其下属子公司资金或采取由海天瑞声及其下属子公司代垫款、代偿债务等方式侵占海天瑞声及其下属子公司资金。

2、对于本人及关联方与海天瑞声及其下属子公司之间必需的一切交易行为，均将严格遵守市场原则，本着平等互利、等价有偿的一般原则公平合理地进行。交易定价有政府定价的，执行政府定价；没有政府定价的，执行市场公允价格；没有政府定价且无可参考市场价格的，按照成本加可比较的合理利润水平确定成本价执行。

3、本人及关联方与海天瑞声及其下属子公司之间的关联交易将严格遵守海天瑞声章程、关联交易决策制度等规定履行必要的法定程序。在海天瑞声权力机构审议有关关联交易事项时主动依法履行回避义务；对须报经有权机构审议的关联交易事项，在有权机构审议通过后方可执行。

4、本人保证不通过关联交易取得任何不正当的利益或使海天瑞声及其下属子公司承担任何不正当的义务。如果因违反上述承诺导致海天瑞声或其下属子公司损失或利用关联交易侵占海天瑞声或其下属子公司利益的，海天瑞声及其下属子公司的损失由本人承担。

5、上述承诺在本人构成海天瑞声的关联方期间持续有效。”

## 第八节 财务会计信息与管理层分析

本节披露或引用的财务会计数据，非经特别说明，均引自毕马威华振会计师事务所（特殊普通合伙）出具的毕马威华振审字第 2103527 号标准无保留意见的《审计报告》。

本节的财务会计数据及有关的分析说明反映了公司 2018 年 12 月 31 日、2019 年 12 月 31 日和 2020 年 12 月 31 日经审计的合并及母公司资产负债表，2018 年度、2019 年度和 2020 年度经审计的合并及母公司利润表、现金流量表和所有者权益变动表以及财务报表附注的主要内容。

本节对财务报表的重要项目进行了说明，公司提醒投资者，如欲更详细地了解公司报告期的财务状况、经营成果和现金流量，请关注本招股说明书所附财务报告及审计报告全文，以获取全部的财务资料。

### 一、报告期经审计的财务报表

#### （一）审计意见

毕马威华振会计师事务所（特殊普通合伙）审计了公司财务报表，包括 2018 年 12 月 31 日、2019 年 12 月 31 日及 2020 年 12 月 31 日的合并及母公司资产负债表，2018 年度、2019 年度及 2020 年度的合并及母公司利润表、合并及母公司现金流量表、合并及母公司股东权益变动表以及相关财务报表附注。

毕马威华振会计师事务所（特殊普通合伙）认为，公司财务报表在所有重大方面按照中华人民共和国财政部颁布的企业会计准则（以下简称“企业会计准则”）的规定编制，公允反映了海天瑞声 2018 年 12 月 31 日、2019 年 12 月 31 日及 2020 年 12 月 31 日的合并及母公司财务状况以及 2018 年度、2019 年度及 2020 年度的合并及母公司经营成果和现金流量。

毕马威华振会计师事务所（特殊普通合伙）识别出的关键审计事项如下：

<b>收入确认</b>	
请参阅本节“四、报告期内主要采用的会计政策和会计估计/（十七）收入”所述的会计政策及“九、经营成果分析/（一）营业收入分析”。	
<b>关键审计事项</b>	<b>在审计中如何应对该事项</b>
北京海天瑞声科技股份有限公司及子公司	与评价收入确认相关的审计程序中主要包括

(以下简称“海天瑞声集团”) 2018 年度、2019 年度及 2020 年度(以下简称“报告期”) 的收入分别为 19,265.77 万元、23,755.81 万元及 23,337.40 万元, 主要包括提供训练数据定制服务收入及训练数据产品收入。

训练数据定制服务收入主要是指海天瑞声集团根据客户需求为客户提供训练数据定制服务而收取的收入。训练数据产品收入主要是指海天瑞声集团授权客户永久使用本集团已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。

于 2018 年度、2019 年度, 海天瑞声集团以风险报酬转移时点作为收入确认的判断标准。对于训练数据定制服务收入, 采用完工百分比法确认收入, 并按照提供给客户并经客户验收和确认的服务量占服务总量的比例计算和确定完工进度。对于训练数据产品收入, 按照合同约定向客户提交训练数据产品并经客户验收确认后的时点确认收入。

自 2020 年 1 月 1 日起, 海天瑞声集团执行《企业会计准则第 14 号——收入(修订)》(以下简称“新收入准则”), 以控制权转移作为收入确认时点的判断依据, 并综合考虑合同条款和业务安排来评估新收入准则的影响。经评估认为, 对于训练数据定制服务收入, 单元定制化训练数据被识别为单项履约义务, 属于某一时点履行的履约义务, 因此在将单元数据向客户交付并经客户验收确认时确认收入。对于训练数据产品收入, 因为训练数据产品属于某一时点履行的知识产权许可授权, 因此在向客户交付训练数据产品并经客户验收确认后确认收入。

由于收入是海天瑞声集团的关键业绩指标之一, 存在管理层为了达到特定目标或期望而操纵收入确认时点的固有风险或收入被确认于不正确的会计期间的错报风险, 同时新收入准则的实施也需要管理层根据合同条款和业务安排评估对收入确认会计政策的影响, 该评估涉及管理层重大判断, 因此毕马威华振事务所将海天瑞声集团的收入确认识别为关键审计事项。

以下:

了解并评价与收入确认相关的关键财务报告内部控制的设计和运行有效性;

选取销售合同或订单, 识别与风险和报酬转移或控制权转移相关的合同或订单条款, 评价海天瑞声集团各类收入的收入确认政策是否符合企业会计准则的相关要求;

选取各期收入的前十大客户和当期新增主要客户, 获取相关公开信息(如通过“全国企业信用信息公示系统”), 核对客户的股东、董事和监事情况, 并和海天瑞声集团提供的信息进行比对, 检查是否存在关联关系;

在抽样的基础上, 将各期记录的收入核对至相关的销售合同或订单、客户验收文件、发票等相关支持性依据, 以评价相关收入是否按照海天瑞声集团的收入确认政策予以确认;

根据客户的交易特点和性质, 选取样本, 就于资产负债表日的往来款项余额及报告期间的销售交易金额执行函证程序;

选取客户, 实地走访或视频访谈, 询问其与海天瑞声集团的主要交易情况, 及其是否与海天瑞声集团存在关联方关系, 检查客户及交易的真实性, 识别是否存在异常情况;

从资产负债表日前后的收入交易中选取样本, 检查与收入确认相关的销售合同或订单、客户验收文件、发票等相关支持性依据, 评价接近资产负债表日前后记录的收入是否记录于恰当期间;

选取符合特定风险标准的收入会计分录(包括期后重大调整分录), 核对至相关支持性文件。

## (二) 合并财务报表

### 1、合并资产负债表

单位: 元

项目	2020 年 12 月 31 日	2019 年 12 月 31 日	2018 年 12 月 31 日
流动资产:			



项目	2020年12月31日	2019年12月31日	2018年12月31日
货币资金	40,263,750.22	163,255,204.88	46,691,570.86
交易性金融资产	331,089,456.16	156,016,898.63	-
应收账款	62,782,914.66	43,645,482.75	29,389,643.27
预付款项	9,990,748.73	4,199,643.99	835,708.87
其他应收款	1,836,268.98	1,692,351.94	1,728,438.23
存货	3,129,382.82	7,547,453.53	20,632,026.63
合同资产	1,374,738.20	-	-
其他流动资产	1,257,978.48	698,458.00	84,927,609.87
<b>流动资产合计</b>	<b>451,725,238.25</b>	<b>377,055,493.72</b>	<b>184,204,997.73</b>
<b>非流动资产：</b>			
固定资产	23,573,225.28	24,811,531.16	25,787,732.29
无形资产	1,158,175.85	1,484,774.84	1,265,924.66
长期待摊费用	510,349.14	876,511.47	598,389.25
递延所得税资产	383,050.47	311,040.69	470,668.15
<b>非流动资产合计</b>	<b>25,624,800.74</b>	<b>27,483,858.16</b>	<b>28,122,714.35</b>
<b>资产总计</b>	<b>477,350,038.99</b>	<b>404,539,351.88</b>	<b>212,327,712.08</b>

单位：元

项目	2020年12月31日	2019年12月31日	2018年12月31日
<b>流动负债：</b>			
应付账款	7,144,414.85	3,596,916.44	1,934,612.92
预收款项	-	24,806,552.34	34,076,903.14
合同负债	14,957,716.71	-	-
应付职工薪酬	9,714,740.17	12,011,512.73	9,213,249.70
应交税费	7,317,834.04	7,871,837.85	761,112.01
其他应付款	258,960.64	301,094.16	208,711.63
<b>流动负债合计</b>	<b>39,393,666.41</b>	<b>48,587,913.52</b>	<b>46,194,589.40</b>
<b>非流动负债：</b>			
<b>非流动负债合计</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>负债总计</b>	<b>39,393,666.41</b>	<b>48,587,913.52</b>	<b>46,194,589.40</b>
<b>股东权益：</b>			
股本	32,100,000.00	32,100,000.00	30,000,000.00
资本公积	186,209,473.40	186,209,473.40	58,109,473.40
其他综合收益	67,214.25	143,301.94	111,810.75

项目	2020年12月31日	2019年12月31日	2018年12月31日
盈余公积	16,485,510.17	16,485,510.17	8,133,151.51
未分配利润	203,094,174.76	121,013,152.85	69,778,687.02
归属于母公司股东权益合计	437,956,372.58	355,951,438.36	166,133,122.68
少数股东权益	-	-	-
<b>股东权益合计</b>	<b>437,956,372.58</b>	<b>355,951,438.36</b>	<b>166,133,122.68</b>
<b>负债和股东权益合计</b>	<b>477,350,038.99</b>	<b>404,539,351.88</b>	<b>212,327,712.08</b>

## 2、合并利润表

单位：元

项目	2020年度	2019年度	2018年度
<b>一、营业收入</b>	<b>233,373,953.01</b>	<b>237,558,118.15</b>	<b>192,657,698.25</b>
减：营业成本	75,377,910.56	70,675,170.49	67,867,821.67
税金及附加	2,031,646.53	2,008,270.31	1,477,223.24
销售费用	11,133,435.04	9,252,818.31	8,744,780.95
管理费用	25,667,440.94	31,559,583.73	21,806,510.73
研发费用	43,497,496.67	41,696,074.47	27,345,336.92
财务费用	2,151,305.50	1,498,249.24	1,127,213.17
其中：利息费用	-	-	-
利息收入	133,758.57	62,390.86	71,970.21
加：其他收益	8,780,434.27	8,322,390.73	7,478,636.24
投资收益	8,394,396.43	4,556,739.26	3,384,296.11
公允价值变动收益	1,089,456.16	-	-
信用减值损失	-1,313,981.11	-1,378,769.07	-
资产减值损失	-182,760.96	-855,632.89	-3,151,711.04
资产处置损失	-64,687.07	-1,529.53	1,035.00
<b>二、营业利润</b>	<b>90,217,575.49</b>	<b>91,511,150.10</b>	<b>72,001,067.88</b>
加：营业外收入	660,826.82	126,200.37	238,512.66
减：营业外支出	71,784.17	277,867.99	114,498.87
<b>三、利润总额</b>	<b>90,806,618.14</b>	<b>91,359,482.48</b>	<b>72,125,081.67</b>
减：所得税费用	8,725,596.23	9,772,657.99	4,994,391.65
<b>四、净利润</b>	<b>82,081,021.91</b>	<b>81,586,824.49</b>	<b>67,130,690.02</b>
(一) 按经营持续性分类			

项目	2020 年度	2019 年度	2018 年度
1、持续经营净利润	82,081,021.91	81,586,824.49	67,130,690.02
2、终止经营净利润	-	-	-
(二) 按所有权归属分类			
1、归属于母公司股东的净利润	82,081,021.91	81,586,824.49	67,130,690.02
2、少数股东损益	-	-	-
<b>五、其他综合收益的税后净额</b>	<b>-76,087.69</b>	<b>31,491.19</b>	<b>72,052.71</b>
(一) 归属母公司股东的其他综合收益的税后净额	-76,087.69	31,491.19	72,052.71
将重分类进损益的其他综合收益	-76,087.69	31,491.19	72,052.71
外币财务报表折算差额	-76,087.69	31,491.19	72,052.71
(二) 归属少数股东的其他综合收益的税后净额	-	-	-
<b>六、综合收益总额</b>	<b>82,004,934.22</b>	<b>81,618,315.68</b>	<b>67,202,742.73</b>
(一) 归属于母公司股东的综合收益总额	82,004,934.22	81,618,315.68	67,202,742.73
(二) 归属于少数股东的综合收益总额	-	-	-
<b>七、每股收益：</b>			
(一) 基本每股收益	2.56	2.72	2.24
(二) 稀释每股收益	2.56	2.72	2.24

### 3、合并现金流量表

单位：元

项目	2020 年度	2019 年度	2018 年度
<b>一、经营活动产生的现金流量：</b>			
销售商品、提供劳务收到的现金	220,653,871.43	229,806,563.10	213,955,395.38
收到的税费返还	109,921.28	1,135,100.58	2,653,180.15
收到其他与经营活动有关的现金	10,941,339.38	8,560,289.49	8,480,865.75
<b>经营活动现金流入小计</b>	<b>231,705,132.09</b>	<b>239,501,953.17</b>	<b>225,089,441.28</b>
购买商品、接受劳务支付的现金	78,631,509.86	66,683,621.87	81,814,874.33
支付给职工以及为职工支付的现金	59,904,894.06	51,202,435.48	43,587,091.65
支付的各项税费	24,919,888.54	19,609,540.79	22,634,434.31
支付其他与经营活动有关	17,072,180.49	18,643,051.18	12,775,089.87

项目	2020 年度	2019 年度	2018 年度
的现金			
<b>经营活动现金流出小计</b>	<b>180,528,472.95</b>	<b>156,138,649.32</b>	<b>160,811,490.16</b>
<b>经营活动产生的现金流量净额</b>	<b>51,176,659.14</b>	<b>83,363,303.85</b>	<b>64,277,951.12</b>
<b>二、投资活动产生的现金流量：</b>			
收回投资收到的现金	870,010,000.00	617,500,000.00	232,000,000.00
处置固定资产、无形资产和其他长期资产收回的现金净额	8,649.46	12,295.04	113,919.61
取得投资收益收到的现金	8,411,295.06	4,539,840.63	3,384,296.11
收到其他与投资活动有关的现金	-	-	-
<b>投资活动现金流入小计</b>	<b>878,429,944.52</b>	<b>622,052,135.67</b>	<b>235,498,215.72</b>
购建固定资产、无形资产和其他长期资产支付的现金	1,800,728.71	3,089,686.86	26,293,694.89
投资支付的现金	1,044,010,000.00	690,500,000.00	240,000,000.00
支付其他与投资活动有关的现金	-	-	-
<b>投资活动现金流出小计</b>	<b>1,045,810,728.71</b>	<b>693,589,686.86</b>	<b>266,293,694.89</b>
<b>投资活动产生的现金流量净额</b>	<b>-167,380,784.19</b>	<b>-71,537,551.19</b>	<b>-30,795,479.17</b>
<b>三、筹资活动产生的现金流量：</b>			
吸收投资收到的现金	-	130,200,000.00	-
<b>筹资活动现金流入小计</b>	<b>-</b>	<b>130,200,000.00</b>	<b>-</b>
分配股利、利润或偿付利息支付的现金	-	22,000,000.00	-
支付其他与筹资活动有关的现金	5,804,950.29	3,575,000.00	-
<b>筹资活动现金流出小计</b>	<b>5,804,950.29</b>	<b>25,575,000.00</b>	<b>-</b>
<b>筹资活动产生的现金流量净额</b>	<b>-5,804,950.29</b>	<b>104,625,000.00</b>	<b>-</b>
<b>四、汇率变动对现金及现金等价物的影响</b>	<b>-982,379.32</b>	<b>112,881.36</b>	<b>191,803.70</b>
<b>五、现金及现金等价物净增加额/（净减少额）</b>	<b>-122,991,454.66</b>	<b>116,563,634.02</b>	<b>33,674,275.65</b>
加：期初现金及现金等价物余额	163,255,204.88	46,691,570.86	13,017,295.21
<b>六、期末现金及现金等价物余额</b>	<b>40,263,750.22</b>	<b>163,255,204.88</b>	<b>46,691,570.86</b>

## 二、与财务会计信息相关的重大事项的判断标准

公司在本节披露的与财务会计信息相关重大事项标准为金额超过最近一期净资产或净利润的 10%，或金额虽未达到前述标准但公司认为较为重要的相关事项。

## 三、财务报表的编制基础、遵循企业会计准则的声明、合并财务报表范围及变化情况

### （一）财务报表的编制基础

公司以持续经营为基础编制财务报表。

公司自 2019 年 1 月 1 日起执行了中华人民共和国财政部 2017 年度修订的《企业会计准则第 22 号——金融工具确认和计量》等新金融工具准则。

公司自 2020 年 1 月 1 日起执行了中华人民共和国财政部 2017 年度修订的《企业会计准则第 14 号——收入》。

公司尚未执行财政部于 2018 年度修订的《企业会计准则第 21 号——租赁》。

### （二）遵循企业会计准则的声明

公司编制的申报财务报表符合财政部颁布的企业会计准则的要求，真实、完整地反映了公司 2018 年 12 月 31 日、2019 年 12 月 31 日及 2020 年 12 月 31 日的合并财务状况和财务状况，2018 年度、2019 年度及 2020 年度的合并经营成果和经营成果及合并现金流量和现金流量。

此外，公司的财务报表同时符合中国证监会 2014 年修订的《公开发行证券的公司信息披露编报规则第 15 号——财务报告的一般规定》有关财务报表及其附注的披露要求。

### （三）合并财务报表范围及变化情况

#### 1、合并财务报表范围

报告期各期，公司合并财务报表范围内子公司如下：

子公司名称	2020 年度	2019 年度	2018 年度
北京中瑞智科技有限公司	是	是	是
海天瑞声（香港）科技有限公司	是	是	是

## 2、报告期内合并财务报表范围的变化情况

报告期内，公司未发生合并财务报表范围的变更。

## 四、报告期内主要采用的会计政策和会计估计

### （一）会计期间

会计年度自公历 1 月 1 日起至 12 月 31 日止。

### （二）营业周期

公司将从购买用于加工的资产起至实现现金或现金等价物的期间作为正常营业周期。公司主要业务的营业周期通常小于 12 个月。

### （三）记账本位币

公司的记账本位币为人民币，编制财务报表采用的货币为人民币。公司及境内子公司选定记账本位币的依据是主要业务收支的计价和结算币种。公司的境外子公司采用美元为记账本位币，在编制本财务报表时，这些子公司的外币财务报表按照（七）进行了折算。

### （四）同一控制下和非同一控制下企业合并的会计处理方法

#### 1、同一控制下企业合并

参与合并的企业在合并前后均受同一方或相同的多方最终控制且该控制并非暂时性的，为同一控制下的企业合并。合并方在企业合并中取得的资产和负债，按照合并日在最终控制方合并财务报表中的账面价值计量。取得的净资产账面价值与支付的合并对价账面价值（或发行股份面值总额）的差额，调整资本公积中的股本溢价；资本公积中的股本溢价不足冲减的，调整留存收益。为进行企业合并发生的直接相关费用，于发生时计入当期损益。合并日为合并方实际取得对被合并方控制权的日期。

## 2、非同一控制下企业合并

参与合并的各方在合并前后不受同一方或相同的多方最终控制的，为非同一控制下的企业合并。公司作为购买方，为取得被购买方控制权而付出的资产（包括购买日之前所持有的被购买方的股权）、发生或承担的负债以及发行的权益性证券在购买日的公允价值之和，减去合并中取得的被购买方可辨认净资产于购买日公允价值份额的差额，如为正数则确认为商誉；如为负数则计入当期损益。公司为进行企业合并发生的各项直接费用计入当期损益。公司在购买日按公允价值确认所取得的被购买方符合确认条件的各项可辨认资产、负债及或有负债。购买日是指购买方实际取得对被购买方控制权的日期。

通过多次交易分步实现非同一控制企业合并时，对于购买日之前持有的被购买方的股权，公司会按照该股权在购买日的公允价值进行重新计量，公允价值与其账面价值的差额计入当期投资收益或其他综合收益。购买日之前持有的被购买方的股权涉及的权益法核算下的以后可重分类进损益的其他综合收益及其他所有者权益变动于购买日转入当期投资收益；购买日之前持有的被购买方的股权为以公允价值计量且其变动计入其他综合收益的权益工具投资的，购买日之前确认的其他综合收益于购买日转入留存收益。

### （五）合并财务报表的编制方法

#### 1、总体原则

合并财务报表的合并范围以控制为基础予以确定，包括公司及公司控制的子公司。控制，是指公司拥有对被投资方的权力，通过参与被投资方的相关活动而享有可变回报，并且有能力运用对被投资方的权力影响其回报金额。在判断公司是否拥有对被投资方的权力时，公司仅考虑与被投资方相关的实质性权利（包括公司自身所享有的及其他方所享有的实质性权利）。子公司的财务状况、经营成果和现金流量由控制开始日起至控制结束日止包含于合并财务报表中。

当子公司所采用的会计期间或会计政策与公司不一致时，合并时已按照公司的会计期间或会计政策对子公司财务报表进行必要的调整。合并时所有集团内部交易及余额，包括未实现内部交易损益均已抵销。集团内部交易发生的未实现损失，有证据表明该损失是相关资产减值损失的，则全额确认该损失。

## 2、合并取得子公司

对于通过同一控制下企业合并取得的子公司，在编制合并当期财务报表时，以被合并子公司的各项资产、负债在最终控制方财务报表中的账面价值为基础，视同被合并子公司在公司最终控制方对其开始实施控制时纳入公司合并范围，并对合并财务报表的期初数以及前期比较报表进行相应调整。

对于通过非同一控制下企业合并取得的子公司，在编制合并当期财务报表时，以购买日确定的被购买子公司各项可辨认资产、负债的公允价值为基础自购买日起将被购买子公司纳入公司合并范围。

## 3、处置子公司

公司丧失对原有子公司控制权时，由此产生的任何处置收益或损失，计入丧失控制权当期的投资收益。

通过多次交易分步处置对子公司长期股权投资直至丧失控制权的，按下述原则判断是否为一揽子交易：

这些交易是同时或者在考虑了彼此影响的情况下订立的；

这些交易整体才能达成一项完整的商业结果；

一项交易的发生取决于其他至少一项交易的发生；

一项交易单独考虑时是不经济的，但是和其他交易一并考虑时是经济的。

如果各项交易不属于一揽子交易的，则在丧失对子公司控制权以前的各项交易，按照不丧失控制权的情况下部分处置对子公司的股权投资的会计政策进行处理。

如果各项交易属于一揽子交易的，则将各项交易作为一项处置原有子公司并丧失控制权的交易进行处理，在丧失控制权之前每一次处置价款与处置投资对应的享有该子公司自购买日开始持续计算的净资产账面价值的份额之间的差额，在合并财务报表中计入其他综合收益，在丧失控制权时一并转入丧失控制权当期的损益。



#### 4、少数股东权益变动

公司因购买少数股权新取得的长期股权投资成本与按照新增持股比例计算应享有子公司的净资产份额之间的差额，以及在不丧失控制权的情况下因部分处置对子公司的股权投资而取得的处置价款与处置长期股权投资相对应享有子公司净资产的差额，均调整合并资产负债表中的资本公积（股本溢价），资本公积（股本溢价）不足冲减的，调整留存收益。

##### （六）现金及现金等价物的确定标准

现金和现金等价物包括库存现金、可以随时用于支付的存款以及持有期限短、流动性强、易于转换为已知金额现金、价值变动风险很小的投资。

##### （七）外币业务和外币报表折算

公司收到投资者以外币投入资本时按当日即期汇率折合为人民币，其他外币交易在初始确认时按交易发生日的即期汇率折合为人民币。

于资产负债表日，外币货币性项目采用该日的即期汇率折算。除与购建符合资本化条件资产有关的专门借款本金和利息的汇兑差额外，其他汇兑差额计入当期损益。以历史成本计量的外币非货币性项目，仍采用交易发生日的即期汇率折算。

对境外经营的财务报表进行折算时，资产负债表中的资产和负债项目，采用资产负债表日的即期汇率折算，股东权益项目中除未分配利润及其他综合收益中的外币财务报表折算差额项目外，其他项目采用发生时的即期汇率折算。利润表中的收入和费用项目，采用交易发生日的即期汇率折算。按照上述折算产生的外币财务报表折算差额，在其他综合收益中列示。处置境外经营时，相关的外币财务报表折算差额自其他综合收益转入处置当期损益。

##### （八）金融工具

公司的金融工具包括货币资金、应收款项、应付款项及股本等。

财政部于2017年颁布了《企业会计准则第22号——金融工具确认和计量（修订）》、《企业会计准则第23号——金融资产转移（修订）》、《企业会计准则第24号——套期会计（修订）》及《企业会计准则第37号——金融工具列报（修订）》

（统称“新金融工具准则”）。新金融工具准则修订了财政部于 2006 年颁布的《企业会计准则第 22 号——金融工具确认和计量》、《企业会计准则第 23 号——金融资产转移》和《企业会计准则第 24 号——套期保值》以及财政部于 2014 年修订的《企业会计准则第 37 号——金融工具列报》（统称“原金融工具准则”）。

公司自 2018 年 1 月 1 日至 2018 年 12 月 31 日止执行原金融工具准则，自 2019 年 1 月 1 日起执行新金融工具准则。

## 1、原金融工具准则

### （1）金融资产及金融负债的确认和计量

金融资产和金融负债在公司成为相关金融工具合同条款的一方时，于资产负债表内确认。

公司在初始确认时按取得资产或承担负债的目的，把金融资产和金融负债分为不同类别：以公允价值计量且其变动计入当期损益的金融资产和金融负债、贷款及应收款项、持有至到期投资、可供出售金融资产和其他金融负债。

在初始确认时，金融资产及金融负债均以公允价值计量。对于以公允价值计量且其变动计入当期损益的金融资产或金融负债，相关交易费用直接计入当期损益；对于其他类别的金融资产或金融负债，相关交易费用计入初始确认金额。初始确认后，金融资产和金融负债的后续计量如下：

#### ①应收款项

应收款项是指在活跃市场中没有报价、回收金额固定或可确定的非衍生金融资产。

初始确认后，应收款项以实际利率法按摊余成本计量。

#### ②可供出售金融资产

公司将在初始确认时即被指定为可供出售的非衍生金融资产以及没有归类到其他类别的金融资产分类为可供出售金融资产。

对公允价值不能可靠计量的可供出售权益工具投资，初始确认后按成本计量；其他可供出售金融资产，初始确认后以公允价值计量，公允价值变动形成的利得或损失，除减值损失和外币货币性金融资产形成的汇兑差额计入当期损益外，其

他利得或损失计入其他综合收益，在可供出售金融资产终止确认时转出，计入当期损益。可供出售权益工具投资的现金股利，在被投资单位宣告发放股利时计入当期损益。按实际利率法计算的可供出售金融资产的利息，计入当期损益。

### ③其他金融负债

其他金融负债是指除以公允价值计量且其变动计入当期损益的金融负债以外的金融负债。初始确认后采用实际利率法按摊余成本计量。

#### (2) 金融资产及金融负债的列报

金融资产和金融负债在资产负债表内分别列示，没有相互抵销。但是，同时满足下列条件的，以相互抵销后的净额在资产负债表内列示：

公司具有抵销已确认金额的法定权利，且该种法定权利是当前可执行的；

公司计划以净额结算，或同时变现该金融资产和清偿该金融负债。

#### (3) 金融资产和金融负债的终止确认

满足下列条件之一时，公司终止确认该金融资产：

收取该金融资产现金流量的合同权利终止；

该金融资产已转移，且公司将金融资产所有权上几乎所有的风险和报酬转移给转入方；

该金融资产已转移，虽然公司既没有转移也没有保留金融资产所有权上几乎所有的风险和报酬，但是放弃了对该金融资产控制。

金融资产整体转移满足终止确认条件的，公司将下列两项金额的差额计入当期损益：

所转移金融资产的账面价值；

因转移而收到的对价，与原直接计入股东权益的公允价值变动累计额之和。

金融负债的现时义务全部或部分已经解除的，公司终止确认该金融负债或其一部分。

#### (4) 金融资产的减值

公司在资产负债表日对以公允价值计量且其变动计入当期损益的金融资产以外的金融资产的账面价值进行检查，有客观证据表明该金融资产发生减值的，计提减值准备。

金融资产发生减值的客观证据，包括但不限于：

发行方或债务人发生严重财务困难；

债务人违反了合同条款，如偿付利息或本金发生违约或逾期等；

债务人很可能倒闭或进行其他财务重组；

因发行方发生重大财务困难，该金融资产无法在活跃市场继续交易；

权益工具发行方经营所处的技术、市场、经济或法律环境等发生重大不利变化，使权益工具投资人可能无法收回投资成本；

权益工具投资的公允价值发生严重下跌（即公允价值下跌超过 50%）或非暂时性下跌（即公允价值下跌持续超过九个月）等。

#### ①应收款项

应收款项按下述原则运用个别方式和组合方式评估减值损失。

运用个别方式评估时，当应收款项的预计未来现金流量（不包括尚未发生的未来信用损失）按原实际利率折现的现值低于其账面价值时，公司将该应收款项的账面价值减记至该现值，减记的金额确认为资产减值损失，计入当期损益。

当运用组合方式评估应收款项的减值损失时，减值损失金额是根据具有类似信用风险特征的应收款项（包括以个别方式评估未发生减值的应收款项）的以往损失经验，并根据反映当前经济状况的可观察数据进行调整确定的。

在应收款项确认减值损失后，如有客观证据表明该金融资产价值已恢复，且客观上与确认该损失后发生的事项有关，公司将原确认的减值损失予以转回，计入当期损益。该转回后的账面价值不超过假定不计提减值准备情况下该金融资产在转回日的摊余成本。

#### A、单项金额重大并单独计提坏账准备的应收款项

<b>单项金额重大的判断依据或金额标准</b>	单项金额大于人民币 100 万元的应收款项视为重大。
<b>单项金额重大并单独计提坏账准备的计提方法</b>	当应收款项的预计未来现金流量（不包括尚未发生的未来信用损失）按原实际利率折现的现值低于其账面价值时，公司对该部分差额确认减值损失，计提应收款项坏账准备。

### B、单项金额不重大但单独计提坏账准备的应收款项

<b>单项金额不重大但单独计提坏账准备的理由</b>	与客户就相关余额存在争议或相关客户信用状况恶化导致偿还能力不足的应收款项。
<b>坏账准备的计提方法</b>	当应收款项的预计未来现金流量（不包括尚未发生的未来信用损失）按原实际利率折现的现值低于其账面价值时，公司对该部分差额确认减值损失，计提应收款项坏账准备。

### C、按信用风险特征组合计提坏账准备的应收款项

对于上述 A 和 B 中单项测试未发生减值的应收款项，公司也会将其包括在具有类似信用风险特征的应收账款组合中再进行减值测试。

<b>按信用风险特征组合计提坏账准备的计提方法</b>	
<b>组合 1：账龄分析组合</b>	公司根据以往的历史经验对应收款项计提比例做出最佳估计，参考应收账款的账龄进行信用风险组合分类。
<b>组合 2：回收风险较低组合</b>	公司根据特定性质就特定对象，认定无信用风险，包括对公司关联方的应收款项等回收风险较低的组合。

组合 1 中，采用账龄分析法计提坏账准备的：

<b>账龄</b>	<b>应收款项计提比例（%）</b>
1 年以内（含 1 年）	5%
1 - 2 年（含 2 年）	10%
2 - 3 年（含 3 年）	30%
3 - 4 年（含 4 年）	50%
4 - 5 年（含 5 年）	80%
5 年以上	100%

### ②可供出售金融资产

可供出售金融资产运用个别方式和组合方式评估减值损失。可供出售金融资产发生减值时，即使该金融资产没有终止确认，公司将原直接计入股东权益的因公允价值下降形成的累计损失从股东权益转出，计入当期损益。

对于已确认减值损失的可供出售债务工具，在随后的会计期间公允价值已上升且客观上与确认原减值损失后发生的事项有关的，公司将原确认的减值损失予

以转回，计入当期损益。可供出售权益工具投资发生的减值损失，不通过损益转回。但是，在活跃市场中没有报价且其公允价值不能可靠计量的权益工具投资，不得转回。

#### (5) 权益工具

公司发行权益工具收到的对价扣除交易费用后，计入股东权益。回购公司权益工具支付的对价和交易费用，减少股东权益。

## 2、自 2019 年 1 月 1 日起执行新金融工具准则下的政策

### (1) 金融资产及金融负债的确认和初始计量

金融资产和金融负债在公司成为相关金融工具合同条款的一方时，于资产负债表内确认。

除不具有重大融资成分的应收账款外，在初始确认时，金融资产及金融负债均以公允价值计量。对于以公允价值计量且其变动计入当期损益的金融资产或金融负债，相关交易费用直接计入当期损益；对于其他类别的金融资产或金融负债，相关交易费用计入初始确认金额。对于不具有重大融资成分的应收账款，公司按照根据（十七）的会计政策确定的交易价格进行初始计量。

### (2) 金融资产的分类和后续计量

#### ① 公司金融资产的分类

公司通常根据管理金融资产的业务模式和金融资产的合同现金流量特征，在初始确认时将金融资产分为不同类别：以摊余成本计量的金融资产、以公允价值计量且其变动计入其他综合收益的金融资产及以公允价值计量且其变动计入当期损益的金融资产。

除非公司改变管理金融资产的业务模式，在此情形下，所有受影响的相关金融资产在业务模式发生变更后的首个报告期间的第一天进行重分类，否则金融资产在初始确认后不得进行重分类。

公司将同时符合下列条件且未被指定为以公允价值计量且其变动计入当期损益的金融资产，分类为以摊余成本计量的金融资产：

公司管理该金融资产的业务模式是以收取合同现金流量为目标；

该金融资产的合同条款规定，在特定日期产生的现金流量，仅为对本金和以未偿付本金金额为基础的利息的支付。

公司将同时符合下列条件且未被指定为以公允价值计量且其变动计入当期损益的金融资产，分类为以公允价值计量且其变动计入其他综合收益的金融资产：

公司管理该金融资产的业务模式既以收取合同现金流量为目标又以出售该金融资产为目标；

该金融资产的合同条款规定，在特定日期产生的现金流量，仅为对本金和以未偿付本金金额为基础的利息的支付。

对于非交易性权益工具投资，公司可在初始确认时将其不可撤销地指定为以公允价值计量且其变动计入其他综合收益的金融资产。该指定在单项投资的基础上作出，且相关投资从发行者的角度符合权益工具的定义。

除上述以摊余成本计量和以公允价值计量且其变动计入其他综合收益的金融资产外，公司将其余所有的金融资产分类为以公允价值计量且其变动计入当期损益的金融资产。在初始确认时，如果能够消除或显著减少会计错配，公司可以将本应以摊余成本计量或以公允价值计量且其变动计入其他综合收益的金融资产不可撤销地指定为以公允价值计量且其变动计入当期损益的金融资产。

管理金融资产的业务模式，是指公司如何管理金融资产以产生现金流量。业务模式决定公司所管理金融资产现金流量的来源是收取合同现金流量、出售金融资产还是两者兼有。公司以客观事实为依据、以关键管理人员决定的对金融资产进行管理的特定业务目标为基础，确定管理金融资产的业务模式。

公司对金融资产的合同现金流量特征进行评估，以确定相关金融资产在特定日期产生的合同现金流量是否仅为对本金和以未偿付本金金额为基础的利息的支付。其中，本金是指金融资产在初始确认时的公允价值；利息包括对货币时间价值、与特定时期未偿付本金金额相关的信用风险、以及其他基本借贷风险、成本和利润的对价。此外，公司对可能导致金融资产合同现金流量的时间分布或金额发生变更的合同条款进行评估，以确定其是否满足上述合同现金流量特征的要求。

## ②公司金融资产的后续计量

#### A、以公允价值计量且其变动计入当期损益的金融资产

初始确认后，对于该类金融资产以公允价值进行后续计量，产生的利得或损失（包括利息和股利收入）计入当期损益，除非该金融资产属于套期关系的一部分。

#### B、以摊余成本计量的金融资产

初始确认后，对于该类金融资产采用实际利率法以摊余成本计量。以摊余成本计量且不属于任何套期关系的一部分的金融资产所产生的利得或损失，在终止确认、按照实际利率法摊销或确认减值时，计入当期损益。

### ③金融负债的分类和后续计量

公司将金融负债分类为以公允价值计量且其变动计入当期损益的金融负债及以摊余成本计量的金融负债。

#### A、以公允价值计量且其变动计入当期损益的金融负债

该类金融负债包括交易性金融负债（含属于金融负债的衍生工具）和指定为以公允价值计量且其变动计入当期损益的金融负债。

初始确认后，对于该类金融负债以公允价值进行后续计量，除与套期会计有关外，产生的利得或损失（包括利息费用）计入当期损益。

#### B、以摊余成本计量的金融负债

初始确认后，对其他金融负债采用实际利率法以摊余成本计量。

### ④抵销

金融资产和金融负债在资产负债表内分别列示，没有相互抵销。但是，同时满足下列条件的，以相互抵销后的净额在资产负债表内列示：

公司具有抵销已确认金额的法定权利，且该种法定权利是当前可执行的；

公司计划以净额结算，或同时变现该金融资产和清偿该金融负债。

### ⑤金融资产和金融负债的终止确认

满足下列条件之一时，公司终止确认该金融资产：



收取该金融资产现金流量的合同权利终止；

该金融资产已转移，且公司将金融资产所有权上几乎所有的风险和报酬转移给转入方；

该金融资产已转移，虽然公司既没有转移也没有保留金融资产所有权上几乎所有的风险和报酬，但是未保留对该金融资产的控制。

金融资产转移整体满足终止确认条件的，公司将下列两项金额的差额计入当期损益：

被转移金融资产在终止确认日的账面价值；

因转移金融资产而收到的对价，与原直接计入其他综合收益的公允价值变动累计额中对应终止确认部分的金额（涉及转移的金融资产为以公允价值计量且其变动计入其他综合收益的债权投资）之和。

金融负债（或其一部分）的现时义务已经解除的，公司终止确认该金融负债（或该部分金融负债）。

#### ⑥减值

公司以预期信用损失为基础，对下列项目进行减值会计处理并确认损失准备：

以摊余成本计量的金融资产；

合同资产；

公司持有的其他以公允价值计量的金融资产不适用预期信用损失模型。

#### A、预期信用损失的计量

预期信用损失，是指以发生违约的风险为权重的金融工具信用损失的加权平均值。信用损失，是指公司按照原实际利率折现的、根据合同应收的所有合同现金流量与预期收取的所有现金流量之间的差额，即全部现金短缺的现值。

在计量预期信用损失时，公司需考虑的最长期限为企业面临信用风险的最长合同期限（包括考虑续约选择权）。

整个存续期预期信用损失，是指因金融工具整个预计存续期内所有可能发生的违约事件而导致的预期信用损失。

未来 12 个月内预期信用损失，是指因资产负债表日后 12 个月内（若金融工具的预计存续期少于 12 个月，则为预计存续期）可能发生的金融工具违约事件而导致的预期信用损失，是整个存续期预期信用损失的一部分。

对于应收账款和合同资产，公司始终按照相当于整个存续期内预期信用损失的金额计量其损失准备，并以逾期天数与预期信用损失率对照表为基础计算其预期信用损失。预期信用损失率基于应收账款/合同资产过去期间的实际信用损失经验计算，并考虑历史数据收集期间的经济状况、当前的经济状况与发行人所认为的预计存续期内的经济状况三者之间的差异。

#### 第一步：确定客户群的构成

根据发行人的历史经验，发行人的客户群体比较稳定，主要为信誉良好的大型科技公司、研究机构等，客户群体发生信用损失的情况没有显著差异，因此在计算减值准备时未进一步区分不同的客户群体。

#### 第二步：确定用于计算历史损失率的历史数据集

资产负债表日，发行人选取截至资产负债表日为止的最近 3 个会计年度末应收账款/合同资产账龄数据用以计算最新的历史损失率。对于截至 2020 年 12 月 31 日止的年度期间而言，发行人选择的是 2018 年至 2020 年这段期间。

报告期内，发行人最近三年末应收账款逾期账龄情况如下表所示：

单位：万元

项目	2020 年 12 月 31 日	2019 年 12 月 31 日	2018 年 12 月 31 日
未逾期	5,477.51	3,493.65	2,699.02
逾期 1 年以内（含 1 年）	1,005.26	1,107.37	338.59
逾期 1 至 2 年（含 2 年）	160.96	2.44	59.14
逾期 2 年以上	56.90	56.90	-
<b>合计</b>	<b>6,700.63</b>	<b>4,660.36</b>	<b>3,096.75</b>

#### 第三步：计算应收账款/合同资产迁徙率

迁徙率是指在一个时间段内没有收回而迁徙至下一个时间段的应收账款/合同资产的比例。发行人根据客户群最近 3 年期间的历史逾期情况，计算应收账款/合同资产迁徙率。

发行人应收账款迁徙率计算过程如下：

单位：万元

项目	2020年 12月31日	2019年 12月31日	2018年 12月31日	年末综合 余额	注释	公式	迁徙率
未逾期	5,477.51	3,493.65	2,699.02	11,670.18	A	B/A	21.00%
逾期1年 以内（含 1年）	1,005.26	1,107.37	338.59	2,451.22	B	C/B	9.08%
逾期1至 2年（含 2年）	160.96	2.44	59.14	222.54	C	D/C	51.14%
逾期2年 以上	56.90	56.90	-	113.80	D	D/D	100.00%
<b>合计</b>	<b>6,700.63</b>	<b>4,660.36</b>	<b>3,096.75</b>	<b>14,457.74</b>			

第四步：计算历史损失率，并结合当前状况以及对未来经济状况的预测，确定预期信用损失率

历史损失率代表在历史上每个时间段中最终不能收回的应收账款/合同资产所占比例。发行根据历史数据经验判断认为逾期2年以上的应收账款/合同资产确认无法收回，损失率为100%，在此基础上，发行人根据历史期间客户群体应收账款/合同资产迁徙率计算得到各账龄段的历史损失率。同时管理层在历史损失率的基础上，基于当前经济状况以及对未来经济状况的预测，对计算出的历史损失率做出调整，确定各账龄段的预期信用损失率。发行人将预期信用损失率乘以应收账款/合同资产余额来建立准备矩阵。

发行人预期信用损失率的计算过程如下：

单位：万元

项目	注释	迁徙率	历史损失 率计算 过程	历史损 失率	调整后的 预期信用 损失率	同行业上市公 司预期信用损 失率区间（注）
未逾期	A	21.00%	$A*B*C*D$	0.98%	3.00%	0.01%-8.00%
逾期1年以 内（含1 年）	B	9.08%	$B*C*D$	4.64%	12.00%	1.92%-50.00%
逾期1至2 年（含2 年）	C	51.14%	$C*D$	51.14%	50.00%	5.00%-100.00%
逾期2年 以上	D	100.00%	D	100.00%	100.00%	6.73%-100.00%

注：同行业上市公司包括，包括荣科科技、彩讯股份、天玑科技、四方精创、恒生电子、用友网络、万兴科技、及泛微网络；对于上述同行业公司按照应收账款账龄划分预期信用损

失率的情形，此处将应收账款账龄区间为 1 年以内的预期信用损失率作为未逾期和逾期 1 年以内（含 1 年）的预期信用损失率，将应收账款账龄区间为 1-2 年的预期信用损失率作为此处逾期 1 至 2 年（含 2 年）的预期信用损失率，将应收账款账龄区间为 2 年以上的预期信用损失率作为逾期 2 年以上的预期信用损失率。

发行人运用迁徙法计算历史损失率，基于当前经济状况以及对未来经济状况的预测，同时对比了同行业可比公司信息后，对计算出的历史损失率做出调整，并基于谨慎性原则采用调整后的预期信用损失率计算应收账款坏账准备。

发行人于每年末根据更新的信息重新计算应收账款的预期信用损失率；如果用于计算预期信用损失率的信息在年内出现重大变化，发行人会重新评估预期信用损失率，必要时进行调整以准确反应应收账款的预期信用损失。

除应收账款和合同资产外，公司对满足下列情形的金融工具按照相当于未来 12 个月内预期信用损失的金额计量其损失准备，对其他金融工具按照相当于整个存续期内预期信用损失的金额计量其损失准备：

该金融工具在资产负债表日只具有较低的信用风险；或

该金融工具的信用风险自初始确认后并未显著增加。

#### B、具有较低的信用风险

如果金融工具的违约风险较低，欠款人在短期内履行其合同现金流量义务的能力很强，并且即便较长时期内经济形势和经营环境存在不利变化但未必一定降低欠款人履行其合同现金流量义务的能力，该金融工具被视为具有较低的信用风险。

由于银行存款主要存放于信用良好的国有及其他大型银行，应收第三方支付平台款项主要存放于信用良好的第三方支付平台，公司认为银行存款和应收第三方支付平台款项具有低信用风险。

#### C、信用风险显著增加

公司通过比较金融工具在资产负债表日发生违约的风险与在初始确认日发生违约的风险，以确定金融工具预计存续期内发生违约风险的相对变化，进而评估金融工具的信用风险自初始确认后是否已显著增加。

在确定信用风险自初始确认后是否显著增加时，公司考虑无须付出不必要的额外成本或努力即可获得合理且有依据的信息，包括前瞻性信息。公司考虑的

信息包括：

债务人未能按合同到期日支付本金和利息的情况；

已发生的或预期的金融工具的外部或内部信用评级（如有）的严重恶化；

已发生的或预期的债务人经营成果的严重恶化；

现存的或预期的技术、市场、经济或法律环境变化，并将对债务人对公司的还款能力产生重大不利影响。

根据金融工具的性质，公司以单项金融工具或金融工具组合为基础评估信用风险是否显著增加。以金融工具组合为基础进行评估时，公司可基于共同信用风险特征对金融工具进行分类，例如逾期信息和信用风险评级。

#### D、已发生信用减值的金融资产

公司在资产负债表日评估以摊余成本计量的金融资产是否已发生信用减值。当对金融资产预期未来现金流量具有不利影响的一项或多项事件发生时，该金融资产成为已发生信用减值的金融资产。金融资产已发生信用减值的证据包括下列可观察信息：

发行方或债务人发生重大财务困难；

债务人违反合同，如偿付利息或本金违约或逾期等；

公司出于与债务人财务困难有关的经济或合同考虑，给予债务人在任何其他情况下都不会做出的让步；

债务人很可能破产或进行其他财务重组；

发行方或债务人财务困难导致该金融资产的活跃市场消失。

#### E、预期信用损失准备的列报

为反映金融工具的信用风险自初始确认后的变化，公司在每个资产负债表日重新计量预期信用损失，由此形成的损失准备的增加或转回金额，应当作为减值损失或利得计入当期损益。对于以摊余成本计量的金融资产，损失准备抵减该金融资产在资产负债表中列示的账面价值。

#### F、核销

如果公司不再合理预期金融资产合同现金流量能够全部或部分收回，则直接减记该金融资产的账面余额。这种减记构成相关金融资产的终止确认。这种情况通常发生在公司确定债务人没有资产或收入来源可产生足够的现金流量以偿还将被减记的金额。但是，被减记的金融资产仍可能受到公司催收到期款项相关执行活动的影响。

已减记的金融资产以后又收回的，作为减值损失的转回计入收回当期的损益。

#### ⑦权益工具

公司发行权益工具收到的对价扣除交易费用后，计入股东权益。回购公司权益工具支付的对价和交易费用，减少股东权益。

回购公司股份时，回购的股份作为库存股管理，回购股份的全部支出转为库存股成本，同时进行备查登记。库存股不参与利润分配，在资产负债表中作为股东权益的备抵项目列示。

库存股注销时，按注销股票面值总额减少股本，库存股成本超过面值总额的部分，应依次冲减资本公积（股本溢价）、盈余公积和未分配利润；库存股成本低于面值总额的，低于面值总额的部分增加资本公积（股本溢价）。

库存股转让时，转让收入高于库存股成本的部分，增加资本公积（股本溢价）；低于库存股成本的部分，依次冲减资本公积（股本溢价）、盈余公积、未分配利润。

### （九）存货

#### 1、存货的分类和成本

公司的存货主要为训练数据定制服务和相关的应用服务发生的项目成本，包括项目开始至项目完成所发生的、与项目相关的直接人工和其他费用。在按照（十七）所确定的收入确认时点确认收入的同时，将与收入相关的项目成本自存货结转至当期成本。

