

海天瑞声

关于北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票 申请文件的审核问询函的回复

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）

二〇二三年九月

上海证券交易所：

贵所 2023 年 8 月 18 日下发的《关于北京海天瑞声科技股份有限公司申请向特定对象发行股票申请文件的审核问询函》（上证科审（再融资）〔2023〕206 号）（以下简称“问询函”）收悉。北京海天瑞声科技股份有限公司仔细阅读了问询函的全部内容，并根据问询函的要求，会同华泰联合证券有限责任公司（以下简称“保荐机构”、“保荐人”）、北京市天元律师事务所（以下简称“律师”、“发行人律师”）、信永中和会计师事务所（特殊普通合伙）（以下简称“会计师”）等中介机构对相关内容和问题进行了核查，对申请材料认真地进行了修改、补充和说明。现对问询函的落实和募集说明书的修改情况逐条书面回复，并提交贵所，请予审核。特别说明：

1、如无特别说明，本回复中使用的简称或名词释义与《北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票并在科创板上市募集说明书》（以下简称“募集说明书”）一致。涉及募集说明书补充披露或修改的内容已在募集说明书中以**楷体加粗**方式列示。

2、本回复部分表格中单项数据加总数与表格合计数可能存在微小差异，均因计算过程中的四舍五入所形成。

目 录

目 录.....	2
1. 关于本次募投项目	3
2. 关于前次募投项目	29
3. 关于融资规模与效益测算	36
4. 关于经营情况	58
5. 关于数据合规性	96
6. 关于境外销售	112
7. 关于财务性投资	125
8. 关于其他	132

1. 关于本次募投项目

根据申报材料，发行人本次发行拟募集资金 78,989.00 万元用于“AI 大模型训练数据集建设项目”和“数据生产垂直大模型研发项目”，场地购置费分别为 1.82 亿元、2,346.00 万元。

请发行人说明：（1）结合市场需求、应用场景及客户、数据采集来源及形成方式说明本次募投项目必要性；（2）结合本次募投项目与发行人现有业务的区别与联系，说明是否存在重复建设，是否涉及新产品、新技术；（3）结合本次募投项目主要客户及在手订单情况、行业发展趋势及同行业竞争情况等，说明本次募投项目的未来商业化前景及对发行人持续经营的作用；（4）本次募投项目场地购置进展及后续安排，是否符合政策及规定要求；结合数据收集所需空间面积的测算情况及已有场地的使用情况，说明本次募投项目中购置房产的合理性和必要性，是否存在变相用于房地产投资的情形，是否符合投向科技创新领域要求；（5）本次募投项目需履行的审批程序及办理进展；募投项目当前建设进展情况、资金预计使用进度安排，本次募集资金是否包含本次发行相关董事会决议日前已投入资金。

请保荐机构核查并发表明确意见。

回复：

一、结合市场需求、应用场景及客户、数据采集来源及形成方式说明本次募投项目必要性

（一）从市场需求来看，本次募投项目具有必要性

随着互联网、云计算、物联网以及穿戴设备的发展，数据要素价值不断释放，数字化转型成为大势所趋。人工智能作为数字经济发展的底层核心技术之一，成为数字经济发展的关键战略抓手。2022 年底 ChatGPT 的出现，掀起又一波人工智能发展热潮，以无监督学习模式为代表的预训练+人类反馈强化学习所构成的大模型技术路线的落地，意味着人工智能开启发展新范式，基础模型能力通过预训练及基于人类反馈的强化学习得到不断解锁，以解决海量开放式任务。

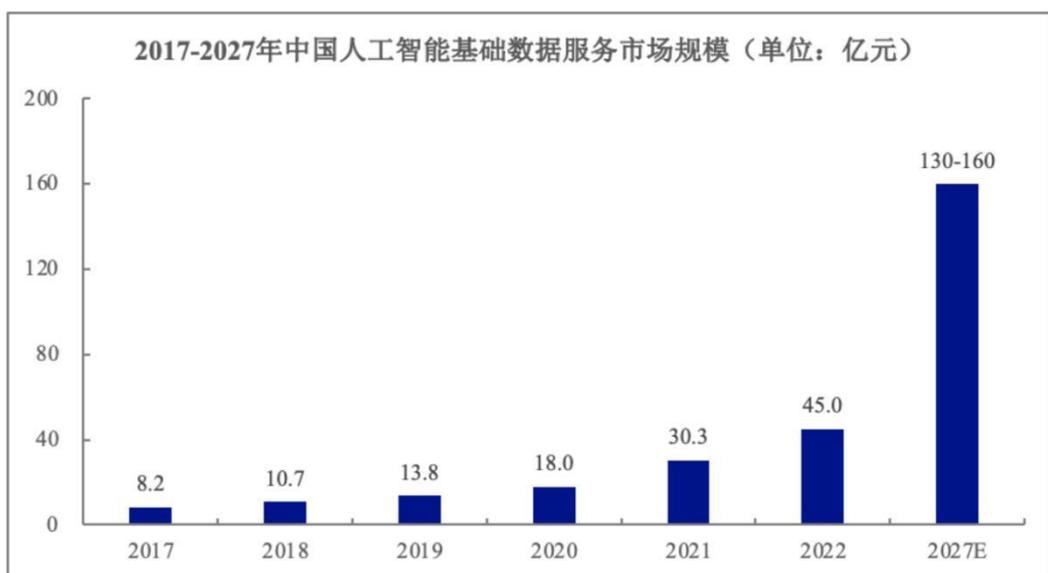
人工智能产业加速发展，基础数据服务业规模增长。根据艾瑞咨询的数据显

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



数据来源：艾瑞咨询

在人工智能产业链中，算法、算力和数据共同构成技术发展的三大核心要素。过去十年，人工智能产业以算法为中心，但随着算法趋于开源，数据的作用愈发重要。人工智能模型从技术理论到应用落地的过程需要依赖大量的训练数据，而且以大量数据为中心的训练方法相较于以模型为中心的训练方法，能够有效提升模型推断结论的可靠性。随着互联网、云计算、物联网、大数据等不断发展，万亿 GB 量级的数据源源不断地产生，但数据质量参差不齐，对海量的复杂数据进行深入挖掘、输出，进而激活和释放数据的深层价值也成为数据资源的发展重点。据德勤统计，2022 年中国人工智能基础数据服务市场规模为 45 亿元，2027 年规模可达到 130-160 亿元，预计 2022 年-2027 年保持较大幅度增长，年复合增长率为 23.6%-28.9%。



数据来源：德勤

综上所述，在人工智能产业尤其是大模型发展的拉动下，各类预训练和强化学习（包括但不限于调优、对齐等）数据集的整体需求攀升，人工智能基础数据服务市场未来需求量较大，本项目的建设具有必要性。

（二）从应用场景及客户来看，本次募投项目具有必要性

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

本次募投项目大模型数据集面向的主要应用场景和客户如下表所示：

数据集类型	数据集明细	应用场景	行业典型客户
通用及特定垂直领域的大语言模型训练数据集	中文高质量语料库、中文及多语种对话语料库、中文及多语种通用场景语料库、中文及多语种指令数据集等	人机交互、问答、信息及知识获取、推理等通用场景	通用场景： 百度、阿里巴巴、科大讯飞、腾讯、字节跳动、鹏城实验室、中国科学院、智谱 AI 等 垂直场景：
	中文特定场景语料库等	金融、教育、医疗等垂直场景	
多模态大模型训练数据集	多语种语音识别大模型数据集、多语言图文识别大模型数据集、多语种语音生成大模型数据	人机交互、问答、创作、信息及知识获取、推理等通用场景	- 金融：度小满、马上消费等； - 教育：网易有道、好未来等；

数据集类型	数据集明细	应用场景	行业典型客户
	集、多语言图文生成大模型数据集等		- 智能驾驶：比亚迪、毫末智行等； - 电商：美团、京东等
	全场景自动驾驶数据集、数字人多模态数据集等	智能驾驶、电商主播、旅游、社交、教育助手等垂直场景	
大模型评测数据集	大模型评测数据集等	大模型评测	百度、科大讯飞、中国信通院等

当前 AI 大模型发展方向主要分为通用大模型和垂直领域大模型两类。其中通用大模型强调模型的通识及泛化能力，要求模型训练的原始数据集要广泛且强化学习调优要好，因此需要超大规模的预训练数据集，以满足通用大模型人机交互、问答、创作、信息及知识获取、推理等基础能力训练，目前国内已有多个通用大模型处在首代产品发布初期/研发阶段，例如百度“文心”大模型、阿里巴巴“通义”大模型、科大讯飞“星火认知”大模型、腾讯“混元”大模型等。

而垂直领域大模型专注于某一特定行业、领域或场景，依托海量、高质量、多样性的行业数据，为大模型构建专业知识壁垒，为特定领域客户提供准确、专业的解决方案，满足用户在特定领域的需求。当前，金融、教育、智能驾驶等领域均有垂直领域大模型布局或落地，例如度小满“轩辕”金融大模型、网易有道“子曰”教育大模型、毫末智行“DriveGPT”自动驾驶大模型等。

综上，本次募投项目在通用和垂直领域具有丰富的应用场景和广泛的客户需求，项目实施具有必要性。

（三）从数据采集来源及形成方式来看，本次募投项目具有必要性

中国境内数据资源丰富，但由于数据挖掘不足，以及大量数据无法在市场上自由流通等原因，优质中文数据集仍然稀缺。以 ChatGPT 为例，其模型训练数据中，中文数据来源不足千分之一。目前国内头部科技企业的数据来源主要基于少量公开数据集以及自身特有的数据进行大模型训练，但由于中文优质数据质量以及数据资源的制约，国内大模型的能力与以 ChatGPT 为代表的国际大模型相比仍存在一定差距。国内缺乏高质量数据集的主要原因包括当前国内数据挖掘和数据治理的力度不足；数据流通与数据安全保障措施不够健全；国内市场缺乏开源意识，大量数据无法在市场上自由流通；国内相关公司成立较晚，数据积累较少；学术领域中文数据集受重视程度低以及国产数据集市场影响力及普及度较低

等。

为解决上述问题，公司作为业内领先的 AI 基础数据服务商，同时具备对人工智能核心算法的理解能力、前瞻性的专业数据集结构设计能力、丰富的语言覆盖能力及场景采集能力、算法辅助数据生产能力以及数据合规管理能力，拟通过本次募投项目的建设，面向社会各类群体（包括但不限于个人信息主体、企业主体、政府部门等）通过采购授权、合规下载或合作获取等方式，汇聚个人数据、企业数据及公共数据等大规模数据资源，经过专业化的数据分类设计、清洗、加工处理，形成覆盖预训练、强化学习及应用拓展阶段的海量、高质量专业数据集，更好的支撑大模型训练，提升客户侧各类大模型产品的输出/交互能力。

此外，为进一步推动国家数字经济发展，发挥数据要素在经济发展中的重要价值，我国推出《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》，从顶层设计角度，在数据产权、流通交易、收益分配、安全治理等方面构建了数据发展的基础制度和规划纲要，以促进数据合规高效流通使用，充分发挥中国海量数据规模和丰富应用场景优势，赋能实体经济，激活数据要素的潜能。北京市则开展数据基础制度“先行先试示范区”建设，提升大规模、高质量数据要素供给能力，归集高质量基础训练数据集。本次募投项目实施地点拟位于北京经济技术开发区“数据基础制度先行先试”区域内，在依托公司自身技术与资源获取能力的基础上，顺应该区域内先行先试的基础制度规划，有助于合规获取更加多元化的大模型数据集所需要的原始数据。

综上所述，本项目建设既是解决国内 AI 行业优质大模型训练数据不足的重要措施，同时也是响应国家和北京市在数据要素领域的政策号召，践行加大高质量数据集供给、为我国大模型领域快速发展提供有效数据支撑的必要举措，具有必要性。

二、结合本次募投项目与发行人现有业务的区别与联系，说明是否存在重复建设，是否涉及新产品、新技术

（一）本次募投项目与发行人现有业务的区别与联系

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

有产品矩阵的拓展和现有技术体系的迭代升级，不涉及新产品、新技术。

1、“AI 大模型训练数据集建设项目”

“AI 大模型训练数据集建设项目”将拓展现有训练数据产品体系，公司顺应人工智能大模型的发展趋势，聚焦人工智能大模型领域进行训练数据的研发、生产和服务，将大幅提升训练数据产品的数据规模，扩大产品覆盖的场景和领域，前瞻性布局大模型领域。“AI 大模型训练数据集建设项目”与公司现有业务的区别与联系具体如下：

(1) 数据集生产模式方面，公司通过设计训练数据集结构、组织原料数据采集、对取得的原料数据进行加工，最终形成可供算法模型训练使用的专业数据集。公司现有训练数据生产过程主要包括四个环节：设计（训练数据集结构设计）、采集（获取原料数据）、加工（数据清洗、标注等）及质检（各环节数据质量、加工质量检测）。



图：训练数据生产过程示意图

本次募投项目亦采用上述生产模式进行建设，但在具体环节上具有一定差异，以更加适配大模型数据集本身的特点和应用效果。例如设计环节需根据大模型数据集的特点进行细致、多维的规则制定（包括但不限于数据清洗规则、伦理/价值观对齐规则等），采集环节的数据类型及来源更为丰富、广泛，需要获取的数据规模更大，加工环节中数据清洗的工作量占比更高、操作规程更复杂，且数据标注的任务类型更多元，质检环节的难度进一步提升。

(2) 服务行业需求方面，公司现有训练数据集产品及服务应用于人工智能算法模型开发训练，即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

用发行人生产的训练数据，对算法模型进行训练、教导，使其表现更贴近人类。本次募投项目所产生的数据集产品，亦服务于上述方向，在此基础上，更专注于大模型“预训练+强化学习”技术路线下的数据需求，例如为在感知基础上增加对认知领域的的能力，需要加强文本类数据集的规模，并增加数据标注中的主观判断部分，如任务泛化、价值观对齐等。

(3)下游客户方面，公司现有业务领域的下游客户主要包括大型科技公司、传统人工智能企业及科研机构等，上述客群当中的相当一部分目前已纷纷开展大模型领域的研发/应用，且其中一些客户的大模型产品已经正式推出上线；同时，随着大模型发展热潮的不断演进，人工智能行业已经涌现出一批新兴大模型技术/应用人工智能科技公司/机构，该类公司/机构自创立之初即以人工智能大模型为核心发展方向，预期也将在大模型产业的发展中扮演愈加重要的角色。上述各类公司、机构构成了公司未来在大模型领域的主要客户群体。

2、“数据生产垂直大模型研发项目”

本次募投项目中的“数据生产垂直大模型”是对现有数据生产平台工具体系的补充和升级，并通过大模型技术能力，进一步研发和打造公司面向未来市场竞争的技术底座，为公司数据处理工具提供更加智能化的技术支撑。

具体而言，项目建成后将丰富公司基础研究层面的能力与种类，并形成“海天瑞声数据生产垂直大模型”技术底座，将其赋予在数据集生产过程中，结合平台工具层各类数据处理模块，包括新增各类大模型数据相关的处理模块，提高公司的数据处理能力及效率，扩展服务范围，提升服务水平。

例如，使用“数据生产垂直大模型”进行文本清洗规则的设计及清洗代码的辅助生成，部分替代由人工进行数据理解与分析、总结概况与提炼、代码编写等过程，预计能够节约研发人员约 30% - 60%的时间¹，提升研发效率。再如，使用“数据生产垂直大模型”对特定文本进行润色处理或改写，可在不改变原有口语化表达特点的基础上，简化对话数据，去除过多的口语词及无意义重复表达等，从而降低预计约 50%-80%的数据处理/校对成本²。

¹ 根据文本的复杂程度不同和清洗粒度要求不同，具体时间有所差异。

² 根据文本内容的不同，节省的成本亦有所差异。

本募投项目建成后，公司新一代一体化数据处理技术支撑平台示意图如下：



图：“数据生产垂直大模型研发项目”与公司现有数据生产平台工具体系（其中灰色部分为公司现有体系，蓝色部分为募投项目新增内容）

（1）基础研究层

在基础研究层，公司已通过在语音、视觉、智能驾驶、文本、数据集结构设计等方面算法与技术的不断探索、研究，赋能一体化数据处理技术支撑平台进行高效研发；本次募投项目中，公司通过开展大模型基础研究，为“数据生产垂直大模型”的研发提供理论基础，具体研究内容包括：预训练数据集设计与处理技术、指令数据集设计与处理技术、任务对齐与泛化技术、强化学习技术、Transformer 技术、大模型训练框架技术、大模型训练相关底层工程技术、大模型评测技术等。

（2）“海天瑞声数据生产垂直大模型”技术底座

立足于大模型基础研究理论，公司将自研“海天瑞声数据生产垂直大模型”基础技术底座。海天瑞声数据生产垂直大模型的研发分为两个阶段，分别为基础大模型研发以及模型调优，面向下游各领域生成不同领域的各类方案设计 & 数据处理结果，包括下载方案、清洗方案、指令方案、指令泛化与迁移、指令数据验证、多模态数据方案、标注成果生成和原始数据生成等。

（3）平台工具层

在平台工具层，公司已构建由语音数据生产系统（ADS, Audio Data System）、视觉数据生产系统（VDS, Visual Data System）及数据中台管理模块构成的平台

工具体系。本次募投项目中，基于数据生产垂直大模型的核心能力，公司将升级现有海天瑞声“一体化数据处理技术支撑平台”，使其具备自动化能力、版本控制能力、可插拔能力及管理监控能力，打造对外模型训练及评测能力，新增工具模块包括预训练数据下载及清洗工具、指令数据集筛选、生成与调优工具、大模型评测数据集评测及质检工具、多模态数据生产工具等。

（二）本次募投项目不存在重复建设，不涉及新产品、新技术

本次募投项目主要采用公司现有生产模式和技术框架进行建设，与公司现有业务在服务行业需求方面具有较强的一致性，公司现有业务领域的部分下游客户亦构成本次募投项目的主要客户主体。“AI 大模型训练数据集建设项目”在设计、采集、加工、质检等具体环节将更加适配大模型数据集本身的特点和应用效果，并更专注于大模型“预训练+强化学习”技术路线下的数据需求；“数据生产垂直大模型研发项目”在现有深度学习技术能力和现有数据处理平台工具的基础上，以研发大模型技术作为核心能力，对一体化数据处理技术支撑平台进行有效补充和升级，为公司数据处理工具提供更加智能化的技术支撑。

因此，本次募投项目内容为大模型领域数据集的建设及垂直领域大模型的研发，不存在重复建设情形；同时，本次募投项目为公司现有产品矩阵的拓展和现有技术体系的迭代升级，与公司现有业务领域在技术层面共通、产品层面互补、客户类型趋同，本项目与公司现有业务高度关联，是现有业务的扩大，均为投向公司目前主营业务领域，不涉及新产品、新技术。

三、结合本次募投项目主要客户及在手订单情况、行业发展趋势及同行业竞争情况等，说明本次募投项目的未来商业化前景及对发行人持续经营的作用

（一）客户需求及在手订单情况

目前，国际巨头纷纷布局以大模型为核心的通用人工智能产业，产业进入加速发展期，国内众多研究机构、企业也在积极研究生成式 AI 大模型技术的最优路径，并进行产品发布。截至本回复出具日，公司已与多家知名大模型厂商/机构签订了大模型训练数据集相关的业务合同或意向订单，合同及订单金额合计约 1,000 万元。其中主要客户情况具体如下：

客户名称	客户介绍	大模型领域布局	合作内容
------	------	---------	------

客户名称	客户介绍	大模型领域布局	合作内容
客户 1	该客户是一家中国领先的科技零售公司，以“零售+科技”为战略，以科技驱动创新发展，引入人工智能等前沿技术提高供应链效率、优化用户体验以及增强竞争力	该客户近期通过股权合作、组建团队等方式在认知智能通用模型研发方向进行布局	垂直领域的大语言模型训练数据集
客户 2	该客户是亚太地区知名的智能语音和人工智能企业，从事智能语音、自然语言理解、计算机视觉等核心技术研究并保持了国际前沿技术水平	该客户已向公众发布其认知大模型，具有多项核心能力	通用领域的大语言模型训练数据集
客户 3	该客户是中国最大的互联网公司之一，在人工智能领域有着积极的布局和投入	该客户已向公众发布其通用大语言模型，具备多项功能	通用领域的大语言模型训练数据集
客户 4	该客户是一家世界领先的互联网科技公司，提供优质数字内容、及云计算、广告、金融科技等一系列企业服务	该客户已发布其通用大语言模型，具备多项高效能力	通用领域的大语言模型训练数据集
客户 5	该客户是国家批准成立的网络通信领域新型科研机构	该客户已推出高级智能算力平台，已支撑起诸多国产 AI 大模型的训练	垂直领域的多模态大模型训练数据集
客户 6	该客户是国内知名互联网公司和技术驱动型公司	该客户已推出其通用大语言&多模态模型，覆盖机器学习、自然语言处理等多项研究领域	通用领域的大语言模型训练数据集、多模态大模型训练数据集
客户 7	该客户是全球领先的软件、服务、设备和解决方案供应商	该客户已在大模型领域进行广泛布局，例如开源大模型优势明显	大模型评测数据集
客户 8	该客户是中国最大的互联网公司之一，在人工智能领域已成为一个拥有强大互联网基础的领先 AI 公司	该客户已发布其通用大语言模型，打造自主创新 AI 底座	通用领域的大语言模型训练数据集

一方面，本次募投项目拟建设的数据集包括通用及特定垂直领域的大语言模型训练数据集、多模态大模型训练数据集及大模型评测数据集，目前已经面世的大模型产品以通用大语言大模型为主，而垂直领域及多模态领域的大模型数量仍相对较少，数据需求尚未充分释放；另一方面，鉴于公司下游大模型相关客户的产品尚处于首代产品发布初期/研发阶段，市场尚未大范围应用，相关数据需求，尤其是投放市场应用后的模型能力优化、能力迁移等需要的数据需求将进一步释放；此外，公司目前已与其他大模型领域的知名机构签署了合作协议，包括中国科学院自动化研究所、澜舟科技、智谱 AI、中国信通院等，未来拟围绕人工智

能大模型数据处理技术、大模型数据集供给和大模型评测等方面进行合作。综上所述，公司大模型业务相关收入未来有望进一步提升。

（二）行业发展趋势

首先，随着新一轮科技革命和产业变革深入发展，数字化转型成为大势所趋，世界主要国家均高度重视发展数字经济，纷纷出台战略规划，重塑数字时代的国际竞争新格局。人工智能作为数字经济发展的底层核心技术之一，正在发挥更加重要的作用。可以预见，人工智能行业将在以 ChatGPT 为代表的大模型技术的推动下进入新一轮产业高速发展期。在市场需求拉动和国家政策的支持引导下，当前我国人工智能产业加速发展，已形成基础底层设施、中层技术以及上层应用的完备的产业链生态，一批创新活跃、特色鲜明的创新企业不断涌现，并联合推动中国人工智能产业实现规模增长。根据艾瑞咨询的数据显示，2022 年中国人工智能产业规模达 1,948 亿元，预计 2027 年市场规模将达到 6,122 亿元，年复合增长率为 25.6%，主要与智算中心建设以及大模型训练等需求拉动的 AI 芯片市场、无接触服务需求拉动的智能机器人及对话式 AI 市场等快速增长相关，并有望在下游制造、交通、金融医疗等多领域不断渗透，实现大规模落地应用。

同时，党的十八大以来，习近平总书记屡次强调建设数字中国以及构建数据要素的重要性，并明确指出数据是新的生产要素，是基础性资源和战略性资源，也是重要生产力。尤其在《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》出台后，我国系统性布局了数据基础制度体系的“四梁八柱”，加速了数据流通交易和数据要素市场发展，进一步推动了个人数据、企业数据、公共数据合规高效流通使用。为更好响应中央号召，北京、上海、广州、深圳、杭州等地数据政策陆续出台，逐步构建了多层次、多元化数据要素市场生态体系。中共北京市委、北京市人民政府 2023 年 7 月印发的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》提出，推动建立供需高效匹配的多层次数据交易市场，充分挖掘数据资产价值，打造数据要素配置枢纽高地。《实施意见》还提出，推进国家数据知识产权试点，探索数据知识产权的制度构建、登记实践、权益保护和交易使用。完善人工智能数据标注库，探索打造数据训练基地，促进研发自然语言、多模态、认知等超大规模智能模型。

此外，国家网信办等七部门于 2023 年 7 月出台了《生成式人工智能服务管

理暂行办法》，该办法从内容合规、数据来源合法性、知识产权及商业秘密保护、虚假信息防范等方面，对生成式人工智能产品提出了全方位的合规要求。该办法明确提出，“提供者应当依法开展预训练、优化训练等训练数据处理活动”、“采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性”。国家政策对数据获取、数据处理提出较高标准，这意味着数据在合规获取和处理的成本将大幅增加，以预训练阶段为例，由于大量数据来源应合法合规，需投入大量成本完成数据获取。因此，出于成本与数据集质量的平衡性考量，在大模型训练中，大模型厂商通常会选择与专业的第三方数据集厂商合作，由专业第三方提供的合规、高质量数据集或相关解决方案将成为践行国家规范生成式人工智能产品要求的重要举措。

综上所述，人工智能产业即将进入新一轮加速发展期，我国对数据要素市场的政策推动已经启动，同时数据合规监管的要求日益提升预期将对优质数据从业企业带来增量市场空间，这些因素将综合作用，驱动行业发展趋势持续向好。

（三）同行业竞争情况

行业内其他主要企业包括 Appen、数据堂、标贝科技等，相关企业均在大模型领域积极布局，大模型领域具有良好的商业化前景。

1、Appen

Appen 公司 1996 年成立于澳大利亚，2015 年 1 月于澳大利亚证券交易所上市。是全球领先的图像、文本、语音、音频、视频等 AI 训练数据服务提供商，拥有业内先进的人工智能辅助数据标注平台和一体化 AI 数据及资源管理平台及全球 100 多万名技能娴熟的众包资源，支持 200 多种语言和方言，为全球科技、汽车、金融服务、零售、医疗和政府等行业赋能。

2、数据堂

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

3、标贝科技

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

发行人行业内主要企业包括 Appen、数据堂、标贝科技等，其基本经营情况、大模型领域投入等情况的对比如下：

基本经营情况				
公司名称	发行人	Appen	数据堂	标贝科技
成立年份	2005 年	1996 年	2010 年	2016 年
市场地位概述	是我国最早从事训练数据研发销售的企业之一；国内首家且是目前唯一一家 A 股上市的人工智能训练数据服务企业	较早从事数据资源开发的数据资源产品服务提供商，经营历史较长，规模、体量较大	新三板挂牌企业，是国内较早从事数据交易、数据采标的服务商之一	-
主要客户/合作伙伴情况	大型科技公司，如阿里巴巴、Meta、腾讯、百度、字节跳动、微软、三星等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学、中国科学技术大学等	微软、亚马逊、谷歌等大型科技公司、汽车厂商及政府	包括百度、腾讯、阿里巴巴、奇虎 360、联想、科大讯飞等国内互联网和高科技企业，微软、NEC、Canon、Intel、Samsung、Fujitsu 等企业及在华研发机构	微软、百度、阿里巴巴、腾讯、京东、滴滴、字节跳动、网易、360、三星、小鹏、美的、中科大、中电科、中国银行等
客户数量	近 900 家	未公开披露	未公开披露	100 余家
应用领域	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言
大模型领域投入情况				
大模型领域动态	公司已开始着手研发用于大模型训练的数据集，如中文千万轮次对话数据集、多语	智能 LLM 大模型开发平台，集大模型数据准备、模型训练、模型推理、模型部署应用于一体	多语言多模态视频文本对齐数据集、中文高质量大模型预训练文本数据集	构建大模型技术的非平衡专业语料，将陆续推出一系列高质量的数据集

	种语音大模型数据集等，目前已完成前期调研及可行性论证；公司已经启动数据生产垂直大模型的调研和早期基础研发工作			
--	--	--	--	--

注：同行业公司大模型领域投入情况来源于官网、新闻动态等公开信息查询。

（四）说明本次募投项目的未来商业化前景及对发行人持续经营的作用

1、本次募投项目具有良好的商业化应用前景

随着国内算法能力以及算力的普及与大规模建设，数据已逐渐成为驱动人工智能大模型发展以及大模型差异化竞争的关键要素。当前国内数据资源虽然规模较大，但优质的大模型训练数据仍然稀缺，在数据规模与数据质量上与全球领先国家仍存在明显差距。因此，公司拟通过本项目的实施，充分利用公司自身的技术与资源获取能力的积累和优势，顺应国家在数据要素基础制度领域的创新性政策，尤其是先行先试示范区在基础制度、数据供给等方面的先行先试政策，采用更加多元化的方式获取大规模原始数据，利用工程化的数据处理技术进行预训练阶段的数据清洗，采用人类反馈强化学习模式，基于微调和奖励模型训练的方法，以人类撰写少量的典型问题和标准答案与深度学习阶段基础性标注相结合的模式，生产用于通用型及各种垂直领域大模型训练的海量、高品质数据集，全力支持我国人工智能大模型领域的发展。

目前，国内大模型主要处于早期开发阶段，公司下游大模型相关客户产品尚处于首代产品发布初期/研发阶段，市场尚未大范围应用，相关数据需求，尤其是投放市场应用后的模型能力优化、能力迁移等需要的数据需求有待进一步释放。作为 AI 领域的前沿研究方向，大模型虽然在国内还未有规模化应用，但目前已有近百个国产大模型处在研发阶段或发布初期，例如百度“文心”大模型、阿里巴巴“通义”大模型、科大讯飞“星火认知”大模型、腾讯“混元”大模型等。未来，随着下游客户的研发进展，对于训练数据的需求有待进一步释放，公司大模型业务相关收入有望进一步提升。

此外，大模型训练数据通常具备数据规模大、数据质量高、数据类型丰富等特点，对数据服务商带来更大挑战，数据服务商往往会被要求在充分理解客户算法的基础上，在合法合规的前提下，具备完整、可持续迭代的大模型数据技术框

架和数据策略，能够对各细分类型数据（例如多模态、多垂向领域），进行包括设计、获取、清洗、标注、安全管理、质控评测等各环节的工程化加工处理。相比于现有数据集制作，大模型训练数据集对数据服务商的综合能力要求更高，更依赖数据服务商丰富的数据处理经验，而目前市场上同时具备多种模态及多垂向领域综合数据理解能力和经验的企业较为集中，属于市场稀缺资源。

综上所述，本次募投项目具有良好的商业化应用前景。

2、对发行人持续经营的作用

（1）AI 大模型训练数据集建设项目

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

（1）通用及特定垂直领域的大语言模型训练数据集，包括但不限于：①中文大模型预训练语料数据集（含通用场景、特定场景、对话场景、指令集等）；②多语言大模型预训练语料数据集（含通用场景、对话场景、指令集等）。

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

（3）大模型评测数据集：可应用于大模型的能力、任务、指标等方面的评测。

本项目实施后将形成新的业绩增长点，为公司持续创造经济价值。根据本次募投项目可行性分析报告，预计项目运营期（含建设期）内新增年均收入约 7,197 万元，年均利润约 3,038 万元。

（2）数据生产垂直大模型研发项目

为应对大模型时代下数据规模量极大、复杂性和多样性高，数据服务规则设计难度指数级提升等诸多问题，保证更高效高质量地完成数据的规模化生产，公司将采用全栈自研的数据生产垂直大模型技术，辅助完成面向多个下游任务的数

据规则设计与处理。同时，为更好实现数据生产垂直大模型的生成能力，公司将研发多项技术，夯实数据生产垂直大模型构建的基础。

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

本项目实施后将构建公司大模型数据处理技术通用化解决方案能力，实现完整、可持续迭代的大模型数据技术框架和数据策略，进一步提高公司在人工智能基础数据服务领域的智能化水平，有效扩展公司生产能力、提高公司生产效率，巩固公司的核心技术壁垒。例如，在使用“数据生产垂直大模型”进行清洗规则的设计及特定文本标注方面，均可依靠提升后的自动化能力，取得效率方面的显著改善。

四、本次募投项目场地购置进展及后续安排，是否符合政策及规定要求；结合数据收集所需空间面积的测算情况及已有场地的使用情况，说明本次募投项目中购置房产的合理性和必要性，是否存在变相用于房地产投资的情形，是否符合投向科技创新领域要求

（一）本次募投项目场地购置进展及后续安排

根据本次发行的方案，发行人各募投项目拟购置的场地的面积、主要内容、当前取得进展及后续安排如下表：

募投项目	场地购置面积	主要内容	当前取得进展及后续安排
AI 大模型训练数据集建设项目	3,450 平米	办公场地、展厅和多模态采集室	公司已对北京经济技术开发区信创园等园区开展了实地考察，并进行了初步询价。目前周边区域满足公司募投相关要求的房源较为充足，公司计划在募集资金到位后六个月内完成房产的选定、购置及装修
数据生产垂直大模型研发项目	460 平米	办公场地	

1、AI 大模型训练数据集建设项目

本项目拟建设办公场地、展厅和多模态采集室，总计 3,450 平米，参考周边

同类房产价格确定场地均价为 4.8 万元/平米,场地购置投资总额为 16,560 万元。
具体明细如下:

投资内容	场地用途	员工人数(人)	人均面积(平米)	总面积(平米)	场地均价(万元/平米)	金额(万元)
场地购置费	办公场地	245	10	2,450	4.8	11,760.00
	展厅	-	-	500	4.8	2,400.00
	多模态采集室	-	-	500	4.8	2,400.00
	合计	-	-	3,450		16,560.00

2、数据生产垂直大模型研发项目

本项目拟购入办公场地 460 平米,场地均价 4.8 万元/平米,总计金额为 2,208 万元。具体明细如下:

投资内容	场地用途	员工人数(人)	人均面积(平米)	总面积(平米)	场地均价(万元/平米)	金额(万元)
场地购置费	办公场地	46	10	460	4.8	2,208.00

(二) 是否符合政策及规定要求

本次募投项目不涉及购置土地,发行人拟在北京经济技术开发区购置场地作为实施地点。北京经济技术开发区内符合条件的房产较多,公司可以灵活选择房产资源。中共北京市委、北京市人民政府印发《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》,明确指出“支持北京经济技术开发区等开展数据基础制度先行先试,打造政策高地、可信空间和数据工场。支持基于信创技术建设数据可信流通体系和‘监管沙盒’,通过物理集中和逻辑汇通相结合的方式,导入工业、金融、能源、科研、商贸、电信、交通、医疗、教育等领域数据资源,促进数据跨行业融合应用,切实激活数据要素资源。”

发行人募投项目相关房产符合政策、城市规划及其他法律法规要求,场地取得不存在重大不确定性风险。

（三）结合数据收集所需空间面积的测算情况及已有场地的使用情况，说明本次募投项目中购置房产的合理性和必要性，是否存在变相用于房地产投资的情形，是否符合投向科技创新领域要求

1、数据收集所需空间面积的测算情况及已有场地的使用情况

公司现有已投入使用的办公场地合计面积共 3,547.71 平方米，以租赁房产为主。2023 年 1-6 月份，公司自有员工及参与数据集生产的需要长期占据办公面积的标注人员规模平均为 425 人，人均办公面积约 8.35 平米。

现有场地规模 (单位：平米)	2023 年 1-6 月平均人员数量 (单位：人)	人均面积 (单位：平米/人)
3,547.71	425	8.35

此外，公司位于北京市东城区的新购自有房产目前正处于装修及改造阶段，暂未达到预定可使用状态，未来该处房产将替代目前已投入使用的主要办公场地，用于满足公司现有业务办公场地需求的增长，包括现有业务相关的销售、项目与研发人员的增加，及现有业务数据采集（包括语音、视觉等）场所的需求扩张等，因此未纳入计算范围。假设上述新购置房产投入使用并替代目前已投入使用的主要办公场地，则公司人均办公面积约 14.57 平米，仍低于软件和信息技术服务业同行业公司募投项目平均人均办公面积 16.24 平米（具体情况请参见本题回复四之“（三）”之“2、发行人本次购置土地房产具有合理性和必要性”）。

2、发行人本次购置土地房产具有合理性和必要性

本次募投项目合计拟购置面积 3,910 平米的房产，其中，AI 大模型训练数据集建设项目拟购置 3,450 平米的房产，将用于建设办公场地、展厅和多模态采集室；数据生产垂直大模型研发项目拟购置 460 平米的房产，将用于建设办公场地。

发行人本次募投项目中，所规划的办公场地购置由发行人结合本次拟投入的人员情况等综合判定，购置房产相较于租赁房产更具有经济性；且根据中共北京市委、北京市人民政府《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》，该场地所处区域是发行人实施本次募投项目的合理选择，因此发行人购置相关房产存在合理性和必要性，具体如下：

(1) 购置办公场地是适应公司人员规模扩大的必要举措，所需办公面积确定依据合理

本次募投项目的实施涉及人员总计 291 人，人员扩充对办公面积提出新需求，假设人均办公面积为 10 平方米，拟需办公场地总面积为 2,910 平方米（除办公场地外，本次募投购置场地面积还包括少量展厅及多模态采集室面积）。

本次募投项目人均办公面积以人均 10 平米测算假设，略低于同行业公司，略高于公司目前人均办公面积，具有合理性。（公司 2023 年 1-6 月份人均办公面积约 8.35 平米，相关分析请参见本题回复四之“（三）”之“1、数据收集所需空间面积的测算情况及已有场地的使用情况”）本项目实施后，公司总体人均办公面积将得到一定提升，有利于提升员工的工作环境及公司研发和办公效率。

项目名称	岗位名称	总计
AI 大模型训练数据集建设项目	项目经理	20
	专家级数据工程师	50
	高级数据标注师	155
	数据清洗工程师	20
数据生产垂直大模型研发项目	大模型算法研发工程师	28
	大模型调优研发工程师	18
合计		291

经对比，软件和信息技术服务业公司募投项目的测算中，人均办公面积的假设情况如下：

公司名称	再融资预案时间	项目名称	购置地点	购置办公面积（平方米）	人员规模	人均办公面积（平方米）
深信服	2021.09	深信服长沙网络安全与云计算研发基地建设项目	长沙	48,600	3,000	16.20
	2020.04	网络信息安全服务与产品研发基地项目	深圳	47,100	3,000	15.70
安恒信息	2020.12	数据安全岛平台研发及产业化项目	上海	10,000	500	20.00
		信创产品研发及产业化项目	杭州	12,000	600	20.00
		网络安全云靶场及教育产业化项目	杭州	3,000	147	20.41
		新一代智能网关产品研发及产业化项目	成都	3,500	175	20.00
万兴科技	2020.08	数字创意资源商城建设项目	长沙	4,356	198	22.00

公司名称	再融资预案时间	项目名称	购置地点	购置办公面积 (平方米)	人员规模	人均办公面积 (平方米)
		AI 数字创意研发中心建设项目		1,628	74	22.00
彩讯股份	2020.07	运营中台建设项目	广州	2,256	250	9.02
		企业协同办公系统项目		2,115	235	9.00
		彩讯云业务产品线研发项目		2,142	238	9.00
软件和信息技术服务业同行业公司募投项目平均人均办公面积						16.24
海天瑞声募投项目人均办公面积						10.00
海天瑞声募投项目人均面积 (含展厅和多模态采集室)						13.44

经与同类型公司对比，人均办公面积在 9-22 平方米范围内，平均为 16.24 平方米。公司本次募投项目人均办公面积小于同行业公司募投项目平均人均办公面积，即使将展厅和多模态采集室的面积纳入测算范围后，公司人均面积仍在合理区间内。

(2) 购置房产相较于租赁房产更具有经济性

本次募投项目所购置的房产计划于募投项目开始实施当年装修完毕并投入使用，采用年限平均法，房屋建筑物可使用年限为 40 年，残值率为 5%，公司购买房产的年折旧费用为 424.51 万元/年。

经网络公开查询，本次拟购买房产标的周边其他同类型场地租金平均单价为 6.0 元/平方米/天，不考虑未来租金价格的上涨因素，则公司租赁等面积的房产每年租金达到 856.29 万元。具体情况如下：

序号	标的位置	单价 (元/平方米/天)
1	荣京东街-朝林广场	7.1
2	荣华南路-国锐广场	5.8
3	荣京东街-大族广场	5.0
平均单价		6.0

经测算，公司购买房产每年新增的折旧费用低于该地区租赁同等面积房产每年所需支付的租金。从中长期看，公司购置房产更具有经济性。

项目	每年新增费用 (万元)
购置房产	424.51
租赁房产	856.29

(3) 顺应“数据基础制度先行先试”区域政策规划，有助于公司获取更多元化的数据资源

根据 2023 年 5 月北京市人民政府办公厅印发的《北京市促进通用人工智能创新发展的若干措施》，北京市开展数据基础制度“先行先试示范区”建设。2023 年 7 月中共北京市委、北京市人民政府发布的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》也指出：

“打造数据基础制度综合改革试验田，支持北京经济技术开发区等开展数据基础制度先行先试，打造政策高地、可信空间和数据工场。支持基于信创技术建设数据可信流通体系和“监管沙盒”，通过物理集中和逻辑汇通相结合的方式，导入工业、金融、能源、科研、商贸、电信、交通、医疗、教育等领域数据资源，促进数据跨行业融合应用，切实激活数据要素资源……示范建设数据服务产业基地，通过开放数据、开放场景和提供算力等，推进各类数据要素型企业入驻数据服务产业基地。”

本次募投项目拟购置房产地点拟位于北京经济技术开发区“数据基础制度先行先试”区域内，顺应该区域内先行先试的基础制度规划，有助于公司合规获取更多元化的大模型数据集所需的原始数据。公司将围绕数据安全和隐私保护进行区域建设，构建数据安全管控体系，对数据存储机房进行物理隔离；探索个人信息保护与分享利用机制，采用安全多方计算、同态加密、数据脱敏等策略，打造安全可控的数据资源获取、处理和流通环境。

因此，本次募投项目建设对办公场地物理空间存在一定的特殊性要求，采用购置自有房产的方式有助于确保公司对于物理空间的改造、建设达到上述安全可控要求，以保证募投项目实施的稳定性。

3、发行人本次购置土地房产不存在变相用于房地产投资的情形

发行人本次募投项目不涉及购置土地，拟购置房产建设办公场地、展厅和多模态采集室，全部用于数据集建设、项目研发等日常经营活动，公司根据项目实施实际需求，制定了具体使用规划，不涉及对外出租或出售的情况。本次使用募集资金购置房产的具体情况如下：

序号	类型	项目	总面积（平米）
----	----	----	---------

1	办公场地	AI 大模型训练数据集建设项目	2,450
2	展厅	AI 大模型训练数据集建设项目	500
3	多模态采集室	AI 大模型训练数据集建设项目	500
4	办公场地	数据生产垂直大模型研发项目	460

本次募投项目所设计的展厅面积共有 500 平米，主要将用于向客户及潜在客户展示公司自身品牌形象，通过视频演绎、海报宣传、展示板展览等形式，集中讲解、宣传、推广公司产品，使得客户了解企业文化、产品类别和服务等各个方面，全方位建设公司品牌形象；多模态采集室面积共有 500 平米，主要将用于满足多模态数据采集需求，为多模态大模型训练所需数据集创造采集环境。

基于上述，结合发行人本次购置房产的相关规划，发行人购置的相关房产中不涉及对外出租或出售的情况，相关面积规划合理，不存在变相投资房地产业务的情形。

4、发行人本次购置土地房产符合募集资金投向科技创新领域的要求

公司购置房产主要用于本次募投项目“AI 大模型训练数据集建设项目”及“数据生产垂直大模型研发项目”建设，系围绕公司主营业务展开。根据国家统计局《战略性新兴产业分类（2018）》，公司所从事的训练数据生产业务属于“新一代信息技术产业—新兴软件和新型信息技术服务—新型信息技术服务—信息处理和存储支持服务—数据加工处理服务”行业，是国家重点支持的“新一代信息技术领域”的战略性新兴产业。因此，本次购置房产符合募集资金投向科技创新领域的要求。

五、本次募投项目需履行的审批程序及办理进展；募投项目当前建设进展情况、资金预计使用进度安排，本次募集资金是否包含本次发行相关董事会决议日前已投入资金。

（一）本次募投项目需履行的审批程序及办理进展

AI 大模型训练数据集建设项目已取得北京市投资项目备案证（备案证号：京技审项（备）[2023]150 号），数据生产垂直大模型研发项目已取得北京市投资项目备案证（备案证号：京技审项（备）[2023]149 号）。

根据对北京经济技术开发区行政审批局环境影响报告书审批窗口的电话咨

询，并查阅《建设项目环境影响评价分类管理名录》北京市实施细化规定（2022年本），本次募集资金投资项目不同于常规生产性项目，不存在废气、废水、废渣等工业污染物，不属于根据《中华人民共和国环境影响评价法》和《建设项目环境影响评价分类管理名录》等相关法律法规需要进行环境影响评价的建设项目。因此本项目无需进行项目环境影响评价，亦不需要取得环保主管部门对项目的审批文件。

（二）募投项目当前建设进展情况、资金预计使用进度安排，本次募集资金是否包含本次发行相关董事会决议日前已投入资金

1、募投项目当前建设进展情况

（1）AI大模型训练数据集建设项目

截至本回复出具日，公司已开始着手研发用于大模型训练的数据集，如中文千万轮次对话数据集、多语种语音大模型数据集等，目前已完成前期调研及可行性论证，且公司已收到 10 余家企业的大模型数据集询盘需求。

（2）数据生产垂直大模型研发项目

发行人已在大模型相关领域开展了相关研究工作，为本次募投项目“数据生产垂直大模型研发项目”的前期论证及实验工作，具体情况如下：

在研项目	进展或阶段性成果	拟达到目标
面向大模型训练、调优及评测服务等的技术研究	针对大模型预训练数据集设计与处理技术进行初步研究和规划，并基于研究成果开展了CommonCrawl、中文书籍等适用于预训练阶段的数据的获取与清洗工作，形成了各项技术的框架方向；开展大模型评测技术调研，完成基础框架设计，形成可行性结论。	完成面向大模型预训练阶段的数据清洗技术、任务对齐阶段的调优、强化学习技术的框架规划和可行性分析。论证面向大模型技术的评测服务能力构建可行性，包括评测方案设计、评测数据获取、自动评测等能力。

截至本回复出具日，公司在预训练数据集设计与处理、指令数据集设计与处理、任务对齐与泛化技术等大模型基础研究方面持续投入，提升基础算法能力。研发团队使用专门针对大模型业务场景建设的数据，成功完成了多次模型微调和优化升级，提升了模型在对话自然度方面的表现，并使大模型初步获得指令方案设计、指令数据验证以及标注成果生成等初阶能力。同时，一体化数据处理技术

支撑平台中与指令数据集、大模型评测数据集相关的工具也已处于调研和启动开发阶段。

2、资金预计使用进度安排，本次募集资金是否包含本次发行相关董事会决议日前已投入资金

本次募集资金不包含本次发行相关董事会决议日前已投入资金，预计使用进度安排具体情况如下：

(1) AI 大模型训练数据集建设项目

本项目总投资 38,337.36 万元，分 3 年投入，具体使用进度安排如下：

单位：万元

序号	项目	金额	占比	第一年	第二年	第三年
1	场地购置及装修	18,195.00	47.46%	18,195.00	-	-
2	设备购置费用	2,563.50	6.69%	955.35	883.30	724.85
3	软件购置费用	2,048.20	5.34%	643.72	777.48	627.00
4	技术人员工资	1,584.00	4.13%	368.00	576.00	640.00
5	数据资源采购	6,690.00	17.45%	1,132.79	2,268.96	3,288.25
6	铺底流动资金	7,256.66	18.93%	1,466.47	2,440.89	3,349.30
合计		38,337.36	100.00%	22,761.33	6,946.63	8,629.40

此外，按照数据来源划分，本次募投项目拟建设的数据集中原始数据资源采购来源主要包括个人数据、企业数据、公共数据等，其中公共数据的获取主要与北京市“数据基础制度先行先试”区域的建设相关。根据公司本次募投项目建设规划，本次募投项目数据来源中公共数据获取成本金额占数据资源采购总金额的比例约为 27.50%；此外，包含公共数据的数据集拟于募投项目建设期第二年开始投入资金建设，按照资金投入计划，第二年公共数据投入金额约为整体公共数据拟投入金额的 27.17%，并于第三年完成建设。

而根据 2023 年 5 月北京市人民政府办公厅印发的《北京市促进通用人工智能创新发展的若干措施》、2023 年 7 月中共北京市委、北京市人民政府印发的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》，北京市开展数据基础制度先行先试，提升大规模、高质量数据要素供给能力，归集高质量基础训练数据集。目前，公司募投项目数据集的建设进度规划与相关区域的建设

进度基本匹配。

(2) 数据生产垂直大模型研发项目

本项目总投资 40,651.64 万元，分 3 年投入，具体使用进度安排如下：

单位：万元

序号	项目	金额	占比	第一年	第二年	第三年
1	场地购置及装修	2,346.00	5.77%	2,346.00	-	-
2	设备购置费用	29,895.00	73.54%	12,145.20	9,497.60	8,252.20
3	软件购置费用	451.89	1.11%	235.77	117.89	98.24
4	研发人员费用	4,902.50	12.06%	1,050.00	1,667.50	2,185.00
5	设备托管费用	3,056.25	7.52%	588.75	1,042.50	1,425.00
合计		40,651.64	100.00%	16,365.72	12,325.49	11,960.44

六、核查程序及核查意见

(一) 核查程序

保荐机构履行了以下核查程序：

1、查阅了发行人本次募投项目的可行性研究报告、相关行业分析报告、相关法律法规文件，了解本次募投项目的市场需求、应用场景及客户、数据采集来源及形成方式等；

2、查阅了发行人本次募投项目的项目备案文件，结合发行人现有主营业务情况，了解本次募投项目与发行人现有业务的区别与联系，是否涉及新产品、新技术；

3、获取发行人大模型领域的业务合同及意向订单，了解双方合作内容，通过公开信息检索查询相关客户的基本信息及大模型领域布局，查阅了本次募投项目的可行性研究报告、近期行业相关法律法规，了解行业发展趋势，通过公开信息检索查询同行业可比公司的基本情况及大模型领域投入情况，了解本次募投项目的未来商业化前景及对发行人持续经营的作用；

4、查阅了本次募投项目的可行性研究报告及测算底稿，向发行人了解本次募投项目场地购置进展及后续安排并获取相关说明文件，查阅相关法律法规文件，获取发行人现有办公场地及人员明细，通过网络公开查询拟购买房产标的周边其

他同类型场地租金情况，查询软件和信息技术服务业公司募投项目人均办公面积的测算情况，论证本次募投项目中购置房产的合理性及符合募集资金投向科技创新领域的要求；

5、查阅发行人本次募投项目的可行性研究报告、项目备案文件等，向发行人了解募投项目当前建设进展情况，并对公司主管人员进行访谈并获取相关文件，了解募投项目数据集建设进度与“数据基础制度先行先试”区域的建设进度是否基本匹配。

（二）核查意见

经核查，保荐机构认为：

1、在人工智能产业尤其是大模型发展的拉动下，人工智能基础数据服务市场未来需求量较大，本次募投项目在通用和垂直领域具有丰富的应用场景和广泛的客户需求，本次募投项目建设是响应国家和北京市在数据要素领域的政策号召，践行数据采集来源及形成过程的合规、合法性的必要举措，本次募投项目建设具有必要性；

2、本次募投项目不涉及新产品、新技术，不存在重复投入情形；

3、本次募投项目具有良好的商业化应用前景并将对发行人的持续经营发挥积极作用；

4、本次募投项目场地购置符合政策及规定要求，具有合理性和必要性，不存在变相用于房地产投资的情形，符合投向科技创新领域要求；

5、本次募投项目已履行相关审批程序，募投项目当前建设进展情况正常，本次募集资金不包含本次发行相关董事会决议日前已投入资金。

2. 关于前次募投项目

根据申报材料，1) 发行人 2021 年首发募集资金净额为 33,635.69 万元，使用募集资金进行“自主研发数据产品扩建项目”的投资建设并补充流动资金，使用自有资金对前次原募投项目“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”继续进行投资建设；2) “自主研发数据产品扩建项目”预计达到可使用状态日期由 2022 年 5 月 31 日延长至 2023 年 5 月 31 日；3) 前募结余资金 3,515.73 万元变更用于永久补充流动资金。

请发行人说明：（1）前次募投项目调整投入金额的具体原因、调整后的项目资金投入情况、建设进度及效益情况，与前次预计效益情况是否相符；（2）“自主研发数据产品扩建项目”延期的原因，相关因素是否已消除或改善，结合此次募投项目的实施情况说明是否存在延期的风险；（3）结合“自主研发数据产品扩建项目”各项目投资构成中的非资本性支出及补充流动资金金额，说明前次募投项目变更后实际用于补充流动资金的金额占募集资金的比例是否超过 30%，是否需要扣减。

请保荐机构及申报会计师核查并发表明确意见。

回复：

一、前次募投项目调整投入金额的具体原因、调整后的项目资金投入情况、建设进度及效益情况，与前次预计效益情况是否相符

（一）前次募投项目调整投入金额的具体原因、调整后的项目资金投入情况

1、调整募集资金投资项目投入募集资金金额的情况

根据公司 2021 年 9 月披露的《北京海天瑞声科技股份有限公司关于调整部分募集资金投资项目资金来源的公告》，由于公司首次公开发行实际募集资金净额为人民币 33,635.69 万元，低于《北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市招股说明书》中拟投入的募集资金金额人民币 76,593.88 万元，为进一步提高募集资金使用效率、保障募集资金投资项目的顺利实施，结合各募集资金投资项目的实际情况，公司对募集资金投资项目投入金额进行相应调整，具体调整如下：

单位：万元

序号	项目名称	总投资额	调整前拟投入募集资金金额	调整后募集资金投入情况
1	自主研发数据产品扩建项目	24,921.18	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	8,714.51
合计		76,593.88	76,593.88	33,635.69

注：上述第 2、3 项募集资金投资项目拟调整资金来源为自有资金，已经公司第一届董事会第三十四次会议及第一届监事会第二十八次会议审议通过，独立董事发表同意意见，已经公司股东大会审议通过。

为进一步提升公司核心竞争力，增厚自有知识产权训练数据产品储备，满足市场对高质量、结构化、大规模训练数据的需求，公司需要持续加强对训练数据产品的投入，同时，公司需要充足的流动资金，支撑训练数据定制服务的提供，经过权衡，决定使用募集资金进行自主研发数据产品扩建项目，并补充流动资金，使用自有资金对一体化平台项目、研发中心升级建设项目继续进行投资建设，以实现募集资金的最优使用。

2、调整募集资金投资项目“自主研发数据产品扩建项目”内部投资结构的情况

根据公司 2021 年 9 月披露的《北京海天瑞声科技股份有限公司关于调整募集资金投资项目拟投入金额及内部投资结构的公告》，结合“自主研发数据产品扩建项目”的实际情况，公司对其内部投资结构进行相应调整，具体调整如下：

单位：万元

募投项目名称	序号	项目	原计划投资金额	增减情况	调整后募集资金投入情况
自主研发数据产品扩建项目	1	场地费用	8,242.90	3,000.00	11,242.90
	2	设备购置费用	2,092.25	-600	1,492.25
	3	软件购置费用	100	-	100
	4	技术人员工资	3,399.50	-	3,399.50
	5	原料数据采集及加工费用	7,946.40	-	7,946.40
	6	基本预备费	435.62	-400	35.62
	7	铺底流动资金	2,704.51	-2,000.00	704.51
合计		24,921.18			24,921.18

为进一步提升募集资金使用效率，根据公司募集资金投资项目实施规划和实际运营需要，应对业务量增长的需求，公司增加“自主研发数据产品扩建项目”所需的人均场地面积，加大场地购置支出的投入，提供更为充足的产品建设场地、更加便利的办公环境，同时减少公司租赁支出。针对设备购置支出，根据采购询价情况，单位设备的价格较预计降低，节省了设备购置费用，实际购买时单位设备的价格出现上涨的情况，公司根据实际情况利用自有资金解决。基本预备费和铺底流动资金是项目规划时预留应对项目在启动和执行阶段的各种变化因素，根据公司产品建设体系和各项实施规划，项目可控性较强，无需前期额外投入，也无场地、设备、人员、原料数据采集及加工费用等以外的其他实施项目，因此，公司使用该等项目补充场地类实施项目不足的部分。

（二）前次募投项目建设进度及效益情况，与前次预计效益情况是否相符

公司前次募投项目中实际投入募集资金的项目为“自主研发数据产品扩建项目”和“补充流动资金项目”。其中“自主研发数据产品扩建项目”计划开发自有知识产权的训练数据产品，分别覆盖智能语音领域、计算机视觉领域和发音词典领域，以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，加快公司对客户需求的响应速度。

目前公司前次募集资金投资项目已经结项，上述训练数据产品已经按计划完成建设，提高了公司训练数据产品的规模和种类，丰富完善了公司的产品体系，满足了下游市场对于训练数据产品延伸开发的更多需求，增强了公司的竞争力，达到了上述预期效果，促进了公司整体产品收入稳步增长。

公司未在招股说明书中对前次募集资金投资项目的使用效益作出任何承诺。截至 2023 年 6 月，前次募集资金投资项目实际产生的收入为 10,784.83 万元，利润总额为 99.60 万元。与公司规划的前次募投项目可行性研究报告预计项目建设期效益相比，收入实现情况基本相符，略有增长；利润总额实现情况高于预计情况，除受收入较预计情况有所增长的影响外，公司还在建设期过程中根据下游行业和客户的技术进展与需求情况，优化数据采集方案、提升自动预标注、实时交互式辅助标注与自动质检，提升了数据集产品生产开发的效率，通过持续降低开发成本，进而使得数据集研发费用的投入低于预计等原因所致。

二、“自主研发数据产品扩建项目”延期的原因，相关因素是否已消除或改善，结合此次募投项目的实施情况说明是否存在延期的风险

根据《北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市招股说明书》（2020年）（以下简称“招股说明书”），“自主研发数据产品扩建项目”预计建设周期为2年，项目建设计划支出中包含的场地购置、原料数据采集加工、设备购置等，由于受到募集资金到账时间晚于初始预期、全球突发性公共卫生事件带来的隔离、交通及人员聚集管制等管控措施对公司原料数据采集等环节的影响，导致项目建设进度较原计划有所滞后，故项目达到预定可使用状态的时间由2022年5月31日延至2023年5月31日。上述导致前次募投项目延期的因素均为偶发性、阶段性因素。本次向特定对象发行股票事项与此前公司首次公开发行相比，进程相对确定，且彼时全球突发性公共卫生事件的影响目前已经消除，在全球数字经济发展以及以ChatGPT为代表的大模型的驱动下，全球人工智能产业即将进入新一轮加速发展期，训练数据作为AI算法发展和演进“燃料”的作用更加凸显。因而此次募投项目的实施不会受前次募投项目延期因素的影响而存在延期的风险。

三、结合“自主研发数据产品扩建项目”各项目投资构成中的非资本性支出及补充流动资金金额，说明前次募投项目变更后实际用于补充流动资金的金额占募集资金的比例是否超过30%，是否需要扣减。

（一）前次IPO募集资金到账后使用计划

公司首次公开发行实际募集资金净额为人民币33,635.69万元，IPO募集资金到账后，募集资金净额33,635.69万元对应的项目投资构成如下，其中资本性支出金额10,435.15万元，占比31.02%，非资本性支出金额23,200.54万元，占比68.98%。

单位：万元

募投项目名称	序号	项目	金额	资本性支出	非资本性支出
自主研发数据产品扩建项目	1	场地费用	8,242.90	8,242.90	-
	2	硬件	2,092.25	2,092.25	-
	3	软件	100	100.00	-
	4	技术人员工资	3,399.50	-	3,399.50

募投项目名称	序号	项目	金额	资本性支出	非资本性支出
	5	数据费	7,946.40	-	7,946.40
	6	基本预备费	435.62	-	435.62
	7	铺底流动资金	2,704.51	-	2,704.51
小计			24,921.18	10,435.15	14,486.03
补充流动资金			8,714.51		8,714.51
合计			33,635.69	10,435.15	23,200.54
占比			100.00%	31.02%	68.98%

(二) 前次 IPO 募集资金实际使用情况

单位：万元

募投项目名称	序号	项目	调整后募集资金投入情况	实际投入募集资金情况	节余情况	资本性支出	非资本性支出
自主研发数据产品扩建项目	1	场地费用	11,242.90	11,242.90	-	11,242.90	-
	2	设备购置费用	1,492.25	664.02	828.23	664.02	828.23
	3	软件购置费用	100	20.80	79.20	20.80	79.20
	4	技术人员工资	3,399.50	2,910.04	489.46	-	3,399.50
	5	原料数据采集及加工费用	7,946.40	6,375.90	1,570.50	-	7,946.40
	6	基本预备费	35.62	-	35.62	-	35.62
	7	铺底流动资金	704.51	689.03	15.48	-	704.51
小计			24,921.18	21,902.69	3,018.49	11,927.72	12,993.46
补充流动资金项目			8,714.51	8,714.51	-	-	8,714.51
合计			33,635.69	30,617.20	3,018.49	11,927.72	21,707.97
占比			100.00%	-	-	35.46%	64.54%