#### 2、存货可变现净值的确定依据及存货跌价准备的计提方法

资产负债表日，存货按照成本与可变现净值孰低计量。可变现净值，是指在日常活动中，存货的估计售价减去至项目完工时估计将要发生的成本、估计的销

售费用以及相关税费后的金额确定。为执行销售合同或者劳务合同而持有的存货，其可变现净值以合同价格为基础计算。

## （十）长期股权投资

### 1、长期股权投资投资成本确定

对于通过企业合并以外的其他方式取得的长期股权投资，在初始确认时，对于以支付现金取得的长期股权投资，公司按照实际支付的购买价款作为初始投资成本；对于发行权益性证券取得的长期股权投资，公司按照发行权益性证券的公允价值作为初始投资成本。

### 2、长期股权投资后续计量及损益确认方法

#### （1）对子公司的投资

在公司个别财务报表中，公司采用成本法对子公司的长期股权投资进行后续计量，除非投资符合持有待售的条件。对被投资单位宣告分派的现金股利或利润由公司享有的部分确认为当期投资收益，但取得投资时实际支付的价款或对价中包含的已宣告但尚未发放的现金股利或利润除外。

对子公司的投资按照成本减去减值准备后在资产负债表内列示。

对子公司投资的减值测试方法及减值准备计提方法参见（十四）。

在公司合并财务报表中，对子公司按（五）进行处理。

## （十一）固定资产

### 1、固定资产确认条件

固定资产指公司为提供劳务或经营管理而持有的，使用寿命超过一个会计年度的有形资产。

外购固定资产的初始成本包括购买价款、相关税费以及使该资产达到预定可使用状态前所发生的可归属于该项资产的支出。

对于构成固定资产的各组成部分，如果各自具有不同使用寿命或者以不同方式为公司提供经济利益，适用不同折旧率或折旧方法的，公司分别将各组成部分确认为单项固定资产。

对于固定资产的后续支出，包括与更换固定资产某组成部分相关的支出，在与支出相关的经济利益很可能流入公司时资本化计入固定资产成本，同时将被替换部分的账面价值扣除；与固定资产日常维护相关的支出在发生时计入当期损益。

固定资产以成本减累计折旧及减值准备后在资产负债表内列示。

## 2、固定资产的折旧方法

公司将固定资产的成本扣除预计净残值和累计减值准备后在其使用寿命内按年限平均法计提折旧。

各类固定资产的使用寿命、残值率和年折旧率分别为：

类别	使用寿命（年）	残值率（%）	年折旧率（%）
房屋及建筑物	20-40	0-5	2.38-5.00
运输工具	3-5	0-5	19.00-33.33
电子设备	3-5	0-5	19.00-33.33
办公家具	3-5	0-5	19.00-33.33

公司至少在每年年度终了对固定资产的使用寿命、预计净残值和折旧方法进行复核。

## 3、减值测试方法及减值准备计提方法参见（十四）。

## 4、固定资产处置

固定资产满足下述条件之一时，公司会予以终止确认。

固定资产处于处置状态；

该固定资产预期通过使用或处置不能产生经济利益。

报废或处置固定资产项目所产生的损益为处置所得款项净额与项目账面金额之间的差额，并于报废或处置日在损益中确认。

## （十二）无形资产

无形资产以成本减累计摊销（仅限于使用寿命有限的无形资产）及减值准备（参见（十四））后在资产负债表内列示。对于使用寿命有限的无形资产，公司将无形资产的成本扣除预计净残值和累计减值准备后按直线法在预计使用寿命期内摊销。



各项无形资产的摊销年限为：

项目	摊销年限（年）
软件	2-5

公司将无法预见未来经济利益期限的无形资产视为使用寿命不确定的无形资产，并对这类无形资产不予摊销。截至资产负债表日，公司没有使用寿命不确定的无形资产。

公司内部研究开发项目的支出分为研究阶段支出和开发阶段支出。

研究阶段的支出，于发生时计入当期损益。开发阶段的支出，如果开发形成的某项产品或工序等在技术和商业上可行，而且公司有充足的资源和意向完成开发工作，并且开发阶段支出能够可靠计量，则开发阶段的支出可以予以资本化。资本化开发支出按成本减减值准备（参见（十四））在资产负债表内列示。其他开发费用则在其产生的期间内确认为费用。

### （十三）长期待摊费用

长期待摊费用在受益期限内分期平均摊销。各项费用的摊销期限分别为：

项目	摊销年限（年）
办公场所装修及其他	2-3

### （十四）除存货及金融资产外的其他资产减值

公司在资产负债表日根据内部及外部信息以确定下列资产是否存在减值的迹象，包括：

长期股权投资；

固定资产；

无形资产；

长期待摊费用。

公司对存在减值迹象的资产进行减值测试，估计资产的可收回金额。

可收回金额是指资产（或资产组、资产组组合，下同）的公允价值（参见（十五））减去处置费用后的净额与资产预计未来现金流量的现值两者之间较高者。

资产组由创造现金流入相关的资产组成，是可以认定的最小资产组合，其产生的现金流入基本上独立于其他资产或者资产组。

资产预计未来现金流量的现值，按照资产在持续使用过程中和最终处置时所产生的预计未来现金流量，选择恰当的税前折现率对其进行折现后的金额加以确定。

可收回金额的估计结果表明，资产的可收回金额低于其账面价值的，资产的账面价值会减记至可收回金额，减记的金额确认为资产减值损失，计入当期损益，同时计提相应的资产减值准备。与资产组或者资产组组合相关的减值损失，先抵减分摊至该资产组或者资产组组合中商誉的账面价值，再根据资产组或者资产组组合中除商誉之外的其他各项资产的账面价值所占比重，按比例抵减其他各项资产的账面价值，但抵减后的各资产的账面价值不得低于该资产的公允价值减去处臵费用后的净额（如可确定的）、该资产预计未来现金流量的现值（如可确定的）和零三者之中最高者。

资产减值损失一经确认，在以后会计期间不会转回。

#### （十五）公允价值的计量

除特别声明外，公司按下述原则计量公允价值：

公允价值是指市场参与者在计量日发生的有序交易中，出售一项资产所能收到或者转移一项负债所需支付的价格。

公司估计公允价值时，考虑市场参与者在计量日对相关资产或负债进行定价时考虑的特征（包括资产状况及所在位置、对资产出售或者使用的限制等），并采用在当前情况下适用并且有足够可利用数据和其他信息支持的估值技术。使用的估值技术主要包括市场法、收益法和成本法。

#### （十六）预计负债

如果与或有事项相关的义务是公司承担的现时义务，且该义务的履行很可能导致经济利益流出公司，以及有关金额能够可靠地计量，则公司会确认预计负债。

预计负债按照履行相关现时义务所需支出的最佳估计数进行初始计量。对于

货币时间价值影响重大的，预计负债以预计未来现金流量折现后的金额确定。在确定最佳估计数时，公司综合考虑了与或有事项有关的风险、不确定性和货币时间价值等因素。所需支出存在一个连续范围，且该范围内各种结果发生的可能性相同的，最佳估计数按照该范围内的中间值确定；在其他情况下，最佳估计数分别下列情况处理：

或有事项涉及单个项目的，按照最可能发生金额确定；

或有事项涉及多个项目的，按照各种可能结果及相关概率计算确定。

公司在资产负债表日对预计负债的账面价值进行复核，并按照当前最佳估计数对该账面价值进行调整。

## （十七）收入

收入是公司在日常活动中形成的、会导致股东权益增加且与股东投入资本无关的经济利益的总流入。

财政部于 2017 年颁布了《企业会计准则第 14 号——收入（修订）》（“新收入准则”）。新收入准则取代了 2006 年颁布的《企业会计准则第 14 号——收入》及《企业会计准则第 15 号——建造合同》（“原收入准则”）。

公司自 2018 年 1 月 1 日至 2019 年 12 月 31 日止执行原收入准则，自 2020 年 1 月 1 日起执行新收入准则。

### 1、原收入准则

#### （1）收入确认的一般原则

收入在其金额及相关成本能够可靠计量、相关的经济利益很可能流入公司并且同时满足以下不同类型收入的其他确认条件时，予以确认。

##### ①销售商品收入

当同时满足上述收入的一般确认条件以及下述条件时，公司确认销售商品收入：

公司将商品所有权上的主要风险和报酬已转移给购货方；

公司既没有保留通常与所有权相联系的继续管理权，也没有对已售出的商品

实施有效控制。

公司按已收或应收的合同或协议价款的公允价值确定销售商品收入金额。

## ②提供劳务收入

公司按已收或应收的合同或协议价款的公允价值确定提供劳务收入金额。

在资产负债表日，劳务交易的结果能够可靠估计的，根据完工百分比法确认提供劳务收入，提供劳务交易的完工进度根据已经提供的劳务占应提供劳务总量的比例确定。

劳务交易的结果不能可靠估计的，如果已经发生的劳务成本预计能够得到补偿的，则按照已经发生的劳务成本金额确认提供劳务收入，并按相同金额结转劳务成本；如果已经发生的劳务成本预计不能够得到补偿的，则不确认提供劳务收入。

## ③让渡资产使用权收入

让渡资产使用权收入包括利息收入、使用费收入等，让渡资产使用权收入同时满足下列条件的，才能予以确认：

相关的经济利益很可能流入公司；

收入的金额能够可靠地计量。

利息收入是按借出货币资金的时间和实际利率计算确定的。

使用费收入按照有关合同或协议约定的收费时间和方法计算确定，如果合同或协议规定一次性收取使用费，且不提供后续服务的，应当视同销售该项资产一次性确认收入；提供后续服务的，应在合同或协议规定的有效期内分期确认收入。

## (2) 收入确认的具体方法

### ①训练数据定制服务收入

训练数据定制服务收入主要是指公司根据客户需求为客户提供训练数据定制服务而收取的收入。公司与客户通过合同或订单约定提供的服务内容、服务总量及交易总额。

公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提

供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。

公司确定完工进度的具体方法为：

(1) 对于一次交付、一次验收的训练数据定制服务，在验收之前，完工进度为零，在验收之后，完工进度为 100%；

(2) 对于多次交付、一次验收的训练数据定制服务，在验收之前，完工进度为零，在验收之后，完工进度为 100%；

(3) 对于多次交付、多次验收的训练数据定制服务，在资产负债表日根据已经验收的服务量占合同或订单要求提供的服务总量的比例来计算完工进度。例如对于某中英文女声语音合成项目，合同要求提供的训练数据定制服务总量为 50 小时，在资产负债表日如果已验收的训练数据定制服务量为 45 小时，则完工进度为  $45 \text{ 小时} / 50 \text{ 小时} = 90\%$ 。

## ②训练数据产品收入

训练数据产品收入主要是指公司授权客户永久使用公司已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。公司与客户签订合同约定授权客户使用的训练数据产品及其交易金额。

公司按照合同约定向客户提交训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。

## ③训练数据相关的应用服务收入

训练数据相关的应用服务收入主要是指公司根据客户需求为其提供应用程序的定制开发服务或者是应用程序的授权使用，以及配套硬件的销售(如有)等，而向客户收取的收入。公司与客户通过签订合同，约定向客户提供的应用程序的定制开发或授权使用，以及配套硬件销售(如有)等服务内容及交易金额。

公司按照合同约定向客户交付开发完成或者授权使用的应用软件和配套硬件(如有)，并经客户验收确认后，按照合同约定的交易金额确认收入。

## 2、新收入准则

### (1) 收入确认的一般原则

公司在履行了合同中的履约义务，即在客户取得相关商品或服务的控制权时，确认收入。

合同中包含两项或多项履约义务的，公司在合同开始日，按照各单项履约义务所承诺商品或服务的单独售价的相对比例，将交易价格分摊至各单项履约义务，按照分摊至各单项履约义务的交易价格计量收入。

交易价格是公司因向客户转让商品或服务而预期有权收取的对价金额，不包括代第三方收取的款项。公司确认的交易价格不超过在相关不确定性消除时累计已确认收入极可能不会发生重大转回的金额。

满足下列条件之一时，公司属于在某一时段内履行履约义务，否则，属于在某一时点履行履约义务：

客户在公司履约的同时即取得并消耗公司履约所带来的经济利益；

客户能够控制公司履约过程中在建的商品；

公司履约过程中所产出的商品具有不可替代用途，且公司在整个合同期间内有权就累计至今已完成的履约部分收取款项。

对于在某一时段内履行的履约义务，公司在该段时间内按照履约进度确认收入。履约进度不能合理确定时，公司已经发生的成本预计能够得到补偿的，按照已经发生的成本金额确认收入，直到履约进度能够合理确定为止。

对于在某一时点履行的履约义务，公司在客户取得相关商品或服务控制权时点确认收入。在判断客户是否已取得商品或服务控制权时，公司会考虑下列迹象：

公司就该商品或服务享有现时收款权利；

公司已将该商品的实物转移给客户；

公司已将该商品的法定所有权或所有权上的主要风险和报酬转移给客户；

客户已接受该商品或服务。

公司已向客户转让商品或服务而有权收取对价的权利（且该权利取决于时间

流逝之外的其他因素)作为合同资产列示,合同资产以预期信用损失为基础计提减值(参见(八)金融工具)。公司拥有的、无条件(仅取决于时间流逝)向客户收取对价的权利作为应收款项列示。公司已收或应收客户对价而应向客户转让商品或服务的义务作为合同负债列示。

## (2) 收入确认的具体方法

### ①训练数据定制服务收入

训练数据定制服务收入主要是指公司根据客户需求为客户提供训练数据定制服务而收取的收入。公司与客户通过合同或订单约定提供的服务内容、服务总量及交易总额。公司在业务执行过程中,以单元定制化训练数据作为交付和验收的最小单位。公司向客户提交的训练数据集包含多个独立的、可明确区分的单元定制化训练数据,即按不同数据形态(音频、视频、图片、文本等)呈现的每计量单位(小时及句或字词、图片张数等)的训练数据(包括数据文件及标注文件)。

发行人提供训练数据定制服务包括三种业务类型,即智能语音、计算机视觉和自然语言。三种业务类型下训练数据定制服务均以软件化的训练数据集为交付物,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。不同业务类型下的具体交付物分别为:(1)智能语音业务的交付物一般包括音频文件、标注文本文件和说明及技术文档;(2)计算机视觉业务的交付物一般包括图像和视频等数据文件、标注文件和说明及技术文档;(3)自然语言业务的交付物一般包括对话文本等数据文件、标注文本文件和说明及技术文档。

上述三种业务类型下的交付物内容一般都包含了数据文档(包含音频文件、图像和视频文件、自然语言对话文本文件等数据文件及标注文件)、说明文档及技术文档。数据文档是核心交付物,也是客户用来进行人工智能模型训练的文件。说明及技术文档一般作为附加文件同时提供给客户,作为数据文档相关标准、规范、使用的说明,不直接应用于客户的人工智能模型训练,客户不能单独受益。

例如,对于发行人向客户提供的10小时中文方言语音识别采集的数据定制服务项目,发行人需要向客户交付的训练数据集包括10小时中文方言音频和其对应的标注文本。

在新收入准则下,公司将每单元定制化训练数据识别为某一时点履行的单项

履约义务。公司评估认为，在将定制化训练数据向客户交付并经客户验收确认的时点，客户已取得相关商品的控制权，并满足收入确认的条件，公司的主要判断依据是基于控制权的定义以及结合以下的迹象：（1）根据合同条款约定或者法律规定，通常情况下，公司对已经交付并经客户验收认可的该部分工作量享有现时收款权，客户就该部分工作量负有现时付款义务；（2）数据在交付验收后，交付物的所有权即已转移给客户；（3）商品的实物在数据交付验收后即已转移给客户；（4）客户的验收表明其已接受该商品；（5）客户对于所交付并验收的定制化训练数据，可以自行决定如何以及何时用于训练人工智能模型并从中受益，数据交付验收后可能发生的减值或毁损等损失也由客户承担，因此客户享有与交付和验收的定制化训练数据相关的主要风险和报酬。

新收入准则下，公司将每单元定制化训练数据识别为单项履约义务。

#### A、服务内容划分为单元定制化训练数据的可行性

发行人提供训练数据定制服务的内容是向客户提供训练数据集，而训练数据集则由多个单元定制化训练数据组成。单元训练数据是指按不同数据形态呈现的每计量单位的训练数据，如 1 小时语音、1 张图片等。以上述举例说明的 10 小时中文方言语音识别采集的数据定制服务项目来进行说明，发行人需要向客户交付的训练数据集包括 10 小时中文方言音频和其对应的标注文本，其中，每 1 小时的中文方言音频及其标注文本构成一个单元定制化训练数据。

发行人与客户签订的合同或订单中对于训练数据的数量、单位均会明确进行单独约定，使得单元定制化训练数据可以被清晰划分。发行人在实际执行训练数据的定制化服务的过程中，也是以单元定制化训练数据作为交付和验收的最小单位。发行人每次交付和验收的数据文档内包含多个单元训练数据，由于每个单元定制化训练数据相对独立，它们之间也不会产生相互影响和重大整合的关系，例如每 1 小时的中文方言音频或每 1 张图片之间均为独立的数据文件。因此客户也不会对分次交付和验收的训练数据进行指定，发行人可以以单元定制化训练数据为基础对定制服务进行分次交付、分次验收，客户也可以将其分别用于人工智能训练。

综上所述，发行人将定制化服务内容划分为单元定制化训练数据是具有可行



性的。

## B、识别单项履约义务的依据

根据新收入准则的相关规定：履约义务，是指合同中企业向客户转让可明确区分商品的承诺。企业向客户承诺的商品同时满足下列条件的，应当作为可明确区分商品：（一）客户能够从该商品本身或从该商品与其他易于获得资源一起使用中受益。（二）企业向客户转让该商品的承诺与合同中其他承诺可单独区分。下列情形通常表明企业向客户转让该商品的承诺与合同中其他承诺不可单独区分：1、企业需提供重大的服务以将该商品与合同中承诺的其他商品整合成合同约定的组合产出转让给客户。2、该商品将对合同中承诺的其他商品予以重大修改或定制。3、该商品与合同中承诺的其他商品具有高度关联性。

根据前述对训练数据定制服务内容的阐释，并结合上述新收入准则的规定，发行人将每单元定制化训练数据识别为单项履约义务的具体依据如下：

a、交付和验收的单元定制化训练数据包含了核心的交付文件，即数据文档，因此客户可以将交付和验收后的每一单元定制化训练数据用于训练人工智能模型，并能够从该单元训练数据的使用中受益；

b、发行人转让每一单元定制化训练数据的承诺在合同中可以明确区分，主要是因为（1）在转让每一单元定制化训练数据之前，发行人无须提供重大服务以对单元定制化训练数据之间进行整合。以智能语音业务为例，在语料文本的设计阶段之后，采集、加工和质检的环节均可以以单元定制化训练数据为基础进行，并以单元定制化训练数据为基础进行数据交付，不论是对于同一发音人采集的不同文本的数据，还是对于不同发音人采集的同一文本的数据，发行人无须再提供重大服务，对生产后的不同数据之间进行重大整合；（2）发行人转让的每一单元定制化训练数据不会影响其他单元定制化训练数据或其他单元定制化训练数据做出重大修改。例如发行人在录制其他文本/发音人的内容时，不会对已录制完成的内容进行重大修改；（3）不同的单元定制化训练数据之间不存在高度关联性，发行人可以对不同的单元定制化训练数据分批提交，客户可以分批验收，而且分批提交已验收完成的数据可以用于客户的模型训练。

综合上述分析，发行人向客户提交的训练数据集包含多个独立的、可明确区

分的单元定制化训练数据，即按不同数据形态（音频、视频、图片、文本等）呈现的每计量单位（小时及句或字词、图片张数等）的训练数据（包括数据文件及标注文件），发行人根据新收入准则的规定将每单元定制化训练数据识别为单项履约义务的依据充分。

### C、交易价格分摊至各单项履约义务的方法

训练数据定制服务的合同或订单中一般包含多项履约义务，因此需要按照上述准则规定将交易价格分摊至各单项履约义务。公司与客户签订的训练数据定制服务的合同或订单中，通常会约定提供的同类型的单元定制化训练数据的数量和总价格，或约定提供的同类型的单元定制化训练数据的单价和数量。由于同类型单项履约义务具有相同的性质和特点，因此对于每一个同类型的单元定制化训练数据，单独售价也相同。公司采用平均分摊的方法将交易价格分摊至单项履约义务，即按照约定的同类型单元定制化训练数据总价格除以数据总量，计算结果作为各个单元定制化训练数据的交易价格。如果约定了单元定制化训练数据的单价，也可以直接将单价作为各个单元定制化训练数据的交易价格。

### ②训练数据产品收入

训练数据产品收入主要是指公司授权客户永久使用公司已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。公司与客户签订合同约定授权客户使用的训练数据产品及其交易金额。

公司不从事对训练数据产品有重大影响的后续活动，因此训练数据产品属于某一时点履行的知识产权许可授权，公司在向客户交付合同中约定的训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。

### ③训练数据相关的应用服务收入

训练数据相关的应用服务收入主要是指公司根据客户需求为其提供应用程序的定制开发服务或者是应用程序的授权使用，以及配套硬件的销售（如有）等，而向客户收取的收入。公司与客户通过签订合同，约定向客户提供的应用程序的定制开发或授权使用，以及配套硬件销售（如有）等服务内容及交易金额。

应用程序的定制开发或者授权许可、配套硬件的销售属于可明确区分的单项履约义务，并且属于在某一时点履行的履约义务；应用程序的定制开发或者授权

使用、以及配套硬件的销售，均是在向客户交付并经客户验收确认时确认收入，并按照合同约定的单项履约义务的交易价格确认收入金额。

### 3、不同交付和验收方式下训练数据定制服务取得的验收确认凭据和收入确认时点，完工进度和收入金额的确定方法以及对应定制服务成本的结转方法

在训练数据定制服务业务类型下，发行人为客户提供训练数据集定制服务，最终形成的训练数据集成品的知识产权由客户享有。该类型业务包含三类不同的交付和验收方式：

方式一：项目完成后向客户一次交付，客户一次性验收；

方式二：项目执行过程中分批交付数据给客户，客户收到最后一批数据后就所有数据整体验收；

方式三：项目执行过程中分批交付数据给客户，客户收到分批交付的数据后分次进行验收。

发行人根据主要合同条款约定来确定与客户签订的合同类型。

对于中小型的或交付难度较低的合同或订单，客户通常会在合同或订单中与发行人约定一次性交付定制化训练数据，并一次性进行验收。合同或订单条款通常表述为“发行人应在合同签订之日起特定时间内向客户一次性交付全部数据，客户自交付之日起在特定时间内进行验收”。因此在这种情况下，发行人与客户签订的合同类型即为“一次交付，一次验收”。

对于交付数据量较大的或交付难度较高的合同或订单，客户通常会在合同或订单中与发行人约定分批次提交定制化数据，并约定在收到分批次提交的定制化数据后分批进行验收。合同或订单条款通常表述为“发行人应在约定的多个日期向客户分别交付各阶段或一定比例的数据，客户在发行人每次交付数据之日起在特定时间内进行验收”。因此在这种情况下，发行人与客户签订的合同类型即为“多次交付，多次验收”。

对于交付数据量较大的或交付难度较高的合同或订单，部分大型客户会考虑到数据体量较大，质检验收耗费时间较长，虽然也会要求发行人分批提交定制化数据，但会在合同或订单中明确约定在收到全部数据后进行一次性验收。合同或

订单条款通常表述为“发行人应在约定的多个日期向客户分别交付各阶段或一定比例的数据，客户自发行人交付全部数据之日起在特定时间内进行验收”。因此在这种情况下，发行人与客户签订的合同类型即为“多次交付，一次验收”。

发行人实施新收入准则后，对于训练数据定制服务，发行人将每单元定制化训练数据识别为某一时点履行的单项履约义务，并评估认为在将定制化训练数据向客户交付并经客户验收确认的时点，客户已取得相关商品的控制权，并满足收入确认的条件。根据上述新收入准则下的收入确认会计政策，发行人对于上述三种类型合同的收入确认时点及方式具体分析如下：

对于“一次交付，一次验收”和“多次交付，一次验收”的合同，不论发行人是一次交付还是多次交付训练数据，客户均是在发行人交付完成全部合同约定的定制化训练数据后，对训练数据进行一次验收。根据发行人新收入准则下的收入确认会计政策，发行人只有在将数据向客户交付并经客户验收确认后，才满足收入确认的条件。因此虽然合同中的每单元定制化训练数据均为单项履约义务，但是在各单元训练数据已交付未验收的情况下，仍是不满足收入确认的条件的。因此对于该两类合同，收入确认的具体时点为合同约定的定制化训练数据全部交付且客户全部验收确认的时点，收入确认金额为合同约定的总交易金额。

对于“多次交付，多次验收”的合同，发行人分次交付训练数据后，客户会进行分次验收。且交付和验收的每批数据均由多个单元定制化训练数据构成，可以识别为多个单项履约义务。根据发行人新收入准则下的收入确认会计政策，在每一次的发行人分次交付数据且经客户验收确认后，该批数据包含的多个单项履约义务即满足了收入确认的条件。因此对于该类合同，收入确认的具体时点为发行人分次提交数据且经客户验收确认的时点，收入确认金额为分摊至该次交付和验收数据的合同金额。

综上所述，发行人在实施新收入准则后，三种类型合同均将每个单元定制化训练数据的提供作为单项履约义务，并属于某一时点履行的履约义务，按照时点法进行收入确认，在单项履约义务层面，三种合同的收入确认时点和方式并无差异；而在合同层面，对于“一次交付，一次验收”和“多次交付，一次验收”的合同，这两类合同的收入确认时点及方式相同，均在将合同约定的定制化数据全部交付客户且经客户验收确认的时点确认收入，且收入确认金额均为合同约定的

总交易金额。对于“多次交付，多次验收”的合同，在将分次提交的数据交付客户并经客户验收确认的时点确认收入，且收入确认金额为分摊至该次交付和验收数据的合同金额，因此收入确认时点及方式与上述两类合同存在差异。此外，公司在新收入准则实施前后对于上述三种类型合同，其各自服务工作量的确认方式并无差异，因此与原收入准则相比，公司在新收入准则下的收入确认时点以及确认金额均无差异。

上述三种交付和验收方式下训练数据定制服务的验收确认凭据、收入确认时点、完工进度、收入金额的确定方法、对应成本的结转方法具体如下表所示：

类型	交付和验收方式	取得的验收确认凭据	收入确认时点	完工进度	收入金额的确定方法	对应定制服务成本的结转方法
方式一	一次交付，一次验收	项目全部完成后，客户对全部交付训练数据进行验收后出具的验收确认文件	客户对提交的训练数据验收确认的时点	100%	收入金额为按照合同或订单约定的服务交易总额	该项目累计发生的支出全部结转至当期营业成本
方式二	多次交付，一次验收					
方式三	多次交付，多次验收(分次验收)	项目执行过程中，客户每次对分批交付的训练数据进行验收后出具的验收文件	客户对分批交付的训练数据每一次验收确认的时点	按照已经提交并验收的服务量占服务总量的比例确定	按照合同或订单约定的服务交易总额乘以完工进度扣除以前会计期间累计已确认提供的服务收入金额，确认当期的服务收入。	按照提供服务预计总成本乘以完工进度扣除以前期间累计已确认服务成本后的金额，结转当期营业成本

发行人以上三种交付和验收方式对应的具体合同条款举例如下：

交付和验收方式	具体合同条款摘录	
	交付	验收
一次交付，一次验收	卖方应在本合同签订之日起【】个工作日内向买方一次性交付全部合同货物。	系统软件全部达到指定现场后【】个工作日内，买方将根据双方确认的验收测试标准进行验收，卖方应在指定日期自费派遣人员到买方现场参加验收。
	乙方应在合同签订后的【】个工作日内完成第一条所约定的语音数据的录制工作并以硬盘或者 FTP 下载形式将数据交付给甲方。	甲方自项目录音交付之日起【】个工作日内根据数据验收规则对本次服务的质量进行验收。
多次交付，一次验收	乙方应按照附件中工期及里程碑中约定的工期，对甲方提供数据服务；并按照附件中的时间点之前将可交付物交付给甲方。	甲方自收到乙方提交的语音库所有数据之日起【】个自然日内对数据的质量进行验收确认。

交付和验收方式	具体合同条款摘录	
	交付	验收
	乙方应按约定的开发成果交付日期（“交付日”）如期交付各阶段的开发成果。	甲方应自乙方所有数据交付之日起【】日内，在甲方所在地或甲方指定的其他地点，依照附件中约定的验收标准和验收方法对开发成果进行验收。
多次交付，多次验收 （分次验收）	交付和验收 A 语种文本转语音数据库（女声），预计完成时间【】； 交付和验收 B 语种文本转语音数据库（女声），预计完成时间【】；……； 交付和验收 C 语种文本转语音数据库（女声）的全部资料，预计完成时间【】	交付和验收 A 语种文本转语音数据库（女声），预计完成时间【】； 交付和验收 B 语种文本转语音数据库（女声），预计完成时间【】；……； 交付和验收 C 语种文本转语音数据库（女声）的全部资料，预计完成时间【】
	交付时间以附件《项目任务书》为准，实际交付周期按实际收到的批次数据量经双方邮件确认为准。	甲方应自收到数据之日起【】个工作日内对数据的质量进行确认和验收。

注：【】为合同条款中具体的日期信息。

#### 4、各类业务的具体交付物及交付方式，各类型收入确认取得的具体验收确认凭据

##### （1）各类业务的具体交付物及交付方式

##### ①训练数据定制服务和训练数据产品

发行人两大主要业务类型训练数据定制服务和训练数据产品覆盖了智能语音、计算机视觉及自然语言处理三大 AI 核心领域。前述服务和产品均以软件化的训练数据集为核心，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。不同业务领域下的具体交付物总结如下：

A、智能语音业务的交付物一般包括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；

B、计算机视觉业务的交付物一般包括图像和视频等数据文件、标注文件和说明及技术文档；

C、自然语言业务的交付物一般包括对话文本等数据文件、标注文本文件和说明及技术文档。

上述业务类型的实际交付方式主要为以 FTP 传输方式交付给客户。

## ②训练数据相关的应用服务

发行人提供算法模型相关训练服务，交付物一般包括开发完成或者授权使用的应用软件和配套硬件（如有），交付方式主要为以 FTP 传输方式交付给客户或交付至客户指定地点并现场安装调试。

### (2) 各类型收入确认取得的具体验收确认凭据

#### ①训练数据定制服务和训练数据产品

发行人训练数据定制服务的各类型收入确认方式，包括三种类型，即（1）一次交付，一次验收；（2）多次交付，一次验收；（3）多次交付，多次验收（分次验收）。训练数据产品已经开发完成，可多次销售并获取授权许可收入，收入确认方式为产品交付验收后一次确认。

上述收入确认方式下，发行人取得的具体验收确认凭据均为客户的验收邮件。项目完成或阶段性工作完成后，发行人项目负责人将服务或产品交付客户，并向客户项目负责人邮件发送验收申请，验收申请附有项目验收报告，列明项目/产品名称及编号、服务/产品内容、本次验收的数据量等具体待验收信息。客户会按照合同约定的技术要求和验收标准对交付的数据进行验收，完成验收后，正式邮件回复确认服务或产品已交付，验收内容无误。

发行人训练数据定制服务和训练数据产品主要客户为长期合作的大型科技公司、人工智能企业及科研机构，客户选择邮件验收主要是考虑到办公习惯及工作效率方面的原因，且邮件验收与签字盖章的验收单同样具有验收作用，符合行业惯例。

## ②训练数据相关的应用服务

应用软件的定制开发或者授权使用、以及配套硬件的销售，由于通常涉及简单的安装调试等现场实施服务工作，一般现场取得盖章纸质验收证明作为收入确认凭据；如无需现场实施服务，则采取邮件验收的方式。

## 5、发行人关于确定合同类型、识别单项履约义务等收入确认相关的内部控制措施情况

发行人关于确定合同类型、识别单项履约义务等收入确认相关的内控措施具

体如下：

#### （1）合同签订环节

发行人的客户会根据自身训练数据服务需求向发行人发起服务采购邀请，经双方商务谈判协商一致后，签订销售合同或订单。销售合同或订单中会详细列示要求发行人提供的定制化训练数据内容、数据量和价格，并对定制化训练数据交付和验收的方式、验收的标准和要求做出明确的规定。合同或订单条款中的交付和验收的方式中会对一次交付或者分次交付，以及一次验收或者分次验收进行具体规定。

发行人的销售合同需要经销售分管负责人、业务部门负责人、财务部、法务人员、总经理逐级审批。重要合同须经总经理办公会审议。

#### （2）项目立项环节

在合同或订单签订后，发行人对训练数据定制服务采用项目制管理，并在业务管理平台上对训练数据定制服务进行项目进度跟踪管理，记录各项目的交付和验收的进展情况，实现对项目的有效监控。

在项目立项阶段，由执行项目的项目负责人在业务管理平台中填写项目信息，除项目组成员、工作内容、实施方案、项目预算等项目计划内容外，还主要包括用于确定合同类型、识别单项履约义务相关的关键信息，如项目合同编号、项目交付物、合同或订单约定的单元定制化训练数据单位及数据量、项目金额、项目周期要求等，并由部门经理、业务总监、财务审核岗依次审核后，项目立项成功。立项过程中，项目负责人会在业务管理平台上传签订的合同或订单作为支持性依据。

#### （3）服务交付环节

在项目执行过程中，项目负责人按照与客户的约定，将定制化训练数据进行分次交付或者全部一次交付，交付方式主要为以 FTP 传输方式交付。项目负责人会在每次交付数据后，同时以邮件的形式通知客户具体交付情况，并会在业务管理平台中登记本次交付时间，并上传交付通知邮件作为交付依据。

#### （4）服务验收环节



对于发行人已经交付的定制化训练数据，项目负责人根据与客户约定的一次验收或者多次验收的时点，向客户发送验收申请的邮件。验收申请附有项目验收报告，列明项目/产品名称及编号、服务/产品内容、本次验收的数据量等具体待验收信息，其中验收数据量是按照单元定制化训练数据为单位进行计算。客户完成验收后，正式邮件回复确认服务已交付，验收内容无误。在客户对产品及服务进行验收确认后，项目负责人在业务管理平台中填写项目验收具体信息，包括验收数据量、验收时间和验收金额等，其中验收数据量是按照单元定制化训练数据为单位进行计算，并需与合同或订单要求的单位一致。同时项目负责人也会将客户的验收确认邮件作为验收依据上传至业务管理平台。

业务部门负责人、财务部相关人员审核业务管理平台记载的验收信息的准确性和完整性，并检查验收信息与上传的验收依据是否一致。

#### (5) 财务审核

财务人员每季度末从业务管理平台导出训练数据定制服务项目的明细，对于系统中状态为执行中的项目，财务人员与业务人员逐一确认项目进度，并与业务管理平台中留存的合同与订单、交付依据、验收依据等支持性依据进行核对，检查是否存在交付与验收进度与支持性依据不符的情形。

综上所述，发行人关于确定合同类型、识别单项履约义务等收入确认相关的内控措施建立健全并有效执行。

### 6、同行业公司类似业务采用的收入确认政策以及与发行人之间的对比情况

代码	信息技术服务类上市公司	收入确认政策
300290.SZ	荣科科技	<p>本公司提供劳务主要包括定制化软件开发收入和技术服务收入，其收入的具体确认原则如下：</p> <p>(1) 定制化软件开发收入是指公司接受客户委托，针对客户提出的软件需求进行研究开发所获得的收入。定制化软件一般是公司基于自主研发的软件平台基础上，按照客户的特定需求，进行定制化开发而形成的应用软件。该类软件一般不具有通用性，该类业务实质上属于提供劳务。</p> <p>对于专业程度较高且能够取得客户确认的完工进度依据的定制化软件开发收入，提供定制软件劳务交易的结果在资产负债表日能够可靠估计的，采用完工百分比法确认提供劳务的收入，并按已经发生的成本占估计总成本的比例确定提供劳务交易的完工进度；提供定制软件劳务交易的结果在资产负债表日不能够可靠估计的，若已经发生的劳务成本预计能够</p>

代码	信息技术服务类上市公司	收入确认政策
		<p>得到补偿，按已经发生的劳务成本金额确认提供劳务收入，并按相同金额结转劳务成本；若已经发生的劳务成本预计不能够得到补偿，将已经发生的劳务成本计入当期损益，不确认劳务收入。</p> <p>对于其他的定制化软件开发收入，本公司在按照合同约定内容实施了开发工作，为购买方完成了安装、调试工作，取得购买方的验收，并同时满足上述提供劳务收入确认条件时确认定制化软件开发收入。</p>
300245.SZ	天玑科技	<p>数据中心 IT 基础设施服务收入确认的一般原则是：根据合同约定提供了相应服务，在相关的经济利益能够流入公司，收入的金额能够可靠地计量时，相关成本能够可靠计量时确认收入。软硬件销售及自有产品销售收入确认的一般原则是：在销售产品时的主要风险和报酬已转移给买方，不再保留与之相联系的管理权和控制权，在相关的经济利益能够流入公司，收入的金额能够可靠地计量时，相关成本能够可靠地计量时确认收入。根据合同约定条款的不同，分别按以下原则确认收入：</p> <p>(1) 合同明确约定服务期限的，在合同约定的服务期限内按时间进度确认收入。</p> <p>(2) 合同明确约定验收条款的，按约定条款验收并确认收入。</p> <p>(3) 合同明确约定开发项目进度的确认条款的，期末按照提供定制软件开发收入的合同金额乘以经客户确认的开发项目进度的量化百分比扣除以前会计期间累计已确认提供劳务收入后的金额，确认当期提供定制软件开发收入；同时，按照提供劳务估计总成本乘以经客户确认的开发项目进度的量化百分比扣除以前会计期间累计已确认劳务成本后的金额，结转当期劳务成本。</p>
300634.SZ	彩讯股份	<p>提供劳务收入确认的原则，本公司在资产负债表日提供劳务交易的结果能够可靠估计的，按照完工百分比法确认提供劳务收入。</p> <p><b>B、定制工程化软件</b></p> <p>合同中明确约定合同总金额中包含质保金的，在确认收入时，先扣除质保金部分，然后按如下方法确认收入；质保期届满后，再将该质保金确认收入。</p> <p>a. 取得初验报告开始确认收入。</p> <p>定制工程化软件一般包括软件功能的开发、现场安装调试、旧系统数据迁移、系统培训、用户测试、初验（系统试运行）和终验等环节，其中初验和终验是客户对系统运行情况作出的评价（用户也可直接终验），也是公司与客户结算的主要依据和环节。在初验阶段完成时，围绕项目的大部分工作均已完成，之后只是在运行期间可能进行部分系统功能的修改，所涉工作内容很少、工作量也不大。因此，公司在取得初验报告后开始确认收入。</p> <p>b. 初验确认收入的比例。</p> <p>由于公司定制工程化软件主要依靠项目实施人员及自有软件技术的整合来进行开发，因此，公司以投入工时为计算基础，以完工进度作为确认收入的依据。根据历年已完工典型合同统计，初验时项目投入工时约为 80%左右，因此，公司按</p>

代码	信息技术服务类上市公司	收入确认政策
		<p>照以下标准孰低的原则确定初验收入的确认比例：</p> <p>I、初验时合同约定的付款比例等于或高于 80%，则以 80% 确认收入；</p> <p>II、初验时合同约定的付款比例低于 80%，则以合同约定的比例确认收入。</p> <p>公司每年年末会根据当年完工合同情况，对初验完工进度进行复核。</p> <p>c. 终验时的收入确认：</p> <p>取得客户终验报告确认时，公司扣除终验前已确认的比例后，将其余部分确认收入，如合同中明确约定合同总金额中包含质保金的，先扣除质保金部分，再按前述办法确认收入。</p>
300468.SZ	四方精创	<p>(1) 软件开发收入的确认原则及方法</p> <p>软件开发收入是指接受客户委托，根据客户的需要，对应用软件技术进行研究开发，由此开发出来的软件为定制软件、不具有通用性。</p> <p>软件开发收入的确认原则及方法分为两大类：</p> <p>①根据合同按每月实际工作量（人/天数）收费，经客户确认实际工作量后确认软件开发收入。</p> <p>②公司软件开发业务按照合同约定提交软件开发成果并经客户验收后确认收入，该业务若存在分阶段多次验收情况，则按该业务完成各阶段工作并经客户验收时，分别确认相应阶段工作成果的收入。</p>
	发行人	<p>公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。</p>

上述内容是对原收入准则下发行人与同行业公司的收入确认政策的对比情况。

通过对比发现，从事信息技术服务类业务的上市公司中，通常将定制化软件开发收入定义为提供劳务收入，在资产负债表日提供劳务交易的结果能够可靠估计的，按照完工百分比法确认提供劳务收入。完工百分比的比例通常采取工作量法、已发生成本占预计总成本的比例确定、已经提供的劳务占应提供劳务总量的比例确定。公司根据自身业务情况确定的验收及收入确认方式符合行业惯例，与同行业公司之间不存在显著差异。

代码	信息技术服务类上市公司	收入确认政策
300290.SZ	荣科科技	<p>本公司提供劳务主要包括定制化软件开发收入和技术服务收入，其收入的具体确认原则如下：</p> <p>A、定制化软件开发收入是指公司接受客户委托，针对客户提出的软件需求进行研究开发所获得的收入。定制化软件一</p>

代码	信息技术服务类上市公司	收入确认政策
		<p>般是公司基于自主研发的软件平台基础上，按照客户的特定需求，进行定制化开发而形成的应用软件。该类软件一般不具有通用性，该类业务实质上属于提供劳务。</p> <p>对于定制化软件开发收入，基于合同条款及适用于合同的法律规定，公司可选择在某一时段内确认收入或者某一时点确认收入，当客户能够控制公司履约过程中在建的商品或者公司履约过程中所产出的商品具有不可替代用途，且公司在整个合同期间内有权就累计至今已完成的履约部分收取款项时，公司选择在某一时段内确认收入，否则，公司选择在某一时点确认收入。</p>
300245.SZ	天玑科技	<p>满足下列条件之一的，属于在某一时段内履行履约义务，否则，属于在某一时点履行履约义务：</p> <ul style="list-style-type: none"> <li>•客户在本公司履约的同时即取得并消耗本公司履约所带来的经济利益。</li> <li>•客户能够控制本公司履约过程中在建的商品。</li> <li>•本公司履约过程中所产出的商品具有不可替代用途，且本公司在整个合同期内有权就累计至今已完成的履约部分收取款项。</li> </ul> <p>对于在某一时段内履行的履约义务，本公司在该段时间内按照履约进度确认收入，但是，履约进度不能合理确定的除外。本公司考虑商品或服务的性质，采用产出法或投入法确定履约进度。当履约进度不能合理确定时，已经发生的成本预计能够得到补偿的，本公司按照已经发生的成本金额确认收入，直到履约进度能够合理确定为止。</p> <p>对于在某一时点履行的履约义务，本公司在客户取得相关商品或服务控制权时点确认收入。在判断客户是否已取得商品或服务控制权时，本公司考虑下列迹象：</p> <p>本公司就该商品或服务享有现时收款权利，即客户就该商品或服务负有现时付款义务。</p> <p>本公司已将该商品的法定所有权转移给客户，即客户已拥有该商品的法定所有权。</p> <ul style="list-style-type: none"> <li>•本公司已将该商品实物转移给客户，即客户已实物占有该商品。</li> <li>•本公司已将该商品所有权上的主要风险和报酬转移给客户，即客户已取得该商品所有权上的主要风险和报酬。</li> <li>•客户已接受该商品或服务。</li> </ul>
300634.SZ	彩讯股份	<p>本公司在履行了合同中的履约义务，即在客户取得相关商品或服务控制权时确认收入。</p> <p>b、定制工程化软件</p> <p>定制工程化软件一般包括软件功能的开发、现场安装调试、旧系统数据迁移、系统培训、用户测试、初验（系统试运行）和终验等环节，其中初验和终验是客户对系统运行情况作出的评价（客户也可直接终验）。</p> <p>公司完成定制工程化软件开发或实施工作后，在取得客户的终验报告时确认为销售收入。</p>
300468.SZ	四方精创	<p>公司在履行了合同中的履约义务，即在客户取得相关商品控制权时确认收入。</p> <p>①软件开发合同</p>

代码	信息技术服务类上市公司	收入确认政策
		公司与客户之间的软件开发服务合同包含为客户进行定制化软件开发的履约义务，由于开发成果在客户验收前，客户无法取得并消耗公司履约所带来的经济利益，无法控制公司履约过程中在建的商品，公司也无权就累计至今已完成的履约部分收取款项，因此公司将其作为在某一时点内履行的履约义务，在客户验收后进行收入确认。
	发行人	每单元定制化训练数据被识别为单项履约义务，属于某一时点履行的履约义务，公司在将定制化训练数据向客户交付并经客户验收确认时确认收入。

发行人与同行业公司自 2020 年 1 月 1 日开始实施新收入准则，上述内容是对新收入准则下发行人与同行业公司的收入确认政策的对比情况。

根据上述对比情况，同行业公司中的荣科科技和天玑科技未明确披露新收入准则下的软件开发服务的收入确认具体方法。同行业公司中的彩讯股份和四方精创明确披露了在新收入准则下其将软件开发服务作为某一时点履行的履约义务，并按照客户验收的时点确认收入，该收入确认的方法与发行人在新准则下采取的收入确认方法不存在显著差异。

## 7、训练数据定制服务成本的核算流程和结转方法

报告期内发行人的训练数据定制服务成本的核算流程和结转方法如下：

报告期内，发行人训练数据定制服务的各项支出，包括职工薪酬、数据服务费、折旧与摊销等，均按照所属项目在“制造费用”中进行归集与核算。在项目分批验收或一次全部验收并相应确认收入后，将该项目归集的项目成本中与确认的收入对应的部分，由制造费用结转至营业成本。

按项目归集成本：发行人严格按照项目为单位进行成本归集，在项目执行之前，每个项目需要编制项目预算，根据训练数据定制服务的具体内容，对项目所需人员及每人投入工时等直接人工内容，采集数量、采集单价、标注工时、标注单价等数据服务采购内容，折旧与摊销、租赁等项目所需场地、设备内容，差旅、交通、办公等其他项目相关支出内容制定详细预算。项目执行过程中，所有支出填写和报销均比照预算类型执行，对预算变更、未及预算、超预算等情况设定完整的审批流程，并制定了严格的项目决算要求，在项目申请决算前，所有支出必须填写和报销完毕，经专人审核后项目进行项目决算。决算后，项目不允许发生额外支出。通过以上内部控制手段，保证每笔支出均可以按项目准确、完整归集。

成本结转：具体表现为两种情形，一种是在项目分批验收并确认收入的情况下，在每次验收并确认收入的同时，按照已经验收的服务量占服务总量的比例乘以预计项目总成本并扣除以前期间累计已结转成本后的金额来结转当期营业成本；另一种是在项目一次全部验收并确认收入的情况下，直接进行项目决算，按照实际发生和归集的项目总成本全部结转至营业成本。制造费用中未结转至营业成本的部分，在财务报表中列示和披露为“存货-在产品”。

报告期内，各相关费用和人工成本在各个项目上的归集和分配的方法如下：

#### （1）职工薪酬

职工薪酬按照工时积分制归集到各个项目。每月业务人员按其实际工作情况将工时积分填写到相应项目上，每人每月填写至所有项目（包括训练数据定制服务项目和训练数据产品开发项目等）的工时积分共计为8分/日\*月工作日。各个项目应分摊的人工成本的计算方式为：各业务人员填写在该项目的积分\*该人员所属级别的人工成本单价=该项目的人工支出。

#### （2）数据服务费

数据服务费指训练数据定制服务项目所需的原料数据采集与标注服务采购支出，业务人员按项目实际发生情况进行支出申请和报销。在进行支出申请和报销时，必须填报项目编号、涉及环节、采购量、采购单价、报销金额、供应商名称等关键信息，保证发生额直接归集到单个项目上，不得与其他项目存在交叉。因此数据服务费支出根据上述项目编号信息直接归集和分配至各个项目上。

#### （3）折旧与摊销

对于各项目使用的设备和软件的折旧和摊销费用，按各项目领用该设备、使用该软件时间的长短，由系统自动计算在各个项目上的分摊金额。计算方式为：项目折旧与摊销支出=项目领用天数/365天\*设备全年折旧额或无形资产全年摊销额。

#### （4）差旅费等其他费用

差旅费等其他费用按项目实际发生情况报销，在报销时，必须选择项目编号及费用类别，费用发生额直接归集到单个项目上，不得与其他项目存在交叉。因

此差旅费等其他费用支出也是根据项目编号信息直接归集和分配至各个项目上。

## （十八）职工薪酬

### 1、短期薪酬

公司在职工提供服务的会计期间，将实际发生或按规定的基准和比例计提的职工工资、奖金、医疗保险费、工伤保险费和生育保险费等社会保险费和住房公积金，确认为负债，并计入当期损益或相关资产成本。