注：上表中节余情况所列金额 3,018.49 万元与募集资金账户节余金额的差额为扣除手续费后的利息收入及理财收益。

前次募投项目结项并永久补流后，自主研发数据产品扩建项目中除场地购置费用、设备购置费用、软件购置费用为资本性支出外，其他用于技术人员工资、原料数据采集及加工费用、基本预备费、铺底流动资金、节余补流等为非资本性支出项目。补充流动资金项目为非资本性支出。综上因素统计，前次募集资金项目实际使用非资本性支出 21,707.97 万元，占前次募集资金净额 33,635.69 万元的 64.54%。

综上所述，公司首次公开发行实际募集资金净额为人民币 33,635.69 万元，

募集资金到账后使用计划中非资本性支出占比 68.98%，前次募集资金项目实际使用非资本性支出占比为 64.54%，超过 30%，未超出募集资金到账后计划使用非资本性支出占比 68.98%。因此，发行人本次融资的募集资金规模不需要调整。

四、核查程序及核查意见

（一）核查程序

保荐机构及申报会计师履行了以下核查程序：

1、获取并查阅发行人前次募投项目调整投入金额的公告及前次募集资金投资项目的可研报告和募集资金专户银行对账单及使用台账等资料，了解前次募集资金使用情况及效益情况；

2、查阅发行人 IPO 招股说明书、募集资金项目延期公告、前次募集资金使用情况报告等公开披露信息文件，分析前次募集资金投入项目延期情况，前次募集资金投入项目相关影响因素对本次募投项目的影响；

3、查阅了发行人前次募投项目调整投入金额的公告、募集资金专户银行对账单及使用台账等资料，分析前次募集资金到账后使用计划中非资本性支出情况以及募投项目结项补流后实际用于非资本性支出的情况。

（二）核查意见

经核查，保荐机构及申报会计师认为：

1、发行人前次募投项目调整投入金额主要系首次公开发行实际募集资金金额低于拟投入募集资金金额所致，前次募投项目目前已经结项，上述训练数据产品已经按计划完成建设，提高了公司训练数据产品的规模、数量和种类，丰富完善了公司的产品体系，前次募投项目实际收入实现情况与规划项目时可研报告预计情况基本相符，略有增长，利润总额情况高于规划项目时可研报告预计，主要与公司生产研发效率提升和收入实现相比预计增长等因素相关；

2、发行人前次募投项目延期由于受到募集资金到账时间晚于初始预期及全球突发性公共卫生事件等因素的影响，彼时影响前次募投项目的因素目前已经消除，此次募投项目的实施不会受前次募投项目延期因素的影响而存在延期的风险；

3、公司首次公开发行募集资金到账后使用计划中非资本性支出占比 68.98%，

前次募集资金项目实际使用非资本性支出占比为 64.54%，超过 30%，未超出募集资金到账后计划使用非资本性支出占比 68.98%。因此，发行人本次融资的募集资金规模不需要调整。

3. 关于融资规模与效益测算

根据申报材料，发行人本次发行拟募集资金 78,989.00 万元，用于“AI 大模型训练数据集建设项目”和“数据生产垂直大模型研发项目”。

请发行人说明：（1）募投项目投资金额的具体内容、测算依据及测算过程，结合本次募投项目中非资本性支出金额情况，测算本次募投项目中实际补充流动资金的具体数额及其占本次拟募集资金总额的比例，是否超过 30%；（2）结合日常营运需要、公司货币资金余额及使用安排、日常经营积累、资金缺口、公司资产负债率与同行业可比公司的对比等情况，分析本次募集资金规模的合理性；（3）项目效益测算的具体情况、测算过程及测算依据，说明增长率、毛利率、预测净利率等收益指标的合理性，结合毛利率、内部收益率、投资回收期等关键指标与同行业同类项目的对比情况说明项目效益测算的谨慎性和合理性；（4）公司针对上述事项履行的决策程序和信息披露情况。

请保荐机构核查并发表明确意见。

回复：

一、募投项目投资金额的具体内容、测算依据及测算过程，结合本次募投项目中非资本性支出金额情况，测算本次募投项目中实际补充流动资金的具体数额及其占本次拟募集资金总额的比例，是否超过 30%

（一）募投项目投资金额的具体内容、测算依据及测算过程

1、AI 大模型训练数据集建设项目

（1）投资数额安排明细及测算依据

本项目投资总额为 38,337.36 万元，分 3 年投入，具体投资数额安排如下：

单位：万元

序号	项目	金额	第一年	第二年	第三年	是否属于资本性支出
1	场地费用	18,195.00	18,195.00	-	-	是
1.1	场地购置费用	16,560.00	16,560.00	-	-	是
1.2	场地装修费用	1,635.00	1,635.00	-	-	是
2	设备购置费用	2,563.50	955.35	883.30	724.85	是

序号	项目	金额	第一年	第二年	第三年	是否属于资本性支出
3	软件购置费用	2,048.20	643.72	777.48	627.00	是
4	技术人员工资	1,584.00	368.00	576.00	640.00	否
5	数据资源采购	6,690.00	1,132.79	2,268.96	3,288.25	否
6	铺底流动资金	7,256.66	1,466.47	2,440.89	3,349.30	否
合计		38,337.36	22,761.33	6,946.63	8,629.40	

(2) 场地费用

1) 场地购置

本项目拟建设办公场地、展厅和多模态采集室，总计 3,450 平方米，场地均价为 4.8 万元/平方米，场地购置投资总额为 16,560 万元。具体明细如下，相关测算依据参见本回复之“1.关于本次募投项目”之“四”之“（三）”：

投资内容	场地用途	项目人员人数 (人)	人均面积 (平方米)	总面积 (平方米)	场地均价 (万元/平方米)	金额 (万元)
场地购置费	办公场地	245	10	2,450	4.8	11,760.00
	展厅	-	-	500	4.8	2,400.00
	多模态采集室	-	-	500	4.8	2,400.00
	合计	-	-	3,450		16,560.00

2) 场地装修

本项目中，场地装修标准参考了公司近年购置场地的装修造价，办公场地装修费用为 0.3 万元/平方米，装修面积为 2,450 平方米；展厅装修费为 0.8 万元/平方米，装修面积为 500 平方米；多模态采集室装修费为 1 万元/平方米，装修面积为 500 平方米，装修费总计 1,635 万元。具体明细如下：

投资内容	场地用途	装修面积 (平方米)	装修单价 (万元/平方米)	金额 (万元)
场地装修费	办公场地	2,450	0.3	735.00
	展厅	500	0.8	400.00
	多模态采集室	500	1.0	500.00
	合计	3,450		1,635.00

(3) 设备购置

“AI 大模型训练数据集建设项目”的设备主要包括数据集产品的存储设备、

采集设备和办公设备，购置数量依据公司现有业务历史经验及新增人员数量等因素预估，具体明细如下：

序号	设备名称	单位	数量	单价 (万元)	金额 (万元)
1	台式机	台	245	1.50	367.50
2	笔记本	台	245	0.80	196.00
3	采集车及相关设备（自动驾驶业务）	台	8	125.00	1,000.00
4	采集及存储设备				
4-1	数据存储服务器	台	8	50.00	400.00
4-2	动作捕捉设备及手套	套	2	106.00	212.00
4-3	手机	个	200	0.60	120.00
4-4	其他：平板电脑、声卡、麦克风、话放、耳机、监听控制器、单反相机、镜头、摄像机、3D 扫描仪、运动相机、3D 扫描相机、高清摄像机、电视、提词器等				268.00
合计					2,563.50

(4) 软件购置

本项目拟投入软件共计 2,048.20 万元，包括办公软件及设计软件等，购置数量根据相关设备的购置数量确定，具体明细如下：

序号	软件名称	单位	数量	单价 (万元)	金额 (万元)
1	Windows 操作系统	套	490	0.30	147.00
2	O365E3	套	245	1.04	254.80
3	Visual Studio	套	245	3.92	960.40
4	Audition /Photoshop 六年版	套	245	2.80	686.00
合计					2,048.20

(5) 技术人员工资

本项目技术人员为项目经理，主要负责项目产品设计、项目进程管理等。根据项目拟建设的数据集规模、需要使用的人员数量，参考公司实际薪资水平、募投项目建设所在地平均薪资水平进行测算，具体明细如下：

职位	项目	第一年	第二年	第三年
项目经理	工资总额 (万元)	368.00	576.00	640.00
	年薪及福利 (万元/人)	32.00	32.00	32.00

	总共人员（人）	12	20	20
--	---------	----	----	----

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

（6）数据资源采购

本次募投项目将提供可供大模型训练和评测的不少于 10 个品类的专业数据集，具体可分为通用及特定垂直领域的大语言模型训练数据集、多模态大模型训练数据集及大模型评测数据集三类。数据资源采购费用根据拟建设的数据集类别及规模，结合公司历史业务数据采集类别、规模及市场定价等因素预估，合计为 6,690.00 万元，具体明细如下：

数据集类型	数据集明细	采购费用（万元）
通用及特定垂直领域的大语言模型训练数据集	中文高质量语料库、中文及多语种对话语料库、中文及多语种通用场景语料库、中文特定场景语料库、中文及多语种指令数据集等	3,565
多模态大模型训练数据集	全场景自动驾驶数据集、多语种语音识别大模型数据集、多语言图文识别大模型数据集、多语种语音生成大模型数据集、多语言图文生成大模型数据集、数字人多模态数据集等	3,025
大模型评测数据集	大模型评测数据集	100

（7）铺底流动资金

本项目铺底流动资金包括数据集标注成本、清洗成本和项目建设期三年新增流动资金需求，其中标注成本为专家级数据工程师和高级数据标注师工资总额，合计为 4,258.94 万元；清洗成本为数据清洗工程师工资总额，合计为 1,237.50 万元；新增流动资金需求合计为 1,760.22 万元。标注及清洗成本中相关人员数量根据项目拟建设的数据集规模、需要使用的人员数量，参考公司实际薪资水平、募投项目建设所在地平均薪资水平进行测算；项目建设期三年新增流动资金需求根据项目预计效益支出情况及历史期间相关科目的周转次数确定，具体明细如下：

1) 标注成本

职位	项目	第一年	第二年	第三年
专家级数据工程师	工资总额（万元）	243.60	584.64	901.32
	年薪及福利（万元/人）	19.49	19.49	19.49
	总共人员（人）	15	35	50
高级数据	工资总额（万元）	324.80	828.24	1,376.34

标注师	年薪及福利（万元/人）	9.74	9.74	9.74
	总共人员（人）	40	100	155

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

2) 清洗成本

职位	项目	第一年	第二年	第三年
数据清洗工程师	工资总额（万元）	262.50	412.50	562.50
	年薪及福利（万元/人）	30.00	30.00	30.00
	总共人员（人）	10	15	20

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

3) 项目建设期三年新增流动资金需求

单位：万元

项目	第一年	第二年	第三年
应收账款（A）	847.18	1,731.83	2,531.55
现金（B）	124.79	226.06	289.10
应付账款（C）	336.40	706.81	1,060.43
所需流动资金（A+B-C）	635.57	1,251.08	1,760.22
各年新增流动资金	635.57	615.51	509.14
三年新增流动资金需求合计	1,760.22		

2、数据生产垂直大模型研发项目

(1) 投资数额安排明细及测算依据

本项目总投资 40,651.64 万元，分 3 年投入，具体投资数额安排如下：

单位：万元

序号	项目	金额	第一年	第二年	第三年	是否属资本性支出
1	场地费用	2,346.00	2,346.00	-	-	是
1.1	场地购置费用	2,208.00	2,208.00	-	-	是
1.2	场地装修费用	138.00	138.00	-	-	是
2	设备购置费用	29,895.00	12,145.20	9,497.60	8,252.20	是
3	软件购置费用	451.89	235.77	117.89	98.24	是
4	技术人员工资	4,902.50	1,050.00	1,667.50	2,185.00	否
5	设备托管费	3,056.25	588.75	1,042.50	1,425.00	否
合计		40,651.64	16,365.72	12,325.49	11,960.44	-

(2) 场地费用

1) 场地购置

本项目拟购入办公场地 460 平方米，场地均价 4.8 万元/平方米，总计金额为 2,208 万元。具体明细如下，相关测算依据参见本回复之“1.关于本次募投项目”之“四”之“（三）”：

投资内容	场地用途	项目人员 人数（人）	人均面积 （平方米）	总面积 （平方米）	场地均价 （万元/平方米）	金额 （万元）
场地购置费	办公场地	46	10	460	4.8	2,208.00

2) 场地装修

本项目中，场地装修标准参考了公司近年购置场地的装修造价，拟投入装修费用 138 万元，装修面积 460 平方米，单价为 0.3 万元/平方米。具体明细如下：

投资内容	场地用途	装修面积 （平方米）	装修单价（万元/平方米）	金额（万元）
场地装修费	办公场地	460	0.3	138.00

(3) 设备购置

“数据生产垂直大模型研发项目”的设备主要包括 GPU 服务器、存储服务器、CPU 服务器等，设备采购单价主要参考同类产品市场价格及其他上市公司采购单价确定，具体明细如下：

序号	设备名称	单位	数量	单价 （万元）	金额 （万元）
1	GPU 服务器	台	220	120.00	26,400.00
2	存储服务器	台	40	50.00	2,000.00
3	CPU 服务器	台	100	10.00	1,000.00
4	其他：台式机、笔记本、智能交换机、共享磁盘阵列等				495.00
合计					29,895.00

①GPU 服务器

GPU 服务器主要用于大模型训练、上线推理、评测验证，具体设备数量根据拟研发模型的数据量、参数规模、研发周期要求等，参考市场案例确定。公司预计投入 220 台 GPU 服务器（对应 1,760 张显卡），并根据模型架构、参数量、Token 数、训练效率等进行实时调整，预计可以完成相关模型训练的实验验证与

模型生产目标。

序号	模型名称	开发商	参数量	Token 数	GPU 型号	GPU 显卡数量	训练效率
1	LLaMa	Meta	65B	1.4T	A100 80G	2,048	21 天/轮
2	ChatGLM	智谱 AI	130B	400B	A100 40G	768	60 天/轮
3	Pangu- α	华为	200B	250B	Ascend910 32G	2,048	未披露

注：模型具体相关信息来源下述论文：“Open and Efficient Foundation Language Models”（LLaMa）、“GLM-130B: AN OPEN BILINGUAL PRE-TRAINED MODEL”（GLM）及“PANGU- α : LARGE-SCALE AUTOREGRESSIVE PRETRAINED CHINESE LANGUAGE MODELS WITH AUTO-PARALLEL COMPUTATION”（Pangu- α ）。

②存储服务器

大模型生产环节需要存储大量原始数据、数据清洗过程中每个模块的阶段性和日志、不同版本的成品数据、模型中间结果和成品模型等，拟需要存储服务器 40 台。

③CPU 服务器

为了满足任务的高效完成和计算资源的要求，根据网络获取/下载、数据清洗和高质量数据质检的需求计算，需要 CPU 服务器数量 100 台。

（4）软件购置

本项目拟购入 451.89 万元软件设备，包括办公软件和设计软件，购置数量根据相关设备的购置数量确定，具体明细如下：

序号	软件名称	单位	数量	单价（万元）	金额（万元）
1	Windows 操作系统	套	92	0.30	27.60
2	O365E3	套	46	1.04	47.84
3	xftp	套	46	0.06	2.94
4	xshell Netsarang	套	46	0.10	4.55
5	Navicat 16 for SQL Server	套	46	0.28	12.74
6	Visual Studio	套	46	3.92	180.32
7	PyCharm Professional Edition	套	46	1.68	77.28
8	Anaconda Pro	套	46	2.14	98.62
总计			451.89		

（5）技术人员工资

本项目技术人员包括大模型算法研发工程师及大模型调优研发工程师，主要负责大模型的算法研发和模型调优等。根据项目拟建设的大模型规模、需要使用的人员数量，参考公司实际薪资水平、募投项目建设所在地平均薪资水平进行测算，具体明细如下：

岗位名称	项目	第一年	第二年	第三年
大模型算法研发工程师	工资总额（万元）	840.00	1,230.00	1,590.00
	年薪及福利（万元/人）	60.00	60.00	60.00
	总共人员（人）	16	22	28
大模型调优研发工程师	工资总额（万元）	210.00	437.50	595.00
	年薪及福利（万元/人）	35.00	35.00	35.00
	总共人员（人）	8	14	18
合计	工资总额（万元）	1,050.00	1,667.50	2,185.00
	人数（人）	24	36	46

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

（6）设备托管费

本项目计划将购置的360台服务器全部托管，设备托管费合计3,056.25万元，设备托管单价主要参考市场价格确定，具体明细如下：

设备名称	台数（台）	单价（万元/年）	第1年托管量（台）	第2年托管量（台）	第3年托管量（台）
GPU 服务器	220	6.00	90	160	220
存储服务器	40	0.75	15	30	40
CPU 服务器	100	0.75	50	80	100
合计	360	-	155	270	360

单位：万元

设备名称	第一年	第二年	第三年
GPU 服务器	540.00	960.00	1,320.00
存储服务器	11.25	22.50	30.00
CPU 服务器	37.50	60.00	75.00
小计	588.75	1,042.50	1,425.00
合计	3,056.25		

综上所述，本次募投项目投资测算依据充分、金额测算合理，投资数额确定具有谨慎性。

(二) 结合本次募投项目中非资本性支出金额情况，测算本次募投项目中实际补充流动资金的具体数额及其占本次拟募集资金总额的比例，是否超过30%

本次募集资金将用于“AI 大模型训练数据集建设项目”及“数据生产垂直大模型研发项目”，总投资额 78,989.00 万元，资本性支出 55,499.59 万元、非资本性支出 23,489.41 万元，非资本性支出占比为 29.74%。本次募投项目中实际补充流动资金的金额为 23,489.41 万元，占本次拟募集资金总额的比例未超过 30%。两个项目具体情况分别如下：

一、AI 大模型训练数据集建设项目

本项目投资所需总资金 38,337.36 万元，全部使用募集资金，其中非资本性支出 15,530.66 万元，具体构成如下：

序号	项目	金额（万元）	是否属于资本性支出
1	场地购置及装修	18,195.00	是
2	设备购置费用	2,563.50	是
3	软件购置费用	2,048.20	是
4	技术人员工资	1,584.00	否
5	数据资源采购	6,690.00	否
6	铺底流动资金	7,256.66	否
合计		38,337.36	-

二、数据生产垂直大模型研发项目

本项目投资所需总资金 40,651.64 万元，全部使用募集资金，其中非资本性支出 7,958.75 万元，具体构成如下：

序号	项目	金额（万元）	是否属于资本性支出
1	场地购置及装修	2,346.00	是
2	设备购置费用	29,895.00	是
3	软件购置费用	451.89	是
4	研发人员费用	4,902.50	否
5	设备托管费用	3,056.25	否
合计		40,651.64	-

二、结合日常营运需要、公司货币资金余额及使用安排、日常经营积累、资金缺口、公司资产负债率与同行业可比公司的对比等情况，分析本次募集资金规模的合理性

（一）公司日常营运需要、公司货币资金余额及使用安排、日常经营积累、资金缺口等情况

截至 2023 年 8 月 31 日，发行人合并报表货币资金余额 5,012.87 万元，交易性金融资产 36,732.70 万元，合计 41,745.56 万元（以上数据未经审计）。综合考虑公司的日常营运需要、公司货币资金余额及使用安排、日常经营积累等，公司目前的资金缺口为 79,081.32 万元，具体测算过程如下：

单位：万元

项目	金额
资金需求：	
1、日常流动资金需求	15,263.49
2、未来三年新增营运资金需求	1,749.01
3、已购置办公场地装修及改造支出	5,000.00
4、未来三年现金分红	5,483.05
5、新增研发方向及未来产业投资布局支出	17,000.00
6、已备案投资项目资金需求	78,989.00
小计	123,484.55
资金来源：	
1、货币资金	5,012.87
2、交易性金融资产	36,732.70
3、未来三年日常经营积累	2,657.67
小计	44,403.23
资金缺口	79,081.32

1、资金需求

（1）日常流动资金需求

公司日常生产经营活动中需要维持一定规模的流动资金，按 2022 年度的经营情况进行测算，发行人月均经营性现金流出为 2,210.58 万元，假设公司日常生产经营需要覆盖 6 个月的经营性现金流出，公司日常流动资金需求约 13,263.49

万元，为应对宏观环境波动及突发事件影响，公司额外预留 2,000 万元日常流动储备资金，基于上述，日常流动资金需求合计约 15,263.49 万元。测算情况如下所示：

财务指标	金额（万元）	计算公式
购买商品、接受劳务支付的现金	10,764.05	A
支付给职工以及为职工支付的现金	11,193.94	B
支付的各项税费	1,497.84	C
支付其他与经营活动有关的现金	3,071.15	D
小计	26,526.98	E=A+B+C+D
月均经营性现金支出	2,210.58	F=E/12
日常流动资金需求（假设覆盖 6 个月）	13,263.49	G=F*6
日常流动储备资金	2,000.00	H
合计	15,263.49	I=G+H

（2）未来三年新增营运资金需求

参考公司过去三年的收入年复合增长率 6.14%，假定未来三年的收入年复合增长率保持同样水平（相关假设仅用于测算新增营运资金需求使用，有关营业收入、资产规模等假设数据不构成对公司的业绩承诺和盈利预测，投资者不应据此进行投资决策，下同），按照收入百分比法预测，发行人未来新增营运资金需求为 1,749.01 万元，具体测算过程如下：

单位：万元

项目	预测期初（2022 年）		预测期末（2025 年）	
	金额	占比	金额	占比
营业收入	26,288.79	100.00%	31,434.59	100.00%
应收账款	10,178.03	38.72%	12,170.29	38.72%
预付款项	258.40	0.98%	308.98	0.98%
存货	461.75	1.76%	552.14	1.76%
合同资产	437.29	1.66%	522.88	1.66%
经营性资产合计	11,335.47	43.12%	13,554.29	43.12%
应付账款	1,826.22	6.95%	2,183.69	6.95%
合同负债	573.92	2.18%	686.26	2.18%
经营性负债合计	2,400.14	9.13%	2,869.95	9.13%

项目	预测期初（2022年）		预测期末（2025年）	
	金额	占比	金额	占比
经营性营运资金占用额	8,935.33	33.99%	10,684.34	33.99%
新增营运资金需求合计（注）	1,749.01			

注：新增营运资金需求合计=预测期末经营性营运资金占用额-预测期初经营性营运资金占用额。

（3）已购置办公场地装修及改造支出

发行人目前正在进行已购置房产的装修及改造，预计后续仍需投入的装修及改造费支出约 5,000 万元。鉴于公司所购置的办公场地建成年代较早，除常规装修支出外，公司还需对楼体进行加固及改造，相关支出主要包括楼体结构及抗震加固、楼体结构相关机电工程、外立面改造工程及单独报建的方案设计费用³等，具体如下：

坐落地址	面积（m ² ）	常规装修支出（万元）	楼体加固及改造支出（万元）	合计支出（万元）
东城区新中街68号8A楼1层公建一层03等21套	6,192.79	2,000	3,000	5,000

（4）未来三年现金分红

公司 2021 年 8 月上市，2021-2022 年累计现金分红金额为 3,655.36 万元，对应年均现金分红金额为 1,827.68 万元。发行人重视股东回报，与投资者共享发展成果，参考未来经营预期、市场情况以及上市后的现金分红情况，预计未来三年的现金分红资金保持之前的分红金额水平，三年合计金额为 5,483.05 万元。

（5）新增研发方向及未来产业投资布局支出

由于人工智能行业整体发展较为迅速，应用场景及下游需求等均不断发生拓展，公司进行数据开发所需的各项技术、产品结构也需要快速迭代以应对市场变化。因此公司计划储备研发资金 12,000 万元用于在合成/仿真数据技术、供应链智能推荐与调配技术、面向具身智能等新兴 AI 应用形态的多模态数据集研发技术及更大范围的行业应用数据集探索等方向进一步加大研发投入，以适应行业及市场发展需要，巩固行业领先地位。此外，公司此前一直坚持自主发展路线，为

³ 发行人已就装修及楼体改造加固履行了招投标程序，并选定北京城建安装集团有限公司为施工承包方，双方签署了施工总承包合同。

更好应对市场变化、提升业务规模及抗风险能力，公司后续计划立足自身战略，适当横拓赛道、纵延深度，围绕人工智能产业链上下游及数据要素产业链进行一定布局投资，培育业绩增长新动能，此部分储备资金预留 5,000 万元。

(6) 已备案投资项目资金需求

截至 2023 年 6 月 30 日，公司已备案的投资项目包括 “AI 大模型训练数据集建设项目” 及 “数据生产垂直大模型研发项目”，总投资额 78,989.00 万元。

2、资金来源

(1) 货币资金及交易性金融资产

截至 2023 年 8 月 31 日，发行人合并报表货币资金余额 5,012.87 万元，交易性金融资产 36,732.70 万元，合计 41,745.56 万元（以上数据未经审计）。

(2) 未来三年日常经营积累

公司未来三年自身经营利润积累以归属于上市公司股东的净利润为基础进行测算。公司 2021 年 8 月上市，2021 年、2022 年及 2023 年 1-6 月（未经审计，已年化）的年均净利润为 885.89 万元，基于谨慎性原则，假设公司未来三年归属于上市公司股东的净利润与上述平均利润水平一致，经测算，公司未来三年预计自身经营利润积累为 2,657.67 万元。

通过上述分析，综合考虑公司的日常营运需要、公司货币资金余额及使用安排、日常经营积累等，公司总体资金缺口为 79,081.32 万元，超过本次计划募集资金总额 78,989.00 万元，公司本次募集资金规模具有合理性。

(二) 公司资产负债率与同行业可比公司的对比情况

报告期内，发行人合并报表资产负债率分别为 8.25%、4.13%、5.41% 及 3.41%，与同行业可比上市公司资产负债率对比情况如下：

项目	证券代码	证券简称	2023.6.30	2022.12.31	2021.12.31	2020.12.31
资产负债率 (合并)	300290.SZ	荣科科技	40.06%	43.61%	42.19%	28.86%
	300245.SZ	天玑科技	11.27%	15.70%	13.91%	10.85%
	300468.SZ	四方精创	5.36%	7.55%	8.15%	6.90%
	300634.SZ	彩讯股份	15.87%	17.99%	15.59%	15.72%

项目	证券代码	证券简称	2023.6.30	2022.12.31	2021.12.31	2020.12.31
	上述 4 家信息技术服务类可比公司平均数		18.14%	21.21%	19.96%	15.58%
	上述 4 家信息技术服务类可比公司中位数		13.57%	16.85%	14.75%	13.29%
	600570.SH	恒生电子	39.63%	42.52%	48.55%	49.48%
	600588.SH	用友网络	50.34%	46.86%	54.26%	49.57%
	300624.SZ	万兴科技	15.57%	42.37%	42.93%	22.18%
	603039.SH	泛微网络	43.41%	46.04%	44.34%	50.35%
	上述 4 家软件产品类可比公司平均数		37.24%	44.45%	47.52%	42.90%
	上述 4 家软件产品类可比公司中位数		41.52%	44.28%	46.45%	49.53%
	可比公司平均数		27.69%	32.83%	33.74%	29.24%
	可比公司中位数		27.75%	42.45%	42.56%	25.52%
	海天瑞声		3.41%	5.41%	4.13%	8.25%

公司资产负债率整体处于较低水平，偿债能力较强。发行人未选择通过银行借款方式进行资金筹措，主要系一方面，授信额度不等同于实际可贷债务，轻资产公司债务融资程序一般均较为繁琐，且借款期限通常较短，无法支撑持续性资本支出需求；另一方面，从节约财务费用、稳健经营的角度考虑，假设本次募投项目全部通过债务融资筹措资金，发行人的合并资产负债率将达到 50.62%，不仅导致公司的资产负债率水平明显高于同行业公司，此外，按照人民银行公布的 1 年内贷款基准利率 4.35% 计算，每年新增财务费用约 3,436.02 万元，将会大幅提高公司的财务风险，因此发行人选择通过股权融资方式进行融资。

综上所述，发行人未来三年资金需求测算谨慎，本次募集资金规模具有合理性。

三、项目效益测算的具体情况、测算过程及测算依据，说明增长率、毛利率、预测净利率等收益指标的合理性，结合毛利率、内部收益率、投资回收期等关键指标与同行业同类项目的对比情况说明项目效益测算的谨慎性和合理性

（一）项目效益测算的具体情况、测算过程及测算依据，说明增长率、毛利率、预测净利率等收益指标的合理性

本次募投项目中“AI 大模型训练数据集建设项目”涉及经济效益，具体经济效益指标如下：

项目	指标
投资回收期（税前，年）	5.45
投资回收期（税后，年）	5.89
内部收益率（税前）	19.36%
内部收益率（税后）	16.82%
财务净现值（税前）	6,444.47 万元
财务净现值（税后）	4,131.15 万元

项目税前及税后财务内部收益率均大于基准收益率 12%，财务净现值大于零，项目财务收益较好，项目整体具有良好的经济效益。各年具体效益情况如下：

单位：万元

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

1、营业收入

本募投项目效益测算中的关键测算指标主要包括产品定价及销售率（销售次数），其中产品定价参考产品开发投入（包括人工费用、数据资源采购费用、数据清洗费用及数据标注费用）进行合理加成确定。根据公司最近 5 年（2018-2022 年）训练数据集产品的开发投入及首次销售收入，计算出加权平均加成率，以此作为参考确定本次募投项目产品定价的加成率。具体过程如下：

单位：万元

现有训练数据集产品首次销售收入	现有训练数据集产品开发投入（注）	加成率①	本次募投开发投入②	本次募投产品总价③=②*（1+①）
10,158.36	8,141.07	25%	13,770.44	17,213.05

注：历史数据统计的开发投入中，除人工费用、数据资源采购费用及数据标注费用外，还包括了少量制作数据集产品使用且归集的设备折旧费用（约占整体开发投入的 5%），为保证

与测算口径一致对该部分进行剔除。

此外，根据公司业务销售模式，鉴于标准化训练数据集产品具有一次性研发生产、可重复多次销售的特点，因此需将产品首次销售时的当期销售率及期后复卖率列入考量，其中当期销售率是指当期新制成的数据集产品在当期实现的销售次数，期后复卖率是指历史期已实现销售的数据集产品在当期实现的销售次数。公司统计了现有业务 2017-2022 年新制成训练数据集产品的销售明细（新制成产品个数合计 963 个，新制成产品当年销售次数合计 886 次），平均当期销售率为 92%，期后 1-5 年平均复卖率区间为 44%-86%。

公司参考上述历史销售数据均值，谨慎确定本次募投大模型数据集产品当期销售率为 90%，建设期内（第 1-3 年）的期后复卖率为 45%，建设期完成后产能销售率的衰减率为 10%⁴，并以此为基础计算产能销售率，具体测算逻辑如下：

（1）第 1 年产能销售率参考当期销售率（90%）；

（2）第 2-3 年产能销售率=（第二年新增产能*当期销售率+之前年度累计产能*期后复卖率）/第二年末累计产能⁵；

（3）建设期完成后（第 4 年开始），产能销售率在前一年度基础上按照 10% 进行衰减计算⁶。

公司募投项目运营期产能销售率及销售收入情况具体如下：

项目	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
累计产能 (万元) (A)	2,914.61	8,752.54	17,213.05	17,213.05	17,213.05	17,213.05	17,213.05	17,213.05
产能 销售率 (B)	90%	75%	60%	54%	49%	44%	39%	35%
销售收入 (万元) (C=A*B)	2,623.15	6,564.40	10,327.83	9,295.05	8,365.54	7,528.99	6,776.09	6,098.48

注：前三年为建设期，累计形成的产能按照投入进度逐年递增，第三年后达到 100%；

⁴ 公司结合数据集历史销售情况（期后 1-5 年平均复卖率区间为 44%-86%），出于谨慎性考虑，将产能销售率在前一年度基础上按照 10% 进行衰减计算，最终第 6-8 年（即建设期开始/结束后 5 年）的销售率为 35%-44%，低于上述期后 1-5 年平均复卖率区间（44%-86%）的最低值 44%。

⁵ 其中，第 2 年产能销售率（75%）=（（8,752.54-2,914.61）*90%+2,914.61*45%）/8,752.54；第 3 年产能销售率（计算结果为 67%，谨慎起见实际选取 60%）=（（17,213.05-8,752.54）*90%+8,752.54*45%）/17,213.05。

⁶ 以第 4 年为例，第 4 年产能销售率（54%）=第 3 年产能销售率（60%）*（1-10%），以此类推。

2、销售费用

销售费用参考公司和同行业水平，占收入比重 5%，公司最近五年（2018-2022 年）平均销售费用率为 5.47%，不具有明显差异。

3、管理费用

管理费用参考公司和同行业平均水平，除募投新增涉及的折旧摊销费以外，其他管理费用占收入比重为 3%。折旧摊销采用年限平均法，具体年限参考公司现有折旧/摊销政策：办公场地及展厅的折旧年限按 40 年计算，残值率为 5%；装修费用摊销年限为 10 年，残值为 0%；软件折旧年限为 5-8 年，残值为 0%。

公司募投项目运营期（含建设期）平均管理费用率为 12.14%，公司最近五年（2018-2022 年）平均管理费用率为 15.50%，整体差异较小；公司最近五年平均管理费用率略高，主要因其包含了上市财经公关及宣传费、房产购置中介服务费 etc 因特定事项产生的中介服务费支出。

4、研发费用

人工费用：本项目人工费用所涉的技术人员为项目经理，主要负责项目产品设计、项目进程管理等。根据项目拟建设的数据集规模、需要使用的人员数量，参考公司实际薪资水平、募投项目建设所在地平均薪资水平进行测算，人员投入的数量在项目建设期完成后以迭代升级为主，投入逐渐减少，具体明细如下：

职位	项目	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
项目经理	工资总额 (万元)	368.00	576.00	640.00	576.00	480.00	384.00	288.00	288.00
	年薪及福利 (万元/人)	32	32	32	32	32	32	32	32
	总共人员(人)	12	20	20	18	15	12	9	9

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

数据资源采购费用：本次募投项目将提供可供大模型训练和评测的不少于 10 个品类的专业数据集，具体可分为通用及特定垂直领域的大语言模型训练数据集、多模态大模型训练数据集及大模型评测数据集三类。数据资源采购费用根据拟建设的数据集类别及规模，结合公司历史业务数据采集类别、规模及市场定价等因素预估，合计为 6,690.00 万元，具体明细如下：

数据集类型	数据集明细	采购费用
-------	-------	------

		(万元)
通用及特定垂直领域的大语言模型训练数据集	中文高质量语料库、中文及多语种对话语料库、中文及多语种通用场景语料库、中文特定场景语料库、中文及多语种指令数据集等	3,565
多模态大模型训练数据集	全场景自动驾驶数据集、多语种语音识别大模型数据集、多语言图文识别大模型数据集、多语种语音生成大模型数据集、多语言图文生成大模型数据集、数字人多模态数据集等	3,025
大模型评测数据集	大模型评测数据集	100

数据标注费用：可按人员级别具体分为专家级数据工程师和高级数据标注师工资总额，合计为 4,258.94 万元，具体明细如下：

职位	项目	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
专家级数据工程师	工资总额(万元)	243.60	584.64	901.32	682.08	477.46	334.22	233.95	163.77
	年薪及福利(万元/人)	19	19	19	19	19	19	19	19
	总共人员(人)	15	35	50	35	25	17	12	8
高级数据标注师	工资总额(万元)	324.80	828.24	1,376.34	584.64	467.71	350.78	233.86	116.93
	年薪及福利(万元/人)	10	10	10	10	10	10	10	10
	总共人员(人)	40	100	155	60	48	36	24	12

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

数据清洗费用：清洗费用为数据清洗工程师工资总额，合计为 1,237.50 万元，具体明细如下：

职位	项目	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
数据清洗工程师	工资总额(万元)	262.50	412.50	562.50	300.00	270.00	240.00	210.00	180.00
	年薪及福利(万元/人)	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00
	总共人员(人)	10	15	20	10	9	8	7	6

注：计算工资总额时，考虑了相关人员的招聘及到岗周期等因素。

折旧摊销：折旧摊销采用年限平均法，具体年限参考公司现有折旧/摊销政策：多模态采集室的折旧年限按 40 年计算，残值率为 5%；硬件设备折旧年限为 3-5 年，残值率均为 5%。

5、退税收入

依据 2011 年 10 月 13 日发布《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，自 2011 年 1 月 1 日起，增值税一般纳税人销售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。

本次募投项目中，公司根据历史产品收入中退税部分的占比进行预估，按照预测产品收入的 80% 享受 10% 的增值税退税款预计退税收入。

（二）结合毛利率、内部收益率、投资回收期等关键指标与同行业同类项目的对比情况说明项目效益测算的谨慎性和合理性

发行人同行业可比公司主要包括 Appen、数据堂及标贝科技，根据公开信息查询，其中 Appen 已推出智能 LLM 大模型开发平台，集大模型数据准备、模型训练、模型推理、模型部署应用于一体；数据堂已推出多语言多模态视频文本对齐数据集、中文高质量大模型预训练文本数据集；标贝科技启动了大模型技术的非平衡专业语料的构建工作，将陆续推出一系列高质量的数据集。上述发行人同行业可比公司均未披露相关大模型数据集产品的销售或毛利情况。

发行人本次募投项目“AI 大模型训练数据集建设项目”主要建设内容为应用于大模型领域的数据集，目前并无完全可比的同行业同类项目披露相关信息，以下选取同为人工智能及大模型领域的建设类募投项目作为同类项目，将公司本次募投项目的毛利率、内部收益率、投资回收期进行对比：投资回收期方面，公司与同行业同类项目相比差异较小；内部收益率方面，公司内部收益率为 16.82%，处于可比区间；毛利率方面，公司数据集产品运营期（含建设期）的投入均计入费用不计入成本，因此毛利率为 100%，同行业同类项目的毛利率均处于较高水平（大于 80%）。因此本次项目效益测算关键指标与同行业同类项目相比，具有谨慎性和合理性。

人工智能公司	证券代码	项目名称	毛利率	内部收益率	投资回收期
云从科技	688327	云从“行业精灵”大模型研发项目	95.08%	15.43%	7.33 年
星环科技	688031	数据分析大模型建设项目	80.36%	13.32%	6.61 年

		智能量化投研一体化平台建设项目	86.89%	26.87%	5.60 年
万兴科技	300624	数字创意资源商城建设项目	88.39%	28.21%	5.29 年
广联达	002410	造价大数据及 AI 应用项目	未披露	20.93%	4.84 年
海天瑞声	688787	AI 大模型训练数据集建设项目	100%	16.82%	5.89 年

四、公司针对上述事项履行的决策程序和信息披露情况

2023 年 6 月 21 日，发行人召开第二届董事会第十七次会议，审议通过了《关于公司符合向特定对象发行 A 股股票条件的议案》《关于公司 2023 年度向特定对象发行 A 股股票方案的议案》《关于公司 2023 年度向特定对象发行 A 股股票预案的议案》《关于公司 2023 年度向特定对象发行 A 股股票方案论证分析报告的议案》《关于公司 2023 年度向特定对象发行 A 股股票募集资金使用可行性分析报告的议案》《关于公司本次募集资金投向属于科技创新领域的说明的议案》《关于公司开立本次向特定对象发行 A 股股票募集资金专项账户的议案》《关于公司前次募集资金使用情况报告的议案》《关于公司 2023 年度向特定对象发行 A 股股票摊薄即期回报的影响与填补回报措施及相关主体承诺的议案》《关于公司未来三年（2023 年—2025 年）股东分红回报规划的议案》《关于提请股东大会授权董事会办理公司本次向特定对象发行 A 股股票相关事宜的议案》等与本次发行相关的议案，并同意将上述与本次发行相关的全部议案提交 2023 年第一次临时股东大会审议，独立董事就本次发行的相关事项发表了事前认可意见和同意的独立意见。

2023 年 7 月 7 日，发行人通过现场表决和网络投票相结合的方式召开 2023 年第一次临时股东大会，审议通过了《关于公司符合向特定对象发行 A 股股票条件的议案》《关于公司 2023 年度向特定对象发行 A 股股票方案的议案》《关于公司 2023 年度向特定对象发行 A 股股票预案的议案》《关于公司 2023 年度向特定对象发行 A 股股票方案论证分析报告的议案》《关于公司 2023 年度向特定对象发行 A 股股票募集资金使用可行性分析报告的议案》《关于公司本次募集资金投向属于科技创新领域的说明的议案》《关于公司前次募集资金使用情况报告的议案》《关于公司 2023 年度向特定对象发行 A 股股票摊薄即期回报的影响与填补回报措施及相关主体承诺的议案》《关于公司未来三年（2023 年—2025

年) 股东分红回报规划的议案》《关于提请股东大会授权董事会办理公司本次向特定对象发行 A 股股票相关事宜的议案》等议案; 就本次发行相关议案, 中小股东表决情况进行单独计票。

上述决议及相关公告已在上海证券交易所指定网站披露。

综上所述, 发行人就上述事项已履行完整内部决策程序并完成信息披露, 符合法律、行政法规及上海证券交易所的相关规定。

五、核查程序及核查意见

(一) 核查程序

保荐机构履行了以下核查程序:

1、查阅了发行人本次募投项目的可行性研究报告及具体投资明细, 测算本次募投项目中实际补充流动资金的具体数额及其占本次拟募集资金总额的比例;

2、复核分析了发行人日常营运需要、公司货币资金余额及使用安排、日常经营积累、资金缺口、公司资产负债率与同行业可比公司的对比等情况, 并判断募集资金规模的合理性;

3、获取并查阅了发行人本次募投项目的具体效益测算明细及对应的历史运营数据依据, 并通过公开信息检索同行业同类型项目相关信息披露, 判断收益指标及项目效益测算的合理性与谨慎性;

4、查阅发行人关于募投项目的决议文件及相关信息披露文件。

(二) 核查意见

经核查, 保荐机构认为:

1、本次募投项目中实际补充流动资金的具体数额其占本次拟募集资金总额的比例不超过 30%;

2、根据日常营运需要、公司货币资金余额及使用安排、日常经营积累、资金缺口、公司资产负债率与同行业可比公司的对比等情况, 本次募集资金规模具有合理性;

3、本次募投项目效益测算主要收益指标具有合理性, 关键指标与同行业同

类项目的对比，具有谨慎性和合理性；

4、发行人本次募投项目已履行有效的决策程序，并及时进行了信息披露。

4. 关于经营情况

根据申报材料，1)最近三年及一期公司营业收入分别为 23,337.40 万元、20,647.65 万元、26,288.79 万元及 2,881.74 万元，净利润分别为 8,208.10 万元、3,160.54 万元、2,945.41 万元和-1,361.63 万元；2)公司的销售毛利率分别为 67.70%、64.01%、64.73%和 47.79%；3)公司期间费用分别为 8,244.97 万元、11,140.30 万元、16,881.20 万元及 3,766.34 万元，占营业收入比重分别为 35.33%、53.95%、64.21%及 130.70%，其中管理费用与研发费用大幅上涨，研发费用主要由职工薪酬及数据服务费构成；4)公司经营活动产生的现金流量净额分别为 5,117.67 万元-1,554.83 万元、3,065.89 万元及-2,268.44 万元；5)最近三年及一期，应收账款账面价值分别为 6,278.29 万元、9,019.65 万元、10,178.03 万元及 8,904.07 万元，1 年以上长账龄占比逐步增加，公司应收账款周转率分别为 4.11、2.51、2.53 和 1.11（已年化）。

请发行人说明：（1）最近三年及一期公司收入与净利润波动趋势差异较大的原因及合理性，结合公司历史收入的季节性分布、同行业可比公司情况等说明公司收入是否受季节性因素影响，2023 年是否面临收入大幅下滑的风险，结合行业竞争格局、技术壁垒及迭代情况、数据监管最新政策及公司对此采取的对策说明公司净利润大幅下滑趋势是否将持续，相关不利因素是否已经消除或改善，公司的持续经营能力是否存在重大不确定性，并完善相关风险提示；（2）结合公司主要产品毛利率的波动趋势及原因、同行业可比公司可比产品毛利率变动趋势、产业链上下游的市场变化情况等，说明公司最近三年及一期毛利率下滑尤其是最近一期末大幅下滑的原因及合理性；（3）结合同行业可比公司各项期间费用情况及变动趋势，说明公司期间费用规模的合理性；最近三年及一期公司管理费用及研发费用大幅增长的原因及合理性；研发费用的明细构成，结合研发人员的具体情况，包括但不限于研发人员数量、学历、人均工资、所从事的研发项目情况，说明研发费用中职工薪酬各期大幅上涨的原因及合理性，研发费用中数据服务费的具体用途，与主营业务的数据服务费是否存在混同情况，数据服务费归集与核算的具体方法，研发费用中数据服务费归集与核算的准确性；（4）报告期内应收账款账面价值大幅上涨、应收账款周转率大幅下滑的原因及合理性，相关不利因素是否已经消除或改善；应收账款的账龄情况，一年以上长账龄的应收

账款占比增长的原因，相关客户的经营情况、回款及逾期情况、相应坏账准备计提情况；结合公司应收账款整体的回款及逾期情况、应收账款坏账准备计提政策、信用政策、同行业可比公司情况等说明公司应收账款坏账准备计提的充分性。

请保荐机构和申报会计师核查并发表明确意见。

一、最近三年及一期公司收入与净利润波动趋势差异较大的原因及合理性，结合公司历史收入的季节性分布、同行业可比公司情况等说明公司收入是否受季节性因素影响，2023 年是否面临收入大幅下滑的风险，结合行业竞争格局、技术壁垒及迭代情况、数据监管最新政策及公司对此采取的对策说明公司净利润大幅下滑趋势是否将持续，相关不利因素是否已经消除或改善，公司的持续经营能力是否存在重大不确定性，并完善相关风险提示

（一）最近三年及一期公司收入与净利润波动趋势差异较大的原因及合理性

报告期内，公司营业收入存在一定波动，同时受到毛利率、研发费用、管理费用等科目波动的影响，导致公司净利润的波动趋势与营业收入差异较大，具体如下：

单位：万元

项目	2023 年 1-6 月	2022 年度	2021 年度	2020 年度
营业收入	7,446.09	26,288.79	20,647.65	23,337.40
营业成本	3,064.03	9,271.05	7,431.19	7,537.79
毛利	4,382.06	17,017.74	13,216.47	15,799.60
毛利率	58.85%	64.73%	64.01%	67.70%
销售费用	1,213.33	1,836.16	1,452.34	1,113.34
管理费用	3,345.07	6,140.46	3,521.80	2,566.74
研发费用	3,221.36	9,427.34	6,052.09	4,349.75
净利润	-1,724.14	2,945.41	3,160.54	8,208.10

2021 年，发行人营业收入较 2020 年降低 2,689.74 万元，净利润较 2020 年降低 5,047.56 万元，净利润降幅较大，主要由于一方面，在境外公共卫生事件出现重大反复的背景下，发行人境外业务遭受一定冲击，境外收入下降 31.53%。由于境外业务系发行人高毛利收入来源，收入降低的同时毛利也减少 2,583.13 万元。另一方面，2021 年发行人期间费用增长较多，其中研发费用增加 1,702.34

万元，主要由于发行人综合判断未来下游客户对多语种训练数据产品、计算机视觉训练数据产品的需求将有所增长，发行人为把握市场趋势，推进了训练数据产品储备，2021 年进行开发的训练数据产品个数从 203 个增长至 353 个，同时，为进一步提高数据处理智能化水平，公司投入大量研发力量，启动建设以智能驾驶为代表的垂直领域技术平台能力；管理费用、销售费用分别增加了 955.06 万元、339.00 万元，主要系公司 2021 年加大管理团队和销售团队人才储备、发放上市专项奖金导致职工薪酬增长较多以及 2020 年社保减免在 2021 年解除等因素所致。

2022 年，发行人营业收入较 2021 年增长 5,641.14 万元，净利润较 2021 年减少 215.13 万元，净利润波动与营业收入差异较大，主要系当年公司管理费用、研发费用、销售费用等期间费用增长幅度大于毛利增长速度，导致净利润水平有所下滑。具体而言，2022 年发行人毛利较上年增长 3,801.27 万元，增幅 28.76%，三项费用增长 6,377.73 万元，增幅 57.84%。其中，研发费用较上年增加 3,375.25 万元，主要系为加速布局和提升智能驾驶数据服务能力、整体数据生产的智能化水平，以及推进前次募集资金规划的数据集产品建设等，发行人持续加大研发投入；管理费用较上年增加 2,618.66 万元，销售费用较上年增加 383.82 万元，主要系发行人 2022 年完成管理架构及薪酬体系升级、全球化营销体系建设、实施限制性股票激励计划、新办公楼购置等重要事项，以上因素共同导致期间费用大幅增长。

2023 年上半年，发行人实现收入 7,446.09 万元，较上年同期下滑 35.13%，实现净利润-1,724.14 万元，较上年同期下滑 188.76%，净利润下滑幅度显著高于营业收入。主要原因为：首先，受到宏观经济波动影响，部分境外客户正在进行业务方向调整、阶段性裁员等，导致其预算释放进度放缓，叠加数据出境相关法规落地实施带来的影响，发行人上半年收入规模较小，而各项费用支出相对刚性；其次，2023 年上半年管理费用、销售费用分别较上年同期增长 1,072.49 万元和 384.82 万元，主要系团队人数增长以及 2022 年第二季度启动的薪酬体系改革导致的职工薪酬增长在 2023 年上半年完整体现，叠加新购置办公楼导致折旧费用同比增加所致，以上因素共同导致发行人净利润下降幅度较大。

关于报告期内收入、毛利率、期间费用波动的具体分析，请参见本题其他部

分回复以及“6. 关于境外销售”的回复。

综上所述，最近三年及一期发行人收入与净利润波动趋势差异较大具备合理性。

（二）结合公司历史收入的季节性分布、同行业可比公司情况等说明公司收入是否受季节性因素影响，2023年是否面临收入大幅下滑的风险

1、公司报告期收入的季节性分布情况

2020年至2022年，发行人分季度收入情况如下：

单位：万元

项目	2022年度		2021年度		2020年度	
	营业收入	占比	营业收入	占比	营业收入	占比
第一季度	4,787.64	18.21%	4,423.89	21.43%	3,126.09	13.40%
第二季度	6,690.80	25.45%	6,176.27	29.91%	7,067.83	30.29%
第三季度	5,237.50	19.92%	2,477.47	12.00%	4,083.44	17.50%
第四季度	9,572.84	36.41%	7,570.03	36.66%	9,060.04	38.82%
合计	26,288.79	100.00%	20,647.65	100.00%	23,337.40	100.00%

根据上表可知，发行人收入存在一定季节性波动特点。受春节假期等因素影响，一季度收入占比普遍较低；由于客户采购节奏等因素，四季度收入占比普遍较高。

2、同行业公司的季节性分布情况

公司所属行业为软件和信息技术服务行业，所在行业的A股可比公司2022年各季度营业收入分布情况如下：

证券简称	第一季度	第二季度	第三季度	第四季度
荣科科技	16.72%	26.21%	35.37%	21.70%
天玑科技	11.47%	13.51%	18.41%	56.61%
四方精创	22.94%	25.92%	24.20%	26.94%
彩讯股份	23.79%	28.92%	25.67%	21.61%
恒生电子	14.97%	21.72%	20.78%	42.53%
用友网络	13.81%	24.38%	22.23%	39.59%
万兴科技	22.26%	23.52%	25.26%	28.96%

泛微网络	12.57%	21.28%	23.82%	42.33%
------	--------	--------	--------	--------

根据上表可知，同行业公司中大多也呈现出一季度收入较小、四季度收入较高的季节性特征，发行人收入的季节性分布情况与行业情况相符。

3、2023 年是否面临收入大幅下滑的风险

2023 年上半年，发行人营业收入规模较小且较去年同期有所下滑，一定程度上为受到季节性影响，主要原因与境内外不利因素的冲击有关，2023 年二季度以来，相关下滑趋势已有所好转，发行人预计 2023 年收入不存在大幅下滑的风险，具体分析如下：

发行人 2023 年上半年收入规模较小，且相比去年同期降低 35.13%，主要受到以下因素影响：一方面由于受宏观经济波动影响，部分境外客户正在进行业务方向调整、阶段性裁员等，导致其预算释放进度放缓，叠加数据出境相关法规落地实施所带来的影响等多重因素，公司境外业务的开展受到影响，境外收入较去年同期收缩明显，上半年境外收入同比下降 55.58%；另一方面，受到宏观经济波动及市场竞争加剧的影响，境内业务收入同比下降 7.86%，除智能驾驶业务需求显著增长外，其他基础数据服务领域客户需求虽有所复苏，但客户投入更为谨慎，导致上半年在手订单执行及新增订单开发不及预期。

自 2023 年二季度以来，发行人境外业务较一季度呈复苏态势，客户的签约及询单量均有所增加，二季度境外业务收入规模相较一季度提升 87.47%，占比由一季度的 35.16% 提升至 41.62%，毛利率由一季度的 65.89% 提升至 81.86%，境外业务规模及盈利质量均有明显提升。境内业务方面，受公司传统语音数据类产品销售量回升及智能驾驶业务拉动，二季度境内收入规模较一季度提升 42.62%，毛利率由一季度的 37.98% 提升至 54.41%，呈现出好转态势。根据发行人管理层预计，随着宏观经济环境回暖，境外客户裁员及预算调整周期逐渐结束，境内外客户需求释放，发行人业绩下滑的外部不利因素将有所减弱，截至目前，发行人在手订单超 7,000 万元，意向性订单超 1.7 亿元，考虑到意向订单转化及项目执行周期，2023 年收入存在一定下滑风险，但下滑幅度预计不超过 30%，因此 2023 年全年收入不存在大幅下滑的风险，发行人已于募集说明书等申报材料中进行风险提示。

（三）结合行业竞争格局、技术壁垒及迭代情况、数据监管最新政策及公司对此采取的对策说明公司净利润大幅下滑趋势是否将持续，相关不利因素是否已经消除或改善，公司的持续经营能力是否存在重大不确定性，并完善相关风险提示

虽然发行人最近一期净利润受到收入下滑叠加费用增长较快的影响，出现较大幅度下滑，但相关不利因素有所改善，公司净利润大幅下滑的趋势不会长期持续，具体分析如下：

1、行业竞争格局方面

公司是我国最早从事训练数据研发销售的企业之一，深耕行业近 20 年，积累了大量的技术产品以及客户服务经验，同时公司是国内目前唯一一家 A 股上市的人工智能训练数据服务企业，具有较高的市场知名度和行业地位，抗风险能力较强。报告期内，境内市场的行业竞争进一步加剧，以低价策略作为主要竞争策略的品牌数据服务商和中小数据供应商，对公司境内业务的收入增长及毛利空间优化均带来了一定的不利影响，发行人相应及时调整策略，针对部分市场供给充分的领域，采取更具竞争力的价格策略，保持市场占有率，同时不断优化技术，降低自身成本，2023 年二季度，发行人境内业务毛利率已呈现复苏态势，由一季度的 37.98% 提升至二季度的 54.41%。由于采取低价策略的中小数据服务商在短期内难以在服务质量和技术水平上赶超发行人，因此长期来看以发行人为代表的综合型数据服务商具备竞争优势。境外市场方面，发行人的产品和服务的境外客户主要集中在美国、日韩等国家或地区，发行人面对的主要竞争对手为 Appen、Defined Crowd 及其他本土数据服务商，发行人与境外本土服务商相比，在语言文字同源、文化同源等方面存在一定劣势，但是发行人在训练数据集丰富度、多语种能力、产品及服务的性价比方面，具备一定的竞争优势。

未来，随着 AI 技术不断革新，应用场景不断增加，各行业、各领域数据安全规范逐渐落地将成为趋势，对于以数据生产为主营业务的数据服务企业，数据安全及合规能力将成为数据供应商新的能力评价维度。发行人一直注重遵守数据安全相关法律法规，在个人数据信息获取和使用过程中始终秉持合规性理念，制定了《数据安全管理制度》《应急响应管理制度》《个人信息出境管理制度》等多项数据合规内控制度，建立了完善的个人信息保护、数据安全管理制度和流程。

此外，公司是行业内为数不多已获得乙级测绘资质的企业，意味着公司具备合规开展智能驾驶采集、标注等业务能力，可帮助公司更好抢占智能驾驶的数据服务市场先机。

2、技术壁垒及迭代情况方面

标准化数据集产品具有可即时获取、价格相对优惠等特点，因此市场需求旺盛。但由于标准化数据集产品是先于客户需求形成，需要数据服务商具备较强的市场趋势前瞻能力以及财务实力进行提前投入，因此行业中具备标准化数据集产品规模化生产能力的服务商较为有限，该方向具备较高竞争壁垒。公司基于近 20 年数据服务经验以及对客户需求的深刻理解，持续研发标准化数据集，自有知识产权的训练数据产品储备近 1,500 个，对于公司而言，标准化训练数据集产品的知识产权由公司享有，具有一次性研发生产、可重复多次销售的特点，可显著提升公司毛利水平，实现训练数据产品的规模化效应，因此也是公司区别于众多竞争对手的一个优势亮点。

公司还拥有较强的综合研发实力，多年来，公司积累了较为完备的综合性、一体化数据处理平台及工具体系，覆盖智能语音、计算机视觉、自然语言等全业态领域，可服务于市面上绝大多数数据处理场景需求，尤其公司于近两年建设的智能驾驶平台目前可覆盖全部舱外场景、主流传感器以及多类前沿的数据处理需求，以上领先的技术平台实力能助力公司适配 AI 行业技术和应用发展的趋势。同时，公司不断优化算法中台中枢能力，通过算法持续赋能语音、图像、文本等各类数据的智能化标注，并实现自动语音识别、光学字符识别、命名实体识别等多类应用场景覆盖，同时重点拓展智能驾驶相关自动预标注能力，截至 2023 年 6 月 30 日，公司已具备 3D 点云分割、鱼眼语义分割、车道线检测、可行驶区域检测等行业前沿算法能力，以上算法的投入，可有效提升公司数据处理效率及降本增效能力，构筑公司数据服务业务的长期护城河。

此外，截至报告期末，公司已经拥有超过 200 个语种/方言的覆盖能力，不仅包括含英、法、德、意、西、日、韩等常见语种，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、中东欧小语种的服务上具备竞争优势。同时，公司已积累超过 120 个语种/方言的发音词典，覆盖希伯来语、乌尔都语、缅甸语、阿姆哈拉语、普什图语、阿尔巴尼亚语、格鲁吉亚语等小语种，

累计词条数超过 1,100 万条,可支撑构建高质量的智能语音以及多模态训练数据,亦构成公司的主要竞争壁垒及核心技术之一。

综上所述,技术驱动并持续迭代技术壁垒已成为公司主营业务运营的核心战略,报告期内公司研发费用总计约 2.3 亿元,2022 年研发费用 0.94 亿元,达到公司历史阶段性高点水平,研发费用率由报告期初的 18.64%增长到 2022 年的 35.86%,在同行业可比公司中处于明显领先水平。

3、数据监管最新政策方面

最近一期,数据出境相关政策的落地实施在短期内对发行人境外业务开展产生了一定影响,公司目前出境数据与境内个人信息关联程度较低,发行人开展境外业务所涉及的数据出境不会对国家安全产生实质影响,随着政策适用范围及实施尺度日渐清晰,逐步度过磨合期,发行人的境外业务预期将逐渐回归正常水平。此外,2023 年 8 月,国家互联网信息办公室等七部门发布实施《生成式人工智能服务管理暂行办法》,针对生成式人工智能产品的预训练数据、优化训练数据来源等的合法性进行了规定,该办法的实施使得训练数据行业尤其是下游客户对于训练数据来源合规、质量等的重视程度进一步加强,进而有利于公司业务的开展。

4、公司对策方面

数据监管最新政策方面,为了降低《评估办法》对公司经营产生的潜在不利影响,发行人采取了以下应对措施:1、公司聘请外部专业顾问团队编制出境风险自评估报告并协助企业办理数据出境安全评估申请事宜,外部顾问团队的核心人员经验丰富,公司将积极与网信部门沟通持续推动出境安全评估申请进程;2、公司正在根据《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》及中共北京市委、北京市人民政府正式印发的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》的要求,积极参与数据要素的跨境流动和业务合作,并充分利用由相关政府部门牵头建设的数字贸易港、数字贸易试验区、数据跨境实验室和数据跨境服务平台等先行先试机制,推进相关工作。上述应对措施有效,《评估办法》对公司境外业务经营产生的潜在不利影响将进一步减弱,公司与境外客户之间保持长期合作关系,未出现主要客户因上述影响而