### 2、离职后福利 - 设定提存计划

公司所参与的设定提存计划是按照中国有关法规要求，公司职工参加的由政府机构设立管理的社会保障体系中的基本养老保险。基本养老保险的缴费金额按国家规定的基准和比例计算。公司在职工提供服务的会计期间，将应缴存的金额确认为负债，并计入当期损益或相关资产成本。

### 3、辞退福利

公司在职工劳动合同到期之前解除与职工的劳动关系，或者为鼓励职工自愿接受裁减而提出给予补偿的建议，在下列两者孰早日，确认辞退福利产生的负债，同时计入当期损益：

公司不能单方面撤回因解除劳动关系计划或裁减建议所提供的辞退福利时；

公司有详细、正式的涉及支付辞退福利的重组计划；并且，该重组计划已开始实施，或已向受其影响的各方通告了该计划的主要内容，从而使各方形成了对公司将实施重组的合理预期时。

## （十九）政府补助

政府补助是公司从政府无偿取得的货币性资产或非货币性资产，但不包括政府以投资者身份向公司投入的资本。

政府补助在能够满足政府补助所附条件，并能够收到时，予以确认。

政府补助为货币性资产的，按照收到或应收的金额计量。政府补助为非货币性资产的，按照公允价值计量。

公司取得的、用于购建或以其他方式形成长期资产的政府补助作为与资产相

关的政府补助。公司取得的与资产相关之外的其他政府补助作为与收益相关的政府补助。与资产相关的政府补助，公司将其确认为递延收益，并在相关资产使用寿命内按照合理、系统的方法分期计入其他收益或营业外收入。与收益相关的政府补助，如果用于补偿公司以后期间的相关成本费用或损失的，公司将其确认为递延收益，并在确认相关成本费用或损失的期间，计入其他收益或营业外收入；否则直接计入其他收益或营业外收入。

## （二十）所得税

除因企业合并和直接计入所有者权益（包括其他综合收益）的交易或者事项产生的所得税外，公司将当期所得税和递延所得税计入当期损益。

当期所得税是按本年度应税所得额，根据税法规定的税率计算的预期应交所得税，加上以往年度应付所得税的调整。

资产负债表日，如果公司拥有以净额结算的法定权利并且意图以净额结算或取得资产、清偿负债同时进行，那么当期所得税资产及当期所得税负债以抵销后的净额列示。

递延所得税资产与递延所得税负债分别根据可抵扣暂时性差异和应纳税暂时性差异确定。暂时性差异是指资产或负债的账面价值与其计税基础之间的差额，包括能够结转以后年度的可抵扣亏损和税款抵减。递延所得税资产的确认以很可能取得用来抵扣可抵扣暂时性差异的应纳税所得额为限。

如果不属于企业合并交易且交易发生时既不影响会计利润也不影响应纳税所得额（或可抵扣亏损），则该项交易中产生的暂时性差异不会产生递延所得税。商誉的初始确认导致的暂时性差异也不产生相关的递延所得税。

资产负债表日，公司根据递延所得税资产和负债的预期收回或结算方式，依据已颁布的税法规定，按照预期收回该资产或清偿该负债期间的适用税率计量该递延所得税资产和负债的账面金额。

资产负债表日，公司对递延所得税资产的账面价值进行复核。如果未来期间很可能无法获得足够的应纳税所得额用以抵扣递延所得税资产的利益，则减记递延所得税资产的账面价值。在很可能获得足够的应纳税所得额时，减记的金额予以转回。



资产负债表日，递延所得税资产及递延所得税负债在同时满足以下条件时以抵销后的净额列示：

纳税主体拥有以净额结算当期所得税资产及当期所得税负债的法定权利；

递延所得税资产及递延所得税负债是与同一税收征管部门对同一纳税主体征收的所得税相关或者是对不同的纳税主体相关，但在未来每一具有重要性的递延所得税资产及负债转回的期间内，涉及的纳税主体意图以净额结算当期所得税资产和负债或是同时取得资产、清偿负债。

#### （二十一）经营租赁

经营租赁租入资产的租金费用在租赁期内按直线法确认为相关资产成本或费用。或有租金在实际发生时计入当期损益。

#### （二十二）股利分配

资产负债表日后，经审议批准的利润分配方案中拟分配的股利或利润，不确认为资产负债表日的负债，在附注中单独披露。

#### （二十三）关联方

一方控制、共同控制另一方或对另一方施加重大影响，以及两方或两方以上同受一方控制、共同控制的，构成关联方。关联方可为个人或企业。仅仅同受国家控制而不存在其他关联方关系的企业，不构成关联方。

此外，公司同时根据中国证监会颁布的《上市公司信息披露管理办法》确定公司或公司的关联方。

#### （二十四）主要会计估计及判断

编制财务报表时，公司管理层需要运用估计和假设，这些估计和假设会对会计政策的应用及资产、负债、收入及费用的金额产生影响。实际情况可能与这些估计不同。公司管理层对估计涉及的关键假设和不确定因素的判断进行持续评估，会计估计变更的影响在变更当期和未来期间予以确认。

除固定资产及无形资产等资产的折旧及摊销和各类资产减值涉及的会计估计外，其他主要的会计估计如下：

- 1、递延所得税资产的确认；
- 2、金融工具公允价值估值。

## 五、重要会计政策、会计估计的变更及其影响

### （一）重要会计政策变更

#### 1、财务报表列报

财政部于 2019 年颁布了《关于修订印发 2019 年度一般企业财务报表格式的通知》（财会[2019]6 号）和《关于修订印发合并财务报表格式（2019 版）的通知》（财会[2019]16 号），公司已根据财会[2019]6 号和财会[2019]16 号规定的财务报表格式编制整个报告期内财务报表。

#### 2、新金融工具准则

财政部于 2017 年颁布了《企业会计准则第 22 号——金融工具确认和计量（修订）》、《企业会计准则第 23 号——金融资产转移（修订）》、《企业会计准则第 24 号——套期会计（修订）》及《企业会计准则第 37 号——金融工具列报（修订）》（统称“新金融工具准则”）。公司自 2019 年 1 月 1 日起执行新金融工具准则。

新金融工具准则修订了财政部于 2006 年颁布的《企业会计准则第 22 号——金融工具确认和计量》、《企业会计准则第 23 号——金融资产转移》和《企业会计准则第 24 号——套期保值》以及财政部于 2014 年修订的《企业会计准则第 37 号——金融工具列报》（统称“原金融工具准则”）。

新金融工具准则将金融资产划分为三个基本分类：（1）以摊余成本计量的金融资产；（2）以公允价值计量且其变动计入其他综合收益的金融资产；及（3）以公允价值计量且其变动计入当期损益的金融资产。在新金融工具准则下，金融资产的分类是基于公司管理金融资产的商业模式及该资产的合同现金流量特征而确定。新金融工具准则取消了原金融工具准则中规定的持有至到期投资、贷款和应收款项及可供出售金融资产三个分类类别。根据新金融工具准则，嵌入衍生工具不再从金融资产的主合同中分拆出来，而是将混合金融工具整体适用关于金融资产分类的相关规定。

新金融工具准则以“预期信用损失”模型替代了原金融工具准则中的“已

发生损失”模型。“预期信用损失”模型要求持续评估金融资产的信用风险，因此在新金融工具准则下，公司信用损失的确认时点早于原金融工具准则。

公司按照新金融工具准则的衔接规定，对新金融工具准则施行日（即 2019 年 1 月 1 日）未终止确认的金融工具的分类和计量（含减值）进行追溯调整。公司未调整比较财务报表数据，将金融工具的原账面价值和在新金融工具准则施行日的新账面价值之间的差额计入 2019 年年初留存收益或其他综合收益。

（1）以按照财会 [2019] 6 号和财会 [2019] 16 号规定追溯调整后的 2018 年 12 月 31 日财务报表为基础，执行新金融工具准则对 2019 年 1 月 1 日合并资产负债表受影响的项目的影响汇总如下：

单位：元

项目	2018 年 12 月 31 日	2019 年 1 月 1 日	调整数
交易性金融资产	-	73,000,000.00	73,000,000.00
其他流动资产	84,927,609.87	11,927,609.87	-73,000,000.00
<b>合计</b>	<b>84,927,609.87</b>	<b>84,927,609.87</b>	<b>-</b>

## （2）金融工具的分类影响

公司持有的保本浮动收益型理财产品，其收益取决于标的资产的收益率，于 2019 年 1 月 1 日，公司将保本浮动收益型理财产品重分类为以公允价值计量且其变动计入当期损益的金融资产，列报为交易性金融资产。

2019 年 1 月 1 日，公司没有将任何金融资产或金融负债指定为以公允价值计量且其变动计入当期损益的金融资产或金融负债，也没有撤销之前的指定。

将金融资产按照原金融工具准则和新金融工具准则的规定进行分类和计量的结果对比如下：

原金融工具准则 (2018 年 12 月 31 日)			新金融工具准则 (2019 年 1 月 1 日)		
项目	计量类别	账面价值 (元)	项目	计量类别	账面价值 (元)
货币资金	摊余成本 (贷款和应收款项)	46,691,570.86	货币资金	摊余成本	46,691,570.86
应收账款	摊余成本 (贷款和应收款项)	29,389,643.27	应收账款	摊余成本	29,389,643.27
其他应收款	摊余成本 (贷款和应收款项)	1,728,438.23	其他应收款	摊余成本	1,728,438.23
其他流动资产	摊余成本 (贷款和应收款项)	10,000,000.00	其他流动资产	摊余成本	10,000,000.00

原金融工具准则 (2018年12月31日)			新金融工具准则 (2019年1月1日)		
项目	计量类别	账面价值(元)	项目	计量类别	账面价值(元)
其他流动资产-可供出售金融资产	以公允价值计量其变动计入其他综合收益	73,000,000.00	交易性金融资产	以公允价值计量且其变动计入当期损益(准则要求)	73,000,000.00

(3) 采用“预期信用损失”模型的影响:

“预期信用损失”模型适用于公司以摊余成本计量的金融资产,采用“预期信用损失”模型对公司2019年1月1日的财务状况和经营成果未产生重大影响。

### 3、新收入准则

财政部于2017年颁布了《企业会计准则第14号——收入(修订)》(“新收入准则”),公司自2020年1月1日起执行新收入准则,对会计政策相关内容进行调整。

新收入准则取代了财政部于2006年颁布的《企业会计准则第14号——收入》及《企业会计准则第15号——建造合同》(统称“原收入准则”)。

在原收入准则下,公司以风险报酬转移作为收入确认时点的判断标准。公司销售商品收入在同时满足下列条件时予以确认,即:商品所有权上的主要风险和报酬已转移给购货方,收入的金额及相关成本能够可靠计量,相关的经济利益很可能流入公司,公司既没有保留通常与所有权相联系的继续管理权,也没有对已售出的商品实施有效控制。提供劳务收入和建造合同收入按照资产负债表日的完工百分比法进行确认。

在新收入准则下,公司以控制权转移作为收入确认时点的判断标准:

公司在履行了合同中的履约义务,即在客户取得相关商品或服务的控制权时,确认收入。在满足一定条件时,公司属于在某一时段内履行履约义务,否则,属于在某一时点履行履约义务。合同中包含两项或多项履约义务的,公司在合同开始日,按照各单项履约义务所承诺商品或服务的单独售价的相对比例,将交易价格分摊至各单项履约义务,按照分摊至各单项履约义务的交易价格计量收入。交易价格是公司因向客户转让商品或服务而预期有权收取的对价金额,不包括代第三方收取的款项。公司确认的交易价格不超过在相关不确定性消除时累计已确认收入极可能不会发生重大转回的金额。合同中存在重大融资成分的,公司按照假

定客户在取得商品或服务控制权时即以现金支付的应付金额确定交易价格。该交易价格与合同对价之间的差额，在合同期间内采用实际利率法摊销。

公司依据新收入准则有关特定事项或交易的具体规定调整了相关会计政策。例如：合同成本、附有销售退回条款的销售、附有质量保证条款的销售、主要责任人和代理人的区分、附有客户额外购买选择权的销售、授予知识产权许可、售后回购、客户未行使权利、无需退回的初始费的处理等。

公司依据新收入准则的规定，根据履行履约义务与客户付款之间的关系在资产负债表中列示合同资产或合同负债。同时，公司依据新收入准则对与收入相关的信息披露要求提供更多披露，例如相关会计政策、有重大影响的判断（可变对价的计量、交易价格分摊至各单项履约义务的方法、估计各单项履约义务的单独售价所用的假设等）、与客户合同相关的信息（本期收入确认、合同余额、履约义务等）、与合同成本有关的资产的信息等。

采用新收入准则对公司会计政策的具体影响如下：

对于训练数据定制服务收入，主要是指公司根据客户需求为客户提供训练数据定制服务而收取的收入，公司与客户通过合同或订单约定提供的服务内容、服务总量及交易总额。在原收入准则下，公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。在新收入准则下，每单元定制化训练数据被识别为单项履约义务，属于某一时点履行的履约义务，公司在将定制化训练数据向客户交付并经客户验收确认时确认收入。采用新收入准则之后，对于公司收入确认时点和收入确认金额没有影响。

对于训练数据产品收入，主要是指公司授权客户永久使用公司已经开发完成并拥有所有权的训练数据产品而向客户收取的收入，公司与客户签订合同约定授权客户使用的训练数据产品及其交易金额。在原收入准则下，公司按照合同约定向客户提交训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。在新收入准则下，公司不从事对训练数据产品有重大影响的后续活动，因此训练数据产品属于某一时点履行的知识产权许可授权，公司在向客

户交付合同中约定的训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。采用新收入准则之后，对于公司收入确认时点和收入确认金额没有影响。

对于训练数据相关的应用服务收入，主要是指公司根据客户需求为其提供应用软件的定制开发服务或者是应用软件的授权使用，以及配套硬件的销售（如有）等，而向客户收取的收入，公司与客户通过签订合同，约定向客户提供的应用软件的定制开发或授权使用，以及配套硬件销售（如有）等服务内容及交易金额。在原收入准则下，公司按照合同约定向客户交付开发完成或者授权使用的应用软件和配套硬件（如有），并经客户验收确认后，按照合同约定的交易金额确认收入。在新收入准则下，应用软件的定制开发或者授权许可、配套硬件的销售属于可明确区分的单项履约义务，并且属于在某一时刻履行的履约义务；应用软件的定制开发或者授权使用、以及配套硬件的销售，均是在向客户交付并经客户验收确认时确认收入，并按照合同约定的单项履约义务的交易价格确认收入金额。采用新收入准则之后，对于公司收入确认时点和收入确认金额没有影响。

公司执行新收入准则后，未对比较财务报表进行调整，对 2020 年 1 月 1 日至 2020 年 12 月 31 日止期间的合并利润表项目无影响，对 2020 年 1 月 1 日的合并资产负债表产生的影响分析如下：

报表项目	采用变更后会计政策增加 / (减少) 报表项目金额
资产：	
应收账款	-259.28
合同资产	259.28
负债：	
预收账款	-2,480.66
合同负债	2,480.66

对 2020 年 12 月 31 日的合并资产负债表产生的影响分析如下：

报表项目	采用变更后会计政策增加 / (减少) 报表项目金额
资产：	
应收账款	-137.47
合同资产	137.47

报表项目	采用变更后会计政策增加 / (减少) 报表项目金额
负债：	
预收账款	-1,495.77
合同负债	1,495.77

#### 4、准则 7 号（2019）

准则 7 号（2019）细化了非货币性资产交换准则的适用范围，明确了换入资产的确认时点和换出资产的终止确认时点并规定了两个时点不一致时的会计处理方法，修订了以公允价值为基础计量的非货币性资产交换中同时换入或换出多项资产时的计量原则，此外新增了对非货币资产交换是否具有商业实质及其原因的披露要求。

准则 7 号（2019）自 2019 年 6 月 10 日起施行，对 2019 年 1 月 1 日至准则施行日之间发生的非货币性资产交换根据该准则规定进行调整，对 2019 年 1 月 1 日之前发生的非货币性资产交换，不再进行追溯调整。采用该准则未对公司的财务状况和经营成果产生重大影响。

#### 5、准则 12 号（2019）

准则 12 号（2019）修改了债务重组的定义，明确了该准则的适用范围，并规定债务重组中涉及的金融工具的确认、计量和列报适用金融工具相关准则的规定。对于以资产清偿债务方式进行债务重组的，准则 12 号（2019）修改了债权人受让非金融资产初始确认时的计量原则，并对于债务人在债务重组中产生的利得和损失不再区分资产转让损益和债务重组损益两项损益进行列报。对于将债务转为权益工具方式进行债务重组的，准则 12 号（2019）修改了债权人初始确认享有股份的计量原则，并对于债务人初始确认权益工具的计量原则增加了指引。

准则 12 号（2019）自 2019 年 6 月 17 日起施行，对 2019 年 1 月 1 日至准则施行日之间发生的债务重组根据该准则规定进行调整，对 2019 年 1 月 1 日之前发生的债务重组，不再进行追溯调整。采用该准则未对公司的财务状况和经营成果产生重大影响。

#### 6、解释第 13 号（2019）

解释第 13 号修订了业务构成的三个要素，细化了业务的判断条件，对非同

一控制下企业合并的购买方在判断取得的经营资产或资产的组合是否构成一项业务时，引入了“集中度测试”的选择。

此外，解释第 13 号进一步明确了企业的关联方还包括企业所属企业集团的其他成员单位（包括母公司和子公司）的合营企业或联营企业，以及对企业实施共同控制的投资方的其他合营企业或联营企业等。

解释第 13 号自 2020 年 1 月 1 日起施行，公司采用未来适用法对上述会计政策变更进行会计处理。根据准则 28 号第 15 条的规定，企业需披露会计政策变更的影响。采用该解释未对公司的财务状况、经营成果和关联方披露产生重大影响。

## 7、财会[2020]10 号

财会[2020]10 号对于满足一定条件的，由新冠肺炎疫情直接引发的租金减让提供了简化方法。如果企业选择采用简化方法，则不需要评估是否发生租赁变更，也不需要重新评估租赁分类。

财会[2020]10 号自 2020 年 6 月 24 日起施行，可以对 2020 年 1 月 1 日至该规定施行日之间发生的相关租金减让根据该规定进行调整，采用上述规定未对公司的财务状况和经营成果产生重大影响。

### （二）重要会计估计变更

报告期内，公司会计估计未发生变更。

## 六、非经常性损益情况

### （一）非经常性损益

按照中国证监会《公开发行证券的公司信息披露解释性公告第 1 号——非经常性损益（2008）》的要求，公司最近三年的非经常性损益情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
非流动性资产处置损益	-12.71	-12.73	-7.93
计入当期损益的政府补助	46.30	165.20	14.60
计入当期损益的对非金融企业收取的资金占用费	-	-	-
交易性金融资产产生的公允价值变动收益，以及处置交易性金融	948.39	455.67	338.43



项目	2020 年度	2019 年度	2018 年度
资产和可供出售金融资产时取得的投资收益			
企业所得税退回	-	-	191.53
除上述各项之外的其他营业外收入和支出	25.84	-14.79	5.84
<b>小计</b>	<b>1,007.82</b>	<b>593.35</b>	<b>542.47</b>
所得税影响额	-101.25	-59.34	-34.61
<b>合计</b>	<b>906.57</b>	<b>534.02</b>	<b>507.86</b>
<b>非经常性损益占净利润比重</b>	<b>11.04%</b>	<b>6.55%</b>	<b>7.57%</b>

报告期内，非经常性损益金额占公司净利润的比重分别为 7.57%、6.55% 和 11.04%。报告期各年度非经常性损益金额占净利润比例较小。2020 年度非经常性损益占净利润比重有所上升主要是由于公司为提高资金收益率而购买的银行理财产品取得的投资收益增加所致。

因此，公司净利润不存在依赖于非经常性损益、合并财务报表范围以外的投资收益的情形。

## （二）扣除非经常性损益后的净利润

公司最近三年扣除非经常性损益后的净利润为：

单位：万元

项目	2020 年度	2019 年度	2018 年度
净利润	8,208.10	8,158.68	6,713.07
扣除非经常性损益后的净利润	7,301.54	7,624.66	6,205.21

## 七、报告期内执行的主要税收政策及缴纳的主要税种

### （一）主要税种及税率

税种	具体税率情况
增值税	按税法规定计算的销售货物和应税劳务收入的 17%、16%、13%、10%、9% 或 6% 计算销项税额，在扣除当期允许抵扣的进项税额后，差额部分为应缴增值税
企业所得税	详见下表
城市维护建设税	实际缴纳增值税的 7% 或 1%
教育费附加	实际缴纳增值税的 3%
地方教育费附加	实际缴纳增值税的 2%

注：根据《财政部、国家税务总局关于调整增值税税率的通知》（财税〔2018〕32号），自2018年5月1日起，纳税人发生增值税应税业务，将增值税税率由17%和11%分别调整为16%和10%；根据《财政部、国家税务总局、海关总署关于深化增值税改革有关政策的公告》（财政部、税务总局、海关总署公告2019年第39号），自2019年4月1日起，纳税人发生增值税应税业务，将增值税税率由16%和10%分别调整为13%和9%。

存在不同企业所得税税率的纳税主体情况：

纳税主体名称	所得税税率
北京海天瑞声科技股份有限公司	按应纳税所得额的10%计缴
北京中瑞智科技有限公司	2018年度及2019年度按应纳税所得额的25%计缴, 2020年度按应纳税所得额的15%计缴
海天瑞声（香港）科技有限公司	按应纳税所得额的16.5%计缴

## （二）税收优惠情况

### 1、增值税

（1）依据《财政部、国家税务总局关于软件产品增值税政策的通知》（财税〔2011〕100号）的规定，公司符合规定的软件产品征收增值税后，对其增值税实际税负超过3%的部分实行即征即退政策。

（2）根据《财政部、国家税务总局、海关总署关于深化增值税改革有关政策的公告》（财政部、税务总局、海关总署公告2019年第39号）规定，自2019年4月1日至2021年12月31日期间，允许生产、生活性服务业纳税人按照当期可抵扣进项税额加计10%，抵减应纳税额（即加计抵减政策）。

### 2、企业所得税

公司的法定税率为25%，报告期内执行的优惠税率如下：

2015年11月24日，公司取得北京市科学技术委员会、北京市财政局、北京市国家税务局、北京市地方税务局核发的编号为GR201511003654的《高新技术企业证书》，有效期三年。根据《中华人民共和国企业所得税法》、《企业所得税实施条例》等相关法律法规的规定，公司自2015年至2017年度，减按15%税率缴纳企业所得税。2018年7月19日，公司重新取得高新技术企业认定（证书编号：GR201811000007），继续享受15%所得税优惠税率。

2020年7月31日，中瑞智取得北京市科学技术委员会、北京市财政局、国家税务总局北京市税务局核发的编号为GR202011000590的《高新技术企业证书》，

有效期三年。根据《中华人民共和国企业所得税法》、《企业所得税实施条例》等相关法律法规的规定，中瑞智自 2020 年至 2022 年，减按 15% 税率缴纳企业所得税。

财政部、国家税务总局、发展改革委、工业和信息化部于 2016 年 5 月 4 日制定下发了《关于软件和集成电路产业企业所得税优惠政策有关问题的通知》（财税〔2016〕49 号），根据此通知，享受财税〔2012〕27 号文件规定的税收优惠政策的规划布局内重点软件企业，每年汇算清缴时应按照《国家税务总局关于发布〈企业所得税优惠政策事项办理办法〉的公告》（国家税务总局公告 2015 年第 76 号）规定向税务机关备案。公司于 2018 年 10 月获得 2017 年国家规划布局内的重点软件企业认定，可以对 2017 年的所得税享受 10% 的所得税税率优惠政策，因此本公司按 10% 的优惠税率重新申报了 2017 年度的企业所得税，并于 2018 年 12 月收到退回的 2017 年多缴纳的企业所得税人民币 191.53 万元。

公司于 2019 年 3 月完成了 2018 年国家规划布局内的重点软件企业相关备案手续，因此 2018 年度继续执行 10% 的所得税优惠税率。

公司于 2020 年 5 月完成了 2019 年国家规划布局内的重点软件企业相关备案手续，因此 2019 年度继续执行 10% 的所得税优惠税率。

国务院于 2020 年 7 月 27 日印发了《国务院关于印发新时期促进集成电路产业和软件产业高质量发展若干政策的通知》（国发〔2020〕8 号），根据此通知，国家发展改革委、工业和信息化部会同相关部门制定国家鼓励的重点集成电路设计企业和软件企业清单（以下简称“清单”），列入该清单的企业可享受以下税收优惠政策：自获利年度起，第一年至第五年免征企业所得税，接续年度减按 10% 的税率征收企业所得税。

国家发展改革委、工业和信息化部、财政部、海关总署、税务总局于 2021 年 03 月 29 日制定下发了《关于做好享受税收优惠政策的集成电路企业或项目、软件企业清单制定工作有关要求的通知》（发改高技〔2021〕413 号），根据此通知，清单印发前，企业可依据税务有关管理规定，先行按照企业条件和项目标准享受相关国内税收优惠政策。清单印发后，如企业未被列入清单，应按规定补缴已享受优惠的企业所得税款。公司已于本报告日前提交列入该清单的申请，预

计在 2020 年继续享受对国家鼓励重点软件企业的 10% 的优惠税率。

### 3、研发费用加计扣除

财政部、税务总局和科技部于 2017 年 5 月 2 日制定下发了《关于提高科技型中小企业研究开发费用税前加计扣除比例的通知》（财税〔2017〕34 号），根据此通知，科技型中小企业开展研发活动中实际发生的研发费用，未形成无形资产计入当期损益的，在按规定据实扣除的基础上，在 2017 年 1 月 1 日至 2019 年 12 月 31 日期间，再按照实际发生额的 75% 在税前加计扣除；形成无形资产的，在上述期间按照无形资产成本的 175% 在税前摊销。本公司于 2017 年符合科技型中小企业评价标准（中小科技企业编号：201811010808001550），按照研发费用实际发生额的 75% 在税前加计扣除。

财政部、税务总局和科技部于 2018 年 9 月 20 日制定下发了《关于提高研究开发费用税前加计扣除比例的通知》（财税〔2018〕99 号），根据此通知，企业开展研发活动中实际发生的研发费用，未形成无形资产计入当期损益的，在按规定据实扣除的基础上，在 2018 年 1 月 1 日至 2020 年 12 月 31 日期间，再按照实际发生额的 75% 在税前加计扣除；形成无形资产的，在上述期间按照无形资产成本的 175% 在税前摊销。公司于 2018 年、2019 年及 2020 年度按照研发费用实际发生额的 75% 在税前加计扣除。中瑞智于 2019 年及 2020 年度按照研发费用实际发生额的 75% 在税前加计扣除。

#### （三）税收优惠对公司经营成果的影响情况

报告期内，公司享受的税收优惠对经营成果的影响情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
增值税即征即退影响额	816.60	638.49	747.86
增值税加计抵减影响额	54.44	40.75	-
企业所得税税率优惠影响额	1,257.73	1,296.61	1,231.72
研发费用加计扣除影响额	136.04	121.97	100.71
合计	<b>2,264.81</b>	<b>2,097.82</b>	<b>2,080.29</b>
占当期利润总额的比例	<b>24.94%</b>	<b>22.96%</b>	<b>28.84%</b>

报告期内，公司享受的税收优惠金额占当期公司利润总额的比例分别为

28.84%、22.96%和 24.94%，公司经营成果对税收优惠不存在重大依赖。同时，公司适用的税收优惠政策在报告期内未发生重大变化，不存在对公司生产经营造成重大影响的情形。

## 八、报告期的主要财务指标

### （一）最近三年主要财务指标

项目	2020年 12月31日	2019年 12月31日	2018年 12月31日
流动比率（倍）	11.47	7.76	3.99
速动比率（倍）	11.39	7.60	3.54
资产负债率（母公司）	7.92%	12.17%	21.98%
归属发行人股东的每股净资产 （元/股）	13.64	11.09	5.54
项目	2020年度	2019年度	2018年度
应收账款周转率（次/年）	4.11	6.12	6.64
存货周转率（次/年）	13.38	4.48	3.78
息税折旧摊销前利润（万元）	9,438.33	9,476.77	7,398.80
归属发行人股东的净利润（万 元）	8,208.10	8,158.68	6,713.07
归属于发行人股东扣除非经常 性损益后的净利润（万元）	7,301.54	7,624.66	6,205.21
研发投入占营业收入的比例	18.64%	17.55%	14.19%
每股经营活动产生的现金流量 （元/股）	1.59	2.60	2.14
每股净现金流量（元/股）	-3.83	3.63	1.12
基本每股收益（元/股）	2.27	2.54	2.07
稀释每股收益（元/股）	2.27	2.54	2.07
加权平均净资产收益率	18.39%	37.17%	46.82%

注：指标计算公式如下：

流动比率=流动资产/流动负债

速动比率=速动资产/流动负债

资产负债率=总负债/总资产

归属于发行人股东的每股净资产=期末归属于母公司股东权益合计/期末股本总额

应收账款周转率=营业收入/应收账款平均余额

存货周转率=营业成本/存货平均余额

息税折旧摊销前利润=净利润+所得税+利息费用+固定资产折旧+摊销

归属于发行人股东的净利润=归属于母公司股东的净利润

归属于发行人股东扣除非经常性损益后的净利润=归属于母公司股东的净利润-非经常性损益的影响数

每股经营活动产生的现金流量=经营活动产生的现金流量净额/期末股本总额

每股净现金流量=现金及现金等价物净增加额/期末股本总额

基本每股收益、稀释每股收益和净资产收益率，依据扣除非经常性损益后的净利润计算，按照《公开发行证券的公司信息披露编报规则第9号——净资产收益率和每股收益的计算及披露》的要求进行计算。

## （二）最近三年净资产收益率及每股收益

根据中国证监会《公开发行证券公司信息披露编报规则第9号——净资产收益率和每股收益的计算及披露》（2010年修订）的规定，公司加权平均净资产收益率及每股收益计算如下：

### 1、加权平均净资产收益率

报告期利润	加权平均净资产收益率		
	2020年度	2019年度	2018年度
归属于公司普通股股东的净利润	20.68%	39.78%	50.65%
扣除非经常性损益后归属于公司普通股股东的净利润	18.39%	37.17%	46.82%

### 2、每股收益

单位：元/股

报告期利润	基本每股收益			稀释每股收益		
	2020年度	2019年度	2018年度	2020年度	2019年度	2018年度
归属于公司普通股股东的净利润	2.56	2.72	2.24	2.56	2.72	2.24
扣除非经常性损益后归属于公司普通股股东的净利润	2.27	2.54	2.07	2.27	2.54	2.07

注：计算公式

1、加权平均净资产收益率= $P \div (E_0 + NP \div 2 + E_i \times M_i \div M_0 - E_j \times M_j \div M_0 \pm E_k \times M_k \div M_0)$

其中：P 分别对应于归属于公司普通股股东的净利润、扣除非经常性损益后归属于公司普通股股东的净利润；NP 为归属于公司普通股股东的净利润；E<sub>0</sub> 为归属于公司普通股股东的期初净资产；E<sub>i</sub> 为报告期发行新股或债转股等新增的、归属于公司普通股股东的净资产；E<sub>j</sub> 为报告期回购或现金分红等减少的、归属于公司普通股股东的净资产；M<sub>0</sub> 为报告期月份数；M<sub>i</sub> 为新增净资产下一月份起至报告期期末的月份数；M<sub>j</sub> 为减少净资产下一月份起至报告期期末的月份数；E<sub>k</sub> 为因其他交易或事项引起的净资产增减变动；M<sub>k</sub> 为发生其他净资产增减变动下一月份起至报告期期末的月份数。

2、基本每股收益= $P \div (S_0 + S_1 + S_i \times M_i \div M_0 - S_j \times M_j \div M_0 - S_k)$

3、稀释每股收益= $[P + (\text{已确认为费用的稀释性潜在普通股利息} - \text{转换费用}) \times (1 - \text{所得税率})] \div (S_0 + S_1 + S_i \times M_i \div M_0 - S_j \times M_j \div M_0 - S_k + \text{认股权证、股份期权、可转换债券等增加的普通股加权平均数})$

其中：P 为归属于公司普通股股东的净利润或扣除非经常性损益后归属于普通股股东的净利润；S<sub>0</sub> 为期初股份总数；S<sub>1</sub> 为报告期因公积金转增股本或股票股利分配等增加股份数；S<sub>i</sub> 为报告期因发行新股或债转股等增加股份数；S<sub>j</sub> 为报告期因回购等减少股份数；S<sub>k</sub> 为报告期缩股数；M<sub>0</sub> 为报告期月份数；M<sub>i</sub> 为增加股份下一月份起至报告期期末的月份数；M<sub>j</sub> 为

减少股份下一月份起至报告期期末的月份数。

## 九、经营成果分析

### （一）营业收入分析

#### 1、营业收入构成及变动分析

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
主营业务收入	23,337.40	100%	23,755.81	100%	19,265.77	100%
合计	<b>23,337.40</b>	<b>100%</b>	<b>23,755.81</b>	<b>100%</b>	<b>19,265.77</b>	<b>100%</b>

报告期内，公司营业收入均来源于主营业务收入。公司营业收入的变动情况与主营业务变动一致，请参见下述分析。

#### 2、按业务类型分类的主营业务收入分析

报告期内，公司按业务类型分类的主营业务收入情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
训练数据定制服务	11,729.91	50.26%	14,232.91	59.91%	12,369.55	64.20%
训练数据产品	10,613.71	45.48%	9,176.47	38.63%	6,601.67	34.27%
训练数据相关的应用服务	993.77	4.26%	346.44	1.46%	294.55	1.53%
合计	<b>23,337.40</b>	<b>100%</b>	<b>23,755.81</b>	<b>100%</b>	<b>19,265.77</b>	<b>100%</b>

报告期内，公司主营业务收入主要来源于训练数据定制服务和训练数据产品，而训练数据相关的应用服务对公司的收入贡献度较低，其变动不对公司主营业务收入构成实质性影响。

##### （1）训练数据定制服务

报告期内，公司训练数据定制服务分别取得营业收入 12,369.55 万元、14,232.91 万元和 11,729.91 万元，占主营业务收入的比例分别为 64.20%、59.91% 和 50.26%，2019 年同比增长 15.06%，2020 年同比下降 17.59%。训练数据定制服务作为公司最主要的业务类型，其收入规模在 2019 年保持增长主要有以下原因：

第一，近年来在深度学习技术和开源平台的推动下，利用人工智能技术的创新和应用不断涌现，下游应用需求快速增长，人工智能进入了加速发展的黄金期。根据国际数据公司（IDC）的数据，2020年，全球人工智能产业规模达到501亿美元，预计2024年全球人工智能市场规模将达到1,100亿美元，年复合增长率为21.73%。随着行业整体快速发展和下游应用需求的快速增长，下游客户对于训练数据需求也日益强烈，促进了训练数据定制服务的整体增长趋势。

第二，公司基于自身在客户资源、技术实力、项目管理等方面的竞争优势，不断为下游各类人工智能客户提供贴合其算法模型训练需求的数据服务，从而稳固客户关系，树立国内领先基础数据服务商的品牌形象，持续保持其市场竞争力。报告期内，公司各年前五大客户基本保持稳定，主要为国内外大型科技公司，且前五大客户的收入占比各年均超过50%，表明公司重要客户具备良好的稳定性和集中度。由于该类大型公司对人工智能的重视程度较高，且持续加大在特定领域的布局与投入，也相应带动了公司训练数据定制服务的增长。

2020年，公司训练数据定制服务收入及占主营业务收入的比例均有所下降，主要原因是2020年全球新冠疫情持续爆发，新冠疫情防控期间原料数据现场采集、标注工作受到限制导致项目实施、产品开发、交付验收的进度均受到影响；同时由于客户停工，延迟了客户新增采购需求，导致新增合同订单增速、项目沟通及验收效率均有所降低所致。

## （2）训练数据产品

报告期内，公司训练数据产品分别取得营业收入6,601.67万元、9,176.47万元和10,613.71万元，占主营业务收入的比例分别为34.27%、38.63%和45.48%，2019年和2020年分别同比增长39.00%和15.66%。2019年，训练数据产品增速较快，主要是由于公司外语种训练数据产品实现了较快增长，从下游客户及应用角度看，语言功能方面开始向外语种方向拓展，公司把握该市场变化，发挥训练数据产品资源储备优势，加大外语种训练数据产品的推广，满足市场拓展需求，提高公司训练数据产品的收入和贡献率。2020年，受新冠肺炎疫情影响，部分训练数据定制服务难以大规模开展，公司引导客户购买已有的训练数据产品，使得训练数据产品收入占主营业务收入的比例上升，达到45.48%。



训练数据定制服务与训练数据产品共同构成公司完整的训练数据服务体系。阿里巴巴、腾讯、百度、微软、三星、亚马逊等公司主要客户普遍存在同时采购训练数据定制服务与训练数据产品的情况，训练数据定制服务满足其日益增长的设备系统独特性、应用领域特殊性的需求，训练数据产品则满足其算法模型上线及功能、语种拓展需要的通用训练数据的快速采购需求。在客户自身技术和产品研发的不同阶段，发行人的训练数据定制服务和训练数据产品都可以提供针对不同需求的支持，如新产品或应用首次研发并上市（使用训练数据产品或训练数据定制服务）、持续性能提升（使用训练数据定制服务）、应用领域拓展（使用针对特定领域的训练数据产品或训练数据定制服务）、语种拓展（使用多语种训练数据产品或训练数据定制服务）等。由于人工智能行业正处于爆发式发展阶段，下游技术应用企业多种多样，上述两种训练数据需求将会长期并存。公司将在满足客户定制服务需求的同时，持续研发训练数据产品，丰富自身训练数据库资源，为下游各种不同类型的客户提供全方位的训练数据服务。

### （3）训练数据相关的应用服务

报告期内，公司训练数据相关的应用服务分别取得营业收入 294.55 万元、346.44 万元和 993.77 万元，占主营业务收入的比例分别为 1.53%、1.46% 和 4.26%，2019 年和 2020 年分别同比增长 17.62% 和 186.85%。训练数据相关的应用服务是公司基于自身训练数据研发和技术优势所提供的增值性服务和延伸服务，收入规模较小，对公司整体业务和营业收入不构成显著影响。

### 3、按应用领域分类的主营业务收入分析

公司两大主要业务类型训练数据定制服务和训练数据产品覆盖了智能语音、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。报告期内，公司训练数据定制服务和训练数据产品按应用领域分类的收入情况如下：

单位：万元

项目		2020 年度		2019 年度		2018 年度	
		金额	占比	金额	占比	金额	占比
智能语音	语音识别采集	9,481.80	42.44%	9,547.84	40.79%	8,634.11	45.51%
	语音识别	3,413.03	15.28%	2,562.42	10.95%	2,960.54	15.61%

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
转写						
语音合成	5,213.48	23.33%	7,010.70	29.95%	4,312.55	22.73%
计算机视觉	2,590.61	11.59%	2,707.71	11.57%	1,855.06	9.78%
自然语言	1,644.71	7.36%	1,580.70	6.75%	1,208.97	6.37%
<b>合计</b>	<b>22,343.63</b>	<b>100%</b>	<b>23,409.38</b>	<b>100%</b>	<b>18,971.22</b>	<b>100%</b>

注：上述收入不包括训练数据相关的应用服务收入，因此合计数与主营业务收入有差异。

报告期内，公司产品及服务的主要应用领域为智能语音，三年合计贡献训练数据定制服务和训练数据产品收入的 80% 以上，其中包括语音识别采集、语音识别转写和语音合成三个业务方向。同时，计算机视觉和自然语言是公司产品及服务的另两大应用领域，得益于下游人工智能行业这两大细分领域的高速发展，公司来自于这两个应用领域的训练数据定制服务和训练数据产品收入也保持整体增长趋势。

对于以上各应用领域收入变动情况的分析，可拆分为销量分析和价格分析，具体请参见下述“4、主要产品和服务的销量和价格分析”。

#### 4、主要产品和服务的销量和价格分析

##### (1) 智能语音领域

##### ① 训练数据定制服务

报告期内，公司智能语音类训练数据定制服务的销量和价格情况如下：

业务细分类别	语种	2020 年度		2019 年度		2018 年度	
		销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)
语音识别采集	中文	39,149	230	18,510	350	38,286	355
	外语种	3,489	1,024	14,906	947	20,055	723
	<b>小计</b>	<b>42,638</b>	<b>294</b>	<b>33,417</b>	<b>616</b>	<b>58,341</b>	<b>482</b>
语音识别转写	中文	100,152	282	74,160	306	87,856	304
	外语种	6,889	859	4,549	639	3,067	955
	<b>小计</b>	<b>107,041</b>	<b>319</b>	<b>78,709</b>	<b>326</b>	<b>90,923</b>	<b>326</b>
语音合成	中文	2,721	8,330	1,899	13,553	766	17,510
	外语种	859	17,752	834	45,022	734	37,832

业务细分 类别	语种	2020 年度		2019 年度		2018 年度	
		销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)
	小计	3,580	10,590	2,733	23,153	1,500	27,458
合计		153,259	552	114,859	953	150,765	656

### A、销量变动情况分析

报告期内，公司训练数据定制服务中，语音识别采集的销量在 2019 年大幅下降、2020 年有所回升，语音识别转写的销量在 2019 年有所下降，语音合成销量保持持续增长。

#### a、语音识别采集

中文业务 2019 年销量大幅下降主要由于中文基础数据服务的市场参与者增加，价格竞争愈加激烈，市场需求一定程度被分流，导致销量降低。2020 年公司采用价格调整策略，以较低报价应对市场竞争的加剧，促进了销量的增长。

外语种业务 2019 年销量有所下降，但相应训练数据产品的语音识别采集外语种业务销量大幅上升，两者之间在公司整体销售策略的把控下存在一定的替代关系，总体服务于下游语音识别领域向外语种拓展的市场需求。具体分析请参见后述“②训练数据产品/A、销量变动情况分析/a、语音识别采集”；2020 年销量大幅下降，主要是受到境外新冠肺炎疫情持续爆发的影响，导致外语种项目的采集开展存在较大困难。

#### b、语音识别转写

语音识别转写销量大部分为中文销量，中文销量在 2018 年受个别主要客户转写需求的影响大幅增长，由于该主要客户在 2018 年购置的训练数据可以在一段时间内支撑其产品或应用性能迭代，因此 2019 年的同类中文转写需求量有所减缓，导致 2019 年语音识别转写总体销量下降。语音识别转写销量在 2020 年实现较大增长主要是由于公司为了开拓市场和维护客户关系，主动调整了报价策略所致。

#### c、语音合成

语音合成项目采用小规模受过专业发音训练的发音人，通常一个项目最终选定一个发音人。采集过程中发音人录音时间长达数月、对语音质量要求高、不

同发音人量级差别大，后续处理精度高，通常需要处理至音素级别，因此按小时计算的销量远低于语音识别类项目，同时单位价格远高于语音识别类项目，符合该类细分业务的业务特性。

语音合成销量逐年显著增长的主要原因是受国内外语音市场持续繁荣的影响，深度学习算法对语音合成数据量的需求增加，应用领域从标准合成到多风格多领域合成均有涉及，从手机助手、智能音箱拓展到智能客服、有声小说等。

## B、价格变动情况分析

报告期内，公司智能语音类训练数据定制服务的价格变动主要受到行业市场竞争、资源稀缺程度、客户或项目的特殊要求等原因的影响，具体情况如下：

### a、语音识别采集

中文业务单价在 2019 年保持稳定，2020 年明显下降，单价整体呈现下降趋势的主要原因是 2018 年以来行业快速发展使得中文数据资源日益丰富，且发行人业务相对成熟，同时市场竞争较为充分，价格竞争开始显现，因此价格相应下降所致。其中，2019 年发行人在 2018 年调整后的中文业务价格策略上整体保持稳定，单价较为接近；2020 年发行人继续采用价格调整策略应对市场竞争的加剧，同时存在部分较为简单的中文项目，采用了较低报价，也在一定程度上降低了整体单价。

外语种业务 2019 年和 2020 年单价较高，主要原因是外语种自由对话项目在 2019 年和 2020 年交付较多，在外语种业务中占比提高，自由对话类项目在采集、加工等环节较一般类项目更为复杂，因此单价较高；此外，2020 年外语种业务涉及一些较难语种，如加拿大法语、美国西班牙语、阿拉伯语等，采集成本较高，相应使得单价较高。

外语种业务单价明显高于中文业务，是由于外语种资源相对稀缺，采集和加工难度均高于中文所致，具备合理性。

### b、语音识别转写

报告期内，中文业务单价总体较为平稳，2020 年有小幅下降，是公司为了开拓市场和维护客户关系，主动调整了报价策略所致。

2018年和2019年，外语种业务由于整体收入规模较小（2018年和2019年收入均不超过300万元，占智能语音类训练数据定制服务收入比例均不超过3%），因此易受单个项目影响。2020年，外语种业务单价上升是由于个别主要客户的项目中包含测试集数据，准确率要求较高，因此单价偏高。

外语种业务单价明显高于中文业务，是由于外语种资源相对稀缺，加工难度高于中文所致，具备合理性。

### c、语音合成

由于语音合成项目的定制化程度最高，因此其不同项目之间的单价差异更明显，报告期各期平均单价的合理波动区间更大。相比语音识别项目的价格受语种影响较大，语音合成项目的价格主要取决于发音人价格和标注类型。发音人主要取决于专业性、知名程度、国籍、当地物价水平等因素，即使是同一语言，个体差异也可能较大。标注类型主要包括韵律标注、词性标注、音素边界标注等，不同的项目需要的标注类型组合不同，也会造成单价差异。

中文业务单价逐年下降，主要原因是中文语音合成技术逐渐成熟，公司开展的无需进行音素边界标注的语音合成项目比例逐年提升，同时，中文发音人的选取和录制方式也更加多样化，使部分项目无需进行发音人采集环节。该类项目成本相对较低，因此价格较低。

2018年和2019年，外语种业务单价较高，主要原因部分主要客户对外语种采集要求较高，例如对其发音人音色的代表性、发音的专业性、录音棚规格、录音过程监控等方面都提出了较高的要求。2018年至2019年，发行人根据客户需求启动了欧美多个语言的语音合成项目，此类项目需求因具备前述特征，使得项目难度和成本较高，同时价格也相应较高。2020年，外语种业务由于承接了较大比例的标注环节相对单一（例如无音素边界标注）的项目，因此价格明显下降。

### ②训练数据产品

报告期内，公司智能语音类训练数据产品的销量和价格情况如下：

业务细分 类别	语种	2020 年度		2019 年度		2018 年度	
		销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)
语音识别 采集	中文	91,037	221	55,820	266	68,537	355
	外语种	107,915	576	80,590	745	47,099	720
	小计	<b>198,951</b>	<b>413</b>	<b>136,411</b>	<b>549</b>	<b>115,636</b>	<b>504</b>
语音合成	中文	418	7,518	137	21,011	50	12,514
	外语种	610	18,164	137	28,864	46	28,520
	小计	<b>1,029</b>	<b>13,832</b>	<b>274</b>	<b>24,946</b>	<b>96</b>	<b>20,124</b>
合计	<b>199,980</b>	<b>482</b>	<b>136,685</b>	<b>598</b>	<b>115,732</b>	<b>520</b>	

## A、销量变动情况分析

### a、语音识别采集

中文业务 2019 年销量下滑的主要原因与前述训练数据定制服务的语音识别采集中文业务相似，请参见前述“①训练数据定制服务/A、销量变动情况分析/a、语音识别采集”。

外语种业务销量逐年保持较快增长，一方面是由于公司更多地将多通道采集技术运用到语音识别采集训练数据产品的开发中，该类型训练数据产品所包含的小时数量相对于单通道训练数据产品成倍增加，另一方面是由于 2019 年和 2020 年在一定程度上对训练数据定制服务的语音识别采集外语种业务形成了替代效应所致。从训练数据产品和训练数据定制服务合计的口径来看，2018-2020 年合计实现销量分别为 67,154 小时、95,496 小时和 111,404 小时，2019 年增长 42%，2020 年增长 17%，与下游客户产品及应用所展现的需求趋势基本一致。

公司自 2019 年开始的语音识别采集业务销售策略向训练数据产品有所倾斜，通过加大市场宣传与客户推广，促进客户更多的选择训练数据产品进行采购，原因在于：一方面训练数据产品价格更有竞争力，且产品由于其可反复销售的特征，经过其他客户验证的产品质量更有保障；另一方面能够发挥公司现有产品资源丰富的优势，增加利润空间，带动新产品的研发。