出现流失的情况。根据发行人二季度业务数据，境外业务已出现回暖态势，公司也将持续加大境外业务团队的建设，促进公司境外业务实现，相关不利影响预计在短期内可以缓解。

收入结构改善方面，除了传统语音数据业务，海天瑞声将紧抓智能驾驶行业发展机遇，全方位布局和提升智能驾驶业务核心能力，同时，公司积极加大大模型领域布局，本次募投亦在拓展 AI 大模型训练数据集产品，完善公司的产品矩阵。截至本回复出具日，公司已与多家客户签订了大模型训练数据集相关的业务合同或意向订单，该领域的业务拓展取得良好开端。此外，公司前募项目中数据库项目已建成完毕，为公司增加了大量的数据产品储备，前期研发投入的效果将会通过后续训练数据产品收入的增长不断显现，随着高毛利收入规模提振，发行人盈利能力将进一步增强。在此之上，公司也将尝试探索现有业务延伸的增值数据服务模式，以更好适应不同类型客户的差异化数据需求。

人工投入及期间费用方面，公司为谋求中长期发展，近年来启动了研发、销售、管理领域的加速投入计划，加大了销售团队的建设投入，新引进中高端管理人员用于加大管理团队人才储备，同时引入较多算法及平台研发人员以持续提高数据处理智能化水平、建设智能驾驶垂直领域技术平台能力，大幅新增投入标准化训练数据集产品的研发投入，提高公司产品模式壁垒。虽然上述情况带来了人工投入及期间费用的上升，导致报告期内业绩有所下滑，但人才储备及研发投入所沉淀下来的优质积累将为公司未来 3-5 年快速扩张提供持续强劲的内在动力，虽然会使短期业绩承压，但将会为公司未来的长期发展、收入扩张、业绩释放奠定坚实的基础，相关不利影响会随着发行人收入和业绩释放得到对冲，发行人 2023 年以来加强了各项费用的预算控制并作为考核标准，以减少不必要的支出，同时持续调整优化管理模式，推动各运营环节降本增效。此外，最近一期内，由于发行人新购置房产折旧与原有房租摊销并行，导致场地相关费用显著提升，也对业绩造成压力，相关不利影响预计在发行人搬迁至新办公场所后得以消除。

整体上，发行人所处行业发展前景广阔，发行人是我国最早从事训练数据研发销售的企业之一，在研发实力、产品储备、合规建设等方面均构建了自身壁垒，市场及技术储备良好，为应对新的数据监管行业政策，发行人积极采取了应对措施，行业政策的实施落地不会对发行人生产经营产生重大不利影响，报告期内发

行人加大研发投入将为公司未来 3-5 年快速扩张提供持续强劲的内在动力，公司的持续经营能力不存在重大不确定性。发行人已在募集说明书及其他申报材料中完善关于业绩下滑的风险提示如下：

“（一）业绩下滑或亏损风险

2022 年度，虽受境外业务复苏以及智能驾驶业务驱动，公司整体营收出现较高增长，但同期公司为更好建设长期技术优势和独具特色的标准化产品模式壁垒持续加大研发投入，且在 2022 年完成管理架构及激励体系升级、全球化营销体系建设等重要事项，以上因素共同导致公司人工投入及期间费用出现明显增长，使得公司净利润出现一定比例下滑。2023 年上半年，发行人营业收入较去年同期下滑 35.13%，扣非前后归母净利润均为负数，主要是由于受到宏观经济波动、市场竞争较为激烈以及数据相关法律法规实施等多重不利因素影响，叠加公司持续加大研发投入并推进薪酬体系改革、新购置房产折旧与原有房租摊销并行等因素，导致最近一期收入及利润承压，上述不利因素目前已有所改善，但如果公司收入增长无法覆盖持续加大的人工投入及期间费用支出，公司业绩存在下滑或亏损的风险。”

二、结合公司主要产品毛利率的波动趋势及原因、同行业可比公司可比产品毛利率变动趋势、产业链上下游的市场变化情况等，说明公司最近三年及一期毛利率下滑尤其是最近一期末大幅下滑的原因及合理性

（一）报告期内公司主要产品毛利率的波动趋势及原因

报告期内，公司主营业务分产品毛利率情况如下：

项目	2023 年 1-6 月		2022 年度	
	毛利率	收入占比	毛利率	收入占比
智能语音	63.62%	63.58%	67.21%	66.30%
计算机视觉	44.94%	25.91%	53.76%	23.04%
自然语言	65.28%	10.47%	72.12%	8.63%
训练数据相关的应用服务	-218.80%	0.04%	76.89%	2.03%
合计	58.85%	-	64.73%	-
项目	2021 年度		2020 年度	
	毛利率	收入占比	毛利率	收入占比

智能语音	65.21%	72.33%	71.00%	77.59%
计算机视觉	55.43%	11.54%	50.45%	11.10%
自然语言	61.34%	12.18%	62.22%	7.05%
训练数据相关的应用服务	75.26%	3.95%	61.62%	4.26%
合计	64.01%	-	67.70%	-

报告期内，公司主营业务综合毛利率分别为 67.70%、64.01%、64.73% 和 58.85%，前三年毛利率水平总体保持稳定，2023 年上半年受毛利水平较高的境外业务以及标准化产品收入占比下降影响，综合毛利率水平有所降低。

从盈利模式来看，由于发行人的收入主要来自定制服务和标准化产品两类，其中标准化产品的毛利率接近 100%，而定制服务的毛利率不同项目间存在一定差异，因此各期收入结构中标准化产品与定制服务的占比结构以及定制服务的毛利率水平变动两因素共同影响整体毛利率水平，具体来说，定制服务占比越高，整体毛利率越低；定制服务毛利率越高，整体毛利率越高。报告期内各主要产品中定制服务的毛利率及收入占比情况如下：

项目	2023 年 1-6 月			2022 年度		
	毛利率	其中：定制服务毛利率	其中：定制服务收入占比	毛利率	其中：定制服务毛利率	其中：定制服务收入占比
智能语音	63.62%	30.78%	52.36%	67.21%	27.50%	45.13%
计算机视觉	44.94%	37.91%	88.55%	53.76%	49.69%	91.83%
自然语言	65.28%	37.61%	55.31%	72.12%	48.02%	53.63%
合计	58.96%	34.06%	62.05%	64.48%	37.63%	56.86%
项目	2021 年度			2020 年度		
	毛利率	其中：定制服务毛利率	其中：定制服务收入占比	毛利率	其中：定制服务毛利率	其中：定制服务收入占比
智能语音	65.21%	26.81%	47.53%	71.00%	37.92%	46.72%
计算机视觉	55.43%	29.05%	62.82%	50.45%	41.18%	84.24%
自然语言	61.34%	36.01%	60.41%	62.22%	42.89%	66.16%
合计	63.55%	28.52%	51.00%	67.97%	38.99%	52.50%

报告期内，各主要产品毛利率有所波动，且最近一期各主要产品线毛利率均有所下滑，主要系定制服务收入占比整体提升、定制服务毛利率下降综合影响导致的，定制服务收入占比下降主要由于宏观经济波动等因素导致的境外收入降低

进而导致标准产品收入占比下滑，定制服务毛利率下降主要由于市场竞争激烈导致价格下降以及境外高毛利收入占比下降的影响，具体如下：

报告期内，收入占比最高的智能语音业务整体毛利率在 2021 年、2023 年上半年出现下滑。2021 年，智能语音业务毛利率由 71.00%降低到 65.21%，主要由于受境外公共卫生事件下居家办公等因素的影响，智能语音业务中境外部分收入贡献占比有所下降，Google、Amazon 等美国客户以及 Samsung 等日韩客户当年订单量均呈现下滑态势，由此导致高毛利的定制服务占比相应减少，进而毛利率有所下降。2023 年上半年，智能语音业务毛利率由 67.21%降低到 63.62%，毛利率下滑主要受 100%毛利率的标准化产品收入占比降低（由 54.87%降低至 47.64%）的影响较大，主要原因为受宏观经济波动影响，部分客户进行业务方向调整及阶段性裁员等，导致标准产品的采购预算释放进度有所放缓。

报告期内，计算机视觉业务在前三年的毛利率维持在 50%以上，最近一期下滑至 44.94%，主要由于定制服务收入占比较高且毛利率下滑较大（由 2022 年度的 49.69%下降至 37.91%）所致。公司通过前期较长时间的客户需求挖掘和客户关系维护，2022 年完成了较多境外多语种手写体及 OCR 项目，同时与部分智能驾驶客户形成战略合作，毛利率水平较高；2023 年上半年由于部分高毛利境外项目进程有所放缓，同时为应对智能驾驶行业竞争日益激烈，公司采取报价优惠的策略并进一步开拓中小型客户的业务需求，因此毛利率有所下降。

报告期内，自然语言业务的毛利率分别为 62.22%、61.34%、72.12%、65.28%，2022 年度毛利率增幅较大主要由于定制服务毛利率由 2021 年度的 36.01%增长至 48.02%，该增长主要系 2022 年完成了较多高毛利的多语种文本采集项目，拉高了整体毛利率水平。

报告期内，训练数据相关的应用服务毛利占比不足 5%且逐年降低，毛利率水平受单一项目影响较大。2023 年上半年的毛利率为负值，主要系该业务整体收入较低，受单一项目亏损影响较大所致。

（二）同行业可比公司可比产品毛利率变动趋势

由于公司主要从事训练数据的研发设计、生产及销售业务，所在细分领域为人工智能基础数据服务领域，细分领域和业务较新，目前尚不存在业务可比或相

似程度较高的同类型 A 股上市公司。发行人行业内主要企业包括 Appen、数据堂、标贝科技等，受限于可比公司数据可获取性，仅澳大利亚上市公司 Appen、新三板挂牌公司数据堂（831428）公开披露了财务数据，毛利率对比分析如下：

公司名称	综合毛利率			
	2023 年 1-6 月	2022 年度	2021 年度	2020 年度
Appen	40.29%	38.76%	39.98%	42.78%
数据堂	44.52%	45.75%	43.24%	41.48%
发行人	58.85%	64.73%	64.01%	67.70%

注 1：标贝科技未公开财务数据；

注 2：Appen 公司的毛利率根据定期报告中“Services revenue”和“Crowd labelling services”科目计算得出，即毛利率=（Services revenue - Crowd labelling services）/Services revenue。

报告期内，发行人毛利率整体高于 Appen，主要为收入结构差异所致，发行人的收入构成中，高毛利率（接近 100%）的训练数据产品业务占比高于 Appen；Appen 收入构成中，低毛利率的文本相关度标注业务占比高于发行人，发行人该类业务毛利率水平仅为 20%-30%，因此造成综合毛利率存在一定差异，具备合理性。变动趋势上，报告期内前三年 Appen 的毛利率逐年下滑，其 2021 年主营业务毛利率下降 2.80 个百分点，与发行人毛利率下降幅度 3.69 个百分点基本趋势一致；2023 年 1-6 月，Appen 毛利率企稳，但仍低于发行人同期毛利率水平。

报告期内，发行人毛利率整体高于数据堂，主要因为一方面发行人拥有的标准训练数据集数量、产品覆盖的语种/方言数量均显著多于数据堂，由于成品训练数据集的销售毛利接近 100%，小语种、稀有语种的项目毛利也显著高于一般常规语种，因此发行人整体的毛利率会更高；另一方面，数据堂境外业务占比为 20%-30%，低于发行人，由于境外业务毛利一般均高于境内业务，因此导致数据堂的综合毛利率偏低，具备合理性。变动趋势上，数据堂的毛利率在前三年呈现逐年上升趋势，与发行人存在差异，主要系由于数据堂前期规模较小且产品线相对集中，报告期内随着收入规模快速增加，毛利率相应提升，最近一期，数据堂毛利率有所下降，较去年同期降低 2.50 个百分点，发行人较去年同期下降 7.53 个百分点，数据堂与发行人波动趋势一致。

整体上，结合公司主要产品毛利率的波动、同行业可比公司毛利率、产业链上下游的市场变化情况等进行分析，发行人报告期内的毛利率波动具备合理性。

三、结合同行业可比公司各项期间费用情况及变动趋势，说明公司期间费用规模的合理性；最近三年及一期公司管理费用及研发费用大幅增长的原因及合理性；研发费用的明细构成，结合研发人员的具体情况，包括但不限于研发人员数量、学历、人均工资、所从事的研发项目情况，说明研发费用中职工薪酬各期大幅上涨的原因及合理性，研发费用中数据服务费的具体用途，与主营业务的数据服务费是否存在混同情况，数据服务费归集与核算的具体方法，研发费用中数据服务费归集与核算的准确性

（一）结合同行业可比公司各项期间费用情况及变动趋势，说明公司期间费用规模的合理性

1、期间费用整体情况分析

报告期内，发行人主要期间费用构成为销售费用、管理费用、研发费用，三项费用合计占营业收入比重为 34.41%、53.40%、66.20%、104.48%，报告期内 8 家 A 股可比公司的三项费用合计占比平均数分别为 48.92%、52.52%、51.96%、58.00%，发行人报告期内前三年费用率整体规模位于可比公司合理区间内，增长幅度大于可比公司平均水平，主要由于可比公司的首发上市年度分布于 2001 年至 2018 年期间，且整体体量较大，公司 2021 年为上市当年，报告期内整体处于规模扩张期和加速投入期，在人员储备及技术投入上需要跟上公司制定的整体发展战略，因此费用率水平增速较快，2023 年 1-6 月，发行人三项费用率超过 100%，主要系当期营业收入规模偏小，同时各项费用支出相对刚性所致。

证券简称	2023 年 1-6 月	2022 年度	2021 年度	2020 年度
荣科科技	38.27%	40.04%	38.84%	29.29%
天玑科技	50.13%	29.75%	32.50%	28.82%
四方精创	23.46%	30.20%	30.34%	34.88%
彩讯股份	23.49%	25.86%	24.63%	25.87%
上述 4 家信息技术服务类可比公司平均数	33.83%	31.46%	31.58%	29.72%
恒生电子	70.28%	58.96%	61.47%	57.26%
用友网络	77.22%	55.14%	53.77%	46.39%
万兴科技	88.64%	90.41%	92.99%	82.08%
泛微网络	92.50%	85.33%	85.61%	86.79%

上述 4 家软件产品类可比公司平均数	82.16%	72.46%	73.46%	68.13%
可比公司平均数	58.00%	51.96%	52.52%	48.92%
发行人	104.48%	66.20%	53.40%	34.41%

2、销售费用

报告期内，发行人与可比公司销售费用率变动情况如下表所示：

证券简称	2023 年 1-6 月	2022 年	2021 年	2020 年
荣科科技	14.39%	13.97%	14.07%	9.78%
天玑科技	12.86%	7.68%	7.70%	5.69%
四方精创	1.87%	1.84%	1.60%	1.58%
彩讯股份	3.01%	3.71%	3.79%	3.95%
上述 4 家信息技术服务类可比公司平均数	8.03%	6.80%	6.79%	5.25%
恒生电子	14.69%	9.49%	10.15%	8.48%
用友网络	32.06%	24.14%	22.70%	18.02%
万兴科技	50.18%	49.15%	46.87%	45.14%
泛微网络	68.76%	67.81%	68.90%	67.97%
上述 4 家软件产品类可比公司平均数	41.42%	37.65%	37.16%	34.91%
可比公司平均数	24.73%	22.22%	21.97%	20.08%
发行人	16.29%	6.98%	7.03%	4.77%

报告期前三年，发行人销售费用的变动趋势与同行业基本一致，不存在明显差异，其中发行人 2021 年销售费用率较 2020 年增长 2.26 个百分点，同行业公司平均值提升 1.89 个百分点。最近一期，发行人销售费用率增加较快，主要系上半年收入规模较小的同时为配合整体战略发展及业务拓展目标，发行人在相关营销支出及团队建设方面加大投入所致。

报告期内，发行人的销售费用率水平低于可比公司平均值，而可比公司中信息技术服务类公司和软件产品类公司的销售费用率水平呈现明显差异。四家信息技术服务类公司销售费用率平均在 5%至 9%之间，而四家软件产品类公司销售费用率平均在 34%至 42%之间，公司销售费用率水平与信息技术服务类公司较为可比，低于软件产品类公司的销售费用率，主要系发行人销售模式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

用所致。

3、管理费用

报告期内，发行人与可比公司管理费用率变动情况如下表所示：

证券简称	2023年1-6月	2022年	2021年	2020年
荣科科技	15.36%	14.22%	14.36%	11.85%
天玑科技	24.32%	13.34%	15.70%	13.13%
四方精创	12.09%	14.18%	15.54%	19.42%
彩讯股份	4.09%	4.74%	5.39%	6.30%
上述4家信息技术服务类可比公司平均数	13.97%	11.62%	12.75%	12.67%
恒生电子	14.16%	13.38%	12.40%	12.93%
用友网络	15.11%	12.07%	12.00%	11.26%
万兴科技	11.67%	11.79%	14.23%	14.55%
泛微网络	4.38%	2.94%	4.55%	5.38%
上述4家软件产品类可比公司平均数	11.33%	10.05%	10.79%	11.03%
可比公司平均数	12.65%	10.83%	11.77%	11.85%
发行人	44.92%	23.36%	17.06%	11.00%

报告期内，发行人2020年的管理费用率与可比公司平均值基本一致，2021年以来，发行人管理费用率增速快于可比公司平均水平，2021年发行人管理费用率同比提高6.06个百分点，主要系当年发放一次性上市专项奖金导致职工薪酬增长较快，扣除专项奖金影响，发行人管理费用率为14.84%，略高于可比公司平均水平，但位于合理区间。2022年，发行人管理费用率进一步提升6.30个百分点至23.36%，主要系当年因特定事项产生部分中介服务费支出，包括购置房产相关的中介服务费、战略管理咨询及行业分析咨询等服务费，扣除该项影响后，发行人管理费用率为20.98%，相较2021年增长3.92个百分点，主要系一方面当年公司加大人员储备、升级管理架构及薪酬体系、实施限制性股票激励计划导致职工薪酬整体上涨较多，另一方面公司增加临时性场地租赁面积带来一定的房租增长。最近一期，发行人管理费用率增长21.56个百分点，主要系发行人上半年收入规模较小，同时管理费用中职工薪酬、折旧摊销金额相对刚性导致的。

4、研发费用

报告期内，发行人与可比公司研发费用率变动情况如下表所示：

证券简称	2023年1-6月	2022年	2021年	2020年
荣科科技	8.52%	11.84%	10.41%	7.66%
天玑科技	12.94%	8.74%	9.10%	10.01%
四方精创	9.49%	14.19%	13.21%	13.88%
彩讯股份	16.39%	17.40%	15.44%	15.63%
上述4家信息技术服务类可比公司平均数	11.83%	13.04%	12.04%	11.79%
恒生电子	41.42%	36.08%	38.92%	35.85%
用友网络	30.05%	18.94%	19.07%	17.11%
万兴科技	26.80%	29.47%	31.88%	22.39%
泛微网络	19.36%	14.58%	12.16%	13.44%
上述4家软件产品类可比公司平均数	29.41%	24.77%	25.51%	22.20%
可比公司平均数	20.62%	18.91%	18.77%	17.00%
发行人	43.26%	35.86%	29.31%	18.64%

报告期内，发行人研发费用率整体高于可比公司水平，主要由于一方面发行人作为科创板上市公司，研发投入支出相对较高；另一方面与发行人持续进行标准化产品研发投入的商业模式相关，由于标准化产品在实现销售前的支出均计入研发费用，也会导致公司研发费用率整体偏高。

2021年，发行人研发费用率为29.31%，同比增长10.67个百分点，同行业可比公司研发费用率平均水平为18.77%，同比增长1.78个百分点，公司研发费用率增速高于行业平均水平，主要系当年正值公司启动研发加速投入计划当年，研发费用在绝对金额低于同行业平均水平的基础上同比增长较快，符合公司业务定位和发展期间的特征，具备合理性；2022年，发行人研发费用率进一步提升6.55个百分点，可比公司平均水平变化不大，主要受到发行人当年加速前募建设并在智能驾驶领域及整体数据智能化水平提升方面持续加大投入的影响，研发费用增长较快；2023年1-6月，发行人研发费用率进一步提升，主要受到当期收入规模较小的影响。

（二）最近三年及一期公司管理费用及研发费用大幅增长的原因及合理性

1、最近三年一期管理费用大幅增长的原因及合理性

报告期内，公司管理费用构成情况如下：

单位：万元

项目	2023年1-6月		2022年度		2021年度		2020年度	
	金额	占比	金额	占比	金额	占比	金额	占比
职工薪酬	2,067.16	61.80%	3,692.63	60.14%	2,391.89	67.92%	1,471.80	57.34%
中介服务费	302.54	9.04%	1,118.94	18.22%	229.02	6.50%	131.23	5.11%
租赁费	274.88	8.22%	515.83	8.40%	49.76	1.41%	600.96	23.41%
折旧与摊销	564.42	16.87%	346.01	5.63%	124.77	3.54%	135.73	5.29%
使用权资产折旧	-	-	161.56	2.63%	436.16	12.38%	-	-
业务招待费	32.75	0.98%	89.00	1.45%	95.61	2.71%	45.71	1.78%
水电费	21.28	0.64%	57.55	0.94%	49.92	1.42%	42.30	1.65%
办公费	8.42	0.25%	41.89	0.68%	34.54	0.98%	26.15	1.02%
差旅费	10.38	0.31%	11.28	0.18%	21.45	0.61%	13.42	0.52%
其他	63.25	1.89%	105.78	1.72%	88.69	2.52%	99.44	3.87%
合计	3,345.07	100.00%	6,140.46	100.00%	3,521.80	100.00%	2,566.74	100.00%

报告期各期，公司管理费用分别为 2,566.74 万元、3,521.80 万元、6,140.46 万元及 3,345.07 万元，占营业收入比例分别为 11.00%、17.06%、23.36% 及 44.92%。主要构成为职工薪酬、中介服务费、租赁费、折旧与摊销、使用权资产折旧，是管理费用波动的主要因素。

（1）职工薪酬

报告期内，职工薪酬主要由管理人员薪酬、股份支付、劳务费构成，结合报告期内人数变动情况分析如下：

单位：万元

项目	2023年1-6月	2022年度	2021年度	2020年度
管理人员薪酬	1,825.52	3,202.68	2,245.89	1,357.39
其中：一次性上市专项奖金	-	-	456.95	-
股份支付	184.69	293.06	-	-
劳务费	56.95	196.89	146.00	114.41

职工薪酬合计	2,067.16	3,692.63	2,391.89	1,471.80
管理人员薪酬(扣除一次性上市专项奖金)	1,825.52	3,202.68	1,788.94	1,357.39
员工人数(人)	81.71	72.53	42.83	40.32
人均薪酬(万元/人)	22.34	44.16	41.77	33.67

注 1:员工人数,以当年计入管理费用核算薪酬的员工按照各年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,发行人为配合整体战略发展及业务拓展目标,公司加大人员储备、升级管理架构及激励体系,导致员工人数及人均薪酬均呈现上涨态势,2022 年员工人数较 2021 年增长较多主要系当年设立集中采购部门用于加强上游供应链的建设和优化,相应增加了 20 余人团队,同时公司为进一步加强上市后的规范运作水平、响应国家日益加强的数据领域安全法规要求及各级政府对数据行业的重视,公司证券、财务、法务、政府事务等职能部门也均有人员增加。除了以上因素,2021 年职工薪酬较 2020 年增长较多,一方面受到 2020 年社保减免在 2021 年解除的影响,2021 年公司社保费用较上年度增加 160.20 万元;另一方面因发行人 2021 年上市成功发放一次性专项奖金导致当期职工薪酬增加 456.95 万元。

(2) 中介服务费

2021 年公司中介服务费较上年度增加 97.79 万元,主要为发行人 2021 年成功上市所发生的财经公关咨询服务费以及上市过程中的宣传费用等增加所导致。2022 年,公司中介服务费较上年度增加 889.92 万元,主要系当年因购置新办公楼产生了偶发性的中介服务费用,此外发行人还采购了战略管理咨询服务及行业分析咨询服务,上市后也产生了年度审计费等,均导致中介服务费有所增长。

(3) 租赁费及使用权资产折旧

2022 年,公司租赁费及使用权资产折旧合计较上年度增加 191.47 万元,主要为 2022 年公司为应对人员增长,增加了临时性场地租赁面积带来的房租增长。

(4) 折旧与摊销

2022 年起,折旧与摊销金额较上年大幅增长,主要系发行人于 2022 年购置了新办公楼所致。

综上所述,发行人报告期内管理费用大幅增长具备合理性。

2、最近三年一期研发费用大幅增长的原因及合理性

报告期内，公司研发费用构成情况如下：

单位：万元

项目	2023年1-6月		2022年度		2021年度		2020年度	
	金额	占比	金额	占比	金额	占比	金额	占比
职工薪酬	1,920.42	59.61%	4,829.93	51.23%	2,724.94	45.02%	1,989.22	45.73%
数据服务费	709.64	22.03%	3,611.74	38.31%	2,459.85	40.64%	1,844.91	42.41%
房租费	84.05	2.61%	236.06	2.50%	191.70	3.17%	86.63	1.99%
折旧与摊销	208.03	6.46%	224.10	2.38%	159.76	2.64%	139.19	3.20%
语言研究	99.81	3.10%	217.93	2.31%	290.03	4.79%	144.78	3.33%
其他	199.42	6.19%	307.57	3.26%	225.82	3.73%	145.02	3.33%
合计	3,221.36	100.00%	9,427.34	100.00%	6,052.09	100.00%	4,349.75	100.00%

报告期各期，公司研发费用分别为 4,349.75 万元、6,052.09 万元、9,427.34 万元及 3,221.36 万元，其中职工薪酬、数据服务费合计占比在 80% 以上，是导致研发费用波动的主要因素。

（1）职工薪酬

报告期内，研发费用中职工薪酬大幅增长的原因请参见本题回复之“三/（三）/2、研发费用中职工薪酬各期大幅上涨的原因及合理性”部分。

（2）数据服务费

数据服务费支出主要对应数据集产品的建设，报告期内，发行人持续加大标准训练数据集产品研发，相应的数据服务费也稳步增长，2022 年数据服务费较上年增长 1,151.89 万元，一方面系由于前次募集资金项目规划的“自主研发数据产品扩建项目”于 2022 年加速推进并于 2023 年 5 月结项，完成了训练数据集产品的扩建目标。报告期内，发行人当期新增的数据集产品个数分别为 143 个、170 个、402 个及 164 个，当期在研的数据集产品个数分别为 203 个、353 个、533 个及 397 个，与数据服务费整体增长趋势相匹配，当期在研的数据集产品具体情况参见本题回复之“三/（三）/2、研发费用中职工薪酬各期大幅上涨的原因及合理性/（2）训练数据产品相关的职工薪酬合理性分析”部分。

（三）研发费用的明细构成，结合研发人员的具体情况，包括但不限于研

发人员数量、学历、人均工资、所从事的研发项目情况，说明研发费用中职工薪酬各期大幅上涨的原因及合理性，研发费用中数据服务费的具体用途，与主营业务的数据服务费是否存在混同情况，数据服务费归集与核算的具体方法，研发费用中数据服务费归集与核算的准确性

1、研发费用的明细构成情况

报告期内，公司研发费用构成情况见本题回复之“三/（二）/2、最近三年一期研发费用大幅增长的原因及合理性”部分。

2、研发费用中职工薪酬各期大幅上涨的原因及合理性

报告期内，为支撑公司技术、平台及工具迭代和拓展的研发需要，以及持续增长的训练数据产品的开发需求，发行人持续增加和完善研发和技术人员数量和结构，研发相关的人员数量、人均薪酬均呈现增长趋势，其中人员数量的增长较为显著。

项目	2023年1-6月	2022年度	2021年度	2020年度
职工薪酬（扣除股份支付）（万元）	1,855.68	4,684.18	2,724.94	1,989.22
平均人数（人）	103.03	113.22	68.96	51.59
平均薪酬（万元/人）	18.01	41.37	39.51	38.54

报告期内，发行人的研发费用主要由基础研发支出和训练数据产品开发支出两部分构成，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，主要归集研发中心、语言研究中心等部门的人员工资；训练数据产品开发支出主要为开发训练数据产品相关的支出，主要归集数据业务中心人员按照工时积分制填报并换算的人工薪酬，两者具体分析如下：

(1) 基础研发相关的职工薪酬的合理性分析

项目		2023年1-6月	2022年度	2021年度	2020年度
职工薪酬（万元）		1,619.17	3,927.17	2,232.49	1,644.78
平均人数（人）		79.72	87.81	51.90	38.79
平均薪酬（万元/人）		20.31	44.72	43.02	42.40
学 历 构 成 (人)	硕士研究生及以上学历	24	24	23	18
	本科学历	47	53	50	24
	专科及以下学历	6	5	4	2
在研项目个数		16	25	26	3
研 发 项 目 情 况	在研项目研发重点及技术亮点	<p>(1) 面向大模型的预训练、调优及评测技术研究：围绕公司大模型相关业务能力升级，新增基础大模型预训练技术研究、面向大模型的海量文本获取与清洗技术研究、大模型微调技术以及大模型评测技术等多个课题；重点完成中文大模型数据库构建技术可行性分析，并启动多个大模型数据库建设任务。</p> <p>(2) 语音数据库云-端协同管理平台建设：重点支持语音识别采集数据库的生产，梳</p>	<p>(1) 智能驾驶平台重大升级：配合公司在智能驾驶数据处理方向的战略，充分挖掘智能驾驶数据生产的行业特性，全面升级面向智能驾驶数据处理场景的一体化平台，产品重点建设工作包括：管理模块全面升级，优化数据流转方式、项目管理模块、统计分析模块等；2D场景全面升级，有效支撑包括车道线、障碍物、可行驶区域等7大类智能驾驶场景的数据生产；3D场景全面升级，针对3D点云数据标注需求，引入</p>	<p>(1) 多模态训练数据生产平台：支持包括声纹识别、动作捕捉、视频动态标注等覆盖语音、图像、文本等多种模态数据处理能力，赋能多模态数据库的生产能力；同时，完成平台前后端分离、架构升级等工作，提升相关数据管理和项目管理能力。</p> <p>(2) 语音识别技术升级：进一步扩大算法团队，新增基于端到端的语音识别技术课题，引入 WeNet 和 ESPNet 两大主流端到端语音识别算法框架，新增数十个语音识</p>	<p>(1) OCR 在线生产平台升级：针对 OCR 光学字符识别所需数据库的生产特点，全面优化了 OCR 在线生产平台的数据采集、标注和质检等功能，有效提升生产效率和数据库质量。</p> <p>(2) 语料标注在线平台：在通用文本处理平台的基础上，结合语音数据库中语料设计环节的具体要求，开发了语料在线生产平台，支撑语料编辑、处理和质检等工作，提升语料生产效率。</p>