此外，语音识别采集训练数据产品的销量变动还受到疫情因素影响。2020 年，受新冠肺炎疫情影响，训练数据定制服务难以大规模开展，公司引导客户购买已有的训练数据产品，使得语音识别采集训练数据产品销量有所增长。

## b、语音合成

报告期内，语音合成训练数据产品销量逐年显著增长。

2019年公司加大客户推广，较2018年新增较多国内外客户，承接其语音合成训练数据采购需求，促进了语音合成产品的销售。2020年，受新冠肺炎疫情影响，训练数据定制服务难以大规模开展，公司引导客户购买已有的训练数据产品，使得语音合成训练数据产品销量明显增长。

## B、价格变动情况分析

公司智能语音类训练数据产品的价格变动主要受到市场同类训练数据产品稀缺程度和发行人训练数据产品的具体定位及拓展方向等原因，具体情况如下：

### a、语音识别采集

报告期内，中文和外语种业务单价呈现下降趋势，主要是由于多通道训练数据产品带来的销量提升并不会同比例提升训练数据产品总体价格，因此多通道产品销售量增长，会造成产品销售单价的下降趋势。此外，中文单价持续下降，也反映了中文语音识别市场价格竞争日益激烈的趋势。

外语种业务单价明显高于中文业务，是由于国内外语种资源相对稀缺，采集和加工难度高于中文所致，具备合理性。

### b、语音合成

如前述原因，语音合成项目按小时计算的产出量远低于语音识别采集项目，同时单位价格远高于语音识别采集项目。

此外，语音合成训练数据产品受到发音人定位的不同，价格变动比较明显，报告期各期平均单价的合理波动区间较大。2019年中文价格出现显著增长，是由于公司部分中文、波兰语、土耳其语等语音合成产品具备较好的市场独特性，下游客户语音合成模型研发过程中对这部分训练数据产品需求较强，因此公司议价空间较大。2020年中文价格较低，是由于当期销售情况较好的产品主要属于中文平均音色库，发音人为普通人，要求较低，使得成本偏低，相应售价较低。

## (2) 计算机视觉和自然语言领域

### ①训练数据定制服务

公司报告期内计算机视觉和自然语言类训练数据定制服务的销量和价格情况如下：

细分领域	2020 年度		2019 年度		2018 年度	
	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)
计算机视觉	172	12.69	213	9.76	185	9.02
自然语言	247	4.41	225	5.35	250	3.24

报告期内，计算机视觉和自然语言类训练数据定制服务的数据库销量变动幅度不大，相对保持稳定。

2020 年，计算机视觉类训练数据定制服务的平均售价有所增长，是由于个别主要客户的大体量项目涉及多个国家或大量人员，难度较大，因此单价较高。

自然语言类训练数据定制服务的总销售额较小，易受单一项目影响。2019 年，公司承接了部分主要客户体量相对较大的语义理解、外语种文本标注项目，项目个数虽然小幅下降，但平均售价有所提升。2020 年平均售价与 2019 年相比有所下降，主要是由于外语种文本标注项目占比较 2019 年有所减少，而中文文本标注项目占比增加，因此拉低了平均售价。

## ②训练数据产品

公司报告期内计算机视觉和自然语言类训练数据产品的销量和价格情况如下：

细分领域	2020 年度		2019 年度		2018 年度	
	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)	销量 (库)	平均售价 (万元)
计算机视觉	33	12.37	25	25.17	25	7.45
自然语言	36	15.46	30	12.56	29	13.72

报告期内，计算机视觉和自然语言类训练数据产品的销量基本稳定，销售单价随具体销售的单库不同影响，各期之间存在波动。计算机视觉类训练数据产品 2019 年销售单价较高是由于公司销售了体量较大、单价较高的数据库，包括人车 2D 图片标注数据库，用于客户在无人驾驶方面的研究，以及像素级分割标注数据库，用于物体检测等。

## ③细分类别销量和价格分析



## A、计算机视觉领域

计算机视觉领域的三类主要细分类别为手写体、光学字符识别、行车相关视频。报告期内上述三类细分类别的销量和价格情况如下：

细分类别	2020 年度		2019 年度		2018 年度	
	销量	平均售价 (元)	销量	平均售价 (元)	销量	平均售价 (元)
手写体	8,840.80	707.50	28,005.45	285.78	2,740.00	494.90
光学字符识别	5,282.63	1,433.98	83,532.02	54.42	47,419.86	65.13
行车相关视频	262,344.61	2.90	1,630,134.90	2.50	1,188,754.20	4.77

注：手写体、光学字符识别、行车相关视频的销量单位分别为千词组、百张、十个目标。

### a、手写体

报告期内，公司手写体业务 2019 年销量明显较高，主要来源于某国际消费电子厂商的移动设备滑行输入业务。

报告期内，公司手写体业务平均售价在 2019 年有所下降，2020 年明显上升，主要是由于该类业务的细分类别差异及语种差异所导致的。2019 年，该项业务主要涉及滑行输入数据，较传统手写输入方式的数据采集和标注难度更低，且主要语种开始从外语种转向简体中文，采集难度明显降低，因此价格也有所下降。2020 年，该类业务中一部分为多地区繁体中文手写体，对被采集人年龄、性别、学历、设备使用经验等均有要求，难度较大，且为华语项目，发行人议价能力较强，因此单价较高，另一部分为外语种手写体，采集难度大、资源稀缺，因此单价较高。

### b、光学字符识别

2019 年，随着光学字符识别的应用场景得到了较大规模的发展，公司在该类业务方面也实现了较快拓展，销量持续增长。2020 年，公司光学字符识别业务转向高复杂度和高精细度的项目，其单张图片采集和标注所需的工作量及工作时长大幅增加，使得该类业务的销量大幅降低，但同时单价大幅增长，具体情况请参见下段内容。

报告期内，公司光学字符识别业务在 2018 年和 2019 年平均售价基本保持稳定，2020 年平均售价非常高，是由于业务收入 70% 以上来源于某大型光学字符

采集项目和某财务文档采集标注项目，该类项目相较于一般光学字符识别项目环节更多，且客户指定特定采集设备拍摄，或采集内容为指定领域的实况图片、财务文档等，采集难度大，后期需标注或质检图片多项内容信息，流程复杂，因此单价非常高。

### c、行车相关视频

行车相关视频业务是公司在 2018 年新拓展的业务类别。2019 年，公司继续拓展该类业务，销量有所增长，同时由于竞争较为激烈，为保持市场占有率，平均售价有所降低。2020 年，随着自动驾驶应用场景进一步发展和普及，公司判断该类业务具有较大市场空间，同时公司自身也具备开发大型训练数据库的能力，因此将相关产能集中到行车相关的训练数据产品的开发上，该产品目前尚未开发完毕并实现收入，预计涵盖 5,000 万目标（以十个目标为单位，即 5,000,000），该策略的调整导致行车相关视频在 2020 年的销量有较大幅度的下降。此外，2020 年该类业务中含采集环节的项目对采集场景、时段、设备均有较高要求，因此使得平均售价略有回升。

### B、自然语言领域

自然语言领域的一类主要细分类别为语义理解。报告期内上述三类细分类别的销量和价格情况如下：

细分类别	2020 年度		2019 年度		2018 年度	
	销量 (小时)	平均售价 (元)	销量 (小时)	平均售价 (元)	销量 (小时)	平均售价 (元)
语义理解	66,822.69	52.44	86,505.88	55.28	110,407.88	45.90

报告期内，公司语义理解业务销量在 2018 年较高，主要是由于个别重要客户根据其业务需要采购规模较大所致；2019 年，虽然个别重要客户在 2019 年仍保持了较高的采购规模，但由于其他客户采购规模有所减少，因此 2019 年销量较 2018 年下降；2020 年，客户总体对于该类业务的需求有所降低，因此销量进一步下降。

报告期内，公司语义理解业务平均售价相对保持稳定，在 2019 年和 2020 年略高的原因是新增了医疗专业领域标注、文本编写、粤语分词、词性标注等高难度项目和外语种项目，需要相关专业的标注员进行标注，因此售价较高。

## 5、按区域分类的主营业务收入分析

### (1) 整体情况

报告期内，公司按区域分类的主营业务收入情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度		
	金额	占比	金额	占比	金额	占比	
境内	14,968.83	64.14%	11,973.75	50.40%	13,293.77	69.00%	
境外	韩国、日本	2,993.85	12.83%	5,688.34	23.95%	2,531.45	13.14%
	美国	4,651.52	19.93%	4,628.51	19.48%	2,496.28	12.96%
	其他地区	723.19	3.10%	1,465.21	6.17%	944.27	4.90%
合计	<b>23,337.40</b>	<b>100.00%</b>	<b>23,755.81</b>	<b>100%</b>	<b>19,265.77</b>	<b>100%</b>	

报告期内，公司主营业务收入大部分来源于境内，分别取得境内收入 13,293.77 万元、11,973.75 万元和 14,968.83 万元，占主营业务收入的比例分别为 69.00%、50.40%和 64.14%，2019 年境内收入占比相对较低，主要系公司海外业务拓展力度加强，及外语种业务需求相对较强所致，同时也存在个别大型跨国企业客户将其采购主体变更为境外主体的原因。

同时，公司也存在一定规模的境外收入，其中以美国、韩国、日本等地区为主。报告期内，公司分别取得境外收入 5,972.00 万元、11,782.07 万元和 8,368.56 万元，占主营业务收入的比例分别为 31.00%、49.60%和 35.86%。

### (2) 境外收入来源于新增客户的情况

报告期内各期，发行人境外收入按是否来源于新增客户拆分如下：

境外国家/地区	2020 年		2019 年		2018 年	
	新增客户收入占比	原有客户收入占比	新增客户收入占比	原有客户收入占比	新增客户收入占比	原有客户收入占比
日韩	1.30%	98.70%	5.39%	94.61%	1.11%	98.89%
美国	11.35%	88.65%	9.84%	90.16%	21.00%	79.00%
其他	28.44%	71.56%	36.37%	63.63%	36.11%	63.89%
境外收入合计	<b>9.23%</b>	<b>90.77%</b>	<b>10.99%</b>	<b>89.01%</b>	<b>14.96%</b>	<b>85.04%</b>

报告期内，发行人各期境外收入均主要来源于原有客户，而非新增客户。

2018-2020 年发行人境外收入整体呈现上升趋势，2020 年境外收入较 2019 年有所下降。

①其中，2018-2019 年境外收入上升的原因具体如下：

A、2018 年，部分客户品牌变更采购签约主体，使得同一客户品牌贡献的境外收入上升、境内收入下降

以大客户品牌三星为例，因其自身业务经营和管理需要，其自 2018 年起变更其向发行人执行采购的签约主体，主要签约主体从境内主体中国三星变更为境外主体韩国三星，2018-2019 年境外签约主体收入显著上升。发行人向三星销售收入按签约主体拆分情况如下：

单位：万元

销售收入	2020 年	2019 年	2018 年
三星	1,383.66	4,762.06	2,436.81
境内签约主体	47.46	493.04	603.15
境外签约主体	1,336.20	4,269.02	1,833.67

B、境外客户因自身业务发展需要拓展采购需求

2018-2019 年，来自境外主要客户亚马逊、微软（境外签约主体）、某国际消费电子产品厂商的销售收入增长。报告期内，前述境外主要客户收入情况如下所示：

单位：万元

销售收入	2020 年	2019 年	2018 年
亚马逊（境外签约主体）	199.11	1,492.98	465.00
微软（境外签约主体）	1,079.20	1,460.49	341.98
某国际消费电子产品厂商	928.33	920.39	278.97

2018-2019 年，亚马逊、微软销售收入增长主要是前述客户拓展其算法模型的语种/方言覆盖，加大对外语种训练数据产品需求和采购力度所致。某国际消费电子产品厂商销售收入增长主要是其对滑行输入训练数据需求持续增长所致。

C、发行人增聘了境外销售人员、加强境外业务开拓力度：发行人高度重视境外业务开拓，持续增聘境外销售人员，增强境外销售团队实力，助力公司境外业务开拓、新客户拓展。发行人注重加强与原有客户的沟通联系、稳固原定境外

客户合作关系、拓展业务合作范围；增强新客户开拓和对接，储备业务合作基础。2019 年至今，发行人新建立合作关系的境外客户超过 50 家，覆盖多家知名大型科技公司、人工智能企业及科研院所。

②2020 年境外收入相比 2019 年有所下降，主要原因是：一方面，三星、亚马逊、IMDA 等主要境外客户依据自身研发节奏，在 2018 年、2019 年集中采购了较大规模的训练数据定制服务/训练数据产品，能够满足其在特定领域一段时间内的需求，因此 2020 年采购量有所回落；另一方面，部分境外客户在 2020 年上半年因自身业务发展需要加大了采购规模，但受境外疫情因素影响，下半年采购规模有所回落。前述两方面因素使得 2020 年境外收入相比 2019 年有所下降。

综上，2018-2019 年发行人境外收入增加主要是：部分主要客户变更境内外采购签约主体；境外客户因其自身业务发展需要、算法模型的语种覆盖需求拓展、对新兴品类训练数据需求增加等因素增加训练数据采购以及发行人提升海外业务拓展力度所致。2020 年境外收入相比 2019 年有所下降，主要是境外客户受自身研发节奏影响，需求有所变动以及境外疫情因素对客户训练数据采购带来一定影响所致。

### （3）发行人海外业务拓展措施

针对海外业务拓展，发行人采取了下述措施：

#### ①建设内部团队，增聘境外销售人员

发行人高度重视境外业务开拓，持续增聘境外销售人员，通过境外销售人员进行有规律的信息收集，获得海外市场与客户的一手信息。同时加强对境外销售人员的产品培训力度，增强境外销售团队实力，助力公司境外业务开拓、新客户拓展。

#### ②跟进境外优质客户的训练数据需求，拓展业务增长点

在业务开展过程中，对于原有客户，发行人注重与客户的长期关系维系和持续服务，持续跟进下游客户业务的发展方向，随之布局、拓展相应的训练数据产品服务类型和应用领域，以持续满足客户需求、拓展营收增长点。

③加强在海外市场线上线下的宣传投放力度，持续针对海外客户及潜在客户进行品牌宣传

线上推广部分：发行人持续针对海外目标客户所在国家和地区进行搜索引擎广告投放，吸引新老客户的关注及问询。同时发行人还在全球知名的职场社交平台“领英”进行广告投放，提升品牌的境外知名度、美誉度。同时，发行人还会根据不同的市场活动，在海外学术论坛发布推广信息。

线下推广部分：发行人持续主动参加国际学术会议，增加展位部署，提升在海外客户市场的宣传曝光力度、吸引海外客户。例如：发行人每年都会参加全球人工智能领域的权威学术会议（如：CVPR-计算机视觉/ICASSP-综合/INTERSPEECH-智能语音等），发行人作为参展商向与会的众多人工智能产业机构进行品牌宣传并直接获客；并通过参与学术会议跟进 AI 技术发展中的前沿方向，以此作为后续产品服务开发拓展的重要参考。

通过采取上述措施，发行人对外持续加强针对海外市场的新增客户开拓能力、原有客户新增业务挖掘能力；对内持续提升训练数据产品服务提供能力。同时，下游 AI 行业的持续发展，算法模型的垂直领域、新兴语种覆盖拓展需求预期均将持续带动对相应训练数据需求的持续增长，发行人的海外业务拓展措施、境外收入均具备可持续性。

#### （4）当前疫情、国际形势对发行人经营状况的影响及发行人的应对措施

疫情、国际形势均属发行人业务的短期影响因素：疫情在短期之内对发行人组织原料数据现场集中采集、标注工作造成了一定影响，发行人通过引导客户采购成品训练数据集（即训练数据产品）平抑相关影响；随着我国疫情逐步稳定、各地有序复工，境内疫情影响逐步减弱；但境外疫情仍在蔓延，发行人的境外原料数据采集、标注工作开展会受到一定影响，发行人通过改善和提升工具性能，引导采集、标注人员使用线上远程采集、标注工具平抑相关影响。国际形势在现阶段对发行人所在行业无实质影响，发行人与境外各客户的商业洽谈、合作和拓展均在稳定、持续进行中。

发行人对此采取的应对措施包括：

①对于疫情：疫情影响发行人原料数据现场采集、标注工作的开展，因此也对训练数据定制服务的开展造成影响，发行人通过引导客户采购成品训练数据集、改善和提升工具性能，引导采集标注人员使用远程采集、标注工具平抑前述短期影响。

②对于国际形势：国际形势在现阶段对发行人所在行业无实质影响。为了应对国际形势未来可能造成的影响，发行人采取的措施包括：A、持续跟进服务现有境内外优质客户；B、拓展新兴优质客户（互联网领域新兴活跃企业，如字节跳动系、拼多多系、小米系、美团系客户等），持续跟进服务前述客户对训练数据的需求；C、跟进 AI 前沿领域发展趋势、拓展训练数据产品服务布局覆盖，整体稳固现有客户收入、拓展业务增长点。

此外，发行人已在招股说明书中针对疫情、国际形势等进行了风险提示，具体参见招股说明书“重大事项提示/三、特别风险提示/（三）中美贸易摩擦风险”及“重大事项提示/四、新型冠状病毒肺炎疫情对公司经营情况的影响”、“第四节 风险因素/二、疫情风险及四、中美贸易摩擦风险”。

## 6、合同及订单数量与营业收入的匹配关系

报告期内，发行人执行的合同数量、合同对应的具体订单数量及对应的营业收入情况如下：

项目	2020 年度	2019 年度	2018 年度
合同数量（个）	594	511	466
合同对应具体订单数量（个）	1,845	1,552	1,378
其中：完成数量（个）	1,544	1,341	1,111
未完成数量（个）	301	211	267
营业收入（万元）	23,337.40	23,755.81	19,265.77

报告期各期，发行人收入的变动趋势与完成的具体订单个数变动趋势基本相符，单个完成的具体订单贡献的收入约为 15-18 万元，较为稳定，合同完成数量与营业收入之间具备匹配关系。

## 7、营业收入的季节性分布

报告期内，发行人营业收入的季节性分布如下：

单位：万元

项目	2020年度		2019年度		2018年度	
	金额	占比	金额	占比	金额	占比
第一季度	3,126.09	13.40%	3,124.09	13.15%	2,848.45	14.79%
第二季度	7,067.83	30.29%	6,977.98	29.37%	3,784.80	19.65%
第三季度	4,083.44	17.50%	3,473.16	14.62%	5,483.29	28.46%
第四季度	9,060.04	38.82%	10,180.58	42.86%	7,149.23	37.11%
合计	<b>23,337.40</b>	<b>100%</b>	<b>23,755.81</b>	<b>100%</b>	<b>19,265.77</b>	<b>100%</b>

同行业可比公司 2020 年各季度营业收入分布情况如下：

	荣科科技	彩讯股份	天玑科技	四方精创	恒生电子	用友网络	万兴科技	泛微网络
第一季度	12.31%	20.01%	14.26%	17.25%	12.31%	12.77%	20.20%	13.58%
第二季度	23.37%	23.67%	21.66%	26.66%	26.50%	21.84%	28.41%	22.38%
第三季度	25.32%	27.37%	15.18%	24.92%	17.22%	19.58%	24.99%	22.45%
第四季度	39.00%	28.94%	48.90%	31.16%	43.97%	45.81%	26.41%	41.58%

由上述表格可以看出，发行人作为软件和信息技术服务行业公司，与同行业可比公司在营业收入的季节性分布方面具有类似的特征，即根据下游客户的年度预算制定及结算需求的节奏，第一季度实现收入普遍较低，第四季度实现收入普遍较高，第二、三季度普遍居中。

报告期内，发行人训练数据定制服务和训练数据产品的收入季节性分布如下：

单位：万元

训练数据定制服务						
项目	2020年度		2019年度		2018年度	
	金额	占比	金额	占比	金额	占比
第一季度	1,137.75	9.70%	2,039.75	14.33%	1,451.06	11.73%
第二季度	2,995.03	25.53%	3,847.99	27.04%	2,142.81	17.32%
第三季度	2,830.04	24.13%	2,186.35	15.36%	3,382.80	27.35%
第四季度	4,767.09	40.64%	6,158.82	43.27%	5,392.88	43.60%
合计	<b>11,729.91</b>	<b>100%</b>	<b>14,232.91</b>	<b>100%</b>	<b>12,369.55</b>	<b>100%</b>
训练数据产品						
项目	2020年度		2019年度		2018年度	



	金额	占比	金额	占比	金额	占比
第一季度	1,988.34	18.73%	1,082.90	11.80%	1,370.70	20.76%
第二季度	3,772.14	35.54%	3,129.43	34.10%	1,493.57	22.62%
第三季度	1,253.40	11.81%	1,273.88	13.88%	2,075.88	31.44%
第四季度	3,599.83	33.92%	3,690.27	40.21%	1,661.52	25.17%
合计	<b>10,613.71</b>	<b>100%</b>	<b>9,176.47</b>	<b>100%</b>	<b>6,601.67</b>	<b>100%</b>

发行人部分年度的第四季度收入及其占比较高，主要原因是部分客户的年度预算制定及结算需求和项目研发周期安排，使得其在第四季度会相对更为集中地进行验收。2018年，发行人训练数据定制服务在第四季度的收入金额及其占比较高，是由于部分主要客户的智能语音项目验收所致；2019年，发行人训练数据定制服务在第四季度的收入金额及其占比较高，是由于个别主要客户的语音合成项目验收所致，训练数据产品在第四季度的收入金额及其占比较高，是由于部分客户购买语音识别训练数据产品所致；2020年，发行人训练数据定制服务在第四季度的收入金额及其占比较高，是由于个别主要客户的语音合成项目和计算机视觉项目验收所致。

## 8、验收周期分析

报告期内，发行人交付训练数据定制服务和训练数据产品至客户验收的平均时长（以下简称“验收周期”）情况如下：

单位：天

项目	2020年度	2019年度	2018年度
验收周期	45	45	49

报告期内，发行人验收周期分别为49天、45天和45天。其中，2018年验收周期较长，一方面是由于个别重要客户验收方式变更，由客户项目组直接验收，改为客户项目组验收后提交客户内部平台，由其采购部门评价后，再由其财务部门审核，完成验收。该变更造成该个别重要客户验收周期加长，同时该客户在2018年的项目数量和收入金额均较大，拉长了发行人2018年总体的验收周期；另一方面是由于部分客户同期项目较多或项目规模较大，导致验收周期较长。扣除2018年特殊原因的影响后，报告期内发行人验收周期呈现逐年增长趋势，体现了发行人客户更高的数据质量要求及数据定制化程度对验收周期的影响，具备合理性。

报告期内，发行人验收周期较为稳定。同时，各年第四季度发行人验收周期情况如下：

单位：天

验收周期	2020年度	2019年度	2018年度
第四季度	44	44	51
全年	45	45	49

由上表可以看出，各年第四季度验收周期与全年水平基本一致，各年末不存在验收周期异常的情形。

由于发行人交付的训练数据定制服务或训练数据产品具有一定程度的差异化，因此会存在部分项目的验收周期偏短或偏长的情形，属于正常情况。

对于在各年第四季度客户验收完成，且验收周期短于 10 天的项目，其收入情况及原因如下：

单位：万元

项目	2020年度	2019年度	2018年度
营业收入金额	1,381.22	1,317.48	1,111.04
营业收入占比	5.92%	5.55%	5.77%
验收周期偏短的原因	1、项目数据量较小；2、项目执行过程中与客户持续保持需求沟通，使得客户提前熟悉数据情况。		

对于交付和客户验收处于不同年度，且验收周期长于 90 天的项目，其收入情况及原因如下：

单位：万元

项目	2020年度	2019年度	2018年度
营业收入金额	238.73	736.55	711.10
营业收入占比	1.02%	3.10%	3.69%
验收周期偏长的原因	1、项目数据量较大或难度较高；2、客户内部验收流程较慢。		

上述项目验收周期偏短或偏长是由于项目或客户的特性所致，原因具备合理性，且营业收入占比较低，并非普遍存在的情形。

综上所述，发行人不存在通过验收时点调节收入确认时点的情形。

## （二）营业成本分析

### 1、营业成本构成及变动分析

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
主营业务成本	7,537.79	100.00%	7,067.52	100.00%	6,786.78	100.00%
合计	<b>7,537.79</b>	<b>100.00%</b>	<b>7,067.52</b>	<b>100.00%</b>	<b>6,786.78</b>	<b>100.00%</b>

报告期内，公司营业成本均为主营业务成本，营业成本的变动情况与主营业务变动一致，请参见下述分析。

### 2、按业务类型分类的主营业务成本分析

报告期内，公司按业务类型分类的主营业务成本情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
训练数据定制服务	7,156.37	94.94%	6,957.69	98.45%	6,724.04	99.08%
训练数据产品	-	-	-	-	-	-
训练数据相关的应用服务	381.42	5.06%	109.83	1.55%	62.75	0.92%
合计	<b>7,537.79</b>	<b>100.00%</b>	<b>7,067.52</b>	<b>100.00%</b>	<b>6,786.78</b>	<b>100.00%</b>

报告期内，公司主营业务成本主要为训练数据定制服务所产生的成本。训练数据定制服务所产生的营业成本与该类业务取得的营业收入的变动情况总体较为匹配。由于训练数据产品所发生的各类支出计入研发费用，因此无相应营业成本。训练数据相关的应用服务由于收入规模很小，相应发生的营业成本也很小，对公司整体主营业务成本不构成实质性影响。

### 3、按应用领域分类的主营业务成本分析

报告期内，公司训练数据定制服务按应用领域分类的主营业务成本情况如下：

单位：万元

项目		2020 年度		2019 年度		2018 年度	
		金额	占比	金额	占比	金额	占比
智能语音	语音识别采集	754.56	10.54%	1,157.17	16.63%	1,606.83	23.90%

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
语音识别 转写	2,646.83	36.99%	1,740.04	25.01%	2,131.45	31.70%
语音合成	1,849.90	25.85%	2,408.59	34.62%	1,584.05	23.56%
计算机视觉	1,283.68	17.94%	1,169.35	16.81%	906.74	13.49%
自然语言	621.39	8.68%	482.53	6.94%	494.96	7.36%
<b>合计</b>	<b>7,156.37</b>	<b>100.00%</b>	<b>6,957.69</b>	<b>100.00%</b>	<b>6,724.04</b>	<b>100.00%</b>

注：上述成本不包括训练数据相关的应用服务成本，因此合计数与主营业务成本有差异。

与营业收入类似，公司训练数据定制服务的营业成本以智能语音领域为主。报告期内，公司各应用领域的训练数据定制服务营业成本随营业收入的变动而变动，两者变动趋势总体一致，与营业收入的变动差异体现在毛利率变动方面，具体分析请参见“（三）毛利及毛利率分析/2、毛利率情况分析”。

#### 4、按支出类型分类的主营业务成本分析

报告期内，公司按支出类型分类的主营业务成本情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
数据服务费	6,019.54	79.86%	5,245.49	74.22%	5,399.51	79.56%
职工薪酬	1,073.03	14.24%	1,185.95	16.78%	971.87	14.32%
折旧摊销	75.23	1.00%	141.62	2.00%	58.14	0.86%
差旅费	43.71	0.58%	41.74	0.59%	44.87	0.66%
其他	326.28	4.33%	452.72	6.41%	312.39	4.60%
<b>合计</b>	<b>7,537.79</b>	<b>100.00%</b>	<b>7,067.52</b>	<b>100.00%</b>	<b>6,786.78</b>	<b>100.00%</b>

报告期内，公司按支出类型分类的训练数据定制服务成本情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
数据服务费	5,991.24	83.72%	5,245.49	75.39%	5,395.72	80.25%
职工薪酬	901.87	12.60%	1,161.02	16.69%	940.99	13.99%
折旧摊销	63.96	0.89%	129.79	1.87%	47.27	0.70%
差旅费	26.40	0.37%	41.74	0.60%	43.96	0.65%

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
其他	172.90	2.42%	379.65	5.46%	296.09	4.40%
<b>合计</b>	<b>7,156.37</b>	<b>100.00%</b>	<b>6,957.69</b>	<b>100.00%</b>	<b>6,724.04</b>	<b>100.00%</b>

报告期内，公司主营业务成本绝大部分为训练数据定制服务成本，因此主营业务成本与训练数据定制服务成本的构成及变动情况一致。报告期内，公司主营业务成本主要由数据服务费和职工薪酬构成，另外存在少量折旧摊销、差旅费和其他类型的成本项目。其中，数据服务费是公司最主要的成本项目，是公司在提供训练数据服务过程中所需的、非核心技术环节的原料数据采集、标注服务，与公司主营业务收入变动总体保持一致，占报告期各年主营业务成本的比例分别为 79.56%、74.22%和 79.86%，占比较为稳定。同时，职工薪酬占报告期各年主营业务成本的比例分别为 14.32%、16.78%和 14.24%，占比较为稳定。

报告期内，公司按支出类型分类的各应用领域训练数据定制服务成本情况如下：

单位：万元

应用领域	项目	2020 年度		2019 年度		2018 年度	
		金额	占比	金额	占比	金额	占比
智能语音	数据服务费	4,441.08	84.57%	3,957.92	74.60%	4,286.41	80.54%
	职工薪酬	596.30	11.36%	869.85	16.39%	674.46	12.67%
	折旧摊销	42.71	0.81%	109.46	2.06%	41.39	0.78%
	差旅费	15.29	0.29%	24.80	0.47%	32.97	0.62%
	其他	155.92	2.97%	343.77	6.48%	287.10	5.39%
	<b>合计</b>	<b>5,251.30</b>	<b>100.00%</b>	<b>5,305.80</b>	<b>100.00%</b>	<b>5,322.33</b>	<b>100.00%</b>
计算机视觉	数据服务费	1,032.08	80.40%	919.30	78.62%	734.90	81.05%
	职工薪酬	203.17	15.83%	179.13	15.32%	150.60	16.61%
	折旧摊销	20.94	1.63%	19.69	1.68%	4.36	0.48%
	差旅费	10.99	0.86%	15.74	1.35%	10.63	1.17%
	其他	16.50	1.29%	35.50	3.04%	6.24	0.69%
	<b>合计</b>	<b>1,283.68</b>	<b>100.00%</b>	<b>1,169.35</b>	<b>100.00%</b>	<b>906.74</b>	<b>100.00%</b>
自然语言	数据服务费	518.08	83.37%	368.27	76.32%	374.41	75.64%

应用领域	项目	2020 年度		2019 年度		2018 年度	
		金额	占比	金额	占比	金额	占比
	职工薪酬	102.40	16.48%	112.05	23.22%	115.93	23.42%
	折旧摊销	0.31	0.05%	0.63	0.13%	1.52	0.31%
	差旅费	0.12	0.02%	1.21	0.25%	0.36	0.07%
	其他	0.48	0.08%	0.38	0.08%	2.75	0.55%
	合计	<b>621.39</b>	<b>100.00%</b>	<b>482.53</b>	<b>100.00%</b>	<b>494.96</b>	<b>100.00%</b>

由上表可以看出，公司各应用领域的训练数据定制服务成本构成与训练数据定制服务总体成本构成基本一致，主要为数据服务费和职工薪酬，其中各应用领域各期数据服务费占比基本均处于 75%-85% 之间，是最重要的构成部分。

### （1）智能语音领域

#### ①数据服务费变动分析

报告期内，发行人智能语音领域训练数据定制服务成本中的数据服务费具体情况如下：

成本项目	单位	2020 年度			2019 年度			2018 年度		
		数量	单价 (元)	金额 (元)	数量	单价 (元)	金额 (元)	数量	单价 (元)	金额 (元)
采集	人	26,978	143.38	3,868,085	21,680	208.17	4,513,130	43,895	174.99	7,681,443
	字/词	22,993,893	0.46	10,632,531	16,812,923	0.63	10,519,499	14,327,056	0.53	7,539,168
标注	小时	1,525,826	19.03	29,039,505	1,139,448	20.72	23,608,395	1,381,279	19.21	26,535,366

由于发行人训练数据定制服务“以销定采”的业务逻辑，上述采集数据量和标注数据量与销量相匹配，用于衡量采集数据量、标注有效工时相对水平的采产比和标产比合理。

上述以“人”为单位的采集单价在 2019 年上升，主要是由于：（1）采集类型由相对简单的读文本向相对复杂的自由对话转变，采集难度提升；（2）语种向多元化方向拓展，中文逐步由普通话拓展到方言，外语种逐步由常见语种拓展到稀有语种，资源稀缺度提升。采集单价在 2020 年下降，主要是由于当期完成了部分较为简单的中文项目，例如中文唤醒词项目等，因此采集单价较低。

上述以“字/词”为单位的采集单价在报告期内单价较为稳定，逐年略有变动主要是因为各期执行项目对发音人和领域多样化的要求有所不同造成的。

上述标注单价在报告期内较为稳定，逐年略有变动主要是由于语种拓展、标注精细度提高、以及对标注人员能力要求的变化等因素所带来的标注难度的变化造成的。

### ②职工薪酬变动分析

发行人智能语音领域训练数据定制服务成本中的职工薪酬变动情况，与发行人营业成本中的职工薪酬变动情况基本一致。

报告期内，发行人主营业务成本中的薪酬情况如下：

类型	项目	单位	2020 年度	2019 年度	2018 年度
与主营业务成本相关	薪酬总额	万元	1,073.03	1,185.95	971.87
	员工数量	人	39.86	40.41	35.02
	人均薪酬	万元/人	26.92	29.35	27.75

注 1：计入主营业务成本中的员工数量按照训练数据定制服务项目在当期结转确认的成本金额中所对应的数据中心员工的项目工时占全年总工时的比例加总计算得出。

注 2：人均薪酬反映当期确认成本的项目对应的数据中心员工，在项目周期覆盖的年度内的年化薪酬按项目工时进行加权平均计算后的年化薪酬。

报告期内，发行人计入主营业务成本核算的人员部门为数据业务中心。

数据业务中心的员工由于同时负责训练数据产品开发项目和训练数据定制服务项目，因此在计算报告期各年/期间员工数量时，按照训练数据定制服务项目在当期结转确认的成本金额中所对应的数据中心员工的项目工时占全年总工时的比例加总计算得出。

报告期内，发行人计入主营业务成本中的员工人均薪酬分别为 27.75 万元、29.35 万元和 26.92 万元，2020 年略低的原因是发行人业务受疫情影响，业绩增速未达预期，导致奖金有所下降所致。

### ③差旅费变动分析

报告期内，发行人智能语音领域训练数据定制服务成本中的差旅费与项目数量的对应关系如下：

项目	2020 年度	2019 年度	2018 年度
差旅费（万元）	15.29	24.80	32.97
差旅项目数量（个）	29	49	35

项目	2020 年度	2019 年度	2018 年度
单个差旅项目平均差旅费 (万元)	0.53	0.51	0.94

由于发行人项目执行范围涉及全国和境外部分国家和地区，对于采集要求较为复杂或采集场景、地区等较新的项目，项目负责人需要出差到现场进行监督、指导和沟通。2019 年单个差旅项目平均差旅费较低的原因是大部分差旅项目规模较小，同时远程沟通方式有效性增强，使得差旅周期缩短，因此费用较低。2020 年，受疫情影响，差旅项目很少且差旅周期很短，费用较低。

#### ④成本结构变动分析

对于智能语音领域，2019 年数据服务费占比下降而职工薪酬占比上升，主要原因是：语音识别类业务收入下降，语音合成类业务收入虽然上升但其毛利率较高，综合导致营业成本中的数据服务费有所下降；同时项目数量上升使得发行人人工投入增长，且发行人持续看好智能语音领域业务的发展前景，在行业竞争加剧、业务规模未显著增长的情况下进行了必要的项目人才储备，并提升人均薪酬增强竞争力，以充分把握未来的业务机会，导致人工投入持续增长；2020 年数据服务费占比上升而职工薪酬占比下降，主要原因是发行人业务受疫情影响，业绩增速未达预期，导致奖金有所下降，因此职工薪酬相对较低，拉升了数据服务费的占比。

#### (2) 计算机视觉和自然语言领域

报告期内，发行人计算机视觉和自然语言领域训练数据定制服务成本中的数据服务费与其收入规模基本保持了一致趋势，符合“以销定采”的业务逻辑，其中自然语言领域 2019 年由于稀有语种词典项目增多，毛利率提升，因此较 2018 年训练数据定制服务成本中的数据服务费基本持平而收入规模有所增长。

对于计算机视觉领域，报告期内成本构成无明显变动。

对于自然语言领域，报告期内数据服务费占比逐年上升而职工薪酬占比逐年下降，主要原因是该领域出于自身特点，项目规模普遍偏小，人工投入占比较高，在报告期内随着公司在该领域的深耕与拓展，单个项目规模呈现上升趋势，公司人工投入逐渐形成规模效应，使得职工薪酬占比下降。



### （三）毛利及毛利率分析

#### 1、毛利情况分析

报告期内，公司主营业务的毛利构成情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
训练数据定制服务	4,573.55	28.95%	7,275.22	43.59%	5,645.52	45.24%
训练数据产品	10,613.71	67.18%	9,176.47	54.99%	6,601.67	52.90%
训练数据相关的应用服务	612.34	3.88%	236.61	1.42%	231.80	1.86%
<b>合计</b>	<b>15,799.60</b>	<b>100.00%</b>	<b>16,688.29</b>	<b>100.00%</b>	<b>12,478.99</b>	<b>100.00%</b>

报告期内，公司主营业务的毛利主要来源于训练数据产品和训练数据定制服务，这两类业务均表现出较强的盈利能力，构成了公司持续盈利能力的基础。公司训练数据产品无相应营业成本发生，其产生的营业收入即毛利，毛利贡献高于训练数据定制服务。同时，训练数据定制服务所贡献的毛利呈现总体增长趋势，为公司毛利规模的提升也提供了有力支撑。

#### 2、毛利率情况分析

报告期内，公司主营业务的毛利率情况如下：

项目	2020 年度		2019 年度		2018 年度	
	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比
训练数据定制服务	38.99%	50.26%	51.12%	59.91%	45.64%	64.20%
训练数据产品	100.00%	45.48%	100.00%	38.63%	100.00%	34.27%
训练数据相关的应用服务	61.62%	4.26%	68.30%	1.46%	78.70%	1.53%
<b>综合毛利率</b>	<b>67.70%</b>	<b>100.00%</b>	<b>70.25%</b>	<b>100.00%</b>	<b>64.77%</b>	<b>100.00%</b>

报告期内，公司综合毛利率水平分别为 64.77%、70.25%和 67.70%，其中训练数据产品所发生的支出在当期全部计入研发费用，无相应的营业成本，毛利率为 100%，训练数据相关的应用服务由于收入和毛利规模较小，对公司综合毛利率水平不构成实质性影响，因此公司综合毛利率主要取决于训练数据定制服务的毛利率水平以及公司主营业务收入在服务和产品上的结构性变动。公司主营业务收入的结构变动情况及分析请参见前述“（一）营业收入分析/2、按业务类型

分类的主营业务收入分析”。

报告期内，训练数据定制服务按应用领域区分的毛利率和收入占比情况如下：

项目	2020 年度		2019 年度		2018 年度	
	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比
智能语音	37.92%	72.12%	51.55%	76.94%	46.18%	79.95%
计算机视觉	41.18%	18.60%	43.74%	14.60%	45.66%	13.49%
自然语言	42.89%	9.28%	59.92%	8.46%	38.98%	6.56%
训练数据定制服务	38.99%	100.00%	51.12%	100.00%	45.64%	100.00%

报告期内，训练数据定制服务各应用领域的收入占比总体保持稳定，因此各领域毛利率的变动是训练数据定制服务毛利率的主要影响因素。其中 2020 年，训练数据定制服务各应用领域中收入结构略有变动，智能语音收入占比小幅下降，计算机视觉收入占比小幅上升，因此训练数据定制服务毛利率同时受到应用领域收入结构变动和各领域毛利率变动两方面影响。

#### (1) 智能语音

报告期内，训练数据定制服务各类智能语音业务的毛利率和收入占比情况如下：

项目	2020 年度		2019 年度		2018 年度	
	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比
语音识别采集	39.91%	14.84%	43.83%	18.81%	42.80%	28.41%
语音识别转写	22.45%	40.35%	32.09%	23.40%	28.00%	29.94%
语音合成	51.20%	44.81%	61.94%	57.79%	61.55%	41.66%
智能语音类训练数据定制服务	37.92%	100.00%	51.55%	100.00%	46.18%	100.00%

报告期内，智能语音类训练数据定制服务的毛利率分别为 46.18%、51.55% 和 37.92%，2019 年有所提升，2020 年出现下降。2019 年毛利率提升的主要原因是相对毛利率较低的语音识别采集和语音识别转写在当期收入占比明显下降，而相对毛利率较高的语音合成在当期收入占比由 41.66% 上升至 57.79%。2020 年毛利率下降的主要原因一方面是相对毛利率较低的语音识别转写在当期收入占比有所上升，达到 40.35%，另一方面是各细分类别的毛利率均有所下降。

语音识别采集的毛利率在 2018 年和 2019 年保持稳定，2020 年略有下降，主要原因是 2020 年公司承接的语音识别采集项目中，中文项目占比较高，且由于部分中文项目难度较低、市场竞争激烈，导致毛利率偏低。

语音识别转写业务在 2018 年和 2019 年毛利率波动较小，2020 年明显下降的主要原因是：1) 短音频业务占比上升，受数据复杂度等因素影响其毛利率水平普遍低于长音频业务；2) 个别主要客户的项目包含测试集数据，准确率要求高，使得成本偏高，拉低了整体毛利率。

语音合成业务涉及到高素质发音人及音素、韵律、词性等多种标注类型，相对复杂度较高，所需技术处理环节较多，因此毛利率相对高于语音识别采集业务和语音识别转写业务。语音合成的毛利率在 2018 年和 2019 年较为稳定，2020 年有所下降的原因一方面为当期外语种项目占比有所降低，而外语种项目毛利率普遍高于中文项目，另一方面为当期个别主要客户的项目类型较为新颖，如歌曲类、表达丰富类语音合成业务，对发音人及录音环境要求高，采集难度较大，成本偏高，因此毛利率偏低。

## (2) 计算机视觉

由于发行人在报告期内逐步调整计算机视觉类业务价格策略，适度降低毛利空间，以获取更多与客户合作的机会，从而在计算机视觉领域树立良好的口碑和品牌形象，为未来的深入拓展奠定基础，因此该类业务的毛利率在报告期内呈现逐年下降趋势。

计算机视觉领域细分类别报告期内的收入占比及毛利率情况如下：

细分类别	2020 年度		2019 年度		2018 年度	
	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比
手写体	54.17%	28.66%	48.19%	38.51%	40.02%	8.13%
光学字符识别	43.89%	34.71%	53.02%	21.87%	57.39%	18.51%
行车相关视频	34.78%	3.49%	36.26%	19.57%	41.30%	33.97%

手写体业务在 2018 年和 2019 年主要为中英文滑行输入法项目，2020 年主要为手写体采集项目。相较于滑行输入，手写数据采集难度较大，且 2020 年承接项目属于涉及多国家、多地区的繁体中文、简体中文手写体采集项目，对被采

集人年龄、性别、学历、设备使用经验以及采集场地等均有要求，且为华语项目，发行人议价能力较强，销售单价较高，因此 2020 年手写体业务毛利率较高。2019 年手写体业务毛利率高于 2018 年的主要原因是随着发行人承接并执行中英文滑翔输入法项目，积累了该类项目的执行经验，采集及标注的工作效率有所提升，因此毛利率有所提高。

光学字符识别业务 2018 年和 2019 年毛利率变化不大，2020 年毛利率有所下降的原因是当年承接的英文手写体图片采集标注项目，包括合同、发票收据、手写笔记、药方等内容，数据采集难度高于预期，标注工作也具备较高复杂度，部分数据需要特定语言人士确认，而发行人总体延续之前年度的报价水平，因此使得毛利率下降。

行车相关视频业务在 2019 年和 2020 年部分客户的订单竞争激烈，发行人通过较低报价争取业务机会，因此毛利率较 2018 年有所下降。

### (3) 自然语言

由于自然语言类业务包含的细分项目类型较多，如语义理解、特定领域文本采集、发音词典、机器翻译、相关性标注等众多类型，同时自然语言类训练数据定制服务在报告期内整体业务规模较小，而项目个数则较多，因此毛利率受单个项目需求差异化特征影响，在报告期内呈现上下波动的情况。

自然语言类业务中，发音词典业务是毛利率较高的业务，该类发音词典项目主要解决发音与文本对应关系与规则的问题，需要较强的语音语言学研究基础，由于行业内能够处理该类语言数据的企业较少，而发行人经过多年积累，拥有多语种数据处理能力，较高的资源稀缺性使得该类项目毛利率较高，进而提升了自然语言类业务整体毛利率水平。2019 年，发音词典毛利率较高是由于发行人在当年完成了较多稀缺语种发音词典的开发，包括印尼语、塔加路语、僧伽罗语、阿塞拜疆语等亚洲小语种，及巴西葡萄牙语、阿根廷西班牙语等欧洲语种的词典数据开发和销售，该类语种稀缺性高，市场同类竞品较少，因此毛利率较高。

上述外语种及方言词典业务的毛利率水平如下：

特定业务	年度	毛利率	收入占比
外语种及方言词典业务	2020 年度	62.59%	2.22%

特定业务	年度	毛利率	收入占比
	2019 年度	89.59%	34.03%
	2018 年度	65.12%	3.07%
扣除特殊项目后的自然语言类业务	2020 年度	42.44%	97.78%
	2019 年度	44.62%	65.97%
	2018 年度	38.15%	96.93%

由上表可以看出，特殊词典业务毛利率随着项目具体执行情况具有较大的波动性。扣除该等特殊项目影响后，自然语言类业务毛利率水平在报告期内稳定在 40% 左右，波动较小。

### 3、同行业毛利率对比分析

发行人主要从事训练数据的研发设计、生产及销售业务，所属行业为软件和信息技术服务业，所在细分领域为人工智能基础数据服务领域。由于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，而业务可比的非上市公司由于不存在公开披露的财务数据，因而无法进行财务数据的对比分析。

澳大利亚上市公司 Appen 与发行人业务具有一定可比性，但由于其信息披露中未对定制化数据服务和训练数据产品的财务数据进行区分，因此仅能就其综合毛利率与发行人综合毛利率进行比较。对于分类业务，发行人在软件和信息技术服务行业中选取部分业务模式有一定相似性的 A 股同行业公司进行对比分析。

#### (1) 综合毛利率

公司名称	2020 年度	2019 年度	2018 年度
Appen	42.04%	41.99%	37.32%
海天瑞声	67.70%	70.25%	64.77%

发行人主营业务毛利率显著高于 Appen 的原因主要有以下几方面：

收入结构方面，首先，由于 Appen 并未披露其自有产权训练数据产品和定制化服务的分类收入，因此以自有产权训练数据产品的数量对比而言，目前发行人训练数据产品数量明显多于 Appen 官网展示的训练数据产品数量，因此 Appen 该类毛利率可达到或接近 100% 的训练数据产品业务对其整体毛利率的提升作用预计小于发行人。其次，就训练数据定制服务而言，根据 Appen 公开披露资料

显示,其大部分收入(报告期内平均占比超过 80%)来源于文本相关度标注业务。而发行人该类业务毛利率水平仅为 20-30%,显著低于其他智能语音类和计算机视觉类业务,发行人报告期内很少从事该类业务,因此 Appen 收入结构中低毛利率的文本相关度标注业务占比较高使得发行人毛利率水平高于 Appen。

数据采集、标注相关的劳务成本和企业人工成本方面,Appen 作为海外公司,主营业务所需的人员、场地均主要分布在美国和澳大利亚,其在亚洲的资源布局有限,获得亚洲知识丰富但价格较低的劳动力市场的支撑程度有限,劳务成本和人工成本相对国内均较高,也会在一定程度上拉低其毛利率。

综上所述,发行人主营业务毛利率显著高于 Appen 具备合理性。

## (2) 训练数据定制服务毛利率

选取 4 家软件和信息技术服务行业中主营业务方向为信息技术服务的上市公司(荣科科技、彩讯股份、天玑科技、四方精创)的全部或部分相关业务,与公司训练数据定制服务进行对比分析。

公司名称	比较业务	相似程度
荣科科技	健康数据服务	定制化服务为主
彩讯股份	技术服务业务	采用项目制服务,营业成本中人工成本占比较大,采购模式包含非核心工序的人力资源外包服务采购
天玑科技	IT 服务及自有产品销售等全部业务	采用项目制服务,营业成本中项目服务费和人力成本为主
四方精创	软件开发、系统集成、维护服务等全部业务	定制化服务为主

公司训练数据定制服务与上述相关公司相关业务的毛利率对比情况如下:

公司名称	2020 年度	2019 年度	2018 年度
荣科科技	50.70%	48.59%	52.87%
彩讯股份	44.54%	42.95%	45.43%
天玑科技	33.92%	29.98%	31.53%
四方精创	47.69%	49.35%	39.18%
<b>平均值</b>	<b>44.21%</b>	<b>42.72%</b>	<b>42.25%</b>
<b>海天瑞声</b>	<b>38.99%</b>	<b>51.12%</b>	<b>45.64%</b>

公司训练数据定制服务的毛利率与上述相关公司相关业务相比整体而言较

为接近。公司训练数据定制服务毛利率 2019 年上升及 2020 年下降的原因请参见上述“2、毛利率情况分析”。由于具体业务内容缺乏较强的可比性，因此仅从信息技术服务行业的整体服务模式而言，公司训练数据定制服务的毛利率具备合理性。

### （3）训练数据产品毛利率

选取 4 家软件和信息技术服务行业中主营业务方向为软件产品的上市公司（恒生电子、用友网络、万兴科技、泛微网络）的相关业务，与公司训练数据产品进行对比分析。

公司名称	比较业务	相似程度
恒生电子	软件产品	通用型软件产品授权销售
用友网络	软件产品	通用型软件产品授权销售
万兴科技	消费类软件	通用型软件产品授权销售
泛微网络	软件产品	通用型软件产品授权销售

公司训练数据产品与上述相关公司相关业务的毛利率对比情况如下：

公司名称	2020 年度	2019 年度	2018 年度
恒生电子	78.31%	97.39%	98.78%
用友网络	98.01%	98.32%	98.49%
万兴科技	95.86%	97.68%	97.49%
泛微网络	98.38%	98.21%	97.10%
<b>平均值</b>	<b>92.64%</b>	<b>97.90%</b>	<b>97.97%</b>
<b>海天瑞声</b>	<b>100.00%</b>	<b>100%</b>	<b>100%</b>

公司训练数据产品无相应的营业成本，毛利率为 100%。上述相关公司相关业务毛利率也普遍很高，具有相似性。由于公司训练数据产品后续销售过程仅涉及电子文件的发送和传输，包括提取训练数据产品至传输服务器，并将传输地址发给客户，涉及支出仅为零星人工，可以忽略不计，并不涉及应用类软件的安装、调试等环节和硬件介质，因此不存在该部分能够与收入相匹配的小额成本。

### （4）训练数据相关的应用服务

由于该项业务收入规模、毛利规模及其占比均很低，对公司经营成果不具有重要性，因此不再单独进行对比分析。

#### 4、训练数据定制服务各服务形式毛利率情况

报告期内，发行人在训练数据定制服务各应用领域均同时提供采集加工服务和纯加工服务。前述两种业务类型分别主要服务于客户不同阶段的需求：

①客户算法模型初始上线前的训练阶段——采集加工服务：在此阶段，客户的 AI 技术应用还未对外提供使用，尚处于算法模型的内部训练和模拟阶段，还未取得算法模型对外上线后能收集到的实际数据。为了使得算法模型取得更好的训练效果、更适用于实际场景，客户可向发行人采购采集加工类别的训练数据定制服务，发行人将根据客户算法模型的个性化需求，有针对性地采集原料数据，并进行加工处理，从而更好地模拟实际应用场景数据，使得客户的算法模型获得更佳训练效果。

②客户算法模型上线后的持续优化迭代阶段——纯加工服务：在此阶段，客户的 AI 算法模型/技术应用已正式上线并对外运营，客户在运营过程中可以收集到大量实际场景下的原料数据。为进一步优化算法模型，客户可向发行人采购纯加工服务，将运营过程中收集的实际场景原料数据提供给发行人，由发行人对其执行标注等加工工作，形成源自真实应用场景的训练数据，对算法模型进行进一步的训练、优化、迭代。

③客户算法模型的领域拓展阶段——采集加工服务：在此阶段，客户拟对其 AI 算法模型进行领域拓展（如应用场景拓展、语种/方言拓展等）。与上述阶段①类似，客户需要先对其算法模型进行相应领域的内部训练，暂未取得相应领域下的实际数据，因此可向发行人采购采集加工服务，以对其算法模型加以训练。

报告期内，训练数据定制服务各领域中采集加工和纯加工服务的营业收入、营业成本和毛利率情况具体如下所示：



金额单位：万元

应用 领域	业务类别	营业收入			营业成本			毛利率			
		2020 年度	2019 年	2018 年	2020 年度	2019 年	2018 年	2020 年度	2019 年	2018 年	
智能 语音	语音识别 (即语音识别采集)	采集加工服务	1,255.64	2,059.97	2,809.18	754.56	1,157.17	1,606.83	39.91%	43.83%	42.80%
		纯加工服务 (即语音识别转写)	3,413.03	2,562.42	2,960.54	2,646.83	1,740.04	2,131.45	22.45%	32.09%	28.00%
	语音合成	采集加工服务	3,135.86	5,653.06	3,613.25	1,614.62	2,173.05	1,394.73	48.51%	61.56%	61.40%
		纯加工服务	654.97	675.01	506.71	235.28	235.53	189.32	64.08%	65.11%	62.64%
计算机视觉	采集加工服务	1,757.56	1,051.77	381.37	1,027.96	595.38	157.49	41.51%	43.39%	58.70%	
	纯加工服务	424.76	1,026.70	1,287.35	255.72	573.97	749.25	39.80%	44.09%	41.80%	
自然语言	采集加工服务	497.81	288.82	205.24	271.32	63.45	108.65	45.50%	78.03%	47.06%	
	纯加工服务	590.29	915.16	605.91	350.07	419.08	386.31	40.70%	54.21%	36.24%	
合计		<b>11,729.91</b>	<b>14,232.91</b>	<b>12,369.55</b>	<b>7,156.37</b>	<b>6,957.69</b>	<b>6,724.04</b>	<b>38.99%</b>	<b>51.12%</b>	<b>45.64%</b>	

2018 年、2019 年及 2020 年，发行人训练数据定制服务中，各领域采集加工服务收入合计值分别为 7,009.05 万元、9,053.63 万元及 6,646.87 万元；纯加工服务收入合计值分别为 5,360.51 万元、5,179.28 万元及 5,083.05 万元。采集加工服务和纯加工服务均为发行人训练数据定制服务业务的重要形式，报告期内收入贡献比例分别约为 60%：40%，采集加工服务收入贡献度整体高于纯加工服务。

其中，在智能语音类别中，语音合成领域纯加工服务的毛利率水平高于语音识别领域纯加工服务，主要原因是：语音合成领域的标注工作涉及音素、韵律、词性等多种标注类型，相对复杂度较高，所需技术处理环节较多，因此毛利率相对高于语音识别领域的纯加工服务。

#### （四）期间费用分析

报告期内，公司的期间费用情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占营业收入的比例	金额	占营业收入的比例	金额	占营业收入的比例
销售费用	1,113.34	4.77%	925.28	3.89%	874.48	4.54%
管理费用	2,566.74	11.00%	3,155.96	13.28%	2,180.65	11.32%
研发费用	4,349.75	18.64%	4,169.61	17.55%	2,734.53	14.19%
财务费用	215.13	0.92%	149.82	0.63%	112.72	0.59%
<b>合计</b>	<b>8,244.97</b>	<b>35.33%</b>	<b>8,400.67</b>	<b>35.36%</b>	<b>5,902.38</b>	<b>30.64%</b>

报告期内，公司期间费用合计金额分别为 5,902.38 万元、8,400.67 万元和 8,244.97 万元，占营业收入的比例分别为 30.64%、35.36% 和 35.33%。

##### 1、销售费用分析

报告期内，公司的销售费用构成情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	980.42	88.06%	830.73	89.78%	767.23	87.74%
市场宣传费	101.76	9.14%	44.76	4.84%	52.82	6.04%
差旅费	12.46	1.12%	39.18	4.23%	37.87	4.33%
业务招待费	9.37	0.84%	3.98	0.43%	14.24	1.63%
其他	9.34	0.84%	6.63	0.72%	2.32	0.27%
<b>合计</b>	<b>1,113.34</b>	<b>100.00%</b>	<b>925.28</b>	<b>100%</b>	<b>874.48</b>	<b>100%</b>

报告期内，公司销售费用分别为 874.48 万元、925.28 万元和 1,113.34 万元，销售费用率分别为 4.54%、3.89% 和 4.77%。销售费用主要为职工薪酬，占销售费用的比例分别为 87.74%、89.78% 和 88.06%。市场宣传费主要是用于扩大公司影响力、获取客户所发生的广告宣传费、展览展位费、宣传资料费、市场调研费等支出。

公司获取客户的方式主要是通过口碑传播、参与学术会议、官方网站展示等方式建立品牌知名度以及与客户之间的沟通联系，后续按照客户的采购要求以商

务谈判、招投标等形式获取业务机会。

基于发行人所属行业专业化强、高速发展、客户定制服务需求种类日益丰富等特征，发行人对销售人员素质和能力要求较高，其主要职责为结合行业发展趋势、客户需求及公司特点，为客户提供高附加值的顾问式销售服务，提升公司在基础数据服务领域的口碑。具体包括：熟悉市场趋势、客户需求、发行人的产品与服务的技术特点等，通过行业会议、展会宣传、网络信息宣传等多种途径与客户初步建立合作关系；基于已有经验帮助客户细化需求，并将需求及时、准确的反馈给公司管理和技术人员，协助制定基本解决方案；对整体合同的工期、内容和报价等与客户进行协商、完成合作协议的签署；持续跟进和维护客户长期合作关系等。

报告期内，训练数据定制服务、训练数据产品等相关的合同数量与销售人员数量的情况如下表所示：

各期签订合同数量	2020 年度	2019 年	2018 年
合同数量	535	390	341
销售人员数量	15.58	12.18	11.25
销售人员人均合同数量	34	32	30

注：销售人员数量是按照当年/期营销中心各月份员工的在职情况计算的平均值。

报告期内，发行人的合同数量呈逐年上升的趋势，同时发行人的销售人员平均人数也逐年上涨，趋势相匹配，销售人员年人均合同数量基本维持在 30-35 个之间。

报告期内，公司销售费用率与信息技术服务类、软件产品类同行业公司的对比分析如下：

公司名称	2020 年度	2019 年度	2018 年度
荣科科技	9.78%	9.96%	9.48%
彩讯股份	3.95%	3.60%	3.21%
天玑科技	5.69%	6.10%	5.53%
四方精创	1.58%	0.64%	0.49%
上述 4 家上市公司 (信息技术服务类) 平均值	5.25%	5.08%	4.68%
恒生电子	8.48%	23.93%	27.03%

公司名称	2020 年度	2019 年度	2018 年度
用友网络	18.02%	19.20%	21.40%
万兴科技	45.14%	47.93%	49.02%
泛微网络	67.97%	70.72%	71.59%
上述 4 家上市公司 (软件产品类) 平均值	<b>34.91%</b>	<b>40.45%</b>	<b>42.26%</b>
海天瑞声	<b>4.77%</b>	<b>3.89%</b>	<b>4.54%</b>

从上表中的可比公司可见,信息技术服务类公司和软件产品类公司的销售费用率呈现明显不同。四家信息技术服务类公司销售费用率为平均约 4-5%左右,而四家软件产品类公司销售费用率均在 30%以上,发行人的销售模式和销售费用率与从事信息技术服务类业务的同行业公司较为接近,不存在大量销售人员和大额营销宣传费用。

## 2、管理费用分析

报告期内,公司的管理费用构成情况如下:

单位:万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	1,471.80	57.34%	1,717.52	54.42%	1,317.22	60.40%
租赁费	600.96	23.41%	548.80	17.39%	426.50	19.56%
中介费	131.23	5.11%	465.40	14.75%	155.09	7.11%
折旧与摊销	135.73	5.29%	128.27	4.06%	88.92	4.08%
业务招待费	45.71	1.78%	86.49	2.74%	27.89	1.28%
差旅费	13.42	0.52%	31.67	1.00%	14.92	0.68%
办公费	26.15	1.02%	38.17	1.21%	52.38	2.40%
水电费	42.30	1.65%	36.78	1.17%	31.08	1.43%
其他	99.44	3.87%	102.86	3.26%	66.65	3.06%
<b>合计</b>	<b>2,566.74</b>	<b>100.00%</b>	<b>3,155.96</b>	<b>100%</b>	<b>2,180.65</b>	<b>100%</b>

报告期内,公司管理费用分别为 2,180.65 万元、3,155.96 万元和 2,566.74 万元,管理费用率分别为 11.32%、13.28%和 11.00%。管理费用主要由职工薪酬、租赁费和中介费构成,合计占管理费用的比例分别为 87.08%、86.56%和 85.87%。

报告期内，发行人管理费用中职工薪酬主要包括发行人管理及行政人员、财务人员薪酬，及董事会会费、劳务费等管理活动相关人员薪酬。董事会会费、劳务费不涉及发行人全职员工，且在职工薪酬中占比较低，对职工薪酬变动不构成显著影响。报告期内职工薪酬主要受发行人管理及行政人员、财务人员薪酬总额变动影响，具体情况如下：

项目	单位	2020 年度	2019 年	2018 年
薪酬总额（管理及行政人员、财务人员）	万元	1,327.39	1,587.97	1,205.39
员工人数	人	40.05	37.22	31.82
人均薪酬	万元/人	33.14	42.66	37.88

注：员工人数，按照当年计入管理费用核算薪酬的员工按照各年/期间的工时扣除其在服务项目和研发项目中所填列工时后的数量占总工时的比例加总计算。

报告期内，发行人管理费用中管理及行政人员、财务人员的职工薪酬金额分别为 1,205.39 万元、1,587.97 万元和 1,327.39 万元，管理费用相关的平均人数分别为 31.82 人、37.22 人和 40.05 人。2019 年管理费用中薪酬总额和人均薪酬均有所增长，主要是由于一方面随着公司业务规模的扩张、对内部管理水平要求的提升和上市筹备事宜，管理及行政人员和财务人员数量增加，另一方面管理人员结构中的中层比例上升，且高级管理人员薪酬随公司业绩规模扩大而增长，管理人员平均薪酬有所提高。2020 年管理费用中薪酬总额和人均薪酬均有所降低，主要是由于发行人业务受疫情影响，业绩增速未达预期，导致包括管理及行政人员、财务人员在内的各部门员工奖金较 2019 年明显下降所致。

2019 年，公司管理费用中租赁费增长系新租办公场地所致，中介费显著增长系支付 2019 年申报科创板上市相关费用所致。2020 年，公司管理费用中租赁费增长系办公场地租金上涨所致。

报告期内，公司管理费用率与软件及信息技术服务类公司的对比分析如下：

公司名称	2020 年度	2019 年度	2018 年度
荣科科技	11.85%	12.64%	14.89%
彩讯股份	6.30%	8.20%	6.08%
天玑科技	13.13%	10.01%	10.80%
四方精创	19.42%	16.50%	11.84%
恒生电子	12.93%	12.56%	13.67%

公司名称	2020 年度	2019 年度	2018 年度
用友网络	11.26%	16.32%	19.02%
万兴科技	14.55%	14.62%	16.44%
泛微网络	5.38%	6.79%	8.41%
上述 8 家上市公司 (软件及信息技术服务类) 平均值	<b>11.85%</b>	<b>12.21%</b>	<b>12.64%</b>
海天瑞声	<b>11.00%</b>	<b>13.28%</b>	<b>11.32%</b>

报告期内，发行人管理费用率与同行业软件及信息技术服务类上市公司平均水平较为接近。

### 3、研发费用分析

#### (1) 研发投入确认依据

报告期内发行人对研发投入按照项目进行管理，即把为研发项目投入的直接费用和间接相关费用纳入研发费用核算，公司研发投入主要为实施研发项目所耗费的职工薪酬、数据服务费、语言研究支出和折旧与摊销等。

#### (2) 核算方法

公司根据《企业会计准则第 6 号——无形资产》的相关规定对研发支出进行核算，将研发项目研究阶段的支出全部费用化，计入当期损益（研发费用）；研发项目开发阶段的支出符合条件的资本化，不符合资本化条件的计入当期损益（研发费用）；若无法区分研究阶段的支出和开发阶段的支出的，将其所发生的研发支出全部费用化，计入当期损益（研发费用）。

公司研发项目具备一定特殊性，在研发项目最终完成前存在一定的不确定性，较难判断其达到资本化条件的具体时点。另外，由于公司各研发项目带来的未来收益及期间存在不确定性，无法准确判断其摊销年限。基于谨慎性原则，公司将报告期内的研发支出全部费用化并计入当期损益（研发费用）。

#### (3) 研发费用明细

报告期内，公司的研发费用构成情况如下：

单位：万元

项目	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	1,989.22	45.73%	1,858.94	44.58%	1,374.24	50.25%
数据服务费	1,844.91	42.41%	1,844.97	44.25%	957.04	35.00%
语言研究	144.78	3.33%	182.04	4.37%	175.46	6.42%
折旧与摊销	139.19	3.20%	128.12	3.07%	69.86	2.55%
房租费	86.63	1.99%	74.34	1.78%	75.33	2.75%
其他	145.02	3.33%	81.19	1.95%	82.61	3.02%
<b>合计</b>	<b>4,349.75</b>	<b>100.00%</b>	<b>4,169.61</b>	<b>100%</b>	<b>2,734.53</b>	<b>100%</b>

报告期内，公司研发费用分别为 2,734.53 万元、4,169.61 万元和 4,349.75 万元，研发费用率分别为 14.19%、17.55%和 18.64%。研发费用主要由数据服务费和职工薪酬构成，合计占研发费用的比例分别为 85.25%、88.83%和 88.15%。公司报告期内不存在研发费用资本化的情形。

报告期内，发行人的研发费用主要由训练数据产品开发支出及基础研发支出两部分构成，训练数据产品开发支出主要为开发训练数据产品相关的人工成本、数据服务费及其他相关费用支出。基础研发支出主要为研究开发训练数据所需的算法、技术以及工具、平台等的人工成本和其他费用支出。报告期内这两类支出的金额及占比如下：

单位：万元

分类	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
训练数据产品开发支出	2,222.68	51.10%	2,217.81	53.19%	1,264.60	46.25%
基础研发支出	2,127.07	48.90%	1,951.79	46.81%	1,469.94	53.75%
<b>合计</b>	<b>4,349.75</b>	<b>100.00%</b>	<b>4,169.61</b>	<b>100.00%</b>	<b>2,734.53</b>	<b>100.00%</b>

## ①训练数据产品开发支出

报告期内训练数据产品开发支出的费用构成、金额、占比情况如下：

单位：万元

类别	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	344.44	15.50%	320.01	14.43%	225.37	17.82%

类别	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
数据服务费	1,844.91	83.00%	1,844.97	83.19%	957.04	75.68%
折旧与摊销	13.28	0.60%	19.88	0.90%	15.89	1.26%
房租费	14.96	0.67%	7.70	0.35%	8.69	0.69%
其他	5.10	0.23%	25.24	1.14%	57.60	4.55%
<b>合计</b>	<b>2,222.68</b>	<b>100.00%</b>	<b>2,217.81</b>	<b>100.00%</b>	<b>1,264.60</b>	<b>100.00%</b>

报告期内，随着发行人加大训练数据产品开发的力度，巩固在自有知识产权训练数据产品规模性和多样性上的优势，训练数据产品开发总支出及各项费用构成总体呈上升趋势。其中，2018 年训练数据产品开发支出各项费用构成较 2019 年和 2020 年略有差异，原因如下：（1）2018 年数据服务费的占比略低，职工薪酬占比略高，主要是由于 2018 年发行人在拥有的资源和产能的基础上优先满足客户增加的定制化需求，训练数据产品开发规模的增速有所放缓，导致发行人 2018 年度投入训练数据产品的开发支出有所减少，发生的数据服务费也随之降低，而由于投入的研发职工薪酬受人均薪酬增长的影响并未同比减少，因此也导致了 2018 年职工薪酬占比略高；（2）2018 年其他支出的占比略高，主要是由于 2018 年发行人同乐乐创新签订抵债协议，发行人及附属子公司以人民币 40 万元购买乐乐创新中文女声语音合成数据库的永久使用权和授予第三方使用的权利及相关收益权，发行人应支付的人民币 40 万元直接冲抵乐乐创新所欠公司债务，乐乐创新向公司开具相应发票，公司将不含税金额人民币 37.74 万元计入训练数据产品开发支出的其他支出，因此造成了 2018 年其他支出占比略高。

报告期内，公司持续研发训练数据产品，丰富自身训练数据库资源。报告期内各期处于开发过程中的训练数据产品的个数如下：

项目	2020 年	2019 年	2018 年
训练数据产品开发支出（万元）	2,222.68	2,217.81	1,264.60
当期进行开发的训练数据产品个数（注）	203	149	112
单位训练数据产品开发支出（万元/个）	10.95	14.88	11.29

注：当期进行开发的训练数据产品个数为当期有研发支出的训练数据产品个数。

2019 年训练数据产品开发支出大幅上升，从 2018 年的 1,264.60 万元增长到 2,217.81 万元，增幅为 75%，主要是由于以下两方面原因：一方面是由于下游客



户对外语种训练数据产品的需求有所增长，公司准确把握市场趋势，推进了训练数据产品储备，2019 年进行开发的训练数据产品个数明显增长，从 112 个增长至 149 个，增幅达 33%；另一方面是由于 2019 年训练数据产品的增长主要是来源于外语种智能语音训练数据产品，这类产品需要由母语发音人进行录音，且标注难度较高，相关数据服务费用支出较一般产品更高，因此 2019 年的单位训练数据产品开发支出也有明显增长，进一步导致了训练数据产品开发支出的大幅上升。

2020 年，公司继续推进训练数据产品储备，进行开发的训练数据产品个数进一步增长，达到 203 个，训练数据产品开发支出与 2019 年相比保持稳定，单位训练数据产品开发支出降低主要是由于随着行业技术发展，公司训练数据产品需要进行升级完善，提升产品质量，因此进行了部分训练数据产品的升级完善等开发工作，该等工作所需支出较低，拉低了单位训练数据产品开发支出。

## ②基础研发支出

报告期内基础研发支出的费用构成、金额、占比情况如下：

单位：万元

类别	2020 年度		2019 年度		2018 年度	
	金额	占比	金额	占比	金额	占比
职工薪酬	1,644.78	77.33%	1,538.93	78.85%	1,148.87	78.16%
语言研究	144.78	6.81%	182.04	9.33%	175.46	11.94%
折旧与摊销	125.91	5.92%	108.25	5.55%	53.96	3.67%
房租费	71.67	3.37%	66.63	3.41%	66.63	4.53%
其他	139.92	6.58%	55.94	2.86%	25.02	1.70%
<b>合计</b>	<b>2,127.06</b>	<b>100.00%</b>	<b>1,951.79</b>	<b>100.00%</b>	<b>1,469.94</b>	<b>100.00%</b>

报告期内，基础研发支出各项费用构成的变化及原因如下：（1）职工薪酬占比保持稳定；（2）语言研究支出占比在报告期内逐年降低，主要是由于随着公司自身在语言学研究方面的积累加强，对外采购语言学专家服务方面的支出在各年间基本稳定，未发生显著增长；（3）折旧与摊销占比在报告期内有所增加，主要是由于各年均均有新购入的房产或设备用于从事研发活动，因此导致分摊的折旧费用逐年增加；（4）房租费从 2018 年开始发生，之后保持稳定，主要是由于公司自 2018 年开始为研发部门设立专属办公场地并将房租费用在研发费用中进

行分摊；（5）其他支出占比在报告期内逐年增长，主要是由于2019年开始专利申请和系统平台升级相关的支出逐年增加，因此导致其他支出占比逐年增长。

报告期内，公司的基础研发项目开展情况如下：

单位：万元

序号	项目名称	预算金额	支出金额			实施进度
			2020 年度	2019 年度	2018 年度	
1	语音识别算法平台	200.00	177.21			已完成
2	目标检测与跟踪平台	160.00	133.21			已完成
3	海天瑞声在线点云 3D 标注平台升级	200.00	148.49			进行中
4	一体化数据处理平台系统升级二期	260.00	153.30			进行中
5	基于 PC 端 TTS 文本分析和标注平台升级	220.00	202.39			进行中
6	语音合成训练数据在线生产平台	160.00	168.63			进行中
7	海天瑞声 OCR 在线生产平台	220.00	223.43			进行中
8	海天瑞声手写体在线生产平台	180.00	187.70			进行中
9	海天瑞声中文多音字标注平台	290.00	233.42			进行中
10	无人车场景的图形图像处理平台	310.00	293.18			进行中
11	智能监听管理系统	80.00	59.66			已完成
12	模型训练自动学习智能平台	30.00	36.61			已完成
13	信创版智能庭审记录系统	40.00	48.82			已完成
14	锐听智能拾音设备管理系统	50.00	61.02			已完成
15	语音数据合成效果示例平台	260.00		144.56		已完成
16	多语种词性标注平台	220.00		109.06		已完成
17	海天瑞声问卷调查平台	150.00		135.46		已完成
18	海天瑞声在线点云 3D 标注平台	200.00		180.61		已完成

序号	项目名称	预算金额	支出金额			实施进度
			2020 年度	2019 年度	2018 年度	
19	基于移动端多通道录音软件	220.00		213.45		已完成
20	基于新闻语料多环节标注软件	180.00		174.64		已完成
21	基于 PC 端图片标注工具	140.00		137.44		已完成
22	基于 PC 端语音数据库质检平台	160.00		157.08		已完成
23	基于 PC 端 TTS 文本分析和标注平台	220.00		215.98		已完成
24	一体化数据处理平台系统升级	260.00		255.25		已完成
25	智能语音记录系统	74.00		13.49		已完成
26	智能外呼系统	100.00		18.23		已完成
27	锐听采集设备二期	60.00		73.80		已完成
28	音频转写标注平台-定制版	16.00		16.44		已完成
29	音频预处理软件	25.00		25.69		已完成
30	钉钉会议预约软件	30.00		18.60		已完成
31	会议管理软件	45.00		27.90		已完成
32	语音标注转写软件	20.00		12.40		已完成
33	庭审智能语音助手软件	15.00		9.30		已完成
34	音频检索软件	20.00		12.40		已完成
35	基于 audiobook 的自动切分与文本对齐系统	180.00			175.34	已完成
36	基于 hts 的语音合成系统	245.00			238.65	已完成
37	基于 Kaldi 的语音数据库评测系统	275.00			267.87	已完成
38	基于 PC 端的 OCR 标注系统	135.00			131.50	已完成
39	跨平台多语种语料设计系统	165.00			160.72	已完成
40	跨平台文本易读性评测系统	175.00			170.47	已完成
41	中文语音合成-音字标注系统	225.00			219.17	已完成

序号	项目名称	预算金额	支出金额			实施进度
			2020 年度	2019 年度	2018 年度	
42	潘多拉魔盒软件	80.00			74.17	已完成
43	工作平台、中小企业平台开发项目	45.00			32.04	已完成
44	基于云端的文本标注软件系统	140.00				已完成
45	人工智能图片数据标注软件系统	220.00				已完成
46	人脸关键点及属性信息标注软件系统	190.00				已完成
47	视频内容标注软件系统	180.00				已完成
48	视频信息检索软件系统	260.00				已完成
49	电子书记员系统	120.00				已完成

报告期内，公司研发费用率与前述同行业公司的对比分析如下：

公司名称	2020 年度	2019 年度	2018 年度
荣科科技	7.66%	6.30%	6.21%
彩讯股份	15.63%	14.88%	12.68%
天玑科技	10.01%	9.51%	8.58%
四方精创	13.88%	16.04%	15.44%
恒生电子	35.85%	40.29%	43.05%
用友网络	17.11%	19.16%	16.88%
万兴科技	22.39%	22.11%	17.85%
泛微网络	13.44%	12.94%	12.50%
<b>平均值</b>	<b>17.00%</b>	<b>17.65%</b>	<b>16.65%</b>
<b>海天瑞声</b>	<b>18.64%</b>	<b>17.55%</b>	<b>14.19%</b>

报告期内，公司的研发费用率与其他同行业公司总体较为接近，符合软件和信息技术服务行业注重研发能力，研发费用率普遍较高的特点，具备合理性。

#### 4、财务费用分析

报告期内，公司的财务费用构成情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
存款及应收款项的利息收入	-13.38	-6.24	-7.20
净汇兑损益	188.97	123.00	101.48
其他	39.53	33.06	18.44
合计	<b>215.13</b>	<b>149.82</b>	<b>112.72</b>

报告期内，公司财务费用分别为 112.72 万元、149.82 万元和 215.13 万元，汇兑损益较大的原因系国内外宏观经济变化形成的汇率波动所致。公司为提高资金使用效率，通常将账面闲置资金购买低风险、短周期的银行理财产品，该等理财产品的收益计入投资收益而非利息收入，因此利息收入较低。

报告期内，公司财务费用率与前述同行业公司的对比分析如下：

公司名称	2020 年度	2019 年度	2018 年度
荣科科技	0.66%	1.08%	0.87%
彩讯股份	0.46%	0.58%	-0.11%
天玑科技	-2.55%	-3.71%	-2.30%
四方精创	0.31%	-2.00%	-2.22%
恒生电子	-0.17%	0.08%	0.06%
用友网络	1.29%	1.38%	1.42%
万兴科技	0.02%	-0.11%	-0.16%
泛微网络	-2.07%	-2.06%	-2.32%
平均值	<b>-0.26%</b>	<b>-0.60%</b>	<b>-0.60%</b>
海天瑞声	<b>0.92%</b>	<b>0.63%</b>	<b>0.59%</b>

报告期内，公司的财务费用率较低，与同行业公司无显著差异，符合软件和信息技术服务行业普遍特点。

#### （五）非经常性损益项目对报告期经营成果的影响分析

报告期内，公司非经常性损益情况请参见本节“六、非经常性损益情况/（一）非经常性损益”。

2018 年，非经常性损益主要为企业所得税退回和理财产品收益，金额分别为 191.53 万元和 338.43 万元。企业所得税退回是由于公司在 2018 年被认定为“2017 年度国家规划布局内的重点软件企业”并按照相关程序完成了备案程序，

自 2017 年起享受“国家规划布局内的重点软件企业”的企业所得税率 10% 的优惠政策,于 2018 年 12 月收到退回的 2017 年多缴纳的企业所得税款 191.53 万元。

2019 年,非经常性损益主要为计入当期损益的政府补助和理财产品收益,金额分别为 165.20 万元和 455.67 万元,政府补助的具体情况请参见下述“(六)政府补助对报告期经营成果及未来期间的影响分析”。

2020 年,非经常性损益主要为理财产品收益,金额为 948.39 万元。

报告期内,公司购买银行理财产品取得投资收益是公司为提高资金使用效益的举措,收益金额分别为 338.43 万元、455.67 万元和 948.39 万元,占利润总额的比例分别为 4.69%、4.99% 和 10.44%,不构成公司的主要盈利来源,对公司的盈利能力稳定性和持续经营能力不构成重大影响。

#### (六) 政府补助对报告期经营成果及未来期间的影响分析

报告期内,公司所获得的政府补助均属于与收益相关的政府补助并计入当期损益,具体情况如下:

单位:万元

年度	种类	金额	是否计入非经常性损益
2020 年	增值税即征即退补贴款	816.60	否
	研发补贴款	7.00	是
	稳岗补贴款	22.07	是
	个人所得返还	16.73	是
	信用报告补助	0.50	是
	<b>合计</b>	<b>862.91</b>	
2019 年	增值税即征即退补贴款	638.49	否
	海淀区研发费用补贴专项	90.00	是
	中小企业服务体系建设项目	63.00	是
	稳岗补贴款	7.79	是
	中关村国际创新资源支持资金	4.42	是
	<b>合计</b>	<b>803.69</b>	
2018 年	增值税即征即退补贴款	747.86	否
	稳岗补贴款	4.60	是

年度	种类	金额	是否计入非经常性损益
	改制补贴款	10.00	是
	合计	762.46	

## （七）纳税情况分析

### 1、报告期公司主要税项缴纳情况

#### （1）增值税

单位：万元

项目	2020 年度	2019 年度	2018 年度
期初未缴数	176.70	58.01	74.32
本期应缴数	1,534.13	1,459.49	1,176.59
本期实缴数	1,308.43	1,340.80	1,192.90
期末未缴数	402.40	176.70	58.01

#### （2）企业所得税

单位：万元

项目	2020 年度	2019 年度	2018 年度
期初未缴数	388.66	-114.60	84.50
本期应缴数	788.58	810.36	474.24
本期实缴数	915.51	307.11	673.34
期末未缴数	261.73	388.66	-114.60

注：2018 年应缴企业所得税金额考虑了冲回 2017 年多缴纳的企业所得税的影响，具体情况请参见前述“七、报告期内执行的主要税收政策及缴纳的主要税种/（二）税收优惠情况/2、企业所得税”。

### 2、税收优惠对公司经营成果的影响分析

请参见本节“七、报告期内执行的主要税收政策及缴纳的主要税种/（二）税收优惠情况”及“（三）税收优惠对公司经营成果的影响情况”。

## 十、资产质量分析

报告期内，公司资产结构如下：

单位：万元

项目	2020 年 12 月 31 日		2019 年 12 月 31 日		2018 年 12 月 31 日	
	余额	占比	余额	占比	余额	占比
流动资产	45,172.52	94.63%	37,705.55	93.21%	18,420.50	86.76%

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
非流动资产	2,562.48	5.37%	2,748.39	6.79%	2,812.27	13.24%
资产总计	<b>47,735.00</b>	<b>100.00%</b>	<b>40,453.94</b>	<b>100.00%</b>	<b>21,232.77</b>	<b>100.00%</b>

报告期各期末，公司资产总额分别为 21,232.77 万元、40,453.94 万元和 47,735.00 万元。2019 年末和 2020 年末，公司资产总额较上年末分别增长 90.53% 和 18.00%，资产总额逐年增长。

### （一）流动资产分析

报告期各期末，公司流动资产的具体构成情况如下：

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	金额	占比	金额	占比	金额	占比
流动资产：						
货币资金	4,026.38	8.91%	16,325.52	43.30%	4,669.16	25.35%
交易性金融资产	33,108.95	73.29%	15,601.69	41.38%	-	-
应收账款	6,278.29	13.90%	4,364.55	11.58%	2,938.96	15.95%
预付款项	999.07	2.21%	419.96	1.11%	83.57	0.45%
其他应收款	183.63	0.41%	169.24	0.45%	172.84	0.94%
存货	312.94	0.69%	754.75	2.00%	2,063.20	11.20%
合同资产	137.47	0.30%	-	-	-	-
其他流动资产	125.80	0.28%	69.85	0.19%	8,492.76	46.10%
<b>流动资产合计</b>	<b>45,172.52</b>	<b>100.00%</b>	<b>37,705.55</b>	<b>100.00%</b>	<b>18,420.50</b>	<b>100.00%</b>

公司流动资产项目具体分析如下：

#### 1、货币资金

报告期各期末，公司货币资金明细如下：

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
库存现金	-	-	-	-	-	-
银行存款	4,013.94	99.69%	16,312.06	99.92%	4,657.15	99.74%
其他货币资金	12.43	0.31%	13.46	0.08%	12.01	0.26%



项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
货币资金合计	4,026.38	100.00%	16,325.52	100.00%	4,669.16	100.00%

报告期各期末,公司货币资金分别为4,669.16万元、16,325.52万元和4,026.38万元。公司货币资金主要为银行存款,占比分别为99.74%、99.92%和99.69%。

公司货币资金的变化主要受到以下各方面因素的影响:1)经营活动所产生的现金净流入积累;2)外部增资;3)公司购买银行理财产品并计入交易性金融资产或其他流动资产;4)利润分配。

报告期各期末,公司货币资金与银行理财产品之和分别为12,969.16万元、31,927.21万元和37,135.32万元,2019年末较2018年末大幅增加,主要原因是股权融资和经营活动现金积累所致;2020年末较2019年末增加,主要原因是经营活动现金积累所致。

报告期内,发行人不存在受限资金。

## 2、交易性金融资产

报告期各期末,公司交易性金融资产明细如下:

单位:万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
理财产品	33,108.95	15,601.69	-
交易性金融资产合计	33,108.95	15,601.69	-

根据2019年执行的新金融工具准则,保本浮动型理财产品在2019年及2020年计入交易性金融资产核算。2018年,该部分理财产品计入其他流动资产核算。

## 3、应收账款

### (1) 应收账款账龄分布

报告期各期末,公司的应收账款账龄分布情况如下:

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
<b>应收账款余额：</b>						
1年以内（含1年）	6,482.77	96.75%	4,494.99	96.45%	3,037.75	98.09%
1-2年（含2年）	60.96	0.91%	108.47	2.33%	59.00	1.91%
2-3年（含3年）	100.00	1.49%	56.90	1.22%	-	-
3年以上	56.90	0.85%	-	-	-	-
<b>应收账款余额</b>	<b>6,700.63</b>	<b>100.00%</b>	<b>4,660.36</b>	<b>100%</b>	<b>3,096.75</b>	<b>100%</b>
<b>坏账准备</b>	<b>422.34</b>	-	<b>295.81</b>	-	<b>157.79</b>	-
<b>应收账款账面净额</b>	<b>6,278.29</b>	-	<b>4,364.55</b>	-	<b>2,938.96</b>	-

报告期各期末，公司95%以上的应收账款账龄在一年以内，应收账款质量较好，回收风险较小。截至2020年末，公司应收账款计提坏账准备422.34万元，计提比例为6.30%，该计提比例和公司应收账款账龄结构相适应。

公司制定了销售管理的相关制度，综合考虑体量规模、信誉状况、合作历史等因素给予客户不同的销售信用期。公司应收账款账龄结构符合公司的销售信用政策。

截至2020年12月31日，发行人账龄在1年以上的应收账款构成及形成原因如下：

单位：万元

债务人名称	年末金额	账龄	形成原因	计提坏账准备金额
Verint Systems Ltd.	60.96	1-2年	客户回款较慢	30.48
深圳市天道日新科技有限公司	100.00	2-3年	客户回款较慢	50.00
上海未来伙伴机器人有限公司	56.90	3年以上	客户发生债务危机	56.90

上海未来伙伴机器人有限公司因债务危机已于2020年4月提起破产重整，公司已对应收上海未来伙伴机器人有限公司的款项56.90万元全额计提了坏账准备。

2020年12月末，公司应收账款及合同资产期后回款情况如下：

单位：万元

项目	2020年12月31日
应收账款及合同资产余额	6,415.77
期后回款金额	4,305.55
期后回款金额占应收账款及合同资产余额的比例	67.11%

注：期后回款数据为截至2021年3月31日的回款情况。

公司报告期末的应收账款及合同资产期后回款情况良好，不存在较大的坏账风险。

## (2) 应收账款坏账准备计提

2019年末和2020年末，公司的应收账款坏账准备计提情况如下：

单位：万元

类别	2020年12月31日		
	预期信用损失率	账面余额	减值准备
未逾期	3.00%	5,477.51	164.33
逾期1年以内（含1年）	12.00%	1,005.26	120.63
逾期1至2年（含2年）	50.00%	160.96	80.48
逾期2年以上	100.00%	56.90	56.90
合计	-	<b>6,700.63</b>	<b>422.34</b>
类别	2019年12月31日		
	预期信用损失率	账面余额	减值准备
未逾期	3.00%	3,493.65	104.81
逾期1年以内（含1年）	12.00%	1,107.37	132.88
逾期1至2年（含2年）	50.00%	2.44	1.22
逾期2年以上	100.00%	56.90	56.90
合计	-	<b>4,660.36</b>	<b>295.81</b>

2018年末，公司按账龄分析法的应收账款坏账准备计提情况如下：

单位：万元

类别	2018年12月31日		
	账面余额	坏账准备	计提比例
1年以内	3,037.75	151.89	5.00%
1年至2年	59.00	5.90	10.00%
合计	<b>3,096.75</b>	<b>157.79</b>	<b>5.10%</b>

## (3) 应收账款前五名客户

报告期各期末，公司应收账款前五名客户情况如下：

单位：万元

2020年12月31日					
排序	客户名称	是否与发行人存在关联关系	应收账款余额	占比	账龄
1	腾讯科技（成都）有限公司	否	1,230.20	18.36%	1年以内
2	淘宝（中国）软件有限公司	否	717.89	10.71%	1年以内
3	中国某大型科技公司	否	462.11	6.90%	1年以内
4	北京有竹居网络技术有限公司	否	439.91	6.57%	1年以内
5	阿里云计算有限公司	否	386.36	5.77%	1年以内
合计			<b>3,236.47</b>	<b>48.31%</b>	
2019年12月31日					
排序	客户名称	是否与发行人存在关联关系	应收账款余额	占比	账龄
1	中国某大型科技公司	否	711.73	15.27%	1年以内
2	Microsoft Corporation	否	379.48	8.14%	1年以内
3	Amazon Web Services, Inc.	否	290.52	6.23%	1年以内
4	浙江天猫技术有限公司	否	256.13	5.50%	1年以内
5	Info-communications Media Development Authority	否	254.30	5.46%	1年以内
合计			<b>1,892.15</b>	<b>40.60%</b>	
2018年12月31日					
排序	客户名称	是否与发行人存在关联关系	应收账款余额	占比	账龄
1	淘宝（中国）软件有限公司	否	752.11	24.29%	1年以内
2	Amazon Web Services, Inc.	否	351.94	11.36%	1年以内
3	北京百度网讯科技有限公司	否	289.80	9.36%	1年以内
4	浙江天猫技术有限公司	否	274.96	8.88%	1年以内
5	阿里巴巴（中国）网络技术有限公司	否	217.11	7.01%	1年以内
合计			<b>1,885.93</b>	<b>60.90%</b>	

上述公司报告期各期末应收账款前五大客户大部分均隶属于国内外大型科技公司，信誉良好。

#### 4、预付款项

报告期各期末，公司预付款项情况如下：

单位：万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
数据服务费	1.51	22.86	58.06
上市服务费	885.72	337.26	-
其他	111.84	59.84	25.51
<b>合计</b>	<b>999.07</b>	<b>419.96</b>	<b>83.57</b>

2019年末和2020年末，预付款项主要为公司预付上市相关中介机构337.26万元和885.72万元上市服务相关费用。报告期内其余预付款项主要为业务采购所形成，总体金额较小。

#### 5、其他应收款

报告期各期末，公司其他应收款情况如下：

单位：万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
员工备用金	7.42	2.27	0.22
房租押金	171.94	172.92	180.60
保证金	5.59	1.28	1.15
其他	8.28	1.74	-
<b>小计</b>	<b>193.22</b>	<b>178.21</b>	<b>181.97</b>
坏账准备	9.60	8.97	9.12
<b>合计</b>	<b>183.63</b>	<b>169.24</b>	<b>172.84</b>

报告期各期末，公司其他应收款主要为房租押金。

报告期内，其他应收款的坏账准备计提情况如下：

(1) 2019年和2020年坏账准备计提情况

单位：万元

坏账准备	2020年			
	第一阶段	第二阶段	第三阶段	合计
	未来12个月预期信用损失	整个存续期预期信用损失-未发生信用减值	整个存续期预期信用损失-已发生信用减值	
年初余额	8.97	-	-	8.97
本年计提	0.62	-	-	0.62
本年收回或转回	-	-	-	-
本年核销	-	-	-	-
年末余额	9.60	-	-	9.60
坏账准备	2019年			
	第一阶段	第二阶段	第三阶段	合计
	未来12个月预期信用损失	整个存续期预期信用损失-未发生信用减值	整个存续期预期信用损失-已发生信用减值	
原金融工具准则下的年初余额	9.12	-	-	9.12
首次执行新金融工具准则的调整金额	-	-	-	-
调整后的年初余额	9.12	-	-	9.12
本年计提	-	-	-	-
本年收回或转回	-0.15	-	-	-0.15
本年核销	-	-	-	-
年末余额	8.97	-	-	8.97

## (2) 2018年坏账准备计提情况

单位：万元

坏账准备	2018年
年初余额	7.90
本年计提	1.22
本年转回	-
本年核销	-
年末余额	9.12

## 6、存货

报告期各期末，公司的存货构成情况如下：

单位：万元

一、账面余额			
项目	2020年12月31日	2019年12月31日	2018年12月31日
在产品	345.88	780.99	2,376.64
二、跌价准备			
项目	2020年12月31日	2019年12月31日	2018年12月31日
在产品	32.94	26.25	313.43
三、账面价值			
项目	2020年12月31日	2019年12月31日	2018年12月31日
在产品	312.94	754.75	2,063.20

报告期内，公司存货均为在产品，用于归集正在履行尚未完成验收的训练数据定制服务和训练数据相关的应用服务所发生的各项支出。

2018年末，存货余额较高，与公司训练数据定制服务业务规模相一致。此外，公司2018年承接的部分主要客户的语音合成外语种项目及滑行输入手写体项目由于项目周期长，在2018年底仍处于执行状态，形成了大量存货。

2019年末，存货余额大幅下降，一方面是由于上述外语种语音合成项目及滑行输入手写体项目在2019年完成并验收，相关支出从存货结转计入营业成本；另一方面则是市场竞争日益加剧，市场参与者逐渐增多，订单谈判周期加长，同时客户对项目周期的要求也越来越高，导致公司截至2019年底在手订单下降和整体在执行项目周期缩短，也使得存货出现较大幅度的下降。

2020年末，由于疫情对训练数据定制服务项目的开展造成明显影响，尤其是境外疫情持续爆发的情况下，部分相关在手订单无法开展或无法顺利按期推进，因此未能形成大量存货，导致存货余额有所下降。

公司将存货在资产负债表日按照成本与可变现净值孰低计量。当存货的可变现净值低于成本时，提取存货跌价准备。存货跌价准备通常按单个存货项目的成本高于其可变现净值的差额提取。

报告期内，公司存货跌价准备的计提、转回及核销情况如下：

存货类型	2020 年				
	年初余额	本年计提	本年转回	本年核销	年末余额
在产品	26.25	106.87	-88.60	-11.58	32.94
存货类型	2019 年度				
	年初余额	本年计提	本年转回	本年核销	年末余额
在产品	313.43	104.05	-18.49	-372.75	26.25
存货类型	2018 年度				
	年初余额	本年计提	本年转回	本年核销	年末余额
在产品	30.16	284.03	-0.75	-	313.43

2018 年、2019 年和 2020 年，公司分别计提存货跌价准备 284.03 万元、104.05 万元和 106.87 万元。2018 年和 2019 年计提存货跌价准备对净利润造成一定影响，主要由于日本纽康株式会社的部分合同属于发行人首次接触的试验性项目或特殊领域项目，难度超过预期，未能及时交付，经双方协商一致，该合同终止执行。公司预计此部分合同的收入为 0，另外按照项目组人员的评估，此部分已完成的数据也无法用于其他项目，可变现净值为 0，因此公司按照已发生的成本全额计提跌价准备。

公司与日本纽康株式会社的其他业务合作按照双方合同约定继续执行。

报告期各期末，未完工项目的类别、数量、金额及对应的合同金额如下：

单位：万元

类型	领域	2020 年 12 月 31 日				
		项目数量	存货账面余额	存货跌价准备	存货账面净值	对应的合同金额
训练数据定制服务	智能语音	102	207.87	-11.76	196.11	2,435.21
	计算机视觉	20	41.41	-19.77	21.64	234.10
	自然语言	38	68.67	-1.41	67.26	411.68
训练数据相关的应用服务		6	27.92	-	27.92	1,105.57
合计		<b>166</b>	<b>345.88</b>	<b>-32.94</b>	<b>312.94</b>	<b>4,186.56</b>
类型	领域	2019 年 12 月 31 日				
		项目数量	存货账面余额	存货跌价准备	存货账面净值	对应的合同金额
训练数据定制服务	智能语音	94	537.45	-24.35	513.10	3,729.47
	计算机视觉	20	50.43	-	50.43	763.46



	自然语言	32	83.95	-	83.95	319.55
训练数据相关的应用服务		5	109.17	-1.90	107.27	1,359.82
<b>合计</b>		<b>151</b>	<b>780.99</b>	<b>-26.25</b>	<b>754.74</b>	<b>6,172.30</b>
<b>类型</b>	<b>领域</b>	<b>2018年12月31日</b>				
		<b>项目数量</b>	<b>存货账面余额</b>	<b>存货跌价准备</b>	<b>存货账面价值</b>	<b>对应的合同金额</b>
训练数据定制服务	智能语音	141	1,653.62	-118.31	1,535.31	8,511.57
	计算机视觉	27	469.12	-5.37	463.75	1,251.35
	自然语言	46	210.01	-167.13	42.88	304.77
训练数据相关的应用服务		10	43.88	-22.63	21.25	1,479.49
<b>合计</b>		<b>224</b>	<b>2,376.64</b>	<b>-313.43</b>	<b>2,063.21</b>	<b>11,547.19</b>

## 7、合同资产

2020年12月末，根据新收入准则的规定，公司已向客户转让商品或服务而有权收取对价的权利（且该权利取决于时间流逝之外的其他因素）作为合同资产列示，具体情况如下表：