项目	2023年1-6月	2022年度	2021年度	2020年度
	<p>理并优化既有能力，全面提升移动端数据采集和处理能力、并优化服务端数据管理和任务管理能力。</p> <p>(3) 智能驾驶平台重大升级：重点支撑百万级别大点云数据处理能力，全面优化激光点云下 4D 标注场景和生产效率。</p>	<p>前端高级开发人员，攻克 3D 点云对象分割等技术难题，完成核心功能开发和上线，支撑众多 3D 点云标注任务。</p> <p>(2) 多语种语料设计平台：在原有语料标注在线平台基础上，完成对多语种的扩展支持；在完成通用语料设计和语料标注模块优化的同时，针对特定语种、特定设计需求（如多感情语料设计）新增平台支撑能力。</p>	<p>别模型，全面提升公司的语音识别能力，助力智能语音数据库的数据质量提升和生产效率提升。</p>	

注：平均人员数量为按照人均在岗月份进行加权平均，学历构成数据为结合各期末研发人员情况统计。

报告期内，基础研发相关职工薪酬占整体比例较高且报告期内增速较快，一方面系由于平均薪酬稳步增长，另一方面主要系研发人数扩充较快导致的。2021年较2020年在研项目增长了23个，基础研发平均人数增长33.80%，为了配合公司在2021年制定的进一步在多模态数据处理及智能驾驶等行业应用数据处理方面的战略方向，2021年发行人在多模态数据处理智能化、智能驾驶数据业务技术平台能力建设等方向规划研发工作，针对性地新增了智能语音、计算机视觉领域的算法研发人员、高级软件开发人员，研发出基于WeNet和ESPNet的端到端语音识别算法平台、无人车场景下物体识别能力、动作捕捉数据生产工具等，提升了数据智能化处理水平和多模态数据处理能力。2022年，发行人持续、大幅扩充智能驾驶领域及算法领域相关研发团队并建立独立的智能驾驶事业部，着重在数据处理一体化平台升级、2D和3D场景数据处理能力优化等方向规划研发工作，新增了具有多年行业背景的自动驾驶产品经理、高级前端开发工程师等岗位的专业人才，研发出了智能驾驶数据管理模块、新一代3D点云分割工具，并优化了多类2D场景数据处理技术，全面提升了公司在智能驾驶领域的数据生产能力。截至2021年12月31日，研发人员人数增长至77人，而2021年全年平均人

数受 2021 年底新增人员影响有限，因此 2022 年平均人数增幅较大达到 69.19%。学历构成上，报告期各期末本科以上学历人员占比始终在 90% 以上，波动不大。整体上，发行人报告期内基础研发中职工薪酬费用增长主要由于发行人为落实各期在研项目推进而进行研发团队扩充而导致的，具备合理性。

(2) 训练数据产品相关的职工薪酬合理性分析

项目		2023年1-6月	2022年度	2021年度	2020年度
职工薪酬（万元）		236.51	757.01	492.45	344.44
平均人数（人）		23.31	25.41	17.06	12.80
平均薪酬（万元/人）		10.15	29.79	28.87	26.91
研发项目情况	当期新增训练数据产品库数量（个）	164	402	170	143
	当期在研训练数据产品库数量（个）	397	533	353	203
	产品库研发重点及应用场景	重口音及方言数据库，用于方言识别场景； 多模态（语音、图像）数据集，用于虚拟人场景； 自动驾驶 OMS座舱数据集，用于自动驾驶领域乘客行为识别场景。	多语种多情感语音合成数据集，用于数字娱乐等场景； 多语言对话语音数据集，用于智能客服、智能会议等场景； 人脸及超市卖场场景数据集，用于零售场景下的人物识别与检测。	单人自述式多语言语音识别数据集，用于语音输入、语音助手等场景； 文学作品、音乐等领域多语言多情感语音合成数据集，用于虚拟主播、数字娱乐等场景； 人体头部三维模型数据库，用于三维人像验证场景。	多国语种语音合成数据集，用于智能音箱、智能客服等场景； 多国身份认证图片数据集，用于多场景外国人身份验证。

报告期内，发行人训练数据产品库相关人员平均薪酬波动幅度不大，2023年1-6月平均薪酬偏低主要系未包含年终奖影响。报告期内职工薪酬增速较快主要受到服务人数增长的影响，人数增长趋势与当期新增的产品库数量及在研产品库数量变动趋势基本一致，2021年平均人数较上年增长5人左右，主要为配合下游市场智能硬件领域拓展计划，发行人加大对语音输入、语音助手、虚拟主播、数字娱乐、三维人像验证等场景的研发投入，新增单人自述式多语言语音识别数据集、文学作品、音乐等领域多语言多情感语音合成数据集、人体头部三维模型数据库等数据集产品。2022年平均人数较上年增长8人左右，主要为配合公司在数字娱乐、智能客服、智能会议、零售场景下的人物识别与检测等应用场景的建设，新增多语种多情感语音合成数据集、多语言对话语音数据集、人脸及超市卖场场景数据集等。同时，2022年发行人加速建设前次募集资金项目，也导致

训练数据产品库研发人员投入较多。

3、研发费用中数据服务费的具体用途，与主营业务的数据服务费是否存在混同情况，数据服务费归集与核算的具体方法，研发费用中数据服务费归集与核算的准确性

报告期内，研发费用中数据服务费主要对应训练数据产品开发所需的原料数据采集与标注服务采购支出，而计入成本的数据服务费主要对应训练数据定制服务的相关数据采集与标注服务采购支出，二者不存在混同情况。

发行人严格按照项目为单位进行研发费用以及成本的归集，无论是定制服务项目还是研发项目，每个项目立项之初，即需在业务管理平台中选择项目类型，每个项目只能选择一个类型，选择训练数据定制服务的项目，支出归集入“制造费用”，并按上述原则结转计入“主营业务成本”；选择训练数据产品开发的项目，支出归集入“研发费用”。因此不会出现训练数据定制服务和训练数据产品开发共享一个项目的情形。在项目执行之前，每个项目需要编制项目预算，根据项目的具体内容，对项目所需项目管理人员及每人投入工时等直接人工内容，采集人数、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。

在项目预算环节，一般的项目预算为三级审批制，重大项目为五级审批制。具体为所有项目的预算均须业务部门经理、分管业务副总经理及财务审核岗三级审批，若项目预算金额或项目对应合同/订单金额重大的，或预算毛利率异常的，则需要追加财务经理四级审批，甚至追加财务总监五级审批。对于项目预算中数据服务费部分，采集数据量会参照“以销定采”的业务逻辑，根据定制服务销售合同/订单约定的交付数据量或产品研发计划制定的产品数据量，并考虑合理的采集损耗率后确定；标注量（即标注有效工时）根据采集数据量并参照合理的标产比（即标注产出比，指完成单位数据量的标注所需要的时间）后确定，标产比根据项目启动前的自测标产比或与客户约定的标产比确定；采集单价和标注单价参照公司采购价格体系确定；项目负责人编制预算时需要将数据服务费的采购量、采购单价均按工作环节逐项、单独填报于业务管理平台，经过业务部门、财务部等多层级审批通过后才可执行。

在项目实际支出环节，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，并设定相应的审批流程，与项目预算及采购单进行核对，保证发生额完整、准确归集到对应项目上，不得与其他项目存在交叉。

综上所述，公司针对训练数据定制服务及训练数据集产品开发的支出进行有效归集，在立项、预算、结算各个环节内，支出量、价与业务数据有效勾稽、验证，多部门交叉复核及监督，形成有效的内部控制机制，保障主营业务成本及研发费用核算的真实、准确、完整。报告期内，发行人不存在将归属于成本的数据服务费计入研发费用的情形，研发费用中数据服务费归集与核算准确。

四、报告期内应收账款账面价值大幅上涨、应收账款周转率大幅下滑的原因及合理性，相关不利因素是否已经消除或改善；应收账款的账龄情况，一年以上长账龄的应收账款占比增长的原因，相关客户的经营情况、回款及逾期情况、相应坏账准备计提情况；结合公司应收账款整体的回款及逾期情况、应收账款坏账准备计提政策、信用政策、同行业可比公司情况等说明公司应收账款坏账准备计提的充分性。

（一）报告期内应收账款账面价值大幅上涨、应收账款周转率大幅下滑的原因及合理性，相关不利因素是否已经消除或改善

报告期内，公司应收账款周转率分别为 4.11 次、2.51 次、2.53 次、1.40 次（年化），总体呈下降趋势。2021 年末公司应收账款周转率相较上年末显著下降，主要系境外收入下滑，以及 2021 年公司调整销售人员业绩考核方案，导致催收客户回款力度较上年末有所下降。公司针对此情况，于 2022 年将回款指标重新纳入销售人员的核心考核指标中，促使 2022 年应收账款周转率指标有所好转。2023 年上半年，受宏观经济波动、市场竞争较为激烈以及数据相关法律法规实施等多重因素影响，公司半年度营业收入同比下滑，从而导致年化周转率有所降低。

报告期内，发行人同行业可比公司应收账款周转率情况如下：

证券代码	证券简称	2023 年 1-6 月	2022 年度	2021 年度	2020 年度
------	------	--------------	---------	---------	---------

证券代码	证券简称	2023年1-6月	2022年度	2021年度	2020年度
300290.SZ	荣科科技	1.10	1.19	1.33	1.25
300245.SZ	天玑科技	1.52	3.81	4.57	3.56
300468.SZ	四方精创	2.91	3.01	2.83	2.87
300634.SZ	彩讯股份	5.40	5.93	5.68	3.28
上述4家信息技术服务类可比公司平均数		2.73	3.49	3.60	2.74
上述4家信息技术服务类可比公司中位数		2.22	3.41	3.70	3.08
600570.SH	恒生电子	4.41	6.17	6.70	7.36
600588.SH	用友网络	2.18	3.61	4.74	4.87
300624.SZ	万兴科技	63.64	29.47	35.09	42.41
603039.SH	泛微网络	7.12	10.11	12.77	12.04
上述4家软件产品类可比公司平均数		19.34	12.34	14.83	16.67
上述4家软件产品类可比公司中位数		5.77	8.14	9.74	9.70
可比公司平均数		11.04	7.91	9.21	9.71
可比公司中位数		3.66	4.87	5.21	4.21
海天瑞声		1.40	2.53	2.51	4.11

注：应收账款周转率=营业收入/应收账款平均余额，2023年6月末数据为简单年化。

发行人应收账款周转率水平位于行业合理区间内，与软件类可比公司用友网络、信息技术服务类可比公司中天玑科技等较为接近。应收账款周转率主要与收入结构、客户群体、信用政策、公司发展阶段等多方面因素相关，发行人主要客户均为国内外知名科技企业及大型科研机构，内部付款审批流程相对较长，一定程度上导致应收账款周转率指标低于行业平均值，具备合理性。如本题回复之“一/（二）/3、2023年是否面临收入大幅下滑的风险”部分所述，随着相关不利因素逐步缓解，发行人全年收入预计不会出现同比大幅下滑，导致周转率下滑的因素将逐步减弱。

（二）应收账款的账龄情况，一年以上长账龄的应收账款占比增长的原因，相关客户的经营情况、回款及逾期情况、相应坏账准备计提情况

1、应收账款的账龄情况，一年以上长账龄的应收账款占比及增长原因

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

账龄	2023年6月30日		2022年12月31日	
	金额	比例	金额	比例
1年以内（含1年）	8,586.58	84.56%	9,520.84	86.04%
1年以上	1,567.50	15.44%	1,544.98	13.96%
其中：1-2年（含2年）	750.29	7.39%	1,090.51	9.85%
2-3年（含3年）	704.13	6.93%	297.57	2.69%
3年以上	113.08	1.11%	156.90	1.42%
合计	10,154.08	100.00%	11,065.82	100.00%
账龄	2021年12月31日		2020年12月31日	
	金额	比例	金额	比例
1年以内（含1年）	8,902.48	91.53%	6,482.77	96.75%
1年以上	824.29	8.47%	217.86	3.25%
其中：1-2年（含2年）	667.39	6.86%	60.96	0.91%
2-3年（含3年）	-	-	100.00	1.49%
3年以上	156.90	1.61%	56.90	0.85%
合计	9,726.78	100.00%	6,700.63	100.00%

报告期内,发行人账龄1年以上应收账款期末余额分别为217.86万元、824.29万元、1,544.98万元及1,567.50万元,占应收账款期末余额比例约分别为3.25%、8.47%、13.96%及15.44%。账龄1年以上的应收账款占比持续提升,主要系一方面,部分大型客户付款流程较长、同时少数客户受公共卫生事件影响或者因自身经营原因未按合同约定回款;另一方面,发行人2021年阶段性调整了销售人员考核方案,导致客户回款有所放慢,2022年发行人已调整销售人员考核方案并加强回款催收,应收账款增速有所放缓。

2、相关客户的经营情况、回款及逾期情况、相应坏账准备计提情况

报告期各期末,公司应收账款账龄1年以上的前五名客户情况如下:

单位：万元

2023年6月30日										
客户名称	品牌	应收账款 期末余额	其中：账龄 1年以上	坏账 准备	逾期情况				期后 回款	期后回 款比例
					未逾期	逾期1年内	逾期1-2年	逾期2年以上		
普强时代（珠海横琴）信息技术有限公司	普强	225.41	225.41	112.71	-	-	225.41	-	10.00	4.44%
北京中核久安科技有限公司	中核久安	214.56	214.56	107.28	-	-	214.56	-	-	-
阿里云计算有限公司	阿里巴巴	155.35	143.85	124.69	11.51	-	39.00	104.85	11.51	7.41%
境外客户J	境外品牌E	122.84	114.09	14.20	6.02	116.82	-	-	-	-
科大讯飞股份有限公司	科大讯飞	157.30	94.42	43.06	26.45	60.94	69.90	-	3.80	2.42%
合计		875.46	792.33	401.94	43.98	177.76	548.87	104.85	25.31	2.89%
2022年12月31日										
客户名称	品牌	应收账款 期末余额	其中：账龄 1年以上	坏账 准备	逾期情况				期后 回款	期后回 款比例
					未逾期	逾期1年内	逾期1-2年	逾期2年以上		
普强时代（珠海横琴）信息技术有限公司	普强	235.41	235.27	28.25	-	235.41	-	-	20.00	8.50%
北京中核久安科技有限公司	中核久安	214.56	214.56	25.75	-	214.56	-	-	-	-
阿里云计算有限公司	阿里巴巴	198.24	143.55	75.98	26.14	28.55	143.55	-	54.40	27.44%
深圳市天道日新科技有限公司	天道日新	100.00	100.00	100.00	-	-	-	100.00	100.00	100.00%
北京京东世纪贸易有限公司	京东	84.09	84.09	42.04	-	-	84.09	-	-	-
合计		832.30	777.47	272.02	26.14	478.52	227.64	100.00	174.40	20.95%

2021年12月31日										
客户名称	品牌	应收账款 期末余额	其中：账龄 1年以上	坏账 准备	逾期情况				期后 回款	期后回 款比例
					未逾期	逾期1年内	逾期1-2年	逾期2年以上		
阿里云计算有限公司	阿里巴巴	311.58	200.69	28.60	97.62	213.96	-	-	214.84	68.95%
北京小米智能科技有限公司	小米	314.08	174.67	17.69	222.17	91.91	-	-	314.08	100.00%
深圳市天道日新科技有限公司	天道日新	100.00	100.00	100.00	-	-	-	100.00	100.00	100.00%
杭州海康威视数字技术股份有限公司	海康威视	111.99	63.95	13.44	-	111.99	-	-	111.99	100.00%
上海未来伙伴机器人有限公司	未来伙伴 机器人	56.90	56.90	56.90	-	-	-	56.90	1.50	2.64%
合计		894.55	596.21	216.63	319.79	417.86	-	156.90	742.41	82.99%
2020年12月31日										
客户名称	品牌	应收账款 期末余额	其中：账龄 1年以上	坏账 准备	逾期情况				期后 回款	期后回 款比例
					未逾期	逾期1年内	逾期1-2年	逾期2年以上		
深圳市天道日新科技有限公司	天道日新	100.00	100.00	50.00	-	-	100.00	-	100.00	100.00%
Verint Systems Ltd.	Verint	60.96	60.96	30.48	-	-	60.96	-	60.59	99.39%
上海未来伙伴机器人有限公司	未来伙伴 机器人	56.90	56.90	56.90	-	-	-	56.90	1.50	2.64%
合计		217.86	217.86	137.38	-	-	160.96	56.90	162.09	74.40%

注：期后回款数据截至2023年8月31日。

报告期各期末，除上海未来伙伴机器人有限公司、深圳市天道日新科技有限公司外，公司账龄 1 年以上的应收账款主要相对方为综合实力较强，资信等级较高的知名企业，目前处于正常经营状态，信用风险未发生明显变化，且均与公司开展多年业务合作，合作状况良好，应收账款回收风险较低，且公司根据“预期信用损失”模型计提了坏账准备。这些客户的具体信用状态情况如下：

客户名称	品牌	客户信用状态
北京中核久安科技有限公司	中核久安	客户于 2016 年注册成立，主要对接政府、法院等终端客户。客户综合实力强，资信状况良好，经营状态正常，公司自 2021 年与该客户开展业务合作。
科大讯飞股份有限公司	科大讯飞	客户成立于 1999 年，是亚太地区知名的智能语音和人工智能上市企业。自成立以来，一直从事智能语音、自然语言理解、计算机视觉等核心技术研究并保持了国际前沿技术水平。2008 年，公司在深圳证券交易所挂牌上市（股票代码：002230）。客户综合实力强，资信状况良好，经营状态正常，公司自 2014 年与该客户开展业务合作以来合作状况良好。
北京小米智能科技有限公司	小米	客户成立于 2018 年，是小米通讯技术有限公司旗下子公司，在业内占据稳定的市场地位。客户综合实力强，资信状况良好，经营状态正常，公司自 2018 年与该客户开展业务合作以来合作状况良好。
阿里云计算有限公司	阿里巴巴	客户成立于 2008 年，是全球领先的云计算及人工智能科技公司，客户综合实力强，资信状况良好，经营状态正常，公司自 2014 年与该客户开展业务合作以来合作状况良好。
杭州海康威视数字技术股份有限公司	海康威视	客户成立于 2001 年，是全球领先的以视频为核心的物联网解决方案提供商，营销及服务网络覆盖全球，为 A 股上市公司。客户综合实力强，资信状况良好，经营状态正常，公司自 2017 年与该客户开展业务合作以来合作状况良好。
北京京东世纪贸易有限公司	京东	客户成立于 2007 年，是一家以批发、零售定型包装食品、保健食品等为主营业务的公司，在“2016 中国企业 500 强”中排名第 88 位。客户综合实力强，资信状况良好，经营状态正常，公司自 2017 年与该客户开展业务合作以来合作状况良好。
Verint Systems Ltd.	Verint	客户成立于 1994 年，为全球 Actionable Intelligence® 解决方案领导者，在客户交互和网络情报领域取得了领先地位。客户综合实力强，资信状况良好，经营状态正常，公司自 2014 年与该客户开展业务合作以来合作状况良好。
普强时代（珠海横琴）信息技术有限公司	普强	客户是一家以智能语音识别、语音分析、语音合成和自然语言处理技术为核心的人工智能高新技术企业。客户综合实力强，资信状况良好，经营状态正常，公司自 2019 年与该客户开展业务合作。
境外客户 J	境外品牌 E	客户是美国一家语音自动化解决方案公司，主要提供技术设计、开发、部署和调优服务。客户综合实力强，资信状况良好，经营状态正常，公司自

客户名称	品牌	客户信用状态
		2021 年与该客户开展业务合作。

最近一年一期末，上述客户中部分客户回款比例较低，除因阿里巴巴、科大讯飞、京东、境外客户 J 等客户内部付款流程较长外，北京中核久安采购产品主要用于政府、法院等终端客户，下游回款周期拉长导致其对发行人回款周期同步延长，阿里巴巴逾期 2 年以上的应收余额形成原因与北京中核久安类似；另外，发行人已经向普强时代(珠海横琴)信息技术有限公司发放律师函催收应收账款，该客户已部分回款。

此外，上海未来伙伴机器人有限公司、深圳市天道日新科技有限公司应收账款账龄较长，主要与客户自身经营情况有关。其中，上海未来机器人有限公司因债务危机已于 2020 年 4 月提起破产重整。2021 年 7 月 12 日，上海市徐汇区人民法院裁定批准上海未来伙伴机器人有限公司重整计划。发行人已针对该主体的应收账款全额计提了坏账准备，同时将持续跟进重整计划的实施。因破产重整实施，2023 年 7 月，发行人已收到上海未来伙伴机器人有限公司回款 1.5 万元。深圳天道日新科技有限公司的应收账款形成原因为发行人已经交付标的物软件系统，但客户欠付相应合同款项。发行人已于 2021 年 8 月就该项合同纠纷向法院提起诉讼，根据 2021 年 11 月 5 日广东省深圳市南山区人民法院民事判决书《（2021）粤 0305 民初 18297 号》判决结果，深圳天道日新应于判决生效起十日内向公司支付剩余款项 100.00 万元及违约金 9.00 万元。发行人于 2021 年至 2022 年针对该主体的应收账款全额计提了坏账准备，并于 2023 年 3 月收回了该款项。

（三）结合公司应收账款整体的回款及逾期情况、应收账款坏账准备计提政策、信用政策、同行业可比公司情况等说明公司应收账款坏账准备计提的充分性

1、公司应收账款整体的回款情况及逾期情况

报告期各期末，公司应收账款整体的回款情况及逾期情况如下：

单位：万元

项目		2023年6月30日			2022年12月31日		
		账面余额	坏账准备	计提比例	账面余额	坏账准备	计提比例
逾期情况	未逾期	8,137.68	244.13	3.00%	8,340.72	250.22	3.00%
	逾期1年以内 (含1年)	1,173.77	140.85	12.00%	2,177.78	261.33	12.00%
	逾期1至2年 (含2年)	637.49	318.74	50.00%	342.17	171.09	50.00%
	逾期2年以上	205.15	205.15	100.00%	205.15	205.15	100.00%
合计		10,154.08	908.88	-	11,065.82	887.79	-
回款情况	回款金额	2,221.70			7,423.60		
	回款比例	21.88%			67.09%		
项目		2021年12月31日			2020年12月31日		
		账面余额	坏账准备	计提比例	账面余额	坏账准备	计提比例
逾期情况	未逾期	7,370.59	221.12	3.00%	5,477.51	164.33	3.00%
	逾期1年以内 (含1年)	2,027.72	243.33	12.00%	1,005.26	120.63	12.00%
	逾期1至2年 (含2年)	171.57	85.79	50.00%	160.96	80.48	50.00%
	逾期2年以上	156.90	156.90	100.00%	56.90	56.90	100.00%
合计		9,726.78	707.13	-	6,700.63	422.34	-
回款情况	回款金额	8,429.37			6,370.22		
	回款比例	86.66%			95.07%		

注：期后回款数据截至2023年8月31日

报告期各期末，未逾期应收账款占比分别为81.75%、75.78%、75.37%、80.14%，整体占比较高，发行人于2023年上半年加强了应收账款催收力度后，情况已有所好转。报告期各期末应收账款的期后回款比例分别为95.07%、86.66%、67.09%、21.88%，回款情况整体良好。最近一期末，由于期后时间较短，仅为2个月，而客户信用期多为60天或90天，因此期后回款比例较低。

2、应收账款坏账准备计提政策、信用政策、同行业可比公司情况

报告期内，发行人执行新金融工具准则，适用“预期信用损失”模型，始终按照相当于整个存续期内预期信用损失的金额计量应收账款的减值准备，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。发行人以客户验收后初始确认应收账款的时点为信用期起点，并根据客户的授信期限区分应收账款信用期内外。预期信用损失率基于应收账款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。

发行人 8 家同行业可比公司均适用“预期信用损失”模型。其中，万兴科技采用区分信用期计提坏账准备，与公司坏账计提政策相似度最高；恒生电子等公司则以账龄为基础计算预期信用损失率。经检索其他市场案例，信息技术-软件与服务行业上市公司中，铜牛信息（300895.SZ）、大汉软件（创业板在审）采用区分信用期计提坏账准备；其他信息技术服务类上市公司中，信音电子（301329.SZ）、弘信电子（300657.SZ）、天德钰（688252.SH）、有研硅（688432.SH）也采用同类计提方式坏账准备，因此，区分信用期的坏账计提比例政策符合行业惯例。

发行人的信用期多为 60 天或 90 天，按照谨慎性原则，在进行可比分析时，以可比公司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的预期信用损失率，应收账款账龄区间为 1-2 年的预期信用损失率作为逾期 1 年以内（含 1 年）的预期信用损失率，应收账款账龄区间为 2-3 年的预期信用损失率作为逾期 1 至 2 年（含 2 年）的预期信用损失率，应收账款账龄区间为 3 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，发行人区分信用期的坏账计提比例与同行业上市公司对比情况如下：

证券代码	证券简称	未逾期	逾期 1 年以内 (含 1 年)	逾期 1 至 2 年 (含 2 年)	逾期 2 年以上
300290.SZ	荣科科技	2.00%	5.00%	30.00%	50.00%-100.00%
300245.SZ	天玑科技	0.00%-5.00%	10.00%	20.00%	50.00%-100.00%
300468.SZ	四方精创	1.02%-2.00%	2.00%-6.70%	2.00%-18.11%	2.00%-100.00%
300634.SZ	彩讯股份	5.00%	10.00%	30.00%	50.00%-100.00%
上述 4 家信息技术服务类可比公司平均数		2.75%	7.34%	22.51%	69.00%
上述 4 家信息技术服务类可比公司中位数		2.00%	8.35%	25.00%	75.00%
600570.SH	恒生电子	5.00%	10.00%	30.00%	100.00%
600588.SH	用友网络	8.00%	26.00%	48.00%	64.00%-100%

证券代码	证券简称	未逾期	逾期1年以内 (含1年)	逾期1至2年 (含2年)	逾期2年以上
300624.SZ	万兴科技	0.01%-100.00%	100.00%	100.00%	100.00%
603039.SH	泛微网络	5.00%	10.00%	20.00%	100.00%
上述4家软件产品类可比公司平均数		17.00%	36.50%	49.50%	95.50%
上述4家软件产品类可比公司中位数		5.00%	18.00%	39.00%	100.00%
可比公司平均数		9.88%	21.92%	36.01%	82.25%
可比公司中位数		5.00%	10.00%	30.00%	100.00%
海天瑞声		3.00%	12.00%	50.00%	100.00%

注：可比公司平均数计算时，区间数据取平均值；可比公司中位数计算时，区间数据取上下端点值。

发行人区分信用期的坏账计提比例严谨。其中，发行人未逾期、逾期1年以内（含1年）的坏账计提比例高于信息技术服务类可比公司，略低于软件产品类可比公司，主要系万兴科技影响，万兴科技应收账款的主要对手方为结算周期较短的第三方平台，导致万兴科技应收账款账龄较短，账龄主要在6个月以内，发生坏账可能性较低，因此万兴科技对于逾期的应收账款适用的预期信用损失率偏高。剔除万兴科技影响后，软件产品类可比公司未逾期、逾期1年以内（含1年）的预期信用损失率的平均数分别降至6.00%和15.33%，全部可比公司未逾期、逾期1年以内（含1年）的预期信用损失率的平均数分别降至4.14%和10.76%，与发行人不存在显著差异。此外，发行人逾期1至2年（含2年）和逾期2年以上的坏账计提比例均高于可比公司，体现了公司坏账准备计提的充分性和谨慎性。

五、核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构和申报会计师履行了如下核查程序：

1、取得了发行人报告期内收入、成本明细表，分析报告期内收入、毛利波动的主要原因，结合发行人分季度收入数据及可比公司公开资料，分析发行人是否受到季节性因素影响；

2、访谈发行人管理层，了解报告期内收入及业绩波动的主要原因，了解发行人未来收入增长的持续性及是否存在净利润持续大幅下降的风险；

3、查阅相关行业研究报告，分析行业发展趋势、政策变化及市场规模情况，结合发行人提供的在手订单及收入预测明细，判断发行人的可持续经营能力；

4、核查了发行人报告期内主要客户、供应商的销售及采购合同、验收单据、回款单据等，对主要客户、供应商进行了访谈和函证；

5、核查了发行人报告期内费用明细表，分析期间费用变动的的原因，结合可比公司情况，分析发行人期间费用规模及趋势的合理性；

6、查阅同行业可比公司应收账款周转率的情况及坏账准备计提政策，分析公司应收账款周转率低于同行业平均值的原因及大幅下滑的合理性；获取发行人应收账款账龄明细表以及坏账准备计提表等资料，检查应收账款计提坏账准备的准确性；

7、了解一年以上长账龄的应收账款主要客户的经营情况，获取主要客户的相关销售合同、收入确认依据等资料，分析应收账款账龄和逾期情况，对于账龄时间较长的应收账款，与管理层进行沟通，了解账龄较长的原因。核查报告期各期末公司应收账款整体的回款情况，并重点关注大额应收账款的期后回款情况，抽查部分回款的相关凭证及银行回单，核查交易回款的真实性。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、最近三年及一期公司收入与净利润波动趋势差异较大具备合理性，发行人存在一季度收入偏低、四季度收入偏高的季节性特点，根据发行人的预测，随着发行人经营内外部环境的逐步好转，并结合在手订单等情况，公司 2023 年全年收入存在下滑风险，但预计不存在大幅下滑的风险。影响发行人最近一期净利润大幅度下滑的不利因素已有所改善，公司净利润大幅下滑的趋势预计不会长期持续，公司的持续经营能力不存在重大不确定性，相关风险提示已充分披露。

2、发行人最近三年及一期毛利率下滑尤其是最近一期末大幅下滑具备合理性。

3、发行人报告期内期间费用规模合理，管理费用及研发费用大幅增长具备合理性，研发费用中职工薪酬各期大幅上涨具备合理性，研发费用中数据服务费

与主营业务的数据服务费不存在混同情况，研发费用中数据服务费归集与核算准确。

4、结合公司应收账款整体的回款及逾期情况、应收账款坏账准备计提政策、信用政策、同行业可比公司等情况，发行人应收账款周转率大幅下滑具备合理性，应收账款坏账准备计提具备合理性和充分性。

5. 关于数据合规性

根据申报材料及公开信息，公司所在的细分领域为 AI 基础数据服务领域。2022 年 9 月，国家互联网信息办公室发布实施《数据出境安全评估办法》，对数据出境安全评估的具体要求、评估程序、监督管理制度、法律责任以及合规整改要求进行规定；2023 年 8 月，国家互联网信息办公室等七部门发布实施《生成式人工智能服务管理暂行办法》，针对生成式人工智能产品的预训练数据、优化训练数据来源等的合法性进行了规定。

请发行人说明：（1）上述规定要求对于发行人生产经营、业务发展的影响及发行人合规情况，说明是否存在因审批程序进度或整改要求影响现有业务及募投项目经营的情形，对发行人相关影响的风险提示是否充分；（2）结合发行人获取及使用个人信息的具体流程，说明发行人数据收集及使用是否符合《数据安全法》《个人信息保护法》等规定要求，是否出现过个人信息隐私泄露事件，是否存在相关行政处罚、诉讼或潜在纠纷。

请保荐机构及发行人律师核查并发表明确意见。

回复：

一、上述规定要求对于发行人生产经营、业务发展的影响及发行人合规情况，说明是否存在因审批程序进度或整改要求影响现有业务及募投项目经营的情形，对发行人相关影响的风险提示是否充分

（一）上述规定要求对于发行人生产经营、业务发展的影响及发行人合规情况

1、《数据出境安全评估办法》对于发行人生产经营、业务发展的影响较小

（1）发行人作为数据处理者涉及《数据出境安全评估办法》（以下简称《评估办法》）规定的情形，向境外提供在中国境内运营中收集和产生的个人信息需要履行数据出境安全评估申报程序

根据《评估办法》第二条，“数据处理者向境外提供在中华人民共和国境内运营中收集和产生的重要数据和个人信息的安全评估，适用本办法。”

根据《评估办法》第四条，“数据处理者向境外提供数据，有下列情形之一

的,应当通过所在地省级网信部门向国家网信部门申报数据出境安全评估:(一)数据处理者向境外提供重要数据;(二)关键信息基础设施运营者和处理 100 万人以上个人信息的数据处理者向境外提供个人信息;(三)自上年 1 月 1 日起累计向境外提供 10 万人个人信息或者 1 万人敏感个人信息的数据处理者向境外提供个人信息;(四)国家网信部门规定的其他需要申报数据出境安全评估的情形。”

发行人作为数据处理者涉及上述“(三)自上年 1 月 1 日起累计向境外提供 10 万人个人信息或者 1 万人敏感个人信息的数据处理者向境外提供个人信息”的情形。因此,发行人向境外提供在中国境内运营中收集和产生的个人信息的,应向国家网信部门申报数据出境安全评估。

(2) 《评估办法》的实施对于发行人生产经营、业务开展的影响较小

①发行人境外业务中,向境外销售涉及境内运营中收集产生的个人信息的数据集属于《评估办法》的适用范围,《评估办法》于 2022 年 9 月 1 日施行后,该类业务应履行数据出境安全评估申报程序。

发行人境外业务主要包括以下三种类型:

A、发行人境外业务中计算机视觉方向中的手写体业务、光学字符识别业务以及被采集对象为非自然人的物体或场景类业务,智能语音方向中的语音识别转写类业务,自然语言中的发音词典类业务,上述业务类型涉及的数据不属于《评估办法》所规定的“个人信息或者重要数据”,因此不属于《评估办法》的适用范围。

B、发行人境外业务中部分原料数据来源为境外采集,主要为外国人语音等个人信息数据,发行人进行标注、整理等处理后再将原始数据(如需)和标注结果传输至境外,发行人并不在中国境内运营过程中收集产生个人信息或重要数据,该业务类型涉及的数据不属于《评估办法》所规定的“在中华人民共和国境内运营中收集和产生的重要数据和个人信息”,因此不属于《评估办法》的适用范围。

C、发行人其他境外业务,涉及出境数据为在中国境内运营中收集产生的个人语音等涉及个人信息的数据集,属于《评估办法》的适用范围。根据《评估办法》规定,该类业务发行人应经国家网信部门安全评估审核通过后方能将数据提

供至境外。

②《评估办法》的实施阶段性影响境外客户部分境外业务的需求，但对发行人生产经营、业务发展的影响较小

《评估办法》施行后，符合规定情形的数据处理者向境外提供涉及境内运营中收集产生的个人信息的数据集应履行数据出境安全评估申报程序，受申报文件准备、申报程序及审核周期的影响，同时出于合规性及审慎性考虑，发行人与境外客户就该部分业务普遍处于观望状态，同时叠加境外客户阶段性裁员或业务方向调整，导致其预算释放进度放缓，从而减少了该部分业务需求，发行人与境外客户就该类业务的联系洽谈及预期订单相应减少。

目前数据出境安全评估申报审核政策已明晰，预计境外客户也将逐渐根据业务需求安排采购该部分业务数据集，且客户要求订单交付时间也将相应考虑审核周期进行调整，发行人结合境外客户订单需求情况已申报或将增加申报数据出境安全评估，发行人对数据出境安全评估审核要求逐渐熟悉，申报材料质量也将进一步完善，准备时间更为可控。

《评估办法》自 2022 年 9 月 1 日施行，根据发行人 2021 年及 2022 年 1-8 月的境外收入业务构成情况，2021 年及 2022 年 1-8 月，发行人属于《评估办法》限制范围内的境外业务收入总计占发行人 2021 年及 2022 年 1-8 月的整体业务收入总额的比例为 4.57%，因此对公司未来整体生产经营、业务发展的影响较小。

综上，发行人作为数据处理者具备《评估办法》规定的情形之一，向境外提供境内运营中收集产生的涉及个人信息的数据集需要申报数据出境安全评估；《评估办法》的实施对发行人的生产经营、业务开展产生的影响较小。

2、《生成式人工智能服务管理暂行办法》对于发行人生产经营、业务发展无不利影响

(1) 发行人不属于《生成式人工智能服务管理暂行办法》中规定的生成式人工智能服务提供者，不直接适用其中对“提供者”的具体要求

根据《生成式人工智能服务管理暂行办法》（简称“《管理办法》”）第二条，“利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等服务（以下称生成式人工智能服务），适用本办法。”

发行人的主营业务为 AI 训练数据的研发设计、生产及销售业务，数据种类涵盖智能语音（语音识别、语音合成等）、计算机视觉、自然语言等领域，所销售的训练数据主要由下游客户用于数据预训练、优化训练等数据训练使用。

截至目前，公司业务并不涉及《管理办法》所规定的“利用生成式人工智能技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容服务”，因此公司不属于生成式人工智能服务提供者，故不直接适用《管理办法》中对“提供者”的具体要求。

（2）发行人向相关客户提供的训练数据产品满足《管理办法》对训练数据的相关要求

根据《管理办法》第七条，“生成式人工智能服务提供者（指利用生成式人工智能技术提供生成式人工智能服务（包括通过提供可编程接口等方式提供生成式人工智能服务）的组织、个人）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：

（一）使用具有合法来源的数据和基础模型；

（二）涉及知识产权的，不得侵害他人依法享有的知识产权；

（三）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；

（四）采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；

（五）《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。”

发行人涉及向《管理办法》规定的“提供者”销售训练数据用于其算法预训练、优化训练。发行人业务经营过程中注重遵守数据安全、网络安全、个人信息保护相关法律法规，涉及个人信息采集的已取得个人或法定监护人同意，相关数据产品由发行人或客户根据约定享有知识产权，不存在侵犯他人的知识产权的情形；个人数据采集后直接上传至发行人一体化数据处理技术支撑平台，发行人通

过采取数据安全技术手段能够保障采集数据的真实性、准确性；同时，数据的客观性和多样性是客户对训练数据普遍存在的需求，发行人在训练数据的采集和处理中严格规范筛选数据来源，加强产品和服务的质量控制，也保证实现了数据的客观性和多样性。因此发行人向生成式人工智能服务提供者销售的训练数据产品或提供的定制服务符合上述《管理办法》第七条对“预训练数据、优化训练数据”的具体要求。

综上，发行人不属于《管理办法》规定的生成式人工智能服务提供者，不直接适用其中对“提供者”的具体要求，发行人向相关客户提供的训练数据产品满足《管理办法》对训练数据的相关要求，《管理办法》对于发行人生产经营、业务发展无不利影响。同时《管理办法》的实施使得训练数据行业尤其是下游客户对于训练数据的合规性及质量等的重视程度进一步加强，将有利于发行人业务的开展。

3、发行人合规情况

发行人一直注重遵守数据安全相关法律法规，自2022年9月1日《评估办法》施行后，受申报文件准备、申报程序及审核周期的影响，同时出于合规性及审慎性考虑，发行人及境外客户对根据《评估办法》应履行数据出境安全评估申报程序的该类境外业务处于观望状态，境外客户该部分业务需求以及相应发行人该部分业务开展相应减少。就发行人和境外客户经沟通拟开展的该类业务，发行人已进行或将进行数据出境安全评估申报，境外客户相应予以配合。

经核查发行人2022年9月1日后新签署的境外业务合同及其履行交付情况，发行人不存在新签合同涉及在中国境内运营中收集和产生的个人信息或重要数据出境的情形；发行人已就拟开展的需履行数据出境安全评估申报程序的交易向国家互联网信息办公室进行申报并处于审核过程中，发行人不存在根据《评估办法》应进行数据出境安全评估申报而未进行申报的情形。

根据发行人确认，并经查询国家企业信用信息公示系统（<https://www.gsxt.gov.cn/index.html>）、信用中国（<https://www.creditchina.gov.cn/>）、国家互联网信息办公室网站（<http://www.cac.gov.cn/>），发行人不存在因违反《评估办法》《管理办法》等数据安全相关法律法规而受到处罚的情况。

（二）说明是否存在因审批程序进度或整改要求影响现有业务及募投项目经营的情形

发行人作为数据处理者具备《评估办法》规定的情形之一，向境外提供境内运营中收集产生的涉及个人信息的数据集需要履行数据出境安全评估申报的审批程序。发行人不属于《管理办法》中规定的生成式人工智能服务提供者，不直接适用《管理办法》中对“提供者”的具体要求，不涉及相应审批程序或整改要求。

数据出境安全评估申报审批程序对发行人现有业务及募投项目经营的影响具体如下：

1、发行人就拟开展交易申报的数据出境安全评估正在审理中

因拟向某境外客户销售语音产品/数据集涉及境内运营中收集产生的个人信息数据出境，发行人已通过北京市互联网信息办公室向国家互联网信息办公室申报上述数据出境安全评估，并于2023年7月12日由国家互联网信息办公室受理，目前正在审理过程中，期间已按照国家互联网信息办公室要求提交补充说明或文件。

2、上述数据出境安全评估审批过程中，国家互联网信息办公室未提出整改要求

根据发行人确认及核查，发行人在申报上述数据出境安全评估的过程中，仅被国家互联网信息办公室要求提供补充说明或文件等，不存在被要求进行整改的情形。

3、数据出境安全评估审批程序进度对现有业务的影响较小

涉及境内运营中收集产生个人信息的数据集出境的部分境外业务属于《评估办法》限制范围，发行人开展该类业务需通过数据出境安全评估审核后方可交付出境。如本题第一部分回复所述，从发行人2021年度、2022年1-8月境外业务收入情况来看，该部分境外业务收入在发行人整体收入中占比较小。

另外，《评估办法》实施后，受申报文件准备、申报程序及审核周期的影响，同时出于合规性及审慎性考虑，公司及存在该类业务需求的境外客户普遍处于观

望状态，同时叠加境外客户阶段性裁员或业务方向调整，导致其预算释放进度放缓，从而境外客户该部分业务需求以及相应发行人该部分业务开展也相应减少。目前数据出境安全评估申报审核政策已明晰，预计境外客户将逐渐根据业务需求安排采购该部分业务数据集，且客户要求订单交付时间也将相应考虑审核周期进行调整，发行人已结合境外客户订单需求情况申报或将增加申报数据出境安全评估，且发行人对数据出境安全评估审核要求逐渐熟悉，申报材料质量也将进一步完善，准备时间更为可控。因此，数据出境安全评估审核对发行人现有业务的影响较小。

4、按照当前募集资金投资项目的经营规划，数据出境安全评估审批程序进度对募投项目不产生影响

本次发行的募集资金投资项目为 AI 大模型训练数据集建设项目、数据生产垂直大模型研发项目。

(1) AI 大模型训练数据集建设项目系生产用于通用型及各种垂直领域大模型训练的海量、高品质数据集项目的数据集产品，具体可分为三大类：

①通用及特定垂直领域的大语言模型训练数据集，包括但不限于：

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

B、多语言大模型预训练语料数据集（含通用场景、对话场景、指令集等）。

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

③大模型评测数据集：可应用于大模型的能力、任务、指标等方面的评测。

上述训练数据集涉及部分在《评估办法》限定范围内的涉及境内运营中收集产生的个人信息的中文数据集，系主要为解决境内大模型训练中优质中文数据集不足的问题，发行人预计该类数据集主要向境内头部人工智能大模型科技企业销售，不涉及数据出境因此无需办理数据出境安全评估。其他多语种语音数据集、多语种多模态数据集等，主要涉及境外采集的外国人个人信息数据，不涉及境内

运营中收集产生的个人信息，以及其他场景类、评测类等数据集不涉及个人信息或重要数据，根据《评估办法》规定无需办理数据出境安全评估。

目前，AI 大模型训练数据集建设项目处于已完成前期调研及可行性论证阶段，未涉及数据出境情形，按照当前经营规划，数据出境安全评估申报审批程序对募集资金投资项目不造成影响。如后续项目开展过程中相关数据集确有需销售至境外应履行数据出境安全评估申报的情形的，发行人将按照《评估办法》的规定，切实履行数据出境安全评估申报程序。

(2) 数据生产垂直大模型研发项目系研发发行人数据生产垂直大模型，并以发行人数据生产垂直大模型为核心，升级海天瑞声一体化数据处理技术支撑平台，不涉及向客户销售及数据出境的情况，无需申报数据出境安全评估，因此不会受到数据出境安全评估审批程序的影响。

综上，按照当前募集资金投资项目的经营规划，数据出境安全评估审批程序进度对发行人募集资金投资项目的当前经营不产生影响，如后续项目开展过程中相关数据集确有需销售至境外应履行数据出境安全评估申报的情形的，发行人将按照《评估办法》的规定，切实履行数据出境安全评估申报程序。

(三) 对发行人相关影响的风险提示是否充分

发行人已在《募集说明书》“第七节与本次发行相关的风险因素”就上述规定对发行人相关影响风险提示如下：

“七、数据安全相关风险

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

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

综上,发行人已在《募集说明书》中进行了充分的风险提示。

二、结合发行人获取及使用个人信息的具体流程,说明发行人数据收集及使用是否符合《数据安全法》《个人信息保护法》等规定要求,是否出现过个人信息隐私泄露事件,是否存在相关行政处罚、诉讼或潜在纠纷

(一) 结合发行人获取及使用个人信息的具体流程,说明发行人数据收集及使用是否符合《数据安全法》《个人信息保护法》等规定要求

1、发行人获取及使用个人信息的具体流程

(1) 根据客户项目需求确定需采集的个人信息范围

发行人业务经营中涉及收集、使用的个人信息按照业务用途主要分为个人基本信息、个人资格信息(辅助信息)和项目数据三类。其中:

①个人基本信息用于项目管理、核查被采集人身份以及与其取得联系,主要包括必要信息和选填信息两部分。必要信息为手机号码、姓名、性别、生日、年龄、所在地(国家/省份/城市)等;选填信息为身份证号码、电子邮箱等。

②个人资格信息(辅助信息)用于确认被采集人是否符合参与项目的资格,所涉及的个人信息种类依不同项目需要而有所不同,例如,智能语音业务中,可能会收集被采集人的语种/口音信息。个人资格信息为被采集人参与特定项目的必要信息。

③项目数据为发行人生产的数据产品、数据服务成果的主体部分,用于向客户交付,主要包括语音、人像、人物视频等。

发行人将结合客户项目需求按照最小化的原则收集必要的个人信息。

(2) 发行人向个人信息主体告知相关事项并获得授权同意后自行收集个人信息,或从原料数据采集供应商或客户获取数据

①发行人经个人信息主体授权同意后自行收集个人信息

发行人收集个人信息的方式包括线上、线下两种，具体情况如下：

A、线上方式：被采集人通过发行人自行研发应用的终端人管理系统注册登录及参与项目录制前，必须点选阅读发行人于注册登陆页面公布的《用户协议》《个人信息保护政策》等文件，上述文件向被采集人明确告知授权内容、范围、期限、处理目的、行使法定权利的方式和方法等事项。终端人管理系统也设置了授权许可管理模块，对被采集人采集前的授权程序进行控制，被采集人明确进行授权同意，具体包括勾选并同意《用户协议》《个人信息保护政策》以及以手写签名方式签署《授权书》，授权同意后才能开始采集；涉及个人敏感信息的还会取得单独同意。

B、线下方式：对于需要线下进行数据采集的被采集人，发行人通过事先提供书面《授权书》向其明确告知授权内容、范围、期限、数据处理目的等法律法规要求的必备信息（其中包括个人敏感信息的提示），被采集人亲笔签署《授权书》表示同意后再开始进行信息采集。发行人通过线下方式采集的，仍是通过使用发行人研发的技术、平台、工具等进行数据采集。

被采集人如为未成年人，则由其法定监护人完成上述授权同意程序。

另外，发行人也在其网站公开了个人信息保护政策，以公开方式告知公众，并便于公众查阅。根据个人信息保护政策，被采集人享有查询、更正、删除等权利，并公布投诉联系方式；发行人收到投诉后将分析原因、积极与客户沟通、提出解决方案，并及时将解决方案告知投诉人。

②发行人从原料数据采集供应商获取数据、由客户自行提供数据

原料数据采集供应商按照发行人的采集要求寻找匹配的终端人员，组织终端人员使用发行人研发的技术、平台、工具等进行数据采集。原料数据采集供应商的采集过程和方式与发行人收集个人信息的过程和方式相同。发行人对原料数据采集供应商进行规范管理，通过制定《数据安全管理制度》、与相关供应商签署数据处理协议和保密协议以及在采集前对供应商进行培训等方式，明确要求供应商采集个人信息数据不得侵犯第三方权益，规定保证数据来源合法及保密等相关责任。

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个人信息数据并委托发行人进行标注处理。客户仅向发行人提供语音、图像、视频等业务数据，不提供涉及的个人基本信息。客户向发行人提供数据，主要通过加密传输等方式提供，部分情形客户要求发行人委派人员在客户工作场所的工作平台上直接进行加工。发行人不参与客户数据收集过程，仅经客户授权许可对数据进行加工。发行人主要客户为阿里巴巴、三星、腾讯、微软等人工智能产业链上知名大型企业，该企业均制定有隐私政策等相关个人信息保护制度，重视业务经营中涉及的个人信息数据来源和使用的合法性。

(3) 将获取的个人信息上传至自主研发平台进行存储、处理

除特定项目需要外，大部分终端采集的个人数据通过发行人自行研发的采集软件、工具直接上传至发行人一体化数据处理技术支撑平台进行存储及标注等处理，该平台整合了采集、加工、质检环节所需的软件工具和模块，融入业务过程中的数据安全需求，能够防范未经授权的个人私自对数据进行处理。

发行人对数据存储采取了严格的内部控制措施保障数据安全，避免发生数据泄露等情形：**A.个人基本信息数据和其他数据分开存储；B.数据加密、个人基本信息去标识化处理；C.设置保存期限，按照客户协议约定及管理需要及时删除；D.访问权限、人员设置及内部审批；E、定期对业务系统进行漏洞扫描，及时对业务系统漏洞进行修复；网络环境中部署 IPS、IDS 等安全设备。**

特定项目中，按照个别大客户要求，发行人向客户提供数据转写、标注等纯加工定制服务时，发行人人员直接在客户的工作平台上对客户提供的个人数据进行标注处理，并不留存相关数据。发行人该等客户为人工智能产业链上知名大型企业，该企业均重视个人信息保护事项，采用技术手段保障工作平台的安全性，能够保障数据安全。

(4) 将最终形成的数据集交付给客户

发行人在被采集人签署的授权声明或协议中约定的使用范围、期限、目的等限制范围内使用个人信息，项目数据和个人资格信息（辅助信息）提供给客户用于训练数据产品的算法模型，个人基本信息仅用于确定被采集人是否符合项目要求以及由发行人留存数据采集记录以保障被采集人依法行使撤回授权、删除等个

人信息权利。发行人对数据传输采用加密方式以保护数据安全。发行人与客户的协议也约定相关个人信息数据安全义务和责任。

(5) 发行人已制定相关制度建立个人信息保护、数据安全管理制度和流程

发行人已制定《数据安全管理制度》《IT 安全管理制度》《存储服务器使用管理制度》《网络安全管理制度》《数据备份制度》《数据分级分类指引》《应急响应管理制度》《个人信息出境管理制度》等制度，建立了个人信息保护、数据安全管理制度和流程。发行人设立数据保护官负责全面统筹个人信息数据保护制度的落实，已建立针对各部门所持有的包含个人信息在内的各类数据类型的授权访问策略，定期开展个人信息安全影响评估和合规审计，建立个人数据泄露等应急处理机制。

发行人已取得《信息系统安全等级保护备案证明》（证书编号：11010899932-22001）、《管理体系认证证书》（证书编号：01121PI1001R0M）、《管理体系认证证书》（证书编号：01122IS20102R1M）等认证证明文件，证明发行人信息安全管理体系符合相关等级和标准要求。

2、发行人个人信息数据收集及使用符合《数据安全法》规定要求

《中华人民共和国数据安全法》（以下简称《数据安全法》）中涉及发行人进行个人信息数据收集及使用的主要规定和发行人的符合情况如下：

法条序号	主要内容	个人信息收集及使用方式是否符合规定
第二十七条	开展数据活动应当依照法律、行政法规的规定和国家标准的强制性要求，建立健全全流程数据安全管理制度，组织开展数据安全教育培训，采取相应的技术措施和其他必要措施，保障数据安全。重要数据的处理者应当设立数据安全负责人和管理机构，落实数据安全保护责任。	符合 发行人业务经营不涉及重要数据，但仍设置数据保护官和管理机构
第二十九条	开展数据处理活动应当加强风险监测，发现数据安全缺陷、漏洞等风险时，应当立即采取补救措施；发生数据安全事件时，应当立即采取处置措施，按照规定及时告知用户并向有关主管部门报告。	符合
第三十条	重要数据的处理者应当按照规定对其数据处理活动定期开展风险评估，并向有关主管部门报送风险评估报告。风险评估报告应当包括处理的重要数据的种类、数量，开展数据处理活动的情况，面临的数据安全风险及其应对措施等。	发行人业务经营不涉及重要数据，不适用

第三十一条	关键信息基础设施的运营者在中华人民共和国境内运营中收集和产生的重要数据的出境安全管理，适用《中华人民共和国网络安全法》的规定；其他数据处理者在中华人民共和国境内运营中收集和产生的重要数据的出境安全管理办法，由国家网信部门会同国务院有关部门制定。	发行人不属于关键信息基础设施的运营者，业务经营不涉及重要数据，不适用
第三十二条	任何组织、个人收集数据，应当采取合法、正当的方式，不得窃取或者以其他非法方式获取数据。法律、行政法规对收集、使用数据的目的、范围有规定的，应当在法律、行政法规规定的目的和范围内收集、使用数据。	符合

经核查，发行人获取和使用个人信息符合《数据安全法》的相关要求。

3、发行人个人信息数据收集及使用符合《个人信息保护法》规定要求

《中华人民共和国个人信息保护法》（以下简称《个人信息保护法》）中涉及发行人进行个人信息数据收集及使用的主要规定和发行人的符合情况如下：

法条序号	主要内容	个人信息获取及使用方式是否符合规定
第五条	处理个人信息应当遵循合法、正当、必要和诚信原则，不得通过误导、欺诈、胁迫等方式处理个人信息。	符合
第六条	处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式。收集个人信息，应当限于实现处理目的的最小范围，不得过度收集个人信息。	符合
第七条	处理个人信息应当遵循公开、透明原则，公开个人信息处理规则，明示处理的目的、方式和范围。	符合
第九条	个人信息处理者应当对其个人信息处理活动负责，并采取必要措施保障所处理的个人信息的安全。	符合
第十条	任何组织、个人不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息；不得从事危害国家安全、公共利益的个人信息处理活动。	符合
第十三条	符合下列情形之一的，个人信息处理者方可处理个人信息： （一）取得个人的同意；	符合
第十四条	基于个人同意处理个人信息的，该同意应当由个人在充分知情的前提下自愿、明确作出。法律、行政法规规定处理个人信息应当取得个人单独同意或者书面同意的，从其规定。 个人信息的处理目的、处理方式和处理的个人信息种类发生变更的，应当重新取得个人同意。	符合
第十五条	基于个人同意处理个人信息的，个人有权撤回其同意。个人信息处理者应当提供便捷的撤回同意的方式。 个人撤回同意，不影响撤回前基于个人同意已进行的个人信息处理活动的效力。	符合
第十七条	个人信息处理者在处理个人信息前，应当以显著方式、清	符合

法条序号	主要内容	个人信息获取及使用方式是否符合规定
	<p>晰易懂的语言真实、准确、完整地向个人告知下列事项：</p> <p>（一）个人信息处理者的名称或者姓名和联系方式；</p> <p>（二）个人信息的处理目的、处理方式，处理的个人信息种类、保存期限；</p> <p>（三）个人行使本法规定权利的方式和程序；</p> <p>（四）法律、行政法规规定应当告知的其他事项。</p> <p>前款规定事项发生变更的，应当将变更部分告知个人。</p> <p>个人信息处理者通过制定个人信息处理规则的方式告知第一款规定事项的，处理规则应当公开，并且便于查阅和保存。</p>	
第十九条	除法律、行政法规另有规定外，个人信息的保存期限应当为实现处理目的所必要的最短时间。	符合
第二十一条	<p>个人信息处理者委托处理个人信息的，应当与受托人约定委托处理的目的、期限、处理方式、个人信息的种类、保护措施以及双方的权利和义务等，并对受托人的个人信息处理活动进行监督。</p> <p>受托人应当按照约定处理个人信息，不得超出约定的处理目的、处理方式等处理个人信息；委托合同不生效、无效、被撤销或者终止的，受托人应当将个人信息返还个人信息处理者或者予以删除，不得保留。</p> <p>未经个人信息处理者同意，受托人不得转委托他人处理个人信息。</p>	符合
第二十三条	个人信息处理者向其他个人信息处理者提供其处理的个人信息的，应当向个人告知接收方的名称或者姓名、联系方式、处理目的、处理方式和个人信息的种类，并取得个人的单独同意。接收方应当在上述处理目的、处理方式和个人信息的种类等范围内处理个人信息。接收方变更原先的处理目的、处理方式的，应当依照本法规定重新取得个人同意。	符合
第二十五条	个人信息处理者不得公开其处理的个人信息，取得个人单独同意的除外。	符合
第二十九条	处理敏感个人信息应当取得个人的单独同意；法律、行政法规规定处理敏感个人信息应当取得书面同意的，从其规定。	符合
第三十条	个人信息处理者处理敏感个人信息的，除本法第十七条第一款规定的事项外，还应当向个人告知处理敏感个人信息的必要性以及对个人权益的影响；依照本法规定可以不向个人告知的除外。	符合
第三十一条	<p>个人信息处理者处理不满十四周岁未成年人个人信息的，应当取得未成年人的父母或者其他监护人的同意。</p> <p>个人信息处理者处理不满十四周岁未成年人个人信息的，应当制定专门的个人信息处理规则。</p>	符合
第三十八条	<p>个人信息处理者因业务等需要，确需向中华人民共和国境外提供个人信息的，应当具备下列条件之一：</p> <p>（一）依照本法第四十条的规定通过国家网信部门组织的安全评估；</p> <p>（二）按照国家网信部门的规定经专业机构进行个人信息</p>	发行人已按照《评估办法》对拟开展的所涉业务申报数据出境安全评

法条序号	主要内容	个人信息获取及使用方式是否符合规定
	<p>保护认证；</p> <p>（三）按照国家网信部门制定的标准合同与境外接收方订立合同，约定双方的权利和义务；</p> <p>（四）法律、行政法规或者国家网信部门规定的其他条件。中华人民共和国缔结或者参加的国际条约、协定对向中华人民共和国境外提供个人信息的条件等有规定的，可以按照其规定执行。</p> <p>个人信息处理者应当采取必要措施，保障境外接收方处理个人信息的活动达到本法规定的个人信息保护标准。</p>	估
第三十九条	<p>个人信息处理者向中华人民共和国境外提供个人信息的，应当向个人告知境外接收方的名称或者姓名、联系方式、处理目的、处理方式、个人信息的种类以及个人向境外接收方行使本法规定权利的方式和程序等事项，并取得个人的单独同意。</p>	符合
第五十一条	<p>个人信息处理者应当根据个人信息的处理目的、处理方式、个人信息的种类以及对个人权益的影响、可能存在的安全风险等，采取下列措施确保个人信息处理活动符合法律、行政法规的规定，并防止未经授权的访问以及个人信息泄露、篡改、丢失：</p> <p>（一）制定内部管理制度和操作规程；</p> <p>（二）对个人信息实行分类管理；</p> <p>（三）采取相应的加密、去标识化等安全技术措施；</p> <p>（四）合理确定个人信息处理的操作权限，并定期对从业人员进行安全教育和培训；</p> <p>（五）制定并组织实施个人信息安全事件应急预案；</p> <p>（六）法律、行政法规规定的其他措施。</p>	符合
第五十二条	<p>处理个人信息达到国家网信部门规定数量的个人信息处理者应当指定个人信息保护负责人，负责对个人信息处理活动以及采取的保护措施等进行监督。</p> <p>个人信息处理者应当公开个人信息保护负责人的联系方式，并将个人信息保护负责人的姓名、联系方式等报送履行个人信息保护职责的部门。</p>	符合
第五十四条	<p>个人信息处理者应当定期对其处理个人信息遵守法律、行政法规的情况进行合规审计。</p>	符合
第五十五条	<p>有下列情形之一的，个人信息处理者应当事前进行个人信息保护影响评估，并对处理情况进行记录：</p> <p>（一）处理敏感个人信息；</p> <p>（二）利用个人信息进行自动化决策；</p> <p>（三）委托处理个人信息、向其他个人信息处理者提供个人信息、公开个人信息；</p> <p>（四）向境外提供个人信息；</p> <p>（五）其他对个人权益有重大影响的个人信息处理活动。</p>	符合