单位：万元

项目	2020年12月31日		
	账面余额	减值准备	账面价值
训练数据定制服务合同产生的合同资产	53.30	-1.60	51.70
训练数据产品合同产生的合同资产	41.62	-1.25	40.37
训练数据相关的应用服务合同产生的合同资产	46.81	-1.40	45.41
<b>合计</b>	<b>141.73</b>	<b>-4.25</b>	<b>137.47</b>

2020年，合同资产计提减值准备情况如下表：

项目	2020年		
	本期计提	本期转回	本期转销或核销
训练数据定制服务合同产生的合同资产	-	-6.42	-
训练数据产品合同产生的合同资产	1.25	-	-
训练数据相关的应用服务合同产生的合同资产	1.40	-	-
<b>合计</b>	<b>2.65</b>	<b>-6.42</b>	<b>-</b>

## 8、其他流动资产

报告期各期末，公司其他流动资产情况如下表：

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
银行理财产品	-	-	-	-	8,300.00	97.73%
预缴企业所得税	-	-	-	-	114.60	1.35%
待抵扣及待认证进项税	11.45	9.10%	0.20	0.28%	39.77	0.47%
预付房屋租赁费	114.35	90.90%	69.65	99.72%	38.40	0.45%
<b>合计</b>	<b>125.80</b>	<b>100.00%</b>	<b>69.85</b>	<b>100%</b>	<b>8,492.76</b>	<b>100%</b>

公司2018年末其他流动资产余额较大，主要是公司为提高资金收益率而购买了银行短期理财产品所致。该类银行短期理财产品均属于低风险、短周期的理财产品。2019年末和2020年末，按照新金融工具准则要求，该类银行短期理财产品计入交易性金融资产核算。

### （二）非流动资产分析

报告期各期末，公司非流动资产的具体构成情况如下：

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	账面价值	占比	账面价值	占比	账面价值	占比
<b>非流动资产：</b>						
固定资产	2,357.32	91.99%	2,481.15	90.28%	2,578.77	91.70%
无形资产	115.82	4.52%	148.48	5.40%	126.59	4.50%
长期待摊费用	51.03	1.99%	87.65	3.19%	59.84	2.13%
递延所得税资产	38.31	1.49%	31.10	1.13%	47.07	1.67%
<b>非流动资产合计</b>	<b>2,562.48</b>	<b>100.00%</b>	<b>2,748.39</b>	<b>100.00%</b>	<b>2,812.27</b>	<b>100.00%</b>

公司非流动资产项目具体分析如下：

#### 1、固定资产

报告期各期末，公司的固定资产情况如下：

单位：万元

2020年12月31日				
类别	原值	累计折旧	减值准备	账面价值
房屋及建筑物	2,102.20	99.85	-	2,002.34
运输工具	158.48	134.45	-	24.03
电子设备	805.80	487.10	-	318.69
办公家具	58.86	46.60	-	12.26
<b>合计</b>	<b>3,125.33</b>	<b>768.01</b>	-	<b>2,357.32</b>
2019年12月31日				
类别	原值	累计折旧	减值准备	账面价值
房屋及建筑物	2,102.20	49.93	-	2,052.27
运输工具	158.48	125.35	-	33.13
电子设备	720.31	342.77	-	377.54
办公家具	57.68	39.47	-	18.22
<b>合计</b>	<b>3,038.67</b>	<b>557.52</b>	-	<b>2,481.15</b>
2018年12月31日				
类别	原值	累计折旧	减值准备	账面价值
房屋及建筑物	2,102.20	-	-	2,102.20
运输工具	158.48	116.29	-	42.19
电子设备	593.46	179.27	-	414.19
办公家具	53.05	32.85	-	20.19
<b>合计</b>	<b>2,907.19</b>	<b>328.41</b>	-	<b>2,578.77</b>

公司作为一家训练数据服务提供商，属于轻资产公司，固定资产规模较小，主要类别包括房屋及建筑物、运输工具、电子设备和办公家具，无大规模生产型机器设备。

报告期内，公司固定资产管理得当、运转良好，未发生减值情形。

## 2、无形资产

报告期各期末，公司的无形资产情况如下：

单位：万元

2020年12月31日				
类别	原值	累计摊销	减值准备	账面价值
软件	264.02	148.20	-	115.82

2019年12月31日				
类别	原值	累计摊销	减值准备	账面价值
软件	232.24	83.76	-	148.48
2018年12月31日				
类别	原值	累计摊销	减值准备	账面价值
软件	162.50	35.91	-	126.59

2019年和2020年公司无形资产规模增长主要是由于采购办公软件所致。

公司按照无形资产的预计使用寿命内采用直线法分期平均摊销，并于资产负债表日判断是否存在减值迹象。如存在减值迹象的，则估计其可收回金额，进行减值测试。减值测试结果表明无形资产的可收回金额低于其账面价值的，按其差额计提减值准备并计入减值损失。报告期内，公司无形资产不存在减值迹象，未发生减值情形。

### 3、其他

公司的长期待摊费用主要系房屋装修费；递延所得税资产主要系公司资产减值准备形成的可抵扣暂时性差异产生的递延所得税。

## 十一、偿债能力、流动性与持续经营能力分析

### （一）偿债能力及流动性分析

#### 1、主要债项

报告期各期末，公司负债的具体构成情况如下：

单位：万元

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
应付账款	714.44	18.14%	359.69	7.40%	193.46	4.19%
预收款项	-	-	2,480.66	51.05%	3,407.69	73.77%
合同负债	1,495.77	37.97%	-	-	-	-
应付职工薪酬	971.47	24.66%	1,201.15	24.72%	921.32	19.94%
应交税费	731.78	18.58%	787.18	16.20%	76.11	1.65%
其他应付款	25.90	0.66%	30.11	0.62%	20.87	0.45%
<b>流动负债合计</b>	<b>3,939.37</b>	<b>100.00%</b>	<b>4,858.79</b>	<b>100.00%</b>	<b>4,619.46</b>	<b>100.00%</b>

项目	2020年12月31日		2019年12月31日		2018年12月31日	
	余额	占比	余额	占比	余额	占比
负债合计	3,939.37	100.00%	4,858.79	100.00%	4,619.46	100.00%

报告期各期末，公司负债合计金额分别为 4,619.46 万元、4,858.79 万元和 3,939.37 万元。报告期内，公司的负债主要为预收款项、合同负债、应付职工薪酬、应交税费等经营性负债。其中，预收款项和合同负债占比最大，主要为尚未达到收入确认条件的销售预收款。报告期内公司不存在银行借款。

#### (1) 预收款项及合同负债

2020 年末，根据新收入准则的规定，公司已收或应收客户对价而应向客户转让商品或服务的义务作为合同负债列示，不再作为预收款项列示。

报告期内，公司的预收款项及合同负债情况如下：

单位：万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
预收款项	-	2,480.66	3,407.69
合同负债	1,495.77	-	-

报告期内，公司预收款项及合同负债均为预收产品及服务款。公司部分客户存在预付一定比例项目定金的情形。公司报告期内销售政策未发生重大变化。

报告期内，公司预收款项及合同负债涉及的主要客户为微软和三星，其中来自微软的预收款随双方业务交易规模在 2019 年有所增长，2020 年由于项目完成预收款结转收入，同时采用预付形式的新增订单较少，因此来自微软的合同负债在 2020 年末有所降低，来自三星的预收款主要由 2017 年和 2018 年在执行合同所产生，该等合同大部分在 2019 年执行完毕，相应预收款项结转收入，因此来自三星的预收款在 2019 年末大幅降低。

#### ①前五大预收款项及合同负债的具体构成情况

2020 年度前五大合同负债的具体构成：

单位：万元

客户名称	期末余额	占合同负债期末余额合计数的比例
微软（中国）有限公司	1,125.32	75.23%

客户名称	期末余额	占合同负债期末余额合计数的比例
Samsung Electronics Co., Ltd.	84.54	5.65%
北京三星通信技术研究有限公司	81.92	5.48%
北京联想软件有限公司	32.45	2.17%
科大讯飞股份有限公司	32.06	2.14%
<b>合计</b>	<b>1,356.28</b>	<b>90.67%</b>

2019 年度前五大预收款项的具体构成：

单位：万元

客户名称	年末余额	占预收账款年末余额合计数的比例
微软（中国）有限公司	1,990.59	80.24%
北京大米未来科技有限公司	106.13	4.28%
北京三星通信技术研究有限公司	100.67	4.06%
客户 2	47.03	1.90%
国家计算机网络与信息安全管理中心	44.55	1.80%
<b>合计</b>	<b>2,288.97</b>	<b>92.27%</b>

2018 年度前五大预收款项的具体构成：

单位：万元

客户名称	年末余额	占预收账款年末余额合计数的比例
微软（中国）有限公司	1,429.38	41.95%
SAMSUNG ELECTRONICS CO.,LTD.	707.82	20.77%
北京三星通信技术研究有限公司	480.59	14.10%
SAMSUNG R&D INSTITUTE INDIA—BANGALORE PRIVATE LIMITED	163.36	4.79%
SOUNDHOUNDINC	131.54	3.86%
<b>合计</b>	<b>2,912.69</b>	<b>85.47%</b>

②超过一年以上合同负债的情况及合同未履行完毕的原因

截至 2020 年 12 月 31 日，公司超过一年以上的合同负债 52.02 万元，由于对应合同尚未履行完毕，因此未结转收入，合同目前处于正常履行状态。

(2) 应付职工薪酬

报告期内，公司的应付职工薪酬情况如下：

单位：万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
短期薪酬	971.47	1,165.94	888.90
离职后福利—设定提存计划	-	35.21	32.42
辞退福利	-	-	-
<b>合计</b>	<b>971.47</b>	<b>1,201.15</b>	<b>921.32</b>

随着公司员工人数和收入规模的增长，公司 2019 年末应付职工薪酬余额有所增长。2020 年末，应付职工薪酬余额下降，主要是由于 2020 年公司业务受疫情影响，业绩增速未达预期，导致奖金较 2019 年有所下降所致。应付职工薪酬以短期薪酬为主，短期薪酬主要包括工资、奖金、津贴和补贴，以及社会保险费、住房公积金、工会经费和职工教育经费、劳务费等。

## 2、偿债能力分析

### (1) 资产负债分析

报告期各期末，公司资产负债情况如下表所示：

单位：万元

项目	2020年12月31日	2019年12月31日	2018年12月31日
资产总额	47,735.00	40,453.94	21,232.77
负债总额	3,939.37	4,858.79	4,619.46
货币资金和银行理财	37,135.32	31,927.21	12,969.16
资产负债率（合并）	8.25%	12.01%	21.76%
资产负债率（母公司）	7.92%	12.17%	21.98%
流动比率（倍）	11.47	7.76	3.99
速动比率（倍）	11.39	7.60	3.54

报告期各期末，公司合并资产负债率分别为 21.76%、12.01%和 8.25%，母公司资产负债率分别为 21.98%、12.17%和 7.92%。2019 年末相比 2018 年末资产负债率显著下降的原因是公司在 2019 年底进行了股权融资，使得净资产显著增加。公司货币性资产（货币资金和银行理财）在报告期各期末均超过负债总额，公司具备良好的资金储备，偿债能力具有充分保障。

### (2) 现金流量分析

报告期各期，公司经营活动产生的现金流量净额均为正，分别为 6,427.80 万

元、8,336.33 万元和 5,117.67 万元。良好的盈利能力和回款能力为公司债务的偿付提供了良好保障。

## （二）报告期股利分配的具体实施情况

2019 年 11 月 5 日，公司召开股东大会，同意基于 2019 年 6 月 30 日累计未分配利润进行现金分红 2,200 万元，各股东按持股比例进行分配。

除上述情况外，公司报告期内不存在其他股利分配的情况。

## （三）现金流量分析

报告期内，公司的现金流量主要情况如下表所示：

单位：万元

项目	2020 年度	2019 年度	2018 年度
经营活动产生的现金流量净额	5,117.67	8,336.33	6,427.80
投资活动产生的现金流量净额	-16,738.08	-7,153.76	-3,079.55
筹资活动产生的现金流量净额	-580.50	10,462.50	-
现金及现金等价物净增加额	-12,299.15	11,656.36	3,367.43
期末现金及现金等价物余额	4,026.38	16,325.52	4,669.16

### 1、经营性现金流量分析

报告期内，公司经营活动产生的现金流量情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
销售商品、提供劳务收到的现金	22,065.39	22,980.66	21,395.54
收到的税费返还	10.99	113.51	265.32
收到其他与经营活动有关的现金	1,094.13	856.03	848.09
<b>经营活动现金流入小计</b>	<b>23,170.51</b>	<b>23,950.20</b>	<b>22,508.94</b>
购买商品、接受劳务支付的现金	7,863.15	6,668.36	8,181.49
支付给职工以及为职工支付的现金	5,990.49	5,120.24	4,358.71
支付的各项税费	2,491.99	1,960.95	2,263.44
支付其他与经营活动有关的现金	1,707.22	1,864.31	1,277.51
<b>经营活动现金流出小计</b>	<b>18,052.85</b>	<b>15,613.86</b>	<b>16,081.15</b>
<b>经营活动产生的现金流量净额</b>	<b>5,117.67</b>	<b>8,336.33</b>	<b>6,427.80</b>

报告期内，公司销售商品、提供劳务收到的现金占营业收入的比例分别为



111.05%、96.74%和 94.55%，销售收款情况良好；经营活动产生的现金流量净额占净利润的比例分别为 95.75%、102.18%和 62.35%，2018 年和 2019 年经营活动现金流量与净利润匹配程度良好，公司盈利质量较高，2020 年该比例较低的主要原因是一方面 2020 年确认收入更多为前期预收款结转或形成应收账款但尚未收款，而同时部分客户的新增订单未采取预付方式，另一方面职工奖金 2020 年计提金额少于 2020 年实际发放的 2019 年计提金额，且支出了 2019 年底分红个人所得税。

## 2、投资性现金流量分析

报告期内，公司投资活动产生的现金流量情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
收回投资收到的现金	87,001.00	61,750.00	23,200.00
处置固定资产、无形资产和其他长期资产收回的现金净额	0.86	1.23	11.39
取得投资收益收到的现金	841.13	453.98	338.43
收到其他与投资活动有关的现金	-	-	-
<b>投资活动现金流入小计</b>	<b>87,842.99</b>	<b>62,205.21</b>	<b>23,549.82</b>
购建固定资产、无形资产和其他长期资产支付的现金	180.07	308.97	2,629.37
投资支付的现金	104,401.00	69,050.00	24,000.00
支付其他与投资活动有关的现金	-	-	-
<b>投资活动现金流出小计</b>	<b>104,581.07</b>	<b>69,358.97</b>	<b>26,629.37</b>
<b>投资活动产生的现金流量净额</b>	<b>-16,738.08</b>	<b>-7,153.76</b>	<b>-3,079.55</b>

报告期内，公司投资活动产生的现金流量主要由以下几方面构成：

(1) 收回投资收到的现金、取得投资收益收到的现金、投资支付的现金：公司为提高资金使用效率，在报告期内购买了一定规模的一年内到期的银行理财产品，并取得了相应的投资收益。

(2) 购建固定资产、无形资产和其他长期资产支付的现金：公司于 2018 年 12 月购入位于北京市海淀区知春路 1 号学院国际大厦的商业办公用房两套，采购额 2,102.20 万元。

### 3、筹资性现金流量分析

报告期内，公司筹资活动产生的现金流量情况如下：

单位：万元

项目	2020 年度	2019 年度	2018 年度
吸收投资收到的现金	-	13,020.00	-
<b>筹资活动现金流入小计</b>	<b>-</b>	<b>13,020.00</b>	<b>-</b>
分配股利、利润或偿付利息支付的现金	-	2,200.00	-
支付其他与筹资活动有关的现金	580.50	357.50	-
<b>筹资活动现金流出小计</b>	<b>580.50</b>	<b>2,557.50</b>	<b>-</b>
<b>筹资活动产生的现金流量净额</b>	<b>-580.50</b>	<b>10,462.50</b>	<b>-</b>

报告期内，公司筹资活动产生的现金流量主要由以下几方面构成：

(1) 吸收投资收到的现金：公司 2019 年完成融资，收到股东投资款 13,020 万元。

(2) 分配股利、利润或偿付利息支付的现金：公司于 2019 年向全体股东分配股利 2,200 万元。

(3) 支付其他与筹资活动有关的现金：公司于 2019 年和 2020 年分别向上市相关中介机构支付上市服务款 357.50 万元和 580.50 万元。

#### (四) 重大资本性支出情况

截至本招股说明书签署日，除本次发行募集资金拟投资项目外，公司无重大资本性支出计划。本次募集资金投资项目具体情况，请参见本招股说明书“第九节 募集资金运用与未来发展规划”。

#### (五) 流动性风险分析

报告期内，公司合并资产负债率分别为 21.76%、12.01%和 8.25%，流动比率分别为 3.99、7.76 和 11.47，速动比率分别为 3.54、7.60 和 11.39。报告期内，公司流动比率、速动比率持续提升，资产负债率持续下降，现金流量状况、销售收款情况和盈利情况良好。

截至报告期末，公司不存在银行借款等付息债务，负债主要为预收款项、应付职工薪酬等经营性负债。因此，公司不存在较为显著的流动性风险。未来，公

公司将进一步加强业务经营方面的销售管理和款项催收，持续保持良好的经营活动现金流状况，避免增加公司的流动性风险。

## （六）持续经营能力分析

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。发行人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。

海天瑞声是我国领先的训练数据专业提供商。自 2005 年成立以来，发行人始终致力于为 AI 产业链上的各类机构提供算法模型开发训练所需的数据集。发行人所提供的训练数据覆盖智能语音（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、智能驾驶、智慧城市等多种创新应用场景。其中智能语音为发行人的优势领域，产品线已包含 160 余个主要语种及方言。

海天瑞声的产品和服务已获得阿里巴巴、腾讯、百度、科大讯飞、海康威视、微软、三星、亚马逊等国内外客户的认可，应用于其研发的个人助手、智能音箱、语音导航、搜索服务、智能驾驶、机器翻译等多种人工智能产品中。目前海天瑞声客户累计数量超 500 家，基本覆盖了主要的大型科技公司、人工智能企业及科研机构。

未来发行人拟以数据为核心稳步发展，依靠自身技术积累，通过引入资本和人才等手段升级技术水平、扩大公司业务规模，持续建设、优化、创新训练数据服务，完善公司的产品体系，使其覆盖所有人工智能主流领域，提升通用训练数据产品、训练数据定制服务等主线业务的领先优势，增强数据驱动人工智能行业发展的能力。此外，发行人抓住人工智能快速发展的机遇，紧跟人工智能技术发展趋势，完善研发体系，持续加强新技术的研发力度，扩大以数据为中心的延伸服务范围，增加利润来源，提高公司获利能力。

## 十二、重大投资或资本性支出、重大资产业务重组或股权收购合并事项

报告期内，公司不存在重大投资或资本性支出、重大资产业务重组或股权收购合并事项。

## 十三、资产负债表日后事项、或有事项及其他重要事项

### （一）资产负债表日后事项

截至本招股说明书签署日，公司无重大资产负债表日后事项。

### （二）或有事项及其他重要事项

截至本招股说明书签署日，公司不存在重大或有事项及其他重要事项。

### （三）重大担保、诉讼事项

截至本招股说明书签署日，公司及公司控股股东、实际控制人、控股子公司，公司董事、监事、高级管理人员和核心技术人员均不存在尚未了结的或可预见的对公司财务状况、盈利能力及持续经营产生重大影响的重大担保、诉讼等事项。

### （四）执行新收入准则对公司的预计影响

#### 1、实施新收入准则对业务模式、合同条款的影响

公司实施新收入准则对公司各类型业务的业务模式、合同条款均无重大影响。

#### 2、执行新收入准则对公司收入确认的主要影响

原收入准则下，公司以风险报酬转移作为收入确认时点的判断标准。

新收入准则下，公司以控制权转移作为收入确认时点的判断标准。公司在履行了合同中的履约义务，将相关商品或服务的控制权转移给客户时确认收入。在满足一定条件时，公司属于在某一时段内履行履约义务，否则，属于在某一时点履行履约义务。合同中包含两项或多项履约义务的，公司在合同开始日，按照各单项履约义务所承诺商品或服务的单独售价的相对比例，将交易价格分摊至各单项履约义务，按照分摊至各单项履约义务的交易价格计量收入。交易价格是公司因向客户转让商品或服务而预期有权收取的对价金额，不包括代第三方收取的款项。公司依据新收入准则有关特定事项或交易的具体规定调整了相关会计政策。

总体而言，与原收入准则相比，公司在新收入准则下的收入确认时点和确认金额均无差异。执行新收入准则对公司各类型业务的收入确认会计政策的具体影响如下：

业务类型	原收入准则	新收入准则
训练数据定制服务	训练数据定制服务收入主要是指公司根据客户需求为客户提供训练数据定制服务而收取的收入。公司与客户通过合同或订单约定提供的服务内容、服务总量及交易总额。	
	公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。	每单元定制化训练数据被识别为单项履约义务，属于某一时点履行的履约义务，公司在将定制化训练数据向客户交付并经客户验收确认时确认收入。
训练数据产品	训练数据产品收入主要是指公司授权客户永久使用公司已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。公司与客户签订合同约定授权客户使用的训练数据产品及其交易金额。	
	公司按照合同约定向客户提交训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。	公司不从事对训练数据产品有重大影响的后续活动，因此训练数据产品属于某一时点履行的知识产权许可授权，公司在向客户交付合同中约定的训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。
训练数据相关的应用服务	训练数据相关的应用服务收入主要是指公司根据客户需求为其提供应用软件的定制开发服务或者是应用软件的授权使用，以及配套硬件的销售（如有）等，而向客户收取的收入。公司与客户通过签订合同，约定向客户提供的应用软件的定制开发或授权使用，以及配套硬件销售（如有）等服务内容及交易金额。	
	公司按照合同约定向客户交付开发完成或者授权使用的应用软件和配套硬件（如有），并经客户验收确认后，按照合同约定的交易金额确认收入。	应用软件的定制开发或者授权许可、配套硬件的销售属于可明确区分的单项履约义务，并且属于在某一时点履行的履约义务；应用软件的定制开发或者授权使用、以及配套硬件的销售，均是在向客户交付并经客户验收确认时确认收入，并按照合同约定的单项履约义务的交易价格确认收入金额。

### 3、新收入准则对首次执行日前各年度合并报表主要财务指标的影响

由于公司在新收入准则下的收入确认时点和确认金额和现行收入准则下均无差异，假定自 2018 年 1 月 1 日起开始全面执行新收入准则，对公司的合并财务报表中各年（末）营业收入、归属于公司普通股股东的净利润、资产总额、归属于公司普通股股东的净资产均无影响。

## 十四、财务报告审计截止日后主要财务信息及经营状况

### （一）会计师事务所的审阅意见

公司财务报告审计截止日为 2020 年 12 月 31 日。毕马威华振会计师事务所（特殊普通合伙）对公司 2021 年第一季度财务报表，包括 2021 年 3 月 31 日合

并及母公司的资产负债表，2021 年第一季度合并及母公司的利润表、合并及母公司的现金流量表以及相关中期财务报表附注进行了审阅，并出具了毕马威华振专字第 2101016 号《审阅报告》，发表意见如下：

“根据我们的审阅，我们没有注意到任何事项使我们相信海天瑞声公司上述中期财务报表没有在所有重大方面按照《企业会计准则第 32 号——中期财务报告》的规定编制。”

## （二）发行人的专项声明

公司董事会、监事会及董事、监事、高级管理人员已对公司 2021 年第一季度未经审计的财务报表进行了认真审阅并出具专项声明，保证该等财务报表所载资料不存在虚假记载、误导性陈述或者重大遗漏，并对其内容的真实性、准确性及完整性承担个别及连带责任。

公司法定代表人、主管会计工作的公司负责人及会计机构负责人已对公司 2021 年第一季度未经审计的财务报表进行了认真审阅并出具专项声明，保证该等财务报表的真实、准确、完整。

## （三）审计截止日后主要财务信息及变动分析

公司 2021 年第一季度未经审计但已经审阅的主要财务数据如下：

### 1、合并资产负债表主要数据

单位：万元

项目	2021 年 3 月 31 日	2020 年 12 月 31 日	变动比率
资产合计	48,773.15	47,735.00	2.17%
负债合计	3,342.91	3,939.37	-15.14%
所有者权益	45,430.25	43,795.64	3.73%
归属于母公司所有者权益合计	45,430.25	43,795.64	3.73%

截至 2021 年 3 月 31 日，公司总资产为 48,773.15 万元，较上年末增加 2.17%；总负债为 3,342.91 万元，较上年末减少 15.14%；所有者权益为 45,430.25 万元，较上年末增加 3.73%。

截至 2021 年 3 月 31 日，公司资产规模保持稳定，负债规模较上年末有所下降的主要原因是上年末存在较大规模的应付职工薪酬，系已计提尚未发放的年终

奖金，2021年3月末年终奖金已发放，因此应付职工薪酬出现明显下降。公司所有者权益较上年末增长主要来源于2021年1-3月所实现的净利润。

## 2、合并利润表主要数据

单位：万元

项目	2021年1-3月	2020年1-3月	变动比率
营业收入	4,423.89	3,126.09	41.52%
营业成本	1,189.62	536.87	121.58%
期间费用合计	1,820.09	1,420.99	28.09%
营业利润	1,841.01	1,560.95	17.94%
利润总额	1,852.31	1,558.97	18.82%
净利润	1,634.17	1,365.37	19.69%
归属于母公司所有者权益的净利润	1,634.17	1,365.37	19.69%
扣除非经常性损益后的归属于母公司股东的净利润	1,432.50	1,135.80	26.12%

2021年1-3月，公司实现营业收入4,423.89万元，同比增长41.52%；实现归属于母公司所有者权益的净利润1,634.17万元，同比增长19.69%；实现扣除非经常性损益后的归属于母公司股东的净利润1,432.50万元，同比增长26.12%。

2021年1-3月，公司营业收入和扣除非经常性损益后的归属于母公司股东的净利润同比较快增长，主要原因是一方面公司通过持续的市场拓展和业务开发，进一步提升了业务规模和盈利能力，另一方面2020年初国内新冠疫情爆发，导致公司业务经营相关的项目实施、产品开发、交付验收的进度和难度均受到影响，而2021年第一季度该等影响已经明显减弱。

2021年1-3月，公司营业收入同比增速高于净利润同比增速的主要原因是2020年1-3月由于疫情影响导致训练数据定制服务难以开展，公司在销售策略方面偏重于向客户推介已有的训练数据产品，由于训练数据产品无对应的营业成本，毛利率为100%，因此使得公司2020年1-3月综合毛利率较高，2021年1-3月国内疫情基本稳定，因此训练数据产品的收入占比相较于上年同期有所下降，从而使得2021年1-3月综合毛利率同步下降。

### 3、合并现金流量表主要数据

单位：万元

项目	2021年1-3月	2020年1-3月	变动比率
经营活动产生的现金流量净额	1,845.62	597.59	208.84%
投资活动产生的现金流量净额	-2,073.65	-11,289.15	-81.63%
筹资活动产生的现金流量净额	-276.58	-66.00	319.07%
汇率变动对现金的影响	13.75	19.70	-30.19%
现金净增加额	-490.86	-10,737.85	-95.43%

2021年1-3月，公司经营活动产生的现金流量净额为1,845.62万元，系公司正常经营活动产生；投资活动产生的现金流量净额为-2,073.65万元，主要为公司购买银行理财产品所支付的现金流出所致，同时2021年3月末公司银行理财产品余额较上年末增加2,200.87万元。

### 4、非经常性损益明细表主要数据

单位：万元

项目	2021年1-3月
非流动性资产处置损益	-1.79
计入当期损益的政府补助	0.80
交易性金融资产产生的公允价值变动收益，以及处置交易性金融资产和可供出售金融资产时取得的投资收益	213.49
除上述各项之外的其他营业外收入和支出	11.50
<b>小计</b>	<b>224.00</b>
所得税影响额	-22.32
<b>合计</b>	<b>201.68</b>

2021年1-3月，公司扣除所得税影响后非经常性损益净额为201.68万元，主要为银行理财产品所取得的投资收益，非经常性损益对经营业绩不构成重大影响。

#### （四）审计截止日后主要经营状况

公司财务报告审计截止日至本招股说明书签署日，公司主要经营状况正常，在经营模式、采购规模及采购价格、销售规模及销售价格，主要客户及供应商构成，税收政策以及其他可能影响投资者判断的重大事项方面未发生重大变化。

综上所述，公司财务报告审计截止日后的经营情况与经营业绩较为稳定，总



体运营情况良好，不存在重大异常变动情况。

## 第九节 募集资金运用与未来发展规划

### 一、募集资金情况概述

#### (一) 本次募集资金金额及使用计划

经第一届董事会第二十三次会议和 2020 年第二次临时股东大会审议，发行人本次拟公开发行人民币普通股不超过 1,070 万股。本次募集资金将严格执行《北京海天瑞声科技股份有限公司募集资金管理制度》（以下简称“《募集资金管理制度》”）的相关规定，存放于董事会决定的专项账户集中管理，并与保荐机构、存放募集资金的商业银行签订募集资金专户存储三方监管协议。公司本次募集资金运用围绕主营业务进行，扣除发行费用后，将按轻重缓急顺序投入以下项目：

单位：万元

序号	项目名称	总投资金额	拟投入募集资金
1	自主研发数据产品扩建项目	24,921.18	24,921.18
2	一体化数据处理技术支撑平台升级项目	25,057.72	25,057.72
3	研发中心升级建设项目	16,614.98	16,614.98
4	补充流动资金	10,000.00	10,000.00
合计		<b>76,593.88</b>	<b>76,593.88</b>

#### (二) 募集资金运用项目的资金投入进度计划

除补充流动资金外，本次募集资金投资项目的资金投入时间进度计划如下：

单位：万元

序号	项目名称	第一年投入	第二年投入
1	自主研发数据产品扩建项目	15,395.90	9,525.28
2	一体化数据处理技术支撑平台升级项目	14,635.67	10,422.05
3	研发中心升级建设项目	10,670.63	5,944.35
合计		<b>40,702.20</b>	<b>25,891.68</b>

如本次募集资金不能满足项目资金的需求，公司将以自有资金、自筹资金等途径自行解决资金缺口，从而保证项目的顺利实施；如果本次募集资金超过项目资金的需要，则超募资金将用于其他与主营业务相关的业务。为充分抓住市场机遇，本次发行的募集资金到位之前，若因市场竞争或公司自身经营需要等因素导致部分投资项目必须进行先期投入的，公司可使用自有资金或者自筹资金先行投

入，在募集资金到位之后予以置换。

### （三）募集资金运用项目的环境影响评价及备案情况

本次公开发行募集资金投资项目的备案及环境影响评价情况具体如下：

序号	项目名称	备案情况	环境影响评价情况
1	自主研发数据产品扩建项目	京海科信局备[2020]30号	该项目不属于环保法规规定的建设项目，不需要进行项目环境影响评价
2	一体化数据处理技术支撑平台升级项目	京海科信局备[2020]31号	该项目不属于环保法规规定的建设项目，不需要进行项目环境影响评价
3	研发中心升级建设项目	京海科信局备[2020]29号	该项目不属于环保法规规定的建设项目，不需要进行项目环境影响评价
4	补充流动资金	-	该项目不属于环保法规规定的建设项目，不需要进行项目环境影响评价

### （四）募集资金投资项目与公司现有主要业务、核心技术之间的关系

本次募集资金拟用于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及补充流动资金。募集资金投资项目均围绕公司主营业务进行，符合公司的发展战略。

自主研发数据产品扩建项目将完善公司训练数据产品体系，拓展公司训练数据产品覆盖场景和领域，加快对客户需求的响应速度；一体化数据处理技术支撑平台升级项目将进一步升级与丰富公司提供训练数据定制服务及生产训练数据产品的数据处理技术和工具，提升公司的训练数据开发服务效率，进而增强公司的核心竞争力；研发中心升级建设项目将增加公司的前瞻性技术储备，通过技术手段的运用提升产品质量。

募集资金投资项目实施后，不会和控股股东、实际控制人及其控制的企业产生同业竞争，也不会对公司独立性产生不利影响。

### （五）募集资金使用管理制度和重点投向科技创新领域的具体安排

公司已建立《募集资金管理制度》，募集资金将存放于募集资金专户集中管理，其存放、使用、变更、管理与监督将根据相关法律法规及《公司章程》、《募集资金管理制度》等的规定进行。公司将根据实际经营活动及发展规划，合理投入募集资金。

本次募集资金拟投入于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等，上述项目将完善公司训练数据产品体系，升级公司提供训练数据服务及生产训练数据产品的数据处理技术和工具，提升公司的生产服务效率和技术水平，增加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。

## 二、募集资金运用情况

### （一）自主研发数据产品扩建项目

#### 1、项目概况

经过在基础数据领域的多年深耕，海天瑞声积累了丰富的训练数据产品设计和生产经验，能够准确把握 AI 市场发展动态和下游客户需求，有针对性地设计生产了大量优质的自有知识产权的训练数据产品，逐步树立了品牌形象。随着人工智能产业持续高速发展，市场对高质量、结构化、大规模训练数据的需求量持续增长。凭借在该领域的丰富经验，海天瑞声有能力设计并开发出大规模的满足市场需求的训练数据产品。

本项目即是海天瑞声顺应人工智能应用领域拓展、增厚自有知识产权训练数据产品储备，满足业内新发训练数据需求而规划的。结合公司对市场需求的判断，本项目在原有 600 多个训练数据产品的基础上，进一步拓展公司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。本项目计划新开发共计 151 个自有知识产权的训练数据产品，分别覆盖智能语音领域（80 个）、计算机视觉领域（22 个）和发音词典领域（49 个）。新开发的训练数据产品将服务于语音助手、智能家居、智能驾驶、智慧城市、智能硬件和数字娱乐等前景广阔的 AI 下游行业。

为了更好地完成建设内容，本项目将在人力、设备和场地方面进行投资。公司将购置写字楼，建设专业的语音识别录音室、语音合成录音室和计算机视觉处理工作室，解决公司训练数据产品开发场地及设备紧张的问题。此外，公司将扩大训练数据产品开发团队，在开发能力上给予充分保障。

#### 2、必要性分析

（1）本项目是满足市场共性需要，快速响应客户需求的必要手段

近几年，中国人工智能产业快速发展，应用范围大幅拓展，竞争主体越来越多，大型互联网公司争先布局，唯恐在激烈的竞争中丧失先机。随着竞争的加剧，对同一热点的追逐达到白热化的程度，比如多家厂商推出智能音箱，多家厂商围绕个人助理推出解决方案，多家厂商正针对智能客服加大投入。同时，人工智能技术的发展和应用的拓展均对数据领域覆盖程度及精细度有了更多、更高的要求。这种局面使得同一类型的数据有了更多的用户，训练数据产品的复用率大幅提高，也对训练数据产品的延伸开发产生了更多的需求。海天瑞声作为训练数据供应商，面对市场的发展，有必要提高训练数据产品的规模、数量和种类，加速扩张储备量，加强产品对特定领域的延伸程度，从而满足人工智能行业大多数客户的共性需要。

另一方面，人工智能行业技术发展速度快，新产品和服务上市时间越早，越容易占领市场主导地位，高质量、结构化、大规模的训练数据能够缩短人工智能技术的研发周期，加快企业新产品和服务的入市进程，这就要求基础数据服务商具备在较短的时间周期内提供高质量、结构化、大规模的训练数据的能力。公司建设稳定的数据开发场地，投入软硬件设备和人才团队等资源，有助于提升训练数据产品的设计和生产能力，缩短训练数据的生产周期，提高训练数据质量，以快速、优质地响应客户需求。

## （2）本项目可以完善公司的产品体系

公司目前的训练数据产品包括语音识别、语音合成、文本、图像和词典等领域。从规模上来看，语音识别训练数据产品的规模积累较大，其他类型训练数据产品的储备规模相对较小。一方面，人工智能技术逐步发展落地，并向各行各业不断渗透；另一方面，公司服务的客户数量不断增加，客户需求日益多元化。因而原有的产品结构需要升级和优化，以满足市场的多样化需求，例如：将智能语音训练数据拓展至更多语种、更多应用场景、更多 AI 设备、覆盖更多音色类型；丰富图像数据库的规模储备、产品类型等。

本项目共计划新开发 151 个自有知识产权的训练数据产品，包括 80 个智能语音训练数据产品、22 个计算机视觉训练数据产品和 49 个发音词典训练数据产品，将有效优化公司训练数据产品结构。

### (3) 本项目是提升盈利能力，增强企业竞争力的有效途径

开发自有知识产权的训练数据产品，并通过销售训练数据产品的使用授权许可获取让渡资产使用权收入是公司实现盈利的主要方式之一，也是公司的优势业务。该类产品通常能够满足大多数客户对算法模型进行训练、优化的一般需求，且其知识产权属于公司，开发完成后可以重复销售，具备较高的毛利回报。此外，从行业竞争角度看，训练数据产品规模积累情况也是人工智能基础数据服务企业实力的体现，有助于公司通过销售通用训练数据产品获取客户，进而拓展其他个性化和深层次的训练数据定制业务，实现对公司总体业务的二次拉动。

## 3、可行性分析

(1) 本项目建设是对公司现有自有知识产权训练数据产品规模的拓展，公司具备项目实施的相关基础

### ①丰富的专业人才和管理人才储备

公司拥有十余年的国内外优质客户服务及训练数据产品开发经验，在训练数据产品的设计、采集、加工、质检等各个环节均积累了大量的实践经验和工具，构建了内外部相互补充的语言学家智库，重视培养掌握 AI 技术的科学家队伍，打造了专业化的管理团队和技术团队，拥有实施本项目所需的核心能力，具体包括对下游客户需求和市场方向的理解和预判能力，高质量训练数据产品的设计和生产能力等。

### ②丰富的技术经验

公司现已储备有自有知识产权的训练数据产品 760 余个，并在开发训练数据产品的过程中积累了大量的技术与工具。截至目前，公司拥有已授权专利 26 个、软件著作权 134 项，自行研发了多种核心软件工具和平台类工具并投入训练数据生产过程中，可覆盖智能语音、自然语言、计算机视觉等各领域训练数据的开发，实现了训练数据产品从设计、采集、加工到质检的全流程技术支持。

### ③成熟的原料数据采集渠道

原料数据的成功采集是训练数据产品高质高效生产的基础。公司已经建立了成熟的原料采集流程和渠道，构建了稳定的供应链，形成了体系化的数据采集标

准和应用方法，可满足本项目采集大批量原料数据并加以加工处理的需求。

## （2）新增产能消化的可行性分析

### ①人工智能产业发展对训练数据的需求持续增长

经过多年的发展，人工智能在深度学习、海量数据和高性能计算的支撑下，现已进入产业化应用期。近年来基于深度学习的智能语音、计算机视觉、自然语言处理等技术开始向各个应用领域渗透，全球人工智能产业规模快速增长。为抢占人工智能高地，阿里巴巴、腾讯、百度、微软、三星、亚马逊等国际知名企业均持续增加资本投入。美国、中国、英国、德国、日本等国家也分别出台相关支持政策及国家战略规划，为整个产业的发展创造良好的政策环境。根据国际数据公司（IDC）的数据，2020年，全球人工智能产业规模达到501亿美元，预计2024年将达到1,100亿美元，年复合增长率为21.73%。

数据是人工智能行业发展的三大核心要素之一，优质、丰富的训练数据对算法模型的改进尤为重要，根据Dimensional Research的全球调研报告，高质量、结构化、大规模的训练数据可以大大提高人工智能技术的研发效率，优化机器学习效果。AI行业对训练数据尤其是高质量的训练数据需求日益增多。根据艾瑞咨询的调研报告，2019年中国基础数据服务行业的市场规模达到30.9亿元，预计2025年将突破100亿元，复合年增长率达到21.8%。在这样的发展趋势下，本项目的训练数据产品将有良好的市场前景。

### ②公司积累了丰富的客户资源，可消化本项目新增的产能

海天瑞声在基础数据服务领域深耕多年，目前与产业链上的各类机构都建立了长期的战略合作伙伴关系。主要合作伙伴包括阿里巴巴、腾讯、百度、微软、三星、亚马逊等大型科技公司，科大讯飞、海康威视等人工智能企业，以及中国科学院、清华大学等科研机构。客户领域覆盖广泛、优质客户长期合作既能使公司的训练数据产品有稳定的客户群，也可增加公司产品的美誉度，推动并促成更多新进入企业与公司合作。

## 4、项目实施方案

### （1）项目实施内容

## ①产品开发

本项目拟制作的 151 个训练数据产品将完善公司的训练数据产品体系，加强公司在各应用领域的训练数据提供能力。

表：新开发训练数据产品概况

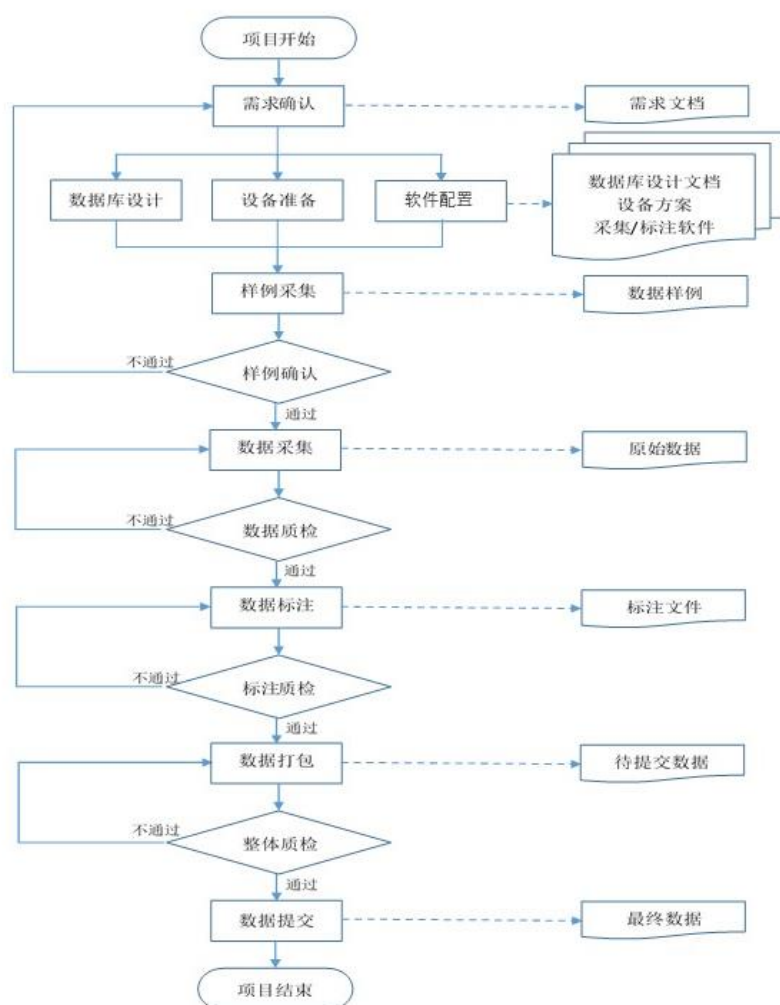
序号	训练数据产品类型	加强领域	数量(个)	规模
1	智能语音	多语种、特殊环境、指定情感，扩大发音人规模；多元化音色、平均音色、有声小说	80	24,300 人、12,750 小时多语种读文本语音识别训练数据产品； 5,500 人、5,500 小时多语种自由对话语音识别训练数据产品； 13,400 人、6,700 小时多语种车载语音识别训练数据产品； 29,000 人、14,500 小时中文多地重口音普通话远场语音识别训练数据产品等； 170,000 句中文成年人语音合成训练数据产品； 40,000 句中文儿童语音合成训练数据产品； 120,000 句多语种语音合成训练数据产品； 15,000 句英文儿童语音合成训练数据产品； 50 人英语平均音色语音合成训练数据产品；
2	计算机视觉	车载用雷达点图库、多表情人脸、扩大单库规模；新增 OCR 和手写体训练数据产品	22	18,000 小时中国、美国路况自动驾驶训练数据产品； 40,000 人人脸训练数据产品； 7,000 人多语种手写体训练数据产品； 多语种多内容 OCR 训练数据产品； 3,000 人动作及步态训练数据产品； 1,000 人手机相册训练数据产品
3	发音词典	小语种拓展	49	约 530 万多语种词条

## ②工作流程图

本项目主要涉及智能语音、计算机视觉、发音词典等类型训练数据产品的开发，针对这些项目的共同特点，公司制定了统一规范的项目流程，以保证项目高效、高质的交付。具体如下：



图：训练数据产品项目开发流程图



## (2) 项目实施目标

本项目的实施后，公司训练数据产品将增加 151 个。

## (3) 项目实施地点与实施周期

### ①项目实施地点

本项目实施地点为北京市海淀区，公司拟在北京市海淀区中关村、上地区域附近购置 1,279.00 平方米房产用于本项目的研发和办公场地。

### ②项目实施周期

本项目建设期 2 年，具体实施规划如下表所示：

表：本项目实施规划

序号	时间安排	Y1						Y2					
		2	4	6	8	10	12	14	16	18	20	22	24
1	场地购置及装修	■	■										
2	设备购置及安装		■	■				■	■	■			
3	人员引进与培训	■	■	■	■	■	■	■	■	■	■	■	■
4	训练数据产品开发	■	■	■	■	■	■	■	■	■	■	■	■
5	测试、试运行		■	■	■	■	■	■	■	■	■	■	■

注：Y1、Y2 代表建设期年份，2、4、6、8 等数字代表月份数

## 5、项目投资估算

### (1) 项目投资进度

本项目总投资 24,921.18 万元，分两年投入，第一年投资 15,395.90 万元，第二年投资 9,525.28 万元。

表：本项目投资规划表

单位：万元

序号	项目	金额	第一年	第二年	投资占比
1	场地费用	8,242.90	8,242.90	0.00	33.08%
1.1	场地购置费用	7,674.00	7,674.00	0.00	30.79%
1.2	场地装修费用	568.90	568.90	0.00	2.28%
2	设备购置费用	2,092.25	1,262.11	830.14	8.40%
3	软件购置费用	100.00	46.76	53.24	0.40%
4	技术人员工资	3,399.50	1,262.40	2,137.10	13.64%
5	原料数据采集及加工费用	7,946.40	2,847.60	5,098.80	31.89%
6	基本预备费	435.62	273.24	162.38	1.75%
7	铺底流动资金	2,704.51	1,460.89	1,243.62	10.85%
	<b>合计</b>	<b>24,921.18</b>	<b>15,395.90</b>	<b>9,525.28</b>	<b>100.00%</b>

### (2) 项目投资具体内容

#### ① 硬件设备

本项目主要硬件设备投资共计 2,092.25 万元。主要包括开发和测试设备、办公设备、车载设备等，具体实施规划如下表所示：

表：本项目硬件投入规划

序号	设备名称	品牌	单位	数量	单价 (万元)	金额 (万元)
1	GPU 服务器	NVIDIA Tesla V100/16G*10	台	8	110.00	880.00
2	数据处理服务器	IBM System X3850 X6	台	20	20.00	400.00
3	数据存储服务器	联想存储服务器 Storage V5030	台	4	55.00	220.00
4	交换机	华为 (HUAWEI) S6700	套	4	10.00	40.00
5	数字录音机	罗兰 R44	个	20	0.60	12.00
6	声卡	Focusrite Scarlett 18i20 等	个	73	1.05	77.00
7	麦克风	Shure/舒尔 UR14D/WL93 等	套	75	1.06	79.25
8	话放	AVALON VT-737SP 晶体管话放	个	8	2.30	18.40
9	耳机放大器	POWERPLAY PRO-XL HA4700	个	5	0.15	0.75
10	耳机	AKG K271MKII	个	10	0.11	1.10
11	监听控制器	Mackie Big Knob	个	5	0.25	1.25
12	手机	苹果、三星、小米、华为、OPPO、VIVO 等	个	80	0.58	46.5
13	台式机	图形工作站, 联想 (ThinkStation) P510	台	20	1.50	30.00
14	台式机	联想	台	20	0.60	12.00
15	笔记本	联想	台	40	0.60	24.00
16	数据采集车	通用型家用/商务用车	台	8	31.25	250.00
<b>合计</b>						<b>2,092.25</b>

### ②软件

本项目主要软件投资共计 100.00 万元。主要包括办公及设计软件等，具体实施规划如下表所示：

表：本项目软件投资情况

序号	软件名称	单位	数量	单价 (万元)	金额 (万元)
1	Window + Office	套	47	1.08	51.00
2	Visual Studio	套	3	8.70	26.10
3	Audition /Photoshop 6 年版	套	14	1.64	22.90
<b>总计</b>					<b>100.00</b>

### ③人员投入

为使本项目顺利完成，计划投入各类岗位人员 65 人。项目实施后，根据实际需求，逐步增加相关人员投入。具体人员配置情况如下表所示。

表：本项目人员规划

岗位名称	利用现有员工	新增员工	总计
高级管理人员	0	1	1
项目管理人员	8	35	43
营销人员	2	2	4
设计人员	3	5	8
语言学专家	5	4	9
合计	18	47	65

## 6、环境评价

本项目属于软件类产品与服务建设项目，对环境无不良影响。项目实施及日常运营所产生的废弃纸张、光盘等各类介质由碎纸机、回收桶等进行回收处理，保证办公环境和周围环境不受污染；本项目引入的设备主要为电脑、开发与测试设备，不产生噪音等污染；项目涉及的能源为办公场所的正常照明用电、电脑等开发设备用电、空调用电等，无特殊工业用电需求，用电亦按规定采取相应安全保护措施。

### （二）一体化数据处理技术支撑平台升级项目

#### 1、项目概况

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务，业务经营涉及到各类客户的 AI 算法研究、产品开发、应用拓展等环节。为了提升训练数据研发的效果和效率，使之可以满足各类应用的训练、拓展需求，公司一直致力于训练数据生产过程相关平台、工具的建设与完善，目前已自主研发了多项核心数据处理工具，涉及训练数据开发包括的设计、采集、加工和质检等环节。

随着近年来 AI 技术的日趋成熟，产业链上各类公司对训练数据的需求持续增长，为了更好地满足日益增长的市场需求、提升产能和生产效率，发行人拟在训练数据生产平台、工具的研发建设上继续增加投入，进一步提升训练数据开发能力、开发效率，保障训练数据质量，加强训练数据的安全性。

本项目拟在公司现有的一体化数据处理平台及其工具集合的基础上，扩充研发团队，提升生产工具、平台系统性开发实力；投入先进的开发设备和软件，推进测试实验室建设，升级建设数据处理工具及平台，为公司的训练数据生产提供

支持。

## 2、必要性分析

现有的一体化数据处理平台未来可能无法完全满足大规模数据开发的需求，因此海天瑞声计划通过升级该平台提高训练数据生产能力，提高公司竞争优势。

(1) 大规模数据开发及延伸服务能力的提升，需要专业高效的处理技术与工具

训练数据的开发和生产，包括设计、采集、加工、质检、安全管理等不同的环节。如果没有系统化的开发平台和专业化的软件处理工具参与到上述环节，大规模数据的开发及延伸服务就需要投入大量的人力，一方面服务的能力、质量和效率均无法保证，另一方面，低效率的处理也加大了服务的成本。随着深度学习算法的普及，高质量数据成为行业发展的重要驱动因素。而高效数据处理平台和工具的研发，可有效提高产能及效率、降低数据开发成本、缩短数据处理时间，打造一站式解决方案，使更多的企业可以享受更及时和更为多样的训练数据及相关配套服务，参与行业竞争，推进行业有序发展。

海天瑞声作为训练数据提供商，建设本项目可以进一步加强数据处理技术，提高数据处理的能力、效率，提升服务范围和水平，推进产业发展。

(2) 客户数据需求的变化要求公司升级现有的数据处理平台与工具

首先，随着人工智能技术在各行业的应用与落地，带来了数据量的大幅增长，需要一体化数据处理平台进一步提升数据处理的能力与效率，及时地应对这一变化。其次，客户对数据的精细化程度和数据质量要求更高且个性化需求更多，尤其是技术较为成熟的大企业，在向新领域拓展的过程中，不断推出新的智能设备，带动了定制化数据需求的不断提升，需要海天瑞声的数据处理平台和工具不断地按照用户的需求进行改进和变化。再次，数据处理各个环节都有大量的人员介入，而且在数据的处理工作中会有终端被采集人员、标注人员、供应商等人员参与，这对数据的安全性提出了更高的要求，对平台和工具提出了更高的要求。最后，由于海天瑞声的客户多为市场领军企业，公司要保持与其同步发展的态势，公司的技术要跟随客户的技术提升和变化，并把相应技术的升级和变化结合到一体化数据处理平台中去。

(3) 一体化数据处理平台是构建护城河，提高公司核心竞争力的必然选择

海天瑞声一直致力于各类数据处理工具的开发，应用各类技术手段，提高产品及服务的质量，一体化数据处理平台上的众多处理工具成为企业发展的有力推手及核心竞争力之一。人工智能行业发展迅猛，越来越多的数据服务企业加入到市场竞争当中，对公司的业务可能形成竞争和挑战。工欲善其事必先利其器，要保持公司的核心竞争力和持续研发能力，还需要大量的资金投入，以支持投资周期较长的平台和工具项目的持续更新和优化。足够的资金才能保证充足优质的人才投入，才能保障一体化数据处理平台的开发效率与质量。

### 3、可行性分析

(1) 海天瑞声的技术储备具备项目的实力

公司原有的关键数据处理工具，涉及数据的设计、采集、加工、质检等方面，处理范围涵盖公司大部分产品及服务。本项目是对原有的一体化数据处理平台进行升级，包括升级平台基础架构部分的全时日志库、建设全生命周期数据仓库，及其相应的交互系统，发挥日志分析对业务的协同与驱动作用，提高数据流转效率，提升平台综合能力；此外，根据实际需要新开发一个采集核心模块和两个标注核心模块。已有模块及功能的应用使得升级目标明确，稳步可靠。已有的模块和功能的设计研发经验会使新模块及功能的设计开发少走弯路，有的放矢。公司拥有稳定的技术团队，该团队长期从事数据处理技术的研发和数据处理工具的开发，具备丰富的研发经验。截止目前，技术团队已经为公司取得了专利 26 件、软件著作权 134 项。除此之外，公司与国内外顶尖人工智能公司保持着紧密的合作，熟知他们对数据的需求和关注点，有利于公司开发出更贴合客户需求的处理工具。

(2) 一站式解决方案有着广阔的市场空间

海天瑞声为全球用户提供高品质的大规模工程化训练数据服务，覆盖语音识别、语音合成、人脸识别、物体识别、轨迹识别、发音词典、机器翻译、网络搜索、语义理解等智能语音、计算机视觉、自然语言处理等技术领域。公司除了拥有丰富的菜单式通用训练数据产品，还可以基于一体化数据处理平台为客户提供从文本设计、数据采集、数据标注、模型训练、引擎评测、多语种语言学咨询到

垂直应用优化在内的训练数据定制解决方案。随着人工智能技术的不断发展和在各个行业的应用落地，更多公司开始有能力开发该领域的技术和产品，他们更需要获取一体化的、全面式的服务，以能够系统性地、高效地满足其在产品设计阶段、模型训练阶段到引擎优化各个阶段的需求。另外，客户在不同发展阶段和业务开发阶段，对数据的需求点也不同，因此对数据服务的个性化需求也会逐渐增多。尤其是随着人工智能某一项应用的深入发展，当通用训练数据产品不能满足特定需求时，必须要采用定制化服务作为有效补充。因此不管是设计定制训练数据，还是提供某一特定化的数据服务，更加全面、高效的一体化数据处理平台都将使公司在服务好客户、提高市场竞争力、推动产业发展等方面的能力大幅提高。

#### 4、项目实施方案

##### (1) 项目实施内容

##### ①产品及技术研发内容

公司现有平台上的多项数据处理工具，涉及训练数据开发包括的设计、采集、加工和质检等环节。本项目将对其中的 26 项数据处理工具进行升级开发，本项目还将在采集环节新开发“视频多通道采集模块”，在标注环节新开发“2D-3D 图像联合标注模块”和“多语种语音合成标注模块”，另外，作为平台基础架构的重要组成部分，本项目将进一步升级“全时日志库”、建设“全生命周期数据仓库”模块，提升日志分析与数据驱动的能力，从而进一步完善一体化数据处理平台。

本项目涉及的新增模块和基础架构，具体开发方向如下：

表：本项目 3 个数据核心模块及 2 个架构模块的开发方向

编号	模块名称	实现的功能
1	视频多通道采集模块	支持 8 通道以上并发同步视频数据采集，从而实现多角度、多光线及多距离条件下的视频采集。
2	2D-3D 图像联合标注模块	基于现有的 2D 图像标注模块与 3D 图像标注模块，实现 2D-3D 图像数据中点、线、框的联合标注功能。增加弹性拉升、多图跟踪等提升标注效率的新功能。
3	多语种语音合成标注模块	新增语音合成数据标注模块，可以实现集成式音字一致性校对、音素边界标注、韵律标注、词性标注等多种功能
4	全时日志库	该日志库目前已支持融合、管理和追踪数据处理平台上所有模块产生的日志文件，并通过数据分析与抽取模块有效地支撑起包含终端人管理、供应商管理、工作量管理、供应链管理等多个重要智能。本项目将升级完善该核心模块，建立独

编号	模块名称	实现的功能
		立的数据挖掘与分析决策引擎，进一步提高与现有业务流转和数据仓库的联动和整合程度，充分发挥智能决策对业务的协同与驱动作用。
5	全生命周期数据仓库	该数据仓库将整合处理平台上所有的训练数据，追踪数据自产生之时到最后交付甚至销毁的所有节点；本项目将实现全业务类型的训练数据覆盖；同时将进一步整合现有的数据保护和隐私保护功能，建立统一的数据安全管理机制，并提高数据全生命周期的实时监控与灾难应急能力。

本项目涉及原有软件类工具和平台类工具的升级，具体升级方向如下：

表：本项目所涉及的现有核心模块的升级方向

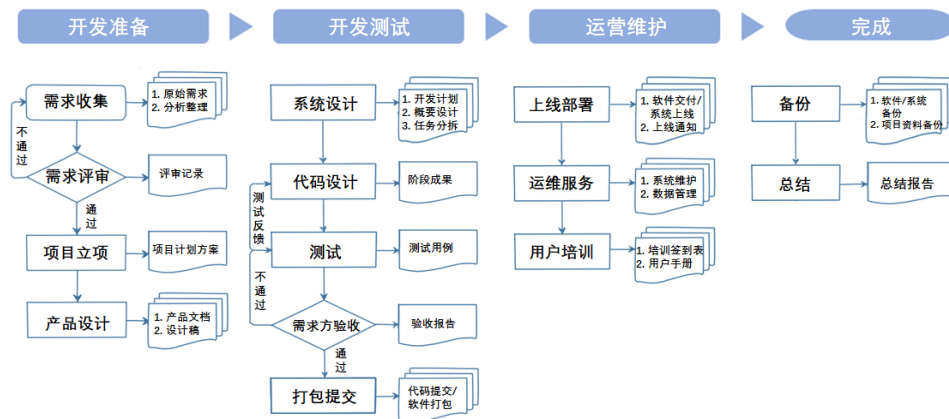
步骤	编号	工具名称	升级的方向及内容
设计	1	多语种特殊语料的设计模块	支持更多语种
	2	多领域语料设计模块	支持客服、儿童读物等特殊领域设计能力
	3	文本易读性模块	新增语速、句长、句型等特征、优化词频特征筛选，提升效果
	4	文本正则化模块	支持更多语种，优化逆正则模块性能
	5	音素集与发音标注模块	支持更多语种
	6	语料清洗模块	支持更多语种，支持更多句型与词法分析子模块
采集	7	手机录音模块	优化用户使用界面，新增极简版、海外定制版等多个子模块
	8	手机图片采集模块	支持更多图像格式与规格，优化数据上传与质检功能
	9	手机视频采集模块	支持多种视频格式，优化数据本地调整功能
	10	手机手写体采集模块	新增任务发布与管理功能，优化数据上传与质检功能
	11	PC端语音识别采集模块	支持更多服务器交互行为，优化多通道采集交互方式和性能
	12	PC端语音合成采集模块	简化合并C++版与Delphi版，优化用户交互方式，提升采集性能
	13	PC端手写体采集模块	新增任务发布与管理功能，优化数据上传与质检功能
	14	PC端视频采集模块	支持深度摄像头视频采集，优化8通道以上的采集性能
加工	15	语音数据处理模块	优化多语种语音数据处理模块以提高工作效率
	16	词典数据处理模块	支持更多语言和字符集；优化交互流程，提高工作效率
	17	语音转写模块	增加终端人分析与推荐功能以提高效率
设计/加工	18	词性标注模块	支持更多语种，并提升主要语种的预测准确率
	19	多语种发音标注模块	支持更多语种，引入发音标注算法反馈机制



步骤	编号	工具名称	升级的方向及内容
	20	命名实体标注模块	支持更多语种，优化任务发布与管理功能
	21	文本注音模块	支持更多语种，引入算法预测与反馈机制
	22	文本分类标注模块	支持更多语种，优化任务发布与管理功能
	23	图片脱敏标注模块	支持车牌、行人、关键路边物的脱敏标注功能
质检	24	集成语音信号质检模块	进一步提升语音信号质检性能
	25	相同说话人检测模块	进一步提升相同说话人检测算法效率与准确率
	26	相同人脸采集人检测模块	新增多个人脸采集算法模块，优化复杂光源下人脸识别与判重的性能

## ②开发流程图

为实现项目的高效开展，公司针对本项目制定了严格的工作流程，将定期召开阶段性会议，对各阶段的进度、质量等进行把控。整个项目的实施主要包括需求分析、产品设计、开发测试和运营维护等阶段，同时在产品设计和开发阶段实施严格的质量控制措施，保证项目的高效高质完成。本项目的开发流程图如下所示：



图：一体化数据处理平台开发流程图

## (2) 项目实施目标

本项目建设目标包括两个方面，一是利用两年的时间，完成已有 26 项数据处理核心模块的升级，并新开发 3 个数据核心模块和升级 2 个基础架构模块，加强各模块与平台间的串联性，把公司数据处理技术的升级转化到工具和平台中，提升产能与工作效率，实现数据处理能力提升 15%-20% 的目标；二是全面提升数据处理的安全性和隐私保护，实现以数据驱动技术的进化，改进优化一体化数

据处理平台，提升公司处理更多类型业务的能力，更全面地为客户提供服务。

### (3) 项目实施地点与实施周期

#### ①项目实施地点

本项目实施地点为北京市海淀区，公司拟在北京市海淀区中关村、上地区域附近购置 1,322.00 平方米房产用于本项目的研发和办公场地。

#### ②项目实施周期

本项目建设期拟定为 2 年，具体实施规划如下表所示：

表：本项目实施规划表

序号	时间安排	Y1						Y2					
		2	4	6	8	10	12	14	16	18	20	22	24
1	场地购置及装修	■	■										
2	设备购置及安装	■	■					■	■				
3	人员引进与培训	■	■	■	■	■	■	■	■				
4	平台和工具的开发与升级	■	■	■	■	■	■	■	■	■	■	■	■
5	测试、运行											■	■

注：Y1、Y2 代表建设期年份，2、4、6、8 等数字代表月份数

## 5、项目投资估算

### (1) 项目投资进度

本项目总投资 25,057.72 万元，分两年投入，其中第一年投入 14,635.67 万元、第二年投入 10,422.05 万元。具体实施规划如下表所示：

表：本项目资金投入规划

单位：万元

序号	项目	金额	第一年	第二年	投资占比
1	场地费用	8,379.20	8,379.20	0.00	33.44%
1.1	场地购置费用	7,932.00	7,932.00	0.00	31.65%
1.2	场地装修费用	447.20	447.20	0.00	1.78%
2	设备购置费用	3,577.00	1,601.20	1,975.80	14.28%
3	软件购置费用	1,162.16	551.55	610.61	4.64%
4	研发人员工资	5,159.00	1,505.00	3,654.00	20.59%
5	基本预备费	365.55	240.74	124.81	1.46%

序号	项目	金额	第一年	第二年	投资占比
6	铺底流动资金	6,414.81	2,357.98	4,056.83	25.60%
	<b>合计</b>	<b>25,057.72</b>	<b>14,635.67</b>	<b>10,422.05</b>	<b>100.00%</b>

## (2) 项目投资具体内容

## ① 硬件设备

本项目主要硬件设备投资共计 3,577.00 万元。主要包括开发和测试设备、办公设备等，具体实施规划如下表所示：

表：本项目硬件投入规划

序号	设备名称	品牌	单位	数量	单位 (万元)	金额 (万元)
1	数据处理服务器	IBM System X3850 X6	台	12	20.00	240.00
2	数据存储服务器	联想存储服务器 Storage V5030	台	16	55.00	880.00
3	GPU 服务器	NVIDIA Tesla V100/32G*10	台	20	110.00	2,200.00
4	机柜交换机等	华为 (HUAWEI) S6700	套	8	10.00	80.00
5	台式机	图形工作站，联想 (ThinkStation) P510	台	30	1.50	45.00
6	台式机	普通	台	30	0.60	18.00
7	笔记本	普通	台	50	0.60	30.00
8	管理平台服务器	浪潮 4830V4 14 核 2.0G CPU 3*480G SSD RAID: 5	台	3	8.00	24.00
9	云服务(云服务器和数据库)	阿里云	个	3	1.00	3.00
10	数据备份硬盘	希捷 4T	个	200	0.10	20.00
11	网络带宽及相关硬件	电信通	M	100	0.10	10.00
12	电脑笔记本	联想、MAC	台	25	1.00	25.00
13	开发测试用手机(云平台测试或租赁)	Ios、Android 主流机型	次	10	0.10	1.00
14	开发测试用电脑(云平台测试或租赁)	Windows、Mac 主流机型	次	10	0.10	1.00
<b>合计</b>						<b>3577.00</b>

## ② 软件

本项目计划投入软件 1,162.16 万元。主要为办公和设计软件，具体情况如下表所示：

表：本项目软件投入规划

序号	软件名称	单位	数量	单价(万元)	金额(万元)
1	Office+Windows	套	29	1.10	31.90
2	windows server2016	套	100	3.92	392.00
3	sql server 2016	套	30	10.00	300.00
4	Visual Studio	套	29	8.70	252.30
5	Audition /Photoshop 6 年版	套	29	2.35	68.15
6	Workstation	License	15	0.17	5.10
7	Microsoft Visual Studio (客户端开发软件)	License	15	4.20	94.50
8	photoshop	License	1	0.16	0.32
9	Navicat (数据库管理和设计工具)	License	3	1.00	6.00
10	ultra edit	License	15	0.07	1.99
11	NuSphere PhpED (PHP 环境开发工具)	License	4	0.36	2.85
12	Phpstorm (PHP 集成开发工具)	License	4	0.43	3.45
13	EmEditor (文本编辑器)	License	15	0.03	0.80
14	Axure RP (产品原型工具)	License	2	0.70	2.80
<b>合计</b>					<b>1162.16</b>

注：上表 6-14 项软件首次购入后需要按年续费，此处列示金额为建设期总投入。

### ③人员投入

为使本项目顺利完成，计划投入各类岗位人员 56 人。项目实施后，根据实际需求，逐步增加相关人员投入。具体人员配置情况如下表所示：

表：本项目人员投入规划

岗位名称	利用现有员工	新增员工	总计
产品经理	1	5	6
IOS 工程师	1	1	2
安卓工程师	1	2	3
UI 设计师	0	2	2
测试工程师	1	5	6
运维工程师	0	4	4
语音工程师	2	1	3
图像工程师	2	1	3
自然语言处理工程师	1	1	2

岗位名称	利用现有员工	新增员工	总计
嵌入式开发工程师	0	2	2
Php 工程师	2	1	3
C 语音工程师	0	2	2
高级研发管理人员	1	1	2
开发工程师	0	2	2
测试工程师	0	3	3
产品经理	0	4	4
项目经理	0	3	3
设计师	0	2	2
运维经理	0	2	2
<b>合计</b>	<b>12</b>	<b>44</b>	<b>56</b>

## 6、环境评价

本项目属于软件类产品与服务建设项目，对环境无不良影响。项目实施及日常运营所产生的废弃纸张、光盘等各类介质由碎纸机、回收桶等进行回收处理，保证办公环境和周围环境不受污染；本项目引入的设备主要为电脑、开发与测试设备，不产生噪音等污染；项目涉及的能源为办公场所的正常照明用电、电脑等开发设备用电、空调用电等，无特殊工业用电需求，用电亦按规定采取相应安全保护措施。

### （三）研发中心升级建设项目

#### 1、项目概况

研发中心建设是提升企业技术创新能力的关键举措，也是企业自我发展、提高竞争力的内在需求和参与市场竞争的必然选择。海天瑞声要为客户提供符合产业发展的优质训练数据与服务，就要在算法与数据、应用产品与数据的结合发展上具备充分的技术实力，以保障在训练数据结构和内容设计上具有核心竞争优势，并通过利用人工智能技术，有效提高公司数据处理的生产效率和产品质量，从而提高公司产品的市场竞争力。海天瑞声经过多年深耕，与其他基础数据服务商比较，在算法技术、数据处理技术与训练数据建设的结合上具备领先优势。伴随人工智能产业快速发展，先进的技术不断推出，需要公司不断地投入大量的人力、物力进行跟进与研发，以保持这方面的技术优势。本项目拟聚焦于语音识别、声

纹识别、语种识别、语音合成、自然语言处理、计算机视觉识别、计算机视觉分割与处理 7 个方向的前沿技术，服务于公司语音识别、语音合成、自然语言处理、计算机视觉 4 个业务方向，提升完善从产品研发到服务客户全流程的各环节，保证公司发展的持续竞争力和领先优势。

## 2、必要性分析

### (1) 公司需要进行前瞻性技术研发以满足日益增长和变化的客户需求

身处高速发展的人工智能领域，公司的健康发展，除了稳固现有主营业务之外，还需要对所在领域进行前瞻性技术研究，形成更广泛和深入的技术储备。首先，AI 技术发展突飞猛进，目前已经有了大量的垂直应用，产生了更多的数据需求，例如：多语种、远场语音、NLP 人物画像、监控视频等。公司需要紧跟技术的发展，涉猎更多的算法技术，才能更好地理解下游行业新的数据需求，清晰的把握下游技术发展方向，设计出与时俱进的产品，为下游行业的产业化提供更好的服务。其次，公司的战略定位是我国领先的、多语种、跨领域的训练数据及数据服务提供商。新的数据需要新的数据处理技术和验证手段，从而需要公司关注全产业链，并加强 AI 技术与数据建设的结合，以满足日益增长和变化的客户需求。

### (2) 公司需要通过技术手段提高产品和服务的质量

目前算法和计算力的阶段性瓶颈已经突破，超大场景数据集成为下一阶段人工智能行业发展的驱动力，其中语音、文本和视频图像数据是发展最快和需求量最大发展方向。除了数据规模的明显扩大，随着人工智能产业的不断深入发展，客户对数据产品及服务的细化与质量提出了更高的要求。数据质量是影响技术发展和应用产品开发效果的核心要素，因此作为专业的训练数据提供商，在数据制作过程中有必要通过对技术手段的合理运用，提升数据制作过程的自动化，并加强对数据质量的验证，提升数据集的一致性、安全性和完整性，进而通过扩大并提高数据应用场景下的基线模型与指标的范围及准确度，为数据使用者提供效果保障，提升海天瑞声的服务质量。

### (3) 公司需要提升盈利能力，培育新的利润增长点

自主创新是公司的核心竞争力，其最终目的是通过对新技术的研究，将其应

用在产品及服务中，提升产品质量，增加产品附加值，培育新的利润增长点。海天瑞声在基础数据服务领域已取得了优异的成绩，发展势头良好。在成本日益上涨、更多企业加入竞争的大环境下，为保持良好的利润空间，缓解持续增长的压力，需要在已经形成的品牌优势下，进一步探索相关领域的发展。本项目的实施，研究方向不仅仅围绕公司现有的产品及服务进行，还将持续关注研究最新的技术和市场需求，各项研究成果转化成新产品及服务进入市场后，将进一步延伸公司的产品线，为公司培育出新的利润增长点。

### 3、可行性分析

(1) 本项目建设内容符合我国产业政策规划方向，受国家政策大力支持

2017年3月，人工智能首次被写入了政府工作报告，李克强总理在报告中指出，要加快培育壮大新兴产业，加快人工智能等技术研发和转化，做大做强产业集群。2017年7月国务院印发《新一代人工智能发展规划》中提出，“到2020年，人工智能总体技术和应用与世界先进水平同步，成为新的重要经济增长点，核心产业规模超过1,500亿元；到2025年，新一代人工智能在智能制造、智能医疗、智慧城市、智能农业、国防建设等领域得到广泛应用，核心产业规模超过4,000亿元”。2019年3月，中央全面深化改革委员会《关于促进人工智能和实体经济深度融合的指导意见》要求：“促进人工智能和实体经济深度融合，要把握新一代人工智能发展的特点，坚持以市场需求为导向，以产业应用为目标，深化改革创新，优化制度环境，激发企业创新活力和内生动力，结合不同行业、不同区域特点，探索创新成果应用转化的路径和方法，构建数据驱动、人机协同、跨界融合、共创分享的智能经济形态。”本项目涉及的业务领域和技术方向，完全符合国家对行业的政策指导方向。

(2) 作为基础数据服务行业的优势企业，海天瑞声拟定的研发方向贴合市场需求

海天瑞声很早就以训练数据提供商的身份进入了基础数据服务领域，产品与服务涉及语音识别/合成、计算机视觉、自然语言处理等主要方向，在训练数据领域取得了领先的市场地位。伴随着核心技术、大量优质数据和算力的支撑，人工智能技术得到的迅猛的发展，在众多垂直领域得到了广泛的商业化应用，例如：

智能驾驶、语音助手、智能客服、人脸识别系统、服务机器人和智能无人机等。本项目围绕人工智能核心领域，甄别出符合下游行业发展需要的重点前沿方向，拟研究的技术方向均能够较好的贴合市场需求。

(3) 海天瑞声较为完备的技术积累以及领先的研发优势，为本项目的成功实施提供了保证

从技术积累方面来看，截至 2021 年 5 月 7 日，海天瑞声已获得 24 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，及 134 项软件著作权，涉及基础算法研究领域、数据处理领域等，技术储备充足。同时，公司拥有类型丰富的自有知识产权的高品质工程化训练数据资源，能够为自主研发提供高质量的训练数据保障。最后，公司十分重视研发团队的知识结构和梯队建设，现有团队皆为相关专业硕士与博士以上成员构成，专业扎实、技术先进、经验丰富，在算法技术研发、数据处理技术研发、模型评测与训练等专业技术方面，已经可以为客户提供高效的技术支持和服务。

#### 4、项目实施方案

##### (1) 项目实施内容

##### ① 技术研发

本研发中心的建设，主要面向语音识别、声纹识别、语种识别、语音合成、自然语言处理、计算机视觉识别、计算机视觉分割与处理 7 个方向的前沿技术进行研究，涉及的关键技术点，如下表列示：

表：本项目技术研发方向

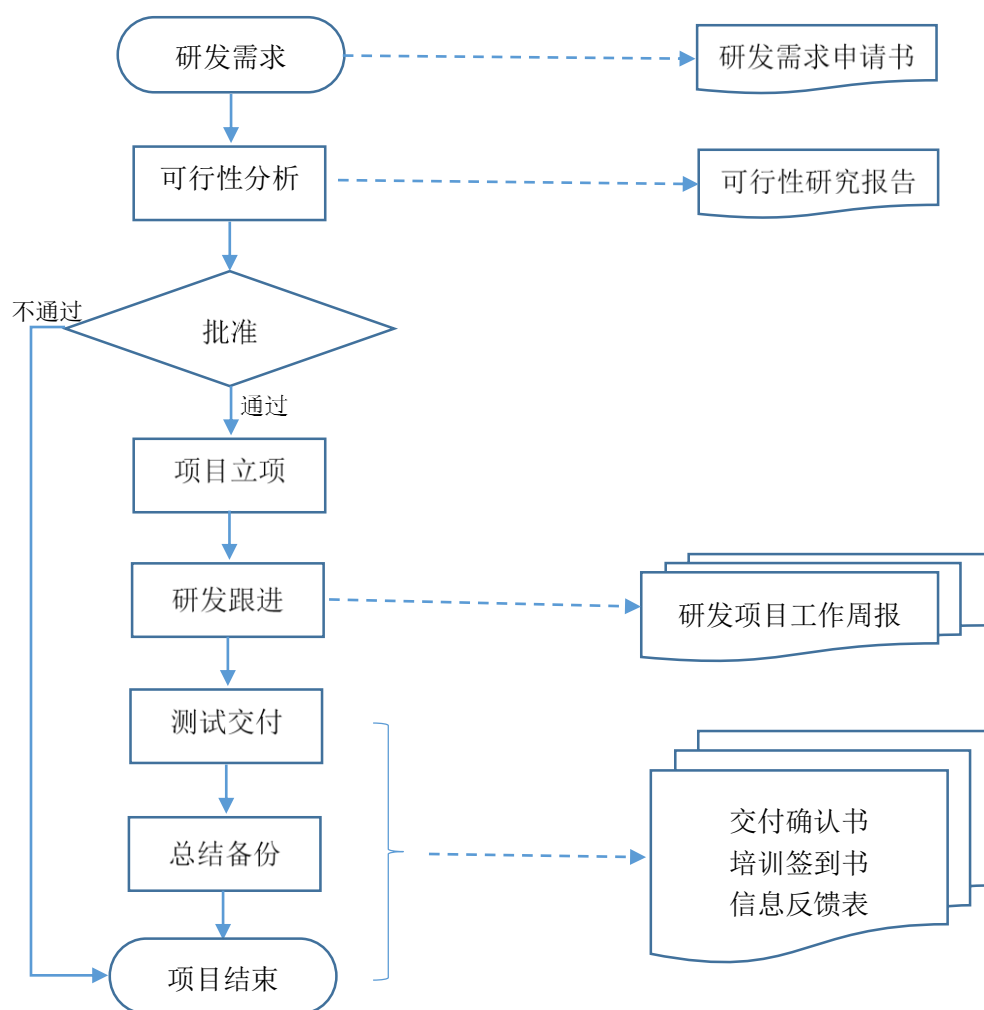
序号	技术方向	关键技术
1	语音识别技术	1.1 升级语音识别解码器技术 1.2 研发噪声去除技术 1.3 升级语音识别后文本规范化处理技术 1.4 升级端到端语音识别技术 1.5 研发低资源语音识别训练数据开发技术
2	声纹识别技术	2.1 研发说话人确认技术 2.2 升级说话人辨识技术 2.3 升级说话人分类技术
3	语种识别技术	3.1 升级语种识别技术 3.2 升级方言识别技术
4	语音合成技术	4.1 升级多语种音素自动切分技术 4.2 升级研发端到端语音合成技术



序号	技术方向	关键技术
5	自然语言处理技术	5.1 研发多语种音素集和发音预测技术 5.2 升级多语种语料清洗技术 5.3 研发中文拼写检查技术 5.4 升级多语种词法分析技术 5.5 升级文本分类技术 5.6 研发中文文本信息抽取技术 5.7 升级知识图谱构建技术 5.8 研发多语种机器翻译技术
6	计算机视觉识别技术	6.1 升级目标检测与定位技术 6.2 升级人脸识别技术 6.3 研发目标跟踪技术
7	计算机视觉分割与处理技术	7.1 升级图像分割技术 7.2 研发图像格式的转换处理技术 7.3 研发图像特效生成技术

## ② 工作流程图

为实现项目的高效开展，公司针对本项目制定了严格的工作流程，将定期召开阶段性会议，对各阶段的进度、质量等进行把控。整个项目的实施主要包括研发需求提出、可行性分析、项目立项、跟进、交付测试、总结备份等阶段，工作流程图如下所示：



图：本项目工作流程图

## (2) 项目实施目标

通过对市场的研究和对用户潜在需求的挖掘，本项目将通过对语音识别、声纹识别、语种识别、语音合成、自然语言处理、计算机视觉识别、计算机视觉分割与处理等前沿技术的研发，为公司提供前瞻性技术储备，并且达到通过技术手段提高产品和服务质量的目的。项目实施后，研究方向不仅仅围绕公司现有产品、服务进行，同时跟进智能驾驶、智慧城市、智能家居等下游最新技术和市场需求动向。研究促进技术转化成新产品与服务进入市场，进一步延伸并拓宽公司的产品线，为公司的未来发展培育新的利润增长点。

表：本项目研发目标

业务方向	研发目的	概括性描述	拟达到的技术效果	研发周期
语音识别	通过技术手段提高产品和服务的	研发多语种语音转写系统，将语音数据自动	提升中、英、法、德等常用语言的识别性能；	2年

业务方向	研发目的	概括性描述	拟达到的技术效果	研发周期
	质量	转换成文本,为项目提供可用的识别结果。	支持 Top-N 识别结果和置信度输出; 优化解码内存占用,支持并行解码; 支持语音识别实时解码; 优化识别文本规范化处理,如加标点、数字正则化、错误词校正等。	
	1、前瞻性技术储备; 2、通过技术手段提高产品和服务的质量; 3、提升盈利能力,培育新利润增长点	提高模型训练的计算力,进行高效的模型计算,对内提供高质量的模型数据,对外扩展新业务方向。	升级现有图像处理单元(GPU)集群规模,提高模型训练计算力; 使用大规模语音数据,训练高质量声学模型和语言模型,应用于多语种语音转写系统,提高语音识别正确率; 升级现有端到端语音识别技术框架,搭建服务器,覆盖常用语言; 针对低资源语音识别训练数据的开发难点,制定可行性分析与实施方案。	1.5 年
	通过技术手段提高产品和服务的质量	在数据采标环节中,应用语种/方言识别技术,优化质检方案,提升数据质量。	升级语种/方言识别基础技术研发,并覆盖更多语言; 提高语种/方言识别算法的实时率; 整合语种/方言批量检测平台至质检环节,提升数据有效性的检测与反馈。	2 年
	通过技术手段提高产品和服务的质量	针对长对话语音训练数据的开发,进行说话人片段自动切分与分类,以便于后续语音转写与标注处理。	进一步提高现有语音片段自动切分技术的准确性; 进一步提高现有说话人分类技术的准确性; 支持复杂录音环境、多说话人分类场景的切分与分类。	1.5 年
语 音 合 成	1、前瞻性技术储备; 2、通过技术手段提高产品和服务的质量	在语音合成训练数据的加工过程中,音素切分依然是一项耗时耗力的工作,且对标注人的专业性有很高要求;音素自动切分技术的准确性一直是重要的技术研究方向。	优化中文音素自动切分服务与性能; 支持英语、韩语、法语等更多语言的自动切分; 扩大音素误标注的质检覆盖范围。	2 年
	通过技术手段提高产品和服务的质量	针对语音合成训练数据,制作语音合成基线系统,得到合成语音样例,给客户提供参考合成样音。	升级现有中文端到端语音合成技术,提升合成效果; 提高端到端合成系统的实时率; 覆盖英语、韩语、法语等更多语种。	1.5 年

业务方向	研发目的	概括性描述	拟达到的技术效果	研发周期
自然语言处理	1、通过技术手段提高产品和服务的质量； 2、提升盈利能力，培育新利润增长点	研发应用于人工智能相关的多语言处理技术，增强公司多语言处理能力，同时扩展可能的业务。	覆盖更多语言和方言的音素集设计，以及发音预测技术的研发； 提供音素集方案； 完成服务器端搭建。	1年
	通过技术手段提高产品和服务的质量	完善语料清洗流程，提高语料设计规范性与语料质量，提升语料设计服务性。	升级语料清洗流程，覆盖更多语种； 完成中文拼写检查技术的研发。	2年
	1、前瞻性技术储备； 2、通过技术手段提高产品和服务的质量	升级文本处理平台，包括词法分析、命名实体识别、句法分析、文本分类、信息抽取、知识图谱构建、机器翻译等自然语言处理技术。	升级中文等现有词法分析技术，语言支持范围扩种至 20 种； 升级文本分类计数； 提高信息抽取技术的准确性，包括对人名、地名、机构名、时间等常用信息的抽取，有效应用于知识图谱训练数据的构建中； 研发中韩、中日、中英等常见平行语言的机器翻译技术。	2年
计算机视觉	1、前瞻性技术储备； 2、通过技术手段提高产品和服务的质量	研究并优化物体分割技术，对图像中的目标物体进行像素级别分割。对自动驾驶，智慧城市，智能家居等领域提供高质量的数据产品。	提升自动分割技术的准确率； 扩大对行车场景，监控场景，家居场景等覆盖； 实现物体轮廓提取； 整合物体检测技术，进一步提高特定种类物体的自动图像分割性能； 提供完整易用的调用接口。	2年
	通过技术手段提高产品和服务的质量	升级目标自动检测和跟踪技术，应对图像和视频不同目标，自动标注，以提高数据加工质量。	支持行人、车辆目标物体； 支持多个目标物体； 实现目标物体自动定位； 提供完整易用的调用接口。	1年
	通过技术手段提高产品和服务的质量	升级物体识别技术，对采集的图像或者视频数据内容进行自动识别，提高物体识别的精度。	提高常见物体、人脸，行人、车辆等目标物体的识别准确率； 支持图像与视频流； 优化现有人脸识别技术，扩展人脸识别技术在更多标注与质检环节中的应用； 支持物体的更多属性自动标注； 提供完整易用的调用接口。	2年
	1、前瞻性技术储备； 2、通过技术手段提高产品和服务	研究图像和视频预处理和后处理技术，应对不同硬件设备、不同数据格式的自动生成。丰	对常见图像和视频源的采集和读写，包括可见光、红外、CT 图像、双目视差图像和 3D 深度图等；	1年

业务方向	研发目的	概括性描述	拟达到的技术效果	研发周期
	的质量	富数据生产的能力,提高训练数据的可扩展性,同时提高客户满意度。	对常见图像格式的自动识别和读写,包括 JPEG、PNG、矢量图、索引图,医疗 DICOM 图像等; 图像配准、拼接、滤波、增强、HDR、去水印、增加特效等图像后处理实现; 提供完整易用的调用接口。	

### (3) 项目实施地点与实施周期

#### ①项目实施地点

本项目实施地点为北京市海淀区,公司拟在北京市海淀区中关村、上地地区购置 850.00 平方米房产用于本项目的研发和办公场地。

#### ②项目实施周期

本项目建设期拟定为 2 年,具体实施规划如下表所示:

表:项目实施规划

序号	时间安排	Y1						Y2					
		2	4	6	8	10	12	14	16	18	20	22	24
1	场地购置、装修	■	■										
2	软硬件购置及安装	■	■					■	■				
3	人员引进与培训		■	■									
4	研究开发		■	■	■	■	■	■	■	■	■	■	■

注:Y1、Y2 代表建设期年份,2、4、6、8 等数字代表月份数

## 5、项目投资估算

### (1) 项目投资进度

本项目总投资 16,614.98 万元,分两年投入,其中第一年投入 10,670.63 万元、第二年投入 5,944.35 万元。具体实施规划如下表所示:

表:项目投资情况

单位:万元

序号	项目	金额	第一年	第二年	投资占比
1	场地费用	5,193.00	5,193.00	0.00	31.25%

序号	项目	金额	第一年	第二年	投资占比
1.1	场地购置费用	5,100.00	5,100.00	0.00	30.70%
1.2	场地装修费用	93.00	93.00	0.00	0.56%
2	设备购置费用	2,549.00	1,107.00	1,442.00	15.34%
3	软件购置费用	1,351.70	651.40	700.30	8.14%
4	技术人员工资	7,195.50	3,510.00	3,685.50	43.31%
5	基本预备费	325.78	209.23	116.55	1.96%
合计		<b>16,614.98</b>	<b>10,670.63</b>	<b>5,944.35</b>	<b>100.00%</b>

## (2) 项目投资具体内容

### ① 硬件设备

本项目主要硬件设备投资共计 2,549.00 万元。主要包括开发和测试设备、办公设备等，具体实施规划如下表所示：

表：硬件投入情况

序号	设备名称	品牌	单位	数量	单价 (万元)	金额 (万元)
1	GPU 服务器	NVIDIA Tesla V100/16G*10	台	13	110.00	1,430.00
2	计算服务器	曙光 (Sugon) 天阔 I980	台	20	40.00	800.00
3	存储服务器	联想存储服务器 Storage V5030	台	3	55.00	165.00
4	机房硬件设备	机柜、交换机、空调、UPS、显示器等	套	1	40.00	40.00
5	质检服务器	联想 (ThinkServer) RD450	台	12	2.00	24.00
6	台式机	普通	台	45	1.00	45.00
7	笔记本	普通	台	45	1.00	45.00
合计						<b>2,549.00</b>

### ② 软件

本项目主要软件投资共计 1,351.70 万元。主要包括开发、测试软件及存储服务等，具体实施规划如下表所示：

表：软件投入情况

序号	软件名称	型号	单位	数量	单价 (万元)	金额 (万元)
1	Windows Server 2016	Standard	套	45	3.92	176.40
2	Office+Windows	--	套	42	1.10	46.20

序号	软件名称	型号	单位	数量	单价 (万元)	金额 (万元)
3	Sql Server 2016	Enterprise	套	60	10.00	600.00
4	Visual Studio 2017	Professional	套	42	8.70	365.40
5	Matlab	2016a	套	10	2.00	20.00
6	Audition /Photoshop 6 年版	--	套	42	2.35	98.70
7	Ali 云服务	--	台	45	1.00	45.00
<b>总计</b>						<b>1,351.70</b>

### ③人员投入

本项目计划组建 45 人的研发团队，其中专职语言学家 6 人，资深科学家 3 人，高级科学家 9 人，研发工程师 27 人，具体人员配置情况如下表所示：

表：人员投入情况

人员类别	项目原有	新增	总人数
专职语言学家	0	6	6
资深科学家	0	3	3
高级科学家	1	8	9
研发工程师	2	25	27
<b>总计</b>			<b>45</b>

## 6、环境评价

本项目属于软件类产品与服务建设项目，对环境无不良影响。项目实施及日常运营所产生的废弃纸张、光盘等各类介质由碎纸机、回收桶等进行回收处理，保证办公环境和周围环境不受污染；本项目引入的设备主要为电脑、开发与测试设备，不产生噪音等污染；项目涉及的能源为办公场所的正常照明用电、电脑等开发测试设备用电、空调用电等，无特殊工业用电需求，用电亦按规定采取相应安全保护措施。

### （四）补充流动资金

#### 1、项目建设背景

以本次发行上市为契机，发行人拟充分发挥公司竞争优势、提升公司整体经营业绩，实现公司的长期可持续发展。为进一步落实公司发展战略、实现公司发展规划，满足公司业务拓展及运营管理需求，结合公司当前的财务经营状况及未

来发展规划，公司拟将本次募集资金中的 10,000 万元用于补充流动资金。

## 2、必要性分析

人工智能行业发展迅速，下游客户对训练数据产品及服务的需求可能持续变化升级，公司将紧跟行业发展趋势，持续关注市场需求变化，着力挖掘和培养自身核心技术优势并拓展自身产品及服务结构，提升公司核心竞争力。随着公司业务规模的持续扩张，发行人对流动资金的需求预计将持续扩大，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。公司补充流动资金具备必要性。

## 3、项目实施的具体安排

综合考虑公司报告期内各项营运资本以及未来业务增长等情况，公司拟将本次募集资金中的 10,000 万元用于补充流动资金。公司已建立《募集资金管理制度》，募集资金将存放于董事会决定的专项账户，公司董事会将保证上述制度的有效实施。募集资金专户不得存放非募集资金或用作其他用途。公司将在募集资金到位后一个月内与保荐机构、存放募集资金的商业银行签订三方监管协议，并严格执行中国证监会及证券交易所有关募集资金使用的规定。

在营运资金的具体使用过程中，公司将根据业务发展进程，在科学测算和合理调度的基础上，合理安排该部分资金投放的进度和金额，保障募集资金的安全和高效使用。在具体资金支付环节，将严格按照公司财务管理制度和资金审批权限进行使用。

### （五）募集资金运用涉及新取得房产的相关说明

发行人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等三个募投项目拟购置房产作为项目研发和办公场地，实施地点为北京市海淀区中关村、上地区域。

截至本招股说明书签署日，公司已经开始就项目研发和办公场地的购置与相关方进行接洽，但尚未确定上述研发和办公场地选址和签署相关购房协议。公司将加快推动相关场地购置工作，但如公司未能如期完成募投项目场地购置，可能会对募集资金投资项目的实施产生一定不利影响。



### 三、未来发展规划

#### （一）公司发展战略

##### 1、整体发展战略

人工智能行业在未来仍将保持快速增长的趋势，行业的快速增长必然导致行业对高质量、结构化、大规模训练数据需求的持续增加。公司将会：①进一步扩大在训练数据领域的优势；②将核心技术和研发生产流程体系应用延伸到行业训练数据的生产、研发中；③紧跟下游需求变化，加深对下游基础技术的理解并开发更多的业务机会。公司希望通过不断拓宽训练数据产品/服务的应用领域覆盖、扩展服务链跨度并升级服务质量，使得公司的训练数据产品/服务的发展能够匹配算法、算力发展进度，共同推动产业发展。

##### 2、发展目标

公司持续为产业链上的各类机构提供训练数据定制服务、训练数据产品和训练数据相关的应用服务。经过多年的经营，海天瑞声已形成了覆盖多领域、多场景、多语种的产品和服务体系。

随着下游行业的高速发展，智能语音、计算机视觉和自然语言技术已经越来越多地应用到更多的场景中，在智能家居、智能驾驶等领域的应用，也对数据的适用性提出了更高的要求。“一带一路”战略使得更多国家能够有机会享受技术发展的成果，这也促使人工智能技术向多语种方向发展。在智能语音领域，公司将加强对西亚、南亚、中东欧，及非洲语言等有文字的语言的训练数据的覆盖能力，并重点、持续支持下游客户在语言维度细化、多语种拓展方面的发展；在计算机视觉和自然语言领域，公司计划覆盖更多细分领域，并提升公司训练产品/服务对应用场景的覆盖能力。

近年来人工智能技术在不同行业中已经得到了更加广泛的应用，产品在行业内的应用效果很大程度上取决于其使用的垂直行业训练数据的数量和质量，快速开发出高质量的垂直行业训练数据的能力是训练数据提供商进入垂直行业应用领域的核心竞争力。公司已经开始研发用于处理某些特定行业相关的智能语音、自然语言领域数据的技术，计划将该类研究持续深化，并逐步扩展到更多的行业领域，提升公司的数据处理能力，赋能人工智能技术与更多行业实体经济的深度

融合。

在过去十几年人工智能技术发展的过程中，公司一直为诸多机构提供训练数据支持，为了适应新算法、新技术的发展需要，公司生产的各个训练数据集往往都有一定的创新点。在研发、生产训练数据的过程中，公司在数据采集、处理、传输、质检等环节积累了多项技术。公司也已经开始与一些提供智能语音技术服务的客户合作，协助其将通用的语音识别等 AI 技术应用到特定场景之中。随着人工智能算法技术逐渐成熟、广泛应用，公司也将与算法公司合作，通过训练数据服务赋能算法、应用的落地，实现算法与数据技术的结合，实现 AI 应用领域的新拓展。

## （二）为实现战略目标已采取的措施

公司已经着手开展与上述规划相关的工作。在训练数据的应用领域拓展方面，公司已经投入研发更多应用场景下的智能语音、计算机视觉训练数据；并已经开始扩充合作的语言学家团队，为公司实现更广泛的语种/方言覆盖部署前期工作。在医疗、自动驾驶等垂直行业的数据处理方面，公司也通过具体训练数据产品/服务定制项目的业务开展，积累相关的资源、工具和经验。在数据处理技术再落地方面，公司已经开始与客户合作，协助实现语音识别应用的场景拓展。

## （三）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、多场景下的语音识别、语音合成、计算机视觉以及自然语言训练数据，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。为了提高完成这一大规模研发项目的工作效率，公司计划升级一体化数据处理平台，扩建研发中心，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，提升公司在训练数据开发工具、平台研发以及数据处理技术能力方面的双重提升。同时，公司也计划与更多的境内外大型科技公司、人工智能公司、科研院所合作，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。

## 第十节 投资者保护

### 一、公司投资者权益保护的情况

#### （一）公司建立了健全的内部信息披露制度和流程

为规范公司信息披露行为，确保信息披露真实、准确、完整、及时，根据《证券法》等相关法律、法规、规范性文件及《公司章程》等的有关规定，公司制定《重大信息内部报告制度》、《信息披露事务管理制度》、《投资者关系工作管理制度》。该等制度明确了重大信息报告、审批、披露程序，明确了公司管理人员在信息披露和投资者关系管理中的责任和义务。该制度有助于加强公司与投资者之间的信息沟通，提升规范运作和公司治理水平，切实保护投资者的合法权益。公司建立并逐步完善公司治理与内部控制体系，组织机构运行良好，经营管理规范，保障投资者的知情权、决策参与权，切实保护投资者的合法权益。