经核查，发行人获取和使用个人信息符合《个人信息保护法》的相关要求。

(二) 是否出现过个人信息隐私泄露事件，是否存在相关行政处罚、诉讼或潜在纠纷

根据发行人确认及中国裁判文书网 (<http://wenshu.court.gov.cn/>)、中国执行信息公开网 (<http://zxgk.court.gov.cn/>)、信用中国 (<https://www.creditchina.gov.cn/>)、国家互联网信息办公室网站 (<http://www.cac.gov.cn/>)、百度 (www.baidu.com) 等公开信息查询检索，发行人未出现过个人信息隐私泄露事件，不存在相关行政处罚、诉讼或潜在纠纷。

三、核查程序及核查意见

(一) 核查程序

保荐机构及发行人律师履行了以下核查程序：

1、研究分析《评估办法》《管理办法》对发行人的适用情况及对发行人业务开展的限制范围；对发行人业务负责人员进行访谈，了解发行人境外业务的分类及《评估办法》《管理办法》对发行人生产经营、业务发展的影响，了解发行人保证训练数据真实性、准确性、客观性、多样性的措施；核查发行人境外收入对应的业务分类情况，核查 2021 年度、2022 年 1-8 月《评估办法》限制范围内境外业务的收入占比；核查 2022 年 9 月 1 日后发行人新签署境外业务合同及履行情况；

查询国家企业信用信息公示系统 (<https://www.gsxt.gov.cn/index.html>)、信用中国 (<https://www.creditchina.gov.cn/>)、国家互联网信息办公室网站 (<http://www.cac.gov.cn/>) 并取得发行人的确认函，核查发行人是否受到相关处罚；

2、取得并查阅发行人数据出境安全自评估报告等数据出境安全评估申报文件，核查发行人数据出境安全评估的受理及审核情况；取得并查阅了国家互联网信息办公室与发行人安全评估申报负责人的往来邮件，核查国家互联网信息办公室在审核数据出境安全评估事项中对发行人的补充意见和要求；对发行人业务负责人员进行访谈，了解发行人现有业务的合规性及审核程序进度对发行人现有业务的影响；查阅发行人募集资金投资项目的主要内容并对发行人业务负责人员进行访谈，了解募集资金投资项目的经营规划、业务内容以及审核程序进度对发行

人募集资金投资项目经营的影响；

3、访谈发行人数据保护官，了解发行人获取及使用个人信息的流程及其合规性；下载并使用发行人线上数据采集软件，核查线上采集的告知程序、告知内容、获得授权的方式等；抽查发行人线下采集签署的授权书或采集合同；查阅发行人官网披露的《个人信息保护政策》；查看发行人数据处理一体化数据处理技术支撑平台及其中数据处理过程的数据保护等措施；取得并查阅发行人《数据安全管理制度》《IT 安全管理制度》《存储服务器使用管理制度》《网络安全管理制度》《数据备份制度》《数据分级分类指引》《应急响应管理制度》《个人信息出境管理制度》等管理制度、发行人信息系统安全等级保护及管理体系认证证书等认证证明文件；查阅发行人与供应商、客户签署的业务协议和保密协议等，核查发行人对供应商、客户的数据安全要求；查阅《数据安全法》《个人信息保护法》中涉及发行人个人信息数据收集及使用的相关内容，对照核查发行人获取及使用个人信息的合规性；

4、查询中国裁判文书网（<http://wenshu.court.gov.cn/>）、中国执行信息公开网（<http://zxgk.court.gov.cn/>）、信用中国（<https://www.creditchina.gov.cn/>）、国家互联网信息办公室网站（<http://www.cac.gov.cn/>）、百度（www.baidu.com）等公开信息查询网站并取得发行人的确认函，核查发行人是否存在出现过个人信息、隐私泄露事件，是否发生过数据安全引发的诉讼、纠纷或行政处罚。

（二）核查意见

经核查，保荐机构及发行人律师认为：

1、发行人作为数据处理者涉及《评估办法》规定的情形之一，部分境外业务涉及向境外提供境内运营中收集产生的个人信息需要申报数据出境安全评估；该部分境外业务因《评估办法》的实施受到阶段性影响，属于《评估办法》限制范围内的境外业务的收入占比较小，《评估办法》对发行人生产经营、业务发展的影响较小；发行人不属于《管理办法》中规定的生成式人工智能服务提供者，不直接适用于《管理办法》中对“提供者”的具体要求，发行人向相关客户提供的训练数据产品满足《管理办法》对训练数据的相关要求，《管理办法》对于发行人生产经营、业务发展无不利影响；发行人已就拟开展的需要履行数据出境安

全评估申报程序的交易向国家互联网信息办公室进行申报并处于审核过程中，发行人不存在根据《评估办法》应进行数据出境安全评估申报而未进行申报的情形，不存在因违反《评估办法》《管理办法》等而受到处罚的情况。

2、发行人就拟进行交易已申报的数据出境安全评估事项目前正在国家互联网信息办公室审核中，国家互联网信息办公室在审核过程中无要求发行人整改的情形；数据出境安全评估申报审批程序进度对现有业务的影响较小，按照募集资金投资项目当前经营规划对募集资金投资项目经营不产生影响；发行人不涉及根据《管理办法》规定需要履行审批程序或进行整改的情形。发行人已在《募集说明书》中就上述规定涉及的风险进行了充分提示。

3、发行人数据收集及使用符合《数据安全法》《个人信息保护法》等规定要求，未出现过个人信息隐私泄露事件，不存在相关行政处罚、诉讼或潜在纠纷。

6. 关于境外销售

根据申报材料，最近三年及一期，公司境外收入为 8,368.56 万元、5,729.72 万元、11,576.20 万元、1,013.36 万元，占营业收入的比重分别为 35.86%、27.75%、44.03% 及 35.16%。

请发行人说明：（1）境外销售具体分布情况，结合外销主要客户情况说明境外收入大幅波动的原因，境内境外销售毛利率是否存在显著差异；（2）境外主要客户的变动情况，新增境外销售客户的情况、销售金额及占比，结合报告期内海关数据、出口退税金额等与外销收入的匹配性说明境外销售的真实性；（3）汇率变动对发行人生产经营的影响，相关风险披露是否充分。

请保荐机构及申报会计师核查并发表明确意见，并说明对境外收入的核查过程、核查比例及核查结论。

回复：

一、境外销售具体分布情况，结合外销主要客户情况说明境外收入大幅波动的原因，境内境外销售毛利率是否存在显著差异

（一）境外销售具体分布情况

发行人按照客户所在地区分境内、境外业务收入，境外业务收入中主要客户分布在美国、日韩等国家，具体区域分布如下：

单位：万元

业务细分类型	2023 年 1-6 月		2022 年度	
	金额	比例	金额	比例
美国	2,070.88	71.09%	9,671.91	83.55%
韩国、日本	688.25	23.63%	1,329.34	11.48%
其他地区	153.95	5.28%	574.95	4.97%
境外合计	2,913.08	100.00%	11,576.20	100.00%
业务细分类型	2021 年度		2020 年度	
	金额	比例	金额	比例
美国	2,174.83	37.96%	4,651.52	55.58%
韩国、日本	2,093.70	36.54%	2,993.85	35.77%
其他地区	1,461.19	25.50%	723.19	8.64%

境外合计	5,729.72	100.00%	8,368.56	100.00%
------	----------	---------	----------	---------

(二) 结合外销主要客户情况说明境外收入大幅波动的原因

报告期内，发行人境外业务收入分别为 8,368.56 万元、5,729.72 万元、11,576.20 万元、2,913.08 万元，波动幅度较大，主要受到宏观经济环境波动、全球公共卫生事件的演进影响以及部分客户在研发预算及周期安排、业务方向和组织结构调整等因素综合导致的。报告期各期境外收入前五大客户情况如下：

单位：万元

2023 年 1-6 月			
客户品牌	营业收入	占境外收入比重	发展年份及客户介绍
Amazon	649.03	22.28%	Amazon.Com Inc.（已于纳斯达克主板、圣地亚哥证券交易所、伦敦证券交易所国际板上市，股票代码分别为：AMZN、AMZN、0R10）成立于 1994 年，是美国最大的电商公司，也是最早经营电商的公司之一，已发展为业务覆盖电商、互联网、云服务的大型科技公司。 发行人自 2012 年开始与该品牌客户合作。
Microsoft	589.69	20.24%	Microsoft Corp（已于纳斯达克主板、伦敦证券交易所国际版、中国香港联交所主板上市，股票代码分别为 MSFT、0QYP、4388）成立于 1975 年，是一家美国的跨国科技公司，是个人计算机软件开发的先驱者，目前是全球最大的电脑软件提供商，业务覆盖电脑软件的研发、制造、授权及相关服务等。 发行人自 2005 年开始与该品牌客户合作。
某国际消费电子产 品厂商	572.54	19.65%	客户是美国的一家大型科技公司，经营范围包括电脑硬件、软件，消费电子产品、数字发布、零售业务等。 发行人自 2010 年开始与该品牌客户合作。
Newcon	387.10	13.29%	日本纽康株式会社是成立于 1991 年的日本企业，主要提供系统解决方案、医疗系统业务、信息服务、客户服务业务等。 发行人自 2015 年开始与该品牌客户合作。
Hyundai	222.28	7.63%	现代集团（Hyundai）成立于 1946 年，是韩国一家以建筑、造船、汽车行业为主，兼营钢铁、机械、贸易、运输、水泥生产、冶金、金融、电子工业等几十个行业的综合性企业集团。旗下子公司有：韩国现代重工集团、现代汽车、现代摩比斯等。 发行人自 2020 年开始与该品牌客户合作。
合计	2,420.64	83.10%	-
2022 年度			
客户品牌	营业收入	占境外收入比重	发展年份及客户介绍
某国际消费电子产	2,805.77	24.24%	见上文

品厂商			
境外品牌 A	2,588.84	22.36%	客户是美国一家大型科技公司，旗下产品覆盖社交媒体平台、即时通讯应用等，还致力于研究人工智能、虚拟现实、增强现实等领域，并开发相应的产品。发行人自 2017 年开始与该品牌客户合作。
Microsoft	2,218.59	19.17%	见上文
境外品牌 B	1,215.46	10.50%	客户是全球可编程图形处理技术领袖，能够提供适用于工作和娱乐应用并且同时支持众多操作系统的全套影院级着色三维图形解决方案的半导体公司，专注于打造能够增强个人和专业计算平台的人机交互体验的产品。发行人自 2021 年开始与该品牌客户合作。
Newcon	725.07	6.26%	见上文
合计	9,553.73	82.53%	-
2021 年度			
客户品牌	营业收入	占境外收入比重	发展年份及客户介绍
Microsoft	923.87	16.12%	见上文
某国际消费电子产品厂商	770.49	13.45%	见上文
Hyundai	687.48	12.00%	见上文
Vin Big Data	611.75	10.68%	Vin Big Data (Vingroup 成员) 成立于 2018 年 8 月，是由越南最大私企 Vingroup 在 2018 年出资成立的专注于计算生物学、医学图像处理、计算机视觉、自然语言处理四个大数据关键领域的研究机构。发行人自 2020 年开始与该品牌客户合作。
境外品牌 C	359.09	6.27%	客户成立于 1970 年，是日本的一家贸易企业，主要从事重型电气设备及工业的进出口贸易，海外材料、零件、设备的采购以及海外新产品、新兴市场制造产品的销售。发行人自 2015 年开始与该品牌客户合作。报告期内，该主体是个别最终需求方的指定代理方，发行人与最终需求方就训练数据产品或服务的具体需求进行直接对接沟通，后续最终需求方出于自身经营管理需要，指定境外品牌 C 为其签约/采购主体。
合计	3,352.67	58.51%	-
2020 年度			
客户品牌	营业收入	占境外收入比重	发展年份及客户介绍
Samsung	1,336.20	15.97%	三星集团成立于 1938 年，是韩国的大型跨国企业集团，旗下拥有三星电子（已于韩国证券交易所、伦敦证券交易所主板上市，股票/存托凭证代码分别为 005930、SMSD 等）、三星物产、三星人寿保险等诸多子公司，业务覆盖电子、金融、机械、化学等众多领域。

			发行人自 2010 年开始与该品牌客户合作。
Microsoft	1,079.20	12.90%	见上文
某国际消费电子产 品厂商	928.33	11.09%	见上文
Google	669.74	8.00%	Alphabet Inc.（已于纳斯达克主板、伦敦证券交易所国际板上市，股票代码分别为 GOOG、OHD6），是美国的一家跨国企业，是全球最大的搜索引擎公司，业务覆盖互联网搜索、云计算、广告业务等，也提供基于互联网的产品与服务。 发行人自 2017 年开始与该品牌客户合作。
境外品牌 D	490.48	5.86%	客户成立于 2010 年，是日本的一家贸易公司，主要从事汽车及机械零件的进出口贸易、海外商业软件代理、行业相关的新技术、新产业信息的贸易和中介等。发行人自 2016 年开始与该品牌客户合作。报告期内，该主体是个别最终需求方的指定代理方，发行人与最终需求方就训练数据产品或服务的具体需求进行直接对接沟通，后续最终需求方出于自身经营管理需要，指定境外品牌 D 为其签约/采购主体。
合计	4,503.94	53.82%	-

注：以上客户的销售金额合并同期同一品牌下不同境外主体的销售金额列示

2021 年，在境外公共卫生事件出现重大反复的大背景下，公司境外业务遭受一定冲击，整体规模较 2020 年下滑显著，主要原因包括：（1）受境外公共卫生事件影响，公司涉及境外开展原料数据采集环节的部分项目进度出现延迟，远程办公使客户办公效率下降。以境外主要客户 Microsoft 为例，其训练数据需求部门、采购部门在 2020 至 2021 年大部分时间改为居家办公的工作模式，导致其内部训练数据采购计划出现一定延迟，整体上致使公司对该客户 2021 年的销售收入有所下降。（2）以 Samsung、Google、Amazon 等为代表的主要境外客户依据自身研发节奏，在 2019、2020 年集中采购了较大规模的智能语音训练数据，能够满足其在特定领域一段时间内的需求，因此 2021 年采购量出现阶段性较大幅度的回落。（3）部分大型客户内部业务方向、研发进度或组织结构出现调整，导致其训练数据采购有所减缓、订单延迟下达。以 Samsung 为例，据发行人了解，伴随其手机在中国市场销量的变化，2021 年开始其自身也在进行业务方向调整，应用于手机端的 AI 相关服务的训练数据需求缩减。

2022 年，随着境外公共卫生事件影响逐渐消退，境外市场需求快速回升，公司境外业务收入达到 1.16 亿元，同比增长 102.04%。主要原因包括：（1）受益于境外企业全球化扩张以及虚拟人、多语言扩展、软硬件升级等研发及业务

方向的发展，多语种多模态数据业务需求呈现快速增长趋势，公司凭借多年来在语言研究、发音词典、语音、计算机视觉以及自然语言方面的技术储备，赢得了境外客户信赖，相关订单数量显著增长，以某国际消费电子产品厂商为例，2022年多语种多模态业务需求量呈现爆发增长，较去年同期增长 264.16%；境外品牌 A 自 2021 年末起，对多语种智能语音、发音词典数据需求量增加，2022 年实现收入占当年境外收入的 22.36%。

2023 年上半年，发行人境外收入较 2022 年同期减少 55.58%，主要原因包括：

(1) 受全球宏观经济波动影响，境外部分客户进行阶段性裁员或业务方向调整，导致其预算释放进度放缓。发行人境外主要客户 Microsoft、Amazon 等均在 2022 年底至 2023 年初期间公布了裁员计划，因宏观经济环境波动，科技型企业裁员的同时也配合着各项预算支出的管控，因此数据服务采购需求有所降低；(2) 《数据出境安全评估办法》于 2023 年一季度进入全面落地实施阶段，公司及境外业务相关客户出于合规性及审慎性考虑普遍处于观望状态，对公司的境外业务开拓产生了一定影响。

(三) 境内境外销售毛利率是否存在显著差异

报告期内，境内境外的销售毛利率情况如下：

项目	2023 年 1-6 月	2022 年度	2021 年度	2020 年度
境内	47.64%	49.71%	56.75%	54.51%
境外	76.30%	83.82%	82.90%	91.29%

由于境外主要客户更认同数据服务商的综合能力及品牌价值，价格敏感度相对较低，定制服务业务毛利率一般高于境内客户，同时境外业务中标准化数据集产品占比更高，综合呈现境外毛利率相比于境内毛利率一贯较高的状态。

二、境外主要客户的变动情况，新增境外销售客户的情况、销售金额及占比，结合报告期内海关数据、出口退税金额等与外销收入的匹配性说明境外销售的真实性

(一) 境外主要客户的变动情况，新增境外销售客户的情况、销售金额及占比

根据前述表格，发行人各期境外前五大客户品牌主要包括 Microsoft、Hyundai、

VinBigData、Samsung、Google 及 Amazon 等人工智能产业链上的知名机构且基本均为发展年份较长的老客户，报告期内前五大客户品牌占比分别为 53.82%、58.51%、82.53%、83.10%，集中度逐年提升。报告期内新增签约境外客户主要包括两类，一类是营销开拓的新品牌客户，另一类为已建立合作关系的客户品牌下新增签约主体，包括品牌内部调整签约主体或者其他主体因其自身业务发展需要新增训练数据采购需求。报告期内，境外新增客户的情况及收入占比如下：

项目	2023年1-6月	2022年度	2021年度	2020年度
当期新增境外客户数量（签约主体口径）（家数）	11	23	34	20
当期新增境外客户实现收入（万元）	1,049.46	548.77	1,122.62	1,478.85
新增收入占比	36.03%	4.74%	19.59%	17.67%

报告期各期新增客户收入规模存在一定波动，但整体占比不高。2023年1-6月，发行人新增客户收入占比达到 36.03%，主要由于当年前五大客户品牌 Amazon、Hyundai 旗下均有新的合作主体且采购金额较高，二者合计占比接近 30%。各期新增收入金额在 50 万元以上的境外客户的具体情况如下：

单位：万元

序号	客户名称	合作开始时间	是否为新增品牌	合作开始当期的营业收入	占当期境外收入比重	销售内容	客户介绍
1	境外客户 A	2023年1-6月	否	649.03	22.28%	智能语音一训练数据产品、自然语言一训练数据产品	客户隶属于美国最大的电商公司，也是最早经营电商的公司之一，已发展为业务覆盖电商、互联网、云服务的大型科技公司。
2	境外客户 B	2023年1-6月	否	222.28	7.63%	智能语音一训练数据产品、自然语言一训练数据产品	客户是韩国一家自主交通即服务（TaaS）商，致力于将尖端技术与量身定制的互联服务相结合，以优化人员和货物的移动流程。
3	境外客户 C	2023年1-6月	是	95.42	3.28%	智能语音一训练数据产品、自然语言一训练数据定制服务	客户是印度一家提供机器学习、计算机视觉、自然语言处理和文本分析等领域的服务商，并根据客户需要提供自定义库。
4	境外客户 D	2022年	是	131.06	1.13%	智能语音一训练数据产品	客户是日本的一家大型科技公司，旗下产品覆盖门户网站、消息应用程序和移动支付应用。

序号	客户名称	合作开始时间	是否为新增品牌	合作开始当期的营业收入	占当期境外收入比重	销售内容	客户介绍
5	境外客户 E	2022 年	是	90.70	0.78%	智能语音数据产品、自然语言训练数据产品	客户是印度一家移动应用程序运营商,为客户提供预订城市出租车、外站旅行、汽车租赁和企业旅行服务。
6	境外客户 F	2021 年	是	181.71	3.17%	智能语音数据产品	客户是欧洲一家面部识别、语音和多模式生物识别系统以及音频和视频记录、处理和分析解决方案的领先开发商。
7	境外客户 G	2021 年	否	162.25	2.83%	智能语音数据定制服务及训练数据产品、自然语言数据定制服务、计算机视觉训练数据定制服务	客户是东南亚一家主要从事软件 and 应用程序(游戏和网络安全除外)开发和广告活动的公司。
8	境外客户 H	2021 年	是	107.10	1.87%	智能语音数据产品	客户是美国一家文字电话生产、软件服务提供商,其主要是为聋人或听力障碍的人提供全国领先的通信服务。
9	境外客户 I	2021 年	否	63.97	1.12%	智能语音数据产品、自然语言-训练数据产品	客户是一家总部位于美国的跨国软件公司,主要开发人工智能(AI)辅助技术。
10	境外客户 J	2021 年	是	57.40	1.00%	智能语音数据产品	客户是美国一家语音自动化解方案公司,主要提供技术设计、开发、部署和调优服务。
11	境外客户 K	2021 年	是	50.15	0.88%	智能语音数据产品	客户是美国一家语音类软件公司,打造了一套基于机器学习的本地快速语音识别 + AI 口音转换解决方案。
12	境外客户 L	2020 年	是	266.88	3.19%	智能语音数据产品	客户是美国一家云软件公司,主要经营客户关系管理(CRM)等平台。
13	境外客户 M	2020 年	是	251.57	3.01%	智能语音数据产品	客户是美国一所研究型大学。
14	境外客户 N	2020 年	是	237.91	2.84%	智能语音数据产品	客户是欧洲一家员工行为监控软件开发商,旨在利用人工智能分析技术帮助企业合规运营。

序号	客户名称	合作开始时间	是否为新增品牌	合作开始当期的营业收入	占当期境外收入比重	销售内容	客户介绍
15	境外客户 O	2020 年	否	235.55	2.81%	智能语音训练数据产品	客户是日本一家全球知名的大型综合性跨国企业集团。
16	Institute of Big Data	2020 年	是	125.73	1.50%	智能语音训练数据产品	客户成立于 2018 年，是由越南最大私企 Vingroup 出资成立的专注于计算生物医学、医学图像处理、计算机视觉、自然语言处理四个大数据关键领域的研究机构。
17	境外客户 P	2020 年	否	120.10	1.44%	智能语音训练数据产品	客户是某知名品牌旗下的先进研发中心。
18	境外客户 Q	2020 年	否	67.35	0.80%	智能语音训练数据定制服务	客户位于中国香港，隶属于某中国知名互联网综合服务提供商之一。

报告期内，新增境外客户向发行人采购的产品以智能语音产品为主，且均为规模体量较大的知名公司，发行人与新增客户的合作均基于真实的业务需求，具备合理性。

(二) 结合报告期内海关数据、出口退税金额等与外销收入的匹配性说明境外销售的真实性

根据《中华人民共和国海关法》第二条规定，“海关依照本法和其他有关法律、行政法规，监管进出境的运输工具、货物、行李物品、邮递物品和其他物品（以下简称进出境运输工具、货物、物品），征收关税和其他税、费，查缉走私，并编制海关统计和办理其他海关业务。”

报告期内，发行人外销业务主要通过网络电子传输方式交付数据服务产品，没有实体进出境的行为，因此不需要向海关进行申报。此外，报告期内发行人未申报出口退税，因此无法通过匹配海关数据、出口退税金额说明境外收入真实性。保荐机构及申报会计师对境外收入履行的核查程序参见本题回复之“四、核查程序及核查意见”部分。

报告期内，发行人境外客户主要为海外知名大型科技公司、人工智能企业及科研院所，与发行人合作稳定，境外收入具备真实性。

三、汇率变动对发行人生产经营的影响，相关风险披露是否充分。

报告期内，发行人境外销售主要通过美元定价结算，还有部分欧元、港币、日元和澳元，报告期内发行人发生外币交易时，按交易发生日的即期汇率将外币金额折算为人民币金额。于资产负债表日，外币货币性交易采用资产负债表日的即期汇率折算为人民币，所产生的折算差额直接计入当期损益（即：“财务费用——汇兑损益”科目）。此外，在发生外币报表折算业务时，外币资产负债表中资产、负债类项目采用资产负债表日的即期汇率折算；所有者权益类项目除“未分配利润”外，均按业务发生时的即期汇率折算；利润表中的收入与费用项目，采用交易发生的即期汇率折算。上述折算产生的外币报表折算差额，在其他综合收益科目中列示。外币现金流量报表采用现金流量发生日的即期汇率折算，汇率变动对现金的影响额，在现金流量表中单独列示。

报告期内发行人汇率变动对发行人生产经营的主要影响情况如下：

单位：万元

项目	2023年1-6月	2022年度	2021年度	2020年度
汇兑损失（“-”代表汇兑收益）	-171.39	-473.67	103.88	188.97
营业收入	7,446.09	26,288.79	20,647.65	23,337.40
占比	-2.30%	-1.80%	0.50%	0.81%

整体上，汇率变动对发行人生产经营存在一定影响，但不存在重大影响，发行人已在募集说明书及其他申报材料中披露风险提示如下：

“九、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 35.86%、27.75%、44.03%及 39.12%，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。”

四、核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构和申报会计师履行了如下核查程序：

1、访谈发行人主要管理人员及财务人员，了解境外销售的模式、流程、收入确认方式以及报告期内发行人境外收入波动的主要原因，主要客户收入金额变动的原因；

2、获取海天瑞声报告期内收入成本明细表，了解境外销售各类别产品和服务的销售收入、产品结构、毛利率等明细及其变化；

3、查阅并取得海天瑞声与主要境外客户签订的销售合同，识别与商品或服务所有权风险和报酬转移相关的合同条款，核查主要境外客户的验收文件，银行回单等支持性证据以验证境外收入的真实性；

4、核查报告期内境外客户的期后回款情况；

5、保荐机构获取了年审会计师 2020-2022 年度审计的往来函证，对境外收入的回函情况及替代性测试进行复核；对报告期内主要境外客户执行访谈、独立函证及替代性程序，具体覆盖比例如下：

单位：万元

项目	2020 年度	2021 年度	2022 年度	2023 年 1-6 月
境外收入	8,368.56	5,729.72	11,576.20	2,913.08
复核会计师函证及替代性程序金额	6,222.20	4,431.17	10,760.31	-
复核覆盖比例	74.35%	77.34%	92.95%	-
访谈金额	4,839.73	1,824.26	5,923.26	1,650.96
访谈覆盖比例	57.83%	31.84%	51.17%	56.67%
独立函证及替代性程序金额	7,581.32	3,604.94	6,964.89	2,403.99
独立函证及替代性程序覆盖比例	90.59%	62.92%	60.17%	82.52%

6、申报会计师对报告期内主要境外客户交易金额执行函证以及访谈程序，具体覆盖比例如下：

单位：万元

项目	2020 年度	2021 年度	2022 年度	2023 年 1-6 月
境外收入	8,368.56	5,729.72	11,576.20	2,913.08
访谈金额	3,348.12	1,824.26	5,923.26	1,650.96
独立函证及替代性程序金额	7,581.32	4,431.17	10,760.31	2,403.99
访谈覆盖比例	40.01%	31.84%	51.17%	56.67%

独立函证及替代性程序覆盖比例	90.59%	77.34%	92.95%	82.52%
----------------	--------	--------	--------	--------

7、获取发行人境外主要子公司的境外法律意见书，了解境外主体经营的合法合规性情况；

8、结合报告期内财务费用等科目明细，分析汇率变动对发行人生产经营的影响。

(二) 核查意见

经核查，保荐机构、申报会计师认为：

报告期内境外收入大幅波动及境外销售毛利率高于境内业务具备合理性，境外业务收入具备真实性，汇率变动对发行人生产经营不存在重大影响，相关风险已充分披露。

7. 关于财务性投资

根据申报材料，截至 2022 年 3 月 31 日发行人不存在财务性投资。

请发行人说明：（1）自本次董事会决议日前六个月至今，公司实施或拟实施财务性投资（含类金融业务）的具体情况，相关财务性投资是否已从本次募集资金总额中扣除；（2）最近一期末公司是否存在持有金额较大、期限较长的财务性投资（包括类金融业务）情形。

请保荐机构及申报会计师核查并发表明确意见，并就发行人是否符合《证券期货法律适用意见第 18 号》第 1 条、《监管规则适用指引——发行类第 7 号》第 1 条的相关规定发表明确意见。

回复：

一、自本次董事会决议日前六个月至今，公司实施或拟实施财务性投资（含类金融业务）的具体情况，相关财务性投资是否已从本次募集资金总额中扣除

（一）财务性投资的认定标准

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

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

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

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

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

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

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

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

1、投资类金融业务

自本次发行相关董事会决议日前六个月起至本回复出具日，公司未投资类金融业务。

2、非金融企业投资金融业务

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

3、股权投资

发行人于 2023 年 8 月 14 日与数安易(北京)科技有限公司签订《投资协议》，约定发行人以 4,000,000 元认购数安易新增注册资本人民币 1,840,000 元，发行人占增资后注册资本的比例为 8.42%，同时约定了后续满足一定条件情况下，发行人有权追加 4,000,000 元投资并新增取得届时增资后 8.42% 的注册资本。发行人已于 2023 年 8 月向数安易支付了上述 4,000,000 元投资款，工商变更登记正在办理中。

数安易成立于 2022 年 5 月，主要从事数据安全相关技术及产品服务，其在数据脱敏、数据加密、数据资产管理平台等方面的能力，可为海天瑞声数据业务的合规开展提供重要技术支撑。同时，数安易在数据安全防护及平台管控服务、

数据库审计、数据库安全运维等方面的业务，是海天瑞声规划的数据要素业务的重要内容板块，数安易在此方面的技术与业务能力预期将成为海天瑞声进一步拓展数据要素业务的重要支撑，公司预计将与数安易联合研发数据服务平台或形成上下游业务合作关系。综上，数安易业务与海天瑞声现有及潜在业务方向具备协同发展前景，本次投资符合海天瑞声的主营业务和战略发展方向。在完成本次投资后，双方未来在各自业务发展中原则上须优先考虑彼此为相关适当业务机会的合作对象，藉此扎实推进实施双方既定的业务协同战略。

本次股权投资与公司主营业务相关，属于“围绕产业链上下游以获取技术、原料或者渠道为目的的产业投资”，不属于财务性投资。

自本次发行相关董事会决议日前六个月起至本回复出具日，公司不存在与主营业务无关的股权投资。

4、投资产业基金、并购基金

本次发行相关董事会决议日前六个月起至本回复出具日，公司不存在投资产业基金、并购基金的情形。

5、拆