#### （二）保护股东合法权益的制度和措施

公司制定了《公司章程》和《股东大会议事规则》等规定，明确了股东享有的权利及履行权利的程序。其中，股东的权利包括：依照其所持有的股份份额获得股利和其他形式的利益分配；依法请求、召集、主持、参加或者委派股东代理人参加股东大会，并行使相应的表决权；对公司的经营进行监督，提出建议或者质询；依照法律、行政法规及章程的规定转让、赠与或质押其所持有的股份；查阅章程、股东名册、公司债券存根、股东大会会议记录、董事会会议决议、监事会会议决议、财务会计报告；公司终止或者清算时，按其所持有的股份份额参加公司剩余财产的分配；对股东大会作出的公司合并、分立决议持异议的股东，要求公司收购其股份；法律、行政法规、部门规章或章程规定的其他权利。

#### （三）完善股东投票机制

公司具有完善的股东大会制度，《公司章程（草案）》和《股东大会议事规则》等制度建立了累积投票制选举公司董事、中小投资者单独计票等机制，对法定事项规定了采取网络投票方式召开股东大会进行审议表决、征集投票权的相关安排等，充分保证了股东权利。

#### （四）信息披露和投资者沟通工作的安排

公司董事会办公室负责公司信息披露、与证券监管部门的联系和解答投资者的有关问题。负责人为董事会秘书吕思遥女士，联系方式如下：

董事会秘书	吕思遥
联系地址	北京市海淀区成府路 28 号 4-801
邮政编码	100083
联系电话	010-62660772
传真号码	010-62660892
电子邮箱	ir@speechocean.com
互联网网址	<a href="http://www.speechocean.com/">http://www.speechocean.com/</a>

#### （五）未来开展投资者关系管理的规划

1、对投资者提出的获取公司资料的要求，在符合法律法规和公司章程的前提下，公司将尽力给予满足；

2、对投资者对公司经营情况和其他情况的咨询，在符合法律法规和公司章程并且不涉及公司商业秘密的前提下，董事会秘书负责尽快给予答复；

3、建立完善的资料保管制度，收集并妥善保管投资者有权获得的资料，保证投资者能够按照有关法律法规的规定，及时获得需要的信息；

4、加强对有关人员的培训工作，从人员上保证服务工作的质量。

## 二、股利分配事项

### （一）发行上市后股利分配政策

#### 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）公司制定或调整利润分配政策的研究论证程序和决策机制

1、公司制定或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、董事会审议制定或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制定或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。

公司制定或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。股东大会审议制定或调整本章程规定的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。

### （三）本次发行前后股利分配政策的差异情况

本次发行前，公司章程未对股利分配做出明确规划；本次发行后生效的公司章程（草案）对股利分配情况进行了详细约定。

## 三、本次发行前滚存利润的分配安排

根据公司 2020 年第二次临时股东大会决议，本次公开发行人民币普通股（A 股）股票并在科创板上市后，公司首次公开发行股票完成前产生的滚存利润由股票发行完成后的新老股东按持股比例共同享有。

## 四、本次公司上市后未来三年分红回报规划

为进一步提高公司股利分配政策的透明度，完善和健全公司分红决策和监督机制，保持利润分配政策的连续性和稳定性，保护投资者的合法权益，便于投资者形成稳定的回报预期，根据相关法律法规及规范性文件以及公司章程等相关规定，并综合考虑企业盈利能力、经营发展规划、股东回报、社会资金成本以及外部融资环境等因素，公司制定了《北京海天瑞声科技股份有限公司上市后未来三年股东分红回报规划》，并由 2020 年第二次临时股东大会审议通过，主要内容如

下：

#### “（一）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。

#### （二）股东分红回报规划制定原则

1、公司的利润分配政策将重视对投资者的合理投资回报，并保持利润分配政策的连续性和稳定性。

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。

#### （三）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修订的利润分配政策，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独立董事过半数同意后提交股东大会审议通过。股东分红回报规划也应经全体监事过半数审议通过。

#### （四）公司上市后前三年股东分红回报具体规划

1、公司在上市后前三年内，可以采取现金分红、股票股利或者现金分红与股票股利相结合的方式进行利润分配，并优先选择现金分红方式进行分配。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。根据《公司法》等有关法律法规及《公司章程》的规定，在符合利润分配原则的前提下，发行上市后未来三年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%。

2、如在公司上市后前三年内公司经营业绩快速增长，董事会可以在现金分红的基础上，根据公司的经营业绩与股本规模的匹配情况择机发放股票股利；也可以根据公司的盈利情况及资金需求状况提议公司进行中期现金分红。

3、公司董事会经综合考虑所处行业特点、发展阶段、自身经营模式、盈利水平及未来重大资金支出安排等因素后认为，公司目前发展阶段属于成长期，资金需求量较大，因此公司上市后前三年进行利润分配时，现金分红在利润分配中所占比例最低达到 20%。”

## 五、主要承诺

### （一）自愿锁定股份的承诺

#### 1、控股股东、实际控制人承诺：

公司控股股东、实际控制人、董事长贺琳承诺：

“自公司股票在证券交易所上市之日起三十六个月内，本人不转让或者委托他人管理本人直接和间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。

在本人担任公司董事、监事或高级管理人员期间，每年转让的股份不超过本人所持有公司股份总数的百分之二十五；离职后半年内，不转让本人持有的公司股份。

本人所持公司首次公开发行股票前已发行的股份在锁定期期满后两年内减持的，减持价格不低于发行价（指公司首次公开发行股票的发行价格，如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理，下同）；公司上市后六个月内如公司股票连续二十个交易日的收盘价均低于发行价，或者上市后六个月期末收盘价低于发行价，本人持有公司股票的锁定期自动延长六个月。上述承诺不因本人职务变更、离职等原因而终止。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

## **2、中瑞安承诺：**

“北京海天瑞声科技股份有限公司（以下简称“公司”）拟申请首次公开发行股票并上市，本方作为公司股东，现就股票锁定期限承诺如下：

自公司股票在证券交易所上市之日起三十六个月内，本方不转让或者委托他人管理本方持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购本方持有的该部分股份。

本方所持公司首次公开发行股票前已发行的股份在锁定期期满后两年内减持的，减持价格不低于发行价（指公司首次公开发行股票的发行价格，如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理，下同）；公司上市后六个月内如公司股票连续二十个交易日的收盘价均低于发行价，或者上市后六个月期末收盘价低于发行价，本方持有公司股票的锁定期限自动延长六个月。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

## **3、中移投资承诺：**

“北京海天瑞声科技股份有限公司（以下简称“公司”）拟申请首次公开发行股票并上市，本方作为公司股东，现就股票锁定期限承诺如下：

自公司股票在证券交易所上市之日起十二个月内，本方不转让或者委托他人管理本方持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购本方持有的该部分股份。

本方在公司申报首次公开发行股票并上市前六个月内通过增资扩股方式取得的公司股份，自公司完成增资扩股工商变更登记手续之日起三年内不转让或者委托他人管理，也不提议由公司回购本方持有的该部分股份。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

#### 4、中网投承诺：

“北京海天瑞声科技股份有限公司（以下简称“公司”）拟申请首次公开发行股票并上市，本方作为公司股东，现就股票锁定期限承诺如下：

自公司股票在证券交易所上市之日起十二个月内，本方不转让或者委托他人管理本方持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购本方持有的该部分股份。

本方在公司申报首次公开发行股票并上市前六个月内通过增资扩股方式取得的公司股份，自公司完成增资扩股工商变更登记手续之日起三年内不转让或者委托他人管理，也不提议由公司回购本方持有的该部分股份。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

#### 5、天津金星承诺：

“北京海天瑞声科技股份有限公司（以下简称“公司”）拟申请首次公开发行股票并上市，本方作为公司股东，现就股票锁定期限承诺如下：

自公司股票在上海证券交易所科创板上市之日起十二个月内，本方不转让或者委托他人管理本方持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购本方持有的该部分股份。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。

若因本方未履行上述承诺（因相关法律法规、政策变化、自然灾害及其他不可抗力等本企业无法控制的客观原因导致的除外），转让相关股份所取得的收益归公司所有。”

#### 6、其他机构股东承诺：

“北京海天瑞声科技股份有限公司（以下简称“公司”）拟申请首次公开发行股票并上市，本方作为公司股东，现就股票锁定期限承诺如下：

自公司股票在证券交易所上市之日起十二个月内，本方不转让或者委托他人管理本方持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购本

方持有的该部分股份。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

**7、持有公司 5%以上股份股东、担任公司董事、高级管理人员、核心技术人员唐涤飞承诺：**

“自公司股票在证券交易所上市之日起十二个月内和离职后六个月内，本人不转让或者委托他人管理本人直接或间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。

本人在担任公司董事、监事或高级管理人员期间，每年转让的股份不超过本人所持有公司股份总数的百分之二十五；离职后半年内，不转让本人持有的公司股份。

本人自所持公司首次公开发行股票前已发行的股份限售期满之日起四年内，每年转让的公司首次公开发行股票前已发行的股份不得超过上市时所持公司首次公开发行股票前已发行的股份总数的百分之二十五，减持比例可以累积使用。

本人所持公司首次公开发行股票前已发行的股份在锁定期期满后两年内减持的，减持价格不低于发行价（指公司首次公开发行股票的发行价格，如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理，下同）；公司上市后六个月内如公司股票连续二十个交易日的收盘价均低于发行价，或者上市后六个月期末收盘价低于发行价，本人持有公司股票的锁定期限自动延长六个月。上述承诺不因本人职务变更、离职等原因而终止。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

**8、间接持有公司股份、公司董事和/或高级管理人员吕思遥、志鹏承诺：**

“自公司股票在证券交易所上市之日起十二个月内，本人不转让或者委托他人管理本人直接或间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。

本人在担任公司董事、监事或高级管理人员期间，每年转让的股份不超过本

人所持有公司股份总数的百分之二十五；离职后半年内，不转让本人持有的公司股份。

本人所持公司首次公开发行股票前已发行的股份在锁定期期满后两年内减持的，减持价格不低于发行价（指公司首次公开发行股票的发行价格，如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理，下同）；公司上市后六个月内如公司股票连续二十个交易日的收盘价均低于发行价，或者上市后六个月期末收盘价低于发行价，本人持有公司股票的锁定期限自动延长六个月。上述承诺不因本人职务变更、离职等原因而终止。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

**9、间接持有公司股份的公司董事、高级管理人员、核心技术人员李科及高级管理人员、核心技术人员郝玉峰承诺：**

“自公司股票在证券交易所上市之日起十二个月内和离职后六个月内，本人不转让或者委托他人管理本人直接或间接持有的公司首次公开发行股票前已发行的股份，也不提议由公司回购该部分股份。

本人在担任公司董事、监事或高级管理人员期间，每年转让的股份不超过本人所持有公司股份总数的百分之二十五；离职后半年内，不转让本人持有的公司股份。

本人自所持公司首次公开发行股票前已发行的股份限售期满之日起四年内，每年转让的公司首次公开发行股票前已发行的股份不得超过上市时所持公司首次公开发行股票前已发行的股份总数的百分之二十五，减持比例可以累积使用。

本人所持公司首次公开发行股票前已发行的股份在锁定期期满后两年内减持的，减持价格不低于发行价（指公司首次公开发行股票的发行价格，如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理，下同）；公司上市后六个月内如公司股票连续二十个交易日的收盘价均低于发行价，或者上市后六个月期末收盘价低于发行价，本人持有公司股票的锁定期限自动延长六个月。上述承诺不因本人职务变更、离职等原因而终止。

如监管规则或监管机构对锁定期有更长期限要求的，按照监管规则或监管机构的要求执行。”

## （二）持股 5%以上股东的持股意向和减持计划

### 1、贺琳、唐涤飞、中瑞安、中瑞立承诺：

“在本方所持公司股票锁定期满后，本方拟减持公司股票的，将严格遵守中国证监会、上海证券交易所关于股东减持的相关规定进行减持。本方减持公司股票具体方式包括但不限于证券交易所集中竞价交易方式、大宗交易方式、协议转让方式等。

本方减持公司股票前，应提前三个交易日予以公告，并按照上海证券交易所规则及时、准确地履行信息披露义务。

在本方所持公司股票锁定期满后两年内拟减持公司股票的，减持价格不低于公司首次公开发行股票的发行人价格（如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理），减持所持有的公司股份数量不超过法律、法规、规范性文件的规定限制。

在本方及一致行动人（如有）持有公司 5% 以上的股份期间，上述承诺持续有效。”

### 2、中移投资承诺：

“在本方所持公司股票锁定期满后，本方拟减持公司股票的，将严格遵守中国证监会、上海证券交易所关于股东减持的相关规定进行减持。本方减持公司股票具体方式包括但不限于证券交易所集中竞价交易方式、大宗交易方式、协议转让方式等。

本方减持公司股票前，应提前三个交易日予以公告，并按照上海证券交易所规则及时、准确地履行信息披露义务。

本方减持所持有的公司股票的价格根据当时的市场价格确定。在锁定期届满后，本方将根据市场情况和投资管理安排决定是否减持公司股份。”

### 3、其他持股 5%以上股东（清德投资、上海丰琬）承诺：

“在本方所持公司股票锁定期满后，本方拟减持公司股票的，将严格遵守中国证监会、上海证券交易所关于股东减持的相关规定进行减持。本方减持公司股票具体方式包括但不限于证券交易所集中竞价交易方式、大宗交易方式、协议转让方式等。

本方减持公司股票前，应提前三个交易日予以公告，并按照上海证券交易所规则及时、准确地履行信息披露义务。

本方减持所持有的公司股票的价格根据当时的二级市场价格确定。在锁定期届满后的两年内，本方将根据市场情况和投资管理安排减持公司股份。”

#### （三）发行人、控股股东、实际控制人及董事、高级管理人员关于稳定股价的预案及承诺

##### 1、启动股价稳定预案的具体条件及停止条件

###### （1）启动条件

自公司股票上市之日起 3 年内，当公司股票连续 20 个交易日的收盘价（如果因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，须按照证券交易所的有关规定作除权、除息处理，下同）均低于公司最近一期经审计的每股净资产时，在不违反相关法律法规规定且不会导致公司不符合上市条件的前提下，公司启动本预案中的股价稳定措施。

###### （2）停止条件

（1）在上述第 1 项启动条件规定的稳定股价具体方案尚未正式实施前，如公司股票连续 5 个交易日收盘价高于每股净资产时，将停止实施股价稳定措施。

（2）在上述第 1 项启动条件规定的稳定股价具体方案的实施期间内，如公司股票连续 5 个交易日收盘价高于每股净资产时，将停止实施股价稳定措施。

（3）继续回购或增持公司股份将导致公司股权分布不符合上市条件。

##### 2、稳定股价具体措施和实施程序

当上述股价稳定预案的启动条件成就时，公司、公司控股股东、实际控制人

贺琳、公司其他董事和高级管理人员（独立董事、不在公司领取薪酬的董事、依法不能持有本公司股票的董事和高级管理人员除外，以下简称“有义务增持的董事、高级管理人员”）将及时采取以下措施稳定公司股价：

（1）公司回购股票

①公司应通过证券交易所集中竞价交易方式、要约方式或证券监督管理部门认可的其他方式回购公司股票以稳定股价。

②公司为稳定股价之目的回购股份，应符合相关法律、行政法规和规范性文件的规定，且不应导致公司股权分布不符合上市条件。

③公司应当在触发股价稳定预案的启动条件后 10 个工作日内召开董事会并提交股东大会审议，且应在 30 个工作日内召开股东大会，审议稳定股价具体方案（方案内容应包括但不限于拟回购本公司股份的种类、数量区间、价格区间、实施期限等内容）。公司股东大会对回购股份方案做出决议，须经出席会议的股东所持表决权的三分之二以上通过；或者依照公司章程的规定或者股东大会的授权直接经三分之二以上董事出席的董事会会议审议通过。

④在股东大会/董事会审议通过股份回购方案后，公司应依法通知债权人，向证券监督管理部门、证券交易所等主管部门报送相关材料，办理审批或备案手续。在完成必需的审批、备案、信息披露等程序后，公司方可实施相应的股份回购方案。若股东大会未通过股份回购方案的，公司应敦促公司控股股东、实际控制人贺琳按照本招股说明书“第十节 投资者保护/五、主要承诺/（三）发行人、控股股东、实际控制人及董事、高级管理人员关于稳定股价的预案及承诺/2、稳定股价具体措施和实施程序/（2）公司控股股东及实际控制人增持股票”部分其出具的承诺履行增持公司股票的义务。

⑤公司为稳定股价之目的进行股份回购的，公司单次用于回购股份的资金总额不超过上一年度经审计的归属于母公司所有者的净利润的 30%，在一个年度内用于回购股份的资金总额不超过上一年度经审计的归属于母公司所有者的净利润的 60%，公司董事会结合公司当时的财务状况和经营状况确定回购股份的资金总额上限。

⑥公司董事会公告回购股份预案后，公司股票若连续 5 个交易日收盘价均超



过每股净资产时，公司董事会可以作出决议终止回购股份事宜。

⑦自履行完毕一次股份回购方案后的 90 个交易日内，公司的回购义务自动暂时解除。自履行完毕一次股份回购方案后的第 91 个交易日起，如稳定股价启动条件再次触发，公司将再次履行股份回购义务。

## (2) 公司控股股东及实际控制人增持股票

①若公司股东大会未通过股份回购方案或公司股份回购方案实施完毕后 90 个交易日内再次触发稳定股价预案启动条件的，公司控股股东、实际控制人贺琳承诺应在符合《上市公司收购管理办法》等法律、行政法规和规范性文件的条件和要求的前提下，通过集中竞价、大宗交易等证券监管机构、证券交易所允许的交易方式增持公司股票。

②公司控股股东、实际控制人贺琳在公司股东大会未通过股份回购方案或者公司股份回购方案实施完毕后 90 个交易日内再次触发稳定股价预案启动条件之日起 10 个工作日内，将其拟增持公司股票的具体计划（内容包括但不限于增持股数区间、计划的增持价格上限、完成时效等）以书面方式通知公司并由公司进行公告。其增持股票计划应保证其股价稳定措施实施后公司的股权分布仍符合上市条件。

③公司控股股东、实际控制人贺琳为稳定股价之目的进行股份增持的，单次用于增持公司股票的资金总额不低于其最近一个会计年度从公司领取的薪酬（税后）或从公司分得的现金股利（税后）的 20%（以孰高为准），在一个年度内用于增持公司股票的资金总额不超过其最近一个会计年度从公司领取的薪酬（税后）的 50% 及从公司分得的现金股利（税后）的 100%。

④公司控股股东、实际控制人贺琳在增持计划完成后 6 个月内不转让所持有的公司股票，包括增持前持有的公司股票。

⑤自履行完毕一次增持方案后的 90 个交易日内，公司控股股东、实际控制人贺琳的增持义务自动暂时解除。自履行完毕一次增持方案后的第 91 个交易日起，如公司控股股东、实际控制人贺琳按照其承诺需履行增持义务的条件再次触发，将再次履行增持义务。

## (3) 其他董事、高级管理人员增持股票

①若自上述第2部分规定的稳定公司股价的措施实施完毕后90个交易日内公司再次触发稳定股价预案启动条件的，有义务增持的董事、高级管理人员承诺应在符合《上市公司收购管理办法》及《上市公司董事、监事和高级管理人员所持本公司股份及其变动管理规则》等法律法规的条件和要求的前提下，通过集中竞价、大宗交易等证券监管机构、证券交易所允许的交易方式增持公司股票。

②有义务增持的董事、高级管理人员在自上述第2部分规定的稳定公司股价的措施实施完毕后90个交易日内公司再次触发稳定股价预案启动条件之日起10个工作日内，将其拟增持股票的具体计划（内容包括但不限于增持股数区间、计划的增持价格上限、完成时效等）以书面方式通知公司并由公司进行公告。其增持股票计划应保证其股价稳定措施实施后公司的股权分布仍符合上市条件。

③有义务增持的公司董事、高级管理人员单次用于增持公司股票的资金不低于该等董事、高级管理人员最近一个会计年度从公司领取的薪酬（税后）或从公司分得的现金股利（税后）的20%（以孰高为准），在一个年度内用于增持公司股票的资金总额不超过该等董事、高级管理人员最近一个会计年度从公司领取的薪酬（税后）的30%及从公司分得的现金股利（税后）的50%。

④有义务增持的公司董事、高级管理人员在增持计划完成后6个月内不转让所持有的公司股票，包括其增持前持有的公司股票。

⑤自履行完毕一次增持方案后的90个交易日内，有义务增持的公司董事、高级管理人员的增持义务自动暂时解除。自履行完毕一次增持方案后的第91个交易日日起，如有义务增持的公司董事、高级管理人员按照其承诺需履行增持义务的条件再次触发，有义务增持的公司董事、高级管理人员将再次履行增持义务。

⑥公司在未来聘任新的董事、高级管理人员前，将要求其签署承诺书，保证其履行公司本次发行上市时董事、高级管理人员已做出的相应承诺。

#### （四）关于欺诈发行上市的股份购回的承诺

##### 1、发行人的相关承诺：

发行人承诺：

“1、本公司首次公开发行股票并在上海证券交易所科创板上市申请文件不存

在虚假记载、误导性陈述或者重大遗漏，本次公开发行股票并在科创板上市不存在任何欺诈发行的情形。

2、如本公司不符合发行上市条件，以欺骗手段骗取发行注册并已经发行上市的，本公司将在中国证监会等有权部门确认后5个工作日内启动股份购回程序，购回本次公开发行的全部新股。”

## **2、控股股东、实际控制人的相关承诺：**

发行人的控股股东、实际控制人贺琳承诺：

“1、公司首次公开发行股票并在上海证券交易所科创板上市申请文件不存在虚假记载、误导性陈述或者重大遗漏，本次公开发行股票并在科创板上市不存在任何欺诈发行的情形。

2、如公司不符合发行上市条件，以欺骗手段骗取发行注册并已经发行上市的，本人将在中国证监会等有权部门确认后5个工作日内启动股份购回程序，购回本次公开发行的全部新股。”

**（五）发行人、控股股东、实际控制人及董事、监事、高级管理人员关于招股说明书及申报文件真实性、准确性和完整性及赔偿的承诺**

### **1、发行人承诺：**

“本公司首次公开发行股票的招股说明书不存在虚假记载、误导性陈述或者重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。若监管部门认定本公司首次公开发行股票的招股说明书有虚假记载、误导性陈述或者重大遗漏，对判断本公司是否符合法律规定的发行条件构成重大、实质影响的，本公司将依法回购首次公开发行的全部新股。在监管部门上述认定后，本公司董事会应及时根据相关法律法规及公司章程规定制定及公告回购计划并提交临时股东大会审议，回购价格按照市场价格且不低于公司首次公开发行股票的发行价格（如果公司上市后因派发现金红利、送股、转增股本、增发新股等原因进行除权、除息的，则按照证券交易所的有关规定作除权除息处理）确定。如公司股票有派息、送股、公积金转增股本等除权、除息事项的，回购的股份包括首次公开发行的全部新股及其派生股份。

若监管部门认定本公司首次公开发行股票时的招股说明书有虚假记载、误导性陈述或者重大遗漏，致使投资者在证券交易中遭受损失的，本公司将依法赔偿投资者损失。具体的赔偿标准、赔偿对象、赔偿金额等详细内容待上述情形实际发生时以有权部门最终确定的赔偿方案为准。”

## **2、控股股东、实际控制人、董事长贺琳承诺：**

“公司首次公开发行股票的招股说明书不存在虚假记载、误导性陈述或者重大遗漏，本人将对其真实性、准确性、完整性承担个别和连带的法律责任。

若监管部门认定公司首次公开发行股票的招股说明书有虚假记载、误导性陈述或者重大遗漏，对判断公司是否符合法律规定的发行条件构成重大、实质影响的，本人承诺将督促公司在监管部门上述认定后履行股份回购事宜的决策程序，并在公司召开董事会、股东大会对回购股份做出决议时，本人承诺就该等回购事宜在董事会、股东大会中投赞成票。

若监管部门认定公司招股说明书有虚假记载、误导性陈述或者重大遗漏，致使投资者在证券交易中遭受损失的，本人将依法赔偿投资者损失。具体的赔偿标准、赔偿对象、赔偿金额等详细内容待上述情形实际发生时以有权部门最终确定的赔偿方案为准。”

## **3、其他董事、监事、高级管理人员承诺：**

“公司首次公开发行股票的招股说明书不存在虚假记载、误导性陈述或者重大遗漏，本人将对其真实性、准确性、完整性承担个别和连带的法律责任。若监管部门认定公司招股说明书有虚假记载、误导性陈述或者重大遗漏，致使投资者在证券交易中遭受损失的，本人将依法赔偿投资者损失。具体的赔偿标准、赔偿对象、赔偿金额等详细内容待上述情形实际发生时以有权部门最终确定的赔偿方案为准。”

### **(六) 关于填补被摊薄即期回报的措施及承诺**

#### **1、发行人承诺：**

“公司首次公开发行股票完成后，公司股本和净资产规模将有较大幅度增加，公司摊薄后的即期及未来每股收益和净资产收益率面临下降的风险。为降低本次

发行摊薄公司即期回报的风险，增强对股东利益的回报，公司拟通过加大研发力度、强化募集资金管理、加快募投项目投资进度、加强经营管理和内部控制、进一步完善利润分配制度等措施，从而提升资产质量，提高销售收入和未来收益，实现可持续发展，以填补回报：

### 1、加大研发力度，不断提高公司市场竞争力和持续盈利能力

公司拥有高水平的研发团队，未来公司将继续加大研发投入，将继续巩固和发挥自身研发、销售等优势，不断丰富和完善产品和服务，提升研发技术水平，持续拓展国内和海外市场，提高公司市场竞争力，增强公司的持续盈利能力，实现公司持续、稳定发展。

### 2、强化募集资金管理，加快募投项目投资进度，争取早日实现项目预期效益

为规范公司募集资金的使用与管理，确保募集资金的使用规范、安全、高效，公司制定了《募集资金管理制度》等相关制度。公司将根据相关法规和《募集资金管理制度》的要求，严格管理募集资金使用，并积极配合监管银行和保荐机构对募集资金使用的检查和监督，以保证募集资金合理规范使用，合理防范募集资金使用风险。

公司本次募集资金投资项目的建成将有利于公司技术研发和主营业务的拓展。本次募集资金到位后，公司将调配内部各项资源、加快推进募投项目建设，提高募集资金使用效率。本次募集资金到位前，公司将通过自有资金和从其他渠道筹措的资金，开展募投项目的前期准备和启动工作，争取尽早实现项目预期收益，增强未来几年的股东回报，降低发行导致的即期回报摊薄风险。

### 3、加强经营管理和内部控制，提升经营效率和盈利能力

公司未来几年将进一步提高经营和管理水平，提升公司的整体盈利能力。公司将努力提高资金的使用效率，完善并强化投资决策程序，节省公司的财务费用支出。公司也将加强企业内部控制，发挥企业管控效能。推进全面预算管理，优化预算管理流程，加强成本管理，强化预算执行监督，全面有效地控制公司经营和管控风险。

### 4、进一步完善利润分配制度，强化投资者回报机制

公司已经按照相关法律法规的规定制订了《公司章程(草案)》(上市后适用)、《公司上市后未来三年股东分红回报规划》，建立了健全有效的股东回报机制。本次发行完成后，将按照法律法规的规定和《公司章程》、《公司上市后未来三年股东分红回报规划》的约定，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，有效维护和增加对股东的回报。”

## **2、控股股东、实际控制人贺琳承诺：**

“为确保北京海天瑞声科技股份有限公司（以下简称“公司”）拟采取的填补因首次公开发行股票而被摊薄即期回报的措施能够切实履行，作为公司的控股股东、实际控制人、董事长，本人承诺如下：

- 1、承诺不越权干预公司经营管理活动，不侵占公司利益；
- 2、承诺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、承诺对董事和高级管理人员的职务消费行为进行约束；
- 4、承诺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、承诺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；
- 6、承诺若公司后续推出股权激励的政策，则拟公布的公司股权激励的行权条件与公司填补回报措施的执行情况相挂钩。”

## **3、其他董事、高级管理人员承诺：**

“为确保北京海天瑞声科技股份有限公司（以下简称“公司”）拟采取的填补因首次公开发行股票而被摊薄即期回报的措施能够切实履行，作为公司的董事、高级管理人员，本人承诺如下：

- 1、承诺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、承诺对董事和高级管理人员的职务消费行为进行约束；
- 3、承诺不动用公司资产从事与其履行职责无关的投资、消费活动；

4、承诺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、承诺若公司后续推出股权激励的政策，则拟公布的公司股权激励的行权条件与公司填补回报措施的执行情况相挂钩。”

#### **（七）发行人、控股股东、实际控制人、董事、监事、高级管理人员及其他股东关于未能履行承诺的约束措施**

##### **1、发行人承诺：**

“本公司将严格履行本公司就首次公开发行股票并上市所作出的所有公开承诺事项，积极接受社会监督。

如本公司并非因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

（1）在本公司股东大会及中国证监会指定信息披露媒体上公开说明未履行相关承诺的具体原因并向股东和社会公众投资者道歉；

（2）向投资者提出补充承诺或替代承诺，以尽可能保护投资者的权益，并将上述补充承诺或替代承诺提交公司股东大会审议；

（3）对本公司该等未履行承诺的行为负有个人责任的董事、监事、高级管理人员调减或停发薪酬或津贴；

（4）给投资者造成损失的，本公司将向投资者依法承担赔偿责任。

如本公司因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

（1）在本公司股东大会及中国证监会指定的披露媒体上公开说明未履行相关承诺的具体原因；

（2）向投资者提出补充承诺或替代承诺，以尽可能保护投资者的权益，并将上述补充承诺或替代承诺提交公司股东大会审议。”

##### **2、控股股东、实际控制人、自然人股东、董事、监事、高级管理人员承诺：**

“本人将严格履行本人就公司首次公开发行股票并上市所作出的所有公开承

诺事项，积极接受社会监督。

如本人并非因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

(1) 在公司股东大会及中国证监会指定信息披露媒体上公开说明未履行相关承诺的具体原因并向公司股东和投资者道歉；

(2) 向公司及其投资者提出补充承诺或替代承诺，以尽可能保护公司及其投资者的权益，并将上述补充承诺或替代承诺提交公司股东大会审议；

(3) 不得转让其持有的公司股份（如有），但因被强制执行、公司重组、为履行保护投资者利益承诺等必须转股的情形除外；

(4) 如其因未履行相关承诺事项而获得收益的，所获收益归公司所有；

(5) 如其未履行相关承诺事项，给公司或投资者造成损失的，依法赔偿公司或投资者损失；

(6) 主动申请调减或停发薪酬或津贴（如有），并将此直接用于执行未履行的承诺或用于赔偿因未履行承诺而给公司或投资者带来的损失；

(7) 将应得的现金分红（如有）由公司直接用于执行未履行的承诺或用于赔偿因未履行承诺而给公司或投资者带来的损失。

如本人因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

(1) 在公司股东大会及中国证监会指定的披露媒体上公开说明未履行相关承诺的具体原因；

(2) 向公司及其投资者提出补充承诺或替代承诺，以尽可能保护公司及其投资者的权益。”

**3、机构股东（中瑞安、中移投资、清德投资、上海丰琬、中瑞立、上海兴富、中网投、杭州银杏数、杭州士兰、芜湖青和、芜湖博信）承诺：**

“本方将严格履行本方就公司首次公开发行股票并上市所作出的所有公开承诺事项，积极接受社会监督。



如本方并非因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

（1）在公司股东大会及中国证监会指定信息披露媒体上公开说明未履行相关承诺的具体原因并向公司股东和投资者道歉；

（2）向公司及其投资者提出补充承诺或替代承诺，以尽可能保护公司及其投资者的权益，并将上述补充承诺或替代承诺提交公司股东大会审议；

（3）不得转让其持有的公司股份（如有），但因继承、被强制执行、公司重组、为履行保护投资者利益承诺等必须转股的情形除外；

（4）如其因未履行相关承诺事项而获得收益的，所获收益归公司所有；

（5）如其未履行相关承诺事项，给公司或投资者造成损失的，依法赔偿公司或投资者损失；

（6）将应得的现金分红（如有）由公司直接用于执行未履行的承诺或用于赔偿因未履行承诺而给公司或投资者带来的损失。

如本方因不可抗力等无法控制的客观原因导致未能履行公开承诺事项的，将接受如下约束措施：

（1）在公司股东大会及中国证监会指定的披露媒体上公开说明未履行相关承诺的具体原因；

（2）向公司及其投资者提出补充承诺或替代承诺，以尽可能保护公司及其投资者的权益。”

#### （八）关于避免同业竞争的承诺

请参见本招股说明书“第七节 公司治理与独立性/三、同业竞争/（二）控股股东、实际控制人为避免同业竞争而出具的承诺”。

#### （九）关于减少和规范关联交易的承诺

请参见本招股说明书“第七节 公司治理与独立性/四、关联方及关联交易/（五）关于规范及减少关联交易的承诺”。

## （十）发行人关于股东信息披露的专项承诺

发行人已经出具《关于北京海天瑞声科技股份有限公司股东信息披露专项承诺》，承诺内容如下：

“本公司股东不存在以下情形：（一）法律法规规定禁止持股的主体直接或间接持有本公司股份；（二）本次发行的中介机构或其负责人、高级管理人员、经办人员直接或间接持有本公司股份；（三）以本公司股权进行不当利益输送。”

## （十一）中介机构的相关承诺

### 1、发行人保荐机构（主承销商）承诺

本公司为本次发行制作、出具的申请文件真实、准确、完整、及时，无虚假记载、误导性陈述或重大遗漏；若因本公司未能勤勉尽责，为本次发行制作、出具的申请文件有虚假记载、误导性陈述或重大遗漏，给投资者造成损失的，本公司将依法赔偿投资者损失。

### 2、发行人律师承诺

如因本所为发行人首次公开发行制作、出具的文件有虚假记载、误导性陈述或者重大遗漏，给投资者造成损失的，将依法赔偿投资者损失。本所将依照相关法律、法规规定承担民事赔偿责任，赔偿投资者损失。该等损失的赔偿金额以投资者因此而实际发生并能举证证实的损失为限，具体的赔偿标准、赔偿主体范围、赔偿金额等详细内容待上述情形实际发生时，以最终确定的赔偿方案为准。

### 3、发行人会计师承诺

本所为北京海天瑞声科技股份有限公司首次公开发行股票出具的报告的真实性和完整性依据有关法律法规的规定承担相应的法律责任，包括如果本所出具的上述报告有虚假记载、误导性陈述或重大遗漏，给投资者造成损失的，将依法赔偿投资者损失。

### 4、发行人验资机构及验资复核机构承诺

因本单位为北京海天瑞声科技股份有限公司首次公开发行制作、出具的文件，若有虚假记载、误导性陈述或者重大遗漏，给投资者造成损失的，本单位将依法赔偿投资者损失。

## 5、发行人资产评估机构承诺

因本单位为北京海天瑞声科技股份有限公司首次公开发行制作、出具的文件有虚假记载、误导性陈述或者重大遗漏，给投资者造成损失的，本单位将依法赔偿投资者损失。

## 第十一节 其他重要事项

### 一、重要合同

报告期内，公司已履行和正在履行的对于公司生产经营、未来发展或财务状况具有重大影响的合同如下：

#### （一）销售合同

公司与主要客户签署框架协议或直接签署具体销售合同，签署框架协议的情形下，主要客户与公司会根据具体需求再签署具体订单或合同。公司重大销售合同的认定标准为：单个合同金额为 500 万元以上的具体销售合同，或年度销售金额为 500 万元以上的客户对应的框架协议。

报告期内，公司已履行和正在履行的重大销售合同情况如下：

1、发行人签署的大额销售合同（单个合同金额为 500 万元以上）如下：

序号	客户名称	合同签署日期	主要内容	合同金额	完成时间	履行状态
1	Samsung Electronics Co., Ltd.	2017年6月	提供训练数据定制服务	121.22 万美元	2018/11/26	履行完毕
2	Google Ireland Limited	2018年1月	提供训练数据产品	131.94 万美元	2018/11/28	履行完毕
3	Microsoft (China) Co. Ltd.	2018年2月	提供训练数据定制服务	549.61 万元	2020/6/29	履行完毕
4	Samsung Electronics Co., Ltd.	2018年2月	提供训练数据定制服务	508.41 万美元	2019/12/17	履行完毕
5	Info-communications Media Development Authority	2018年4月	提供训练数据定制服务	131.50 万新加坡元	2020/3/5	履行完毕
6	阿里巴巴(中国)网络技术有限公司	2018年6月	提供训练数据库产品	753.47 万元	2018/11/5	履行完毕
7	Samsung Electronics Co., Ltd.	2018年8月	提供训练数据定制服务	87.23 万美元	2020/7/1	履行完毕
8	某国际消费电子产品厂商	2018年8月	提供训练数据定制服务	95.00 万美元	2019/3/6	履行完毕
9	Amazon Web Services, Inc.	2018年10月	提供训练数据定制服务及训练数据产品	167.04 万美元	2019/9/23	履行完毕
10	Microsoft Corporation	2019年4月	提供训练数据产品	93.24 万美元	2019/7/5	履行完毕
11	Microsoft (China) Co. Ltd.	2019年5月	提供训练数据定制服务	650.00 万元	-	正在履行

序号	客户名称	合同签署日期	主要内容	合同金额	完成时间	履行状态
12	Microsoft Corporation	2019年6月	提供训练数据产品	90.35 万美元	2020/1/8	履行完毕
13	Microsoft Corporation	2019年11月	提供训练数据产品	76.48 万美元	2020/5/29	履行完毕
14	Google LLC	2019年12月	提供训练数据产品	77.46 万美元	2020/5/11	履行完毕
15	北京有竹居网络技术有限公司	2020年5月	提供训练数据产品	505.00 万元	2021/3/5	履行完毕
16	Samsung Electronics Co., Ltd.	2020年10月	提供训练数据产品	176.45 万美元	2020/12/30	履行完毕
17	Institute of Big Data	2020年12月	提供训练数据产品	109.12 万美元	-	正在履行

2、发行人签署的主要框架协议（年度销售金额为500万元以上）如下：

序号	客户名称	合同签署日期	合同有效期	主要内容	履行状态
1	腾讯科技（深圳）有限公司	2017年1月	2年	提供训练数据定制服务	履行完毕
2	淘宝（中国）软件有限公司	2017年8月	3年	提供训练数据定制服务	履行完毕
3	中国某大型科技公司	2017年12月	2年	提供训练数据定制服务及训练数据产品	履行完毕
4	中国某大型科技公司	2018年12月	-		正在履行
5	腾讯科技（成都）有限公司	2019年1月	1年	提供训练数据定制服务及训练数据产品	履行完毕
6	Microsoft Corporation	2020年1月	5年	提供训练数据定制服务	正在履行
7	腾讯科技（成都）有限公司	2020年3月	至2021年3月31日	提供训练数据定制服务及训练数据产品	正在履行
8	淘宝（中国）软件有限公司	2020年3月	至2021年4月30日	提供训练数据定制服务及训练数据产品	正在履行

注：上述表格中，发行人与 Microsoft Corporation 已经于 2020 年 2 月就 2020 年 1 月签署的框架协议（序号 6）签署补充协议。

发行人与淘宝（中国）软件有限公司已经于 2020 年 6 月及 7 月就 2020 年 3 月签署的框架协议（序号 8）分别签署补充协议。

## （二）采购合同

公司与主要供应商签署采购框架协议，公司会就具体采购需求向主要供应商另行下达采购通知。公司重大采购合同的认定标准为：年度采购金额为 300 万元以上的供应商对应的框架协议。

报告期内，公司已履行和正在履行的重大采购框架协议（年度采购金额为 300 万元以上）如下：

序号	供应商名称	合同名称	服务内容	签订日期	合同期限	费用标准	履行状态
1	芜湖善慧人力资源服务有限公司	《业务服务协议》及《补充协议》	原料数据采集、标注服务	2017年1月	至2019年12月31日	按实际发生结算	履行完毕
2	北京冠华英才国际经济技术有限公司	《数据加工服务协议》	数据标注服务	2017年3月	至2022年12月31日	按实际发生结算	正在履行
		《补充协议》		2018年3月			
		《补充协议》		2019年12月			
3	北京博睿驰信人力资源有限公司	《业务服务协议》及《补充协议》	原料数据采集、标注服务	2017年1月	至2022年12月31日	按实际发生结算	正在履行
		补充协议		2019年12月			
4	善世(广东)企业服务外包有限公司	《业务服务协议》及《补充协议》	原料数据采集、标注服务	2017年1月	至2022年12月31日	按实际发生结算	正在履行
		补充协议		2019年12月			
5	共青城宜邦人力资源服务有限公司	业务服务协议	原料数据采集、标注服务	2019年10月	至2022年12月31日	按实际发生结算	正在履行
		补充协议		2019年12月			
6	才星(广州)科技服务有限公司	业务服务协议	原料数据的采集及处理	2020年11月	至2023年12月31日	按实际发生结算	正在履行

## 二、对外担保情况

报告期内，发行人不存在对外担保事项。

## 三、重大诉讼或仲裁事项

截至本招股说明书签署日，公司不存在尚未了结或可预见的对财务状况、经营成果、声誉、业务活动、未来前景可能产生较大影响的诉讼或仲裁事项。

截至本招股说明书签署日，公司的控股股东、实际控制人、控股子公司、董事、监事、高级管理人员和核心技术人员不存在尚未了结的或可预见的作为一方当事人可能对发行人产生影响的刑事诉讼、重大诉讼或仲裁事项。

#### **四、董事、监事、高级管理人员和核心技术人员最近三年涉及行政处罚、被司法机关立案侦查、被中国证监会立案调查情况**

截至本招股说明书签署日，公司董事、监事、高级管理人员和核心技术人员最近3年不存在行政处罚、被司法机关立案侦查、被中国证监会立案调查情况。

#### **五、控股股东、实际控制人报告期内涉及重大违法行为情况**

截至本招股说明书签署日，公司控股股东、实际控制人在报告期内不存在重大违法行为。

## 第十二节 有关声明

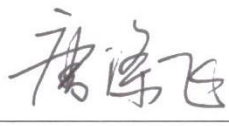
### 一、发行人全体董事、监事、高级管理人员声明

本公司全体董事、监事、高级管理人员承诺本招股说明书不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。

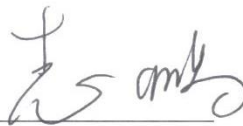
全体董事签名：




贺琳



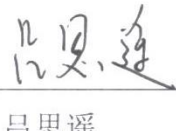
唐涤飞



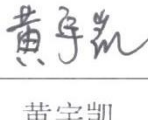
志鹏



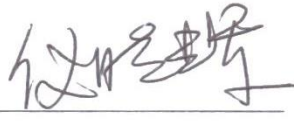
李科



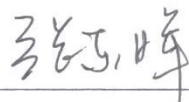
吕思遥



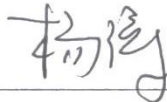
黄宇凯



仪晓辉



张东晖



杨涛

北京海天瑞声科技股份有限公司

2021年8月9日





## 一、发行人全体董事、监事、高级管理人员声明

本公司全体董事、监事、高级管理人员承诺本招股说明书不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。

### 全体监事签名：

 张小龙	 方彦彬	 熊敏
--	--	---

### 其他高级管理人员：

 郝玉峰
--

北京海天瑞声科技股份有限公司



2021年8月9日

## 二、控股股东、实际控制人声明

本人承诺本招股说明书不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担个别和连带的法律责任。

控股股东及实际控制人签名：

  
贺琳

北京海天瑞声科技股份有限公司



### 三、保荐人（主承销商）声明

本公司已对招股说明书进行了核查，确认不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

项目协办人： 廖蔚铭

廖蔚铭

保荐代表人： 张鹏

张鹏

葛青

葛青

总经理： 马骁

马骁

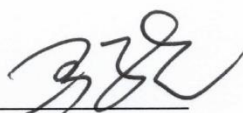
董事长、法定代表人（或授权代表）： 江禹

江禹




#### 四、保荐机构董事长、总经理声明

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

保荐机构总经理： 

马 骁

保荐机构董事长（或授权代表）： 

江 禹

华泰联合证券有限责任公司

2021年8月9日



## 五、发行人律师声明

本所及经办律师已阅读招股说明书，确认招股说明书与本所出具的法律意见书无矛盾之处。本所及经办律师对发行人在招股说明书中引用的法律意见书的内容无异议，确认招股说明书不致因上述内容而出现虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

经办律师： 周世君

周世君

王韶华

王韶华

顾鼎鼎

顾鼎鼎

律师事务所负责人： 朱小辉

朱小辉



北京市天元律师事务所

2021年8月9日

## 六、会计师事务所声明

本所及签字注册会计师已阅读招股说明书，确认招股说明书与本所出具的审计报告、内部控制审核报告及经本所鉴证的非经常性损益明细表等无矛盾之处。本所及签字注册会计师对发行人在招股说明书中引用的审计报告、内部控制审核报告及经本所鉴证的非经常性损益明细表等的内容无异议，确认招股说明书不致因上述内容而出现虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

签字注册会计师： 雷江  
雷江



张杨  
张杨



会计师事务所负责人： 邹俊  
邹俊



毕马威华振会计师事务所（特殊普通合伙）

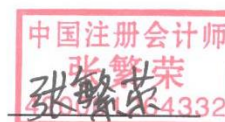
2021年8月9日

## 七、验资机构声明

本机构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招股说明书，确认招股说明书与本机构出具的验资报告（报告号为中兴华验字（2020）第 010005 号、中兴华验字（2020）第 010006 号）无矛盾之处。本机构及签字注册会计师对发行人在招股说明书中引用的验资报告的内容无异议，确认招股说明书不致因上述内容而出现虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

签字注册会计师：

张 箫



张繁荣

会计师事务所负责人：

李尊农

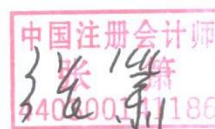
中兴华会计师事务所（特殊普通合伙）



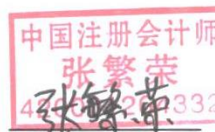
## 八、验资复核机构声明

本机构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招股说明书，确认招股说明书与本机构出具的验资复核报告（报告号为中兴华核字（2020）第 010114 号、中兴华核字（2020）第 010115 号）无矛盾之处。本机构及签字注册会计师对发行人在招股说明书中引用的验资复核报告的内容无异议，确认招股说明书不致因上述内容而出现虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

签字注册会计师：



张箫



张繁荣

会计师事务所负责人：

李尊农

中兴华会计师事务所（特殊普通合伙）





## 九、资产评估机构声明

本机构及签字注册资产评估师已阅读招股说明书，确认招股说明书与本机构出具的资产评估报告无矛盾之处。本机构及签字注册资产评估师对发行人在招股说明书中引用的资产评估报告的内容无异议，确认招股说明书不致因上述内容而出现虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应的法律责任。

签字注册资产评估师：\_\_\_\_\_

毕娇（已离职）

\_\_\_\_\_

李鹏举（已离职）

资产评估机构负责人： 罗俊军

罗俊军

蓝策亚洲（北京）资产评估有限公司



## 关于资产评估机构更名及签字注册资产评估师离职的声明

亚洲（北京）资产评估有限公司于 2017 年 11 月 6 日就北京海天瑞声科技有限公司整体变更设立北京海天瑞声科技股份有限公司事项，出具了《评估报告》（京亚评报字[2017]第 110 号），本机构说明如下：

1、本机构名称已变更为蓝策亚洲（北京）资产评估有限公司。

2、由于签字注册资产评估师毕娇、李鹏举已从本公司离职，故北京海天瑞声科技股份有限公司本次发行申请文件中，资产评估机构声明无毕娇、李鹏举的签字。

特此说明。

资产评估机构负责人： 罗俊军

罗俊军

蓝策亚洲（北京）资产评估有限公司

2021 年 8 月 9 日



## 第十三节 附件

### 一、备查文件

- (一) 发行保荐书；
- (二) 上市保荐书；
- (三) 法律意见书；
- (四) 财务报告及审计报告；
- (五) 公司章程（草案）；
- (六) 发行人及其他责任主体作出的与发行人本次发行上市相关的承诺事项；
- (七) 内部控制审核报告；
- (八) 经注册会计师鉴证的非经常性损益明细表；
- (九) 中国证监会同意发行人本次公开发行注册的文件；
- (十) 其他与本次发行有关的重要文件。

### 二、文件查阅地址及时间

#### (一) 查阅地点及联系方式

1、发行人：北京海天瑞声科技股份有限公司

地址：北京市海淀区成府路 28 号 4-801

电话：010-62660772

传真：010-62660892

2、保荐机构（主承销商）：华泰联合证券有限责任公司

地址：北京市西城区丰盛胡同 22 号丰铭国际大厦 A 座 6 层

电话：010-56839300

传真：010-56839400

## （二）查阅时间

工作日上午 9:00~11:30；下午 13:30~17:00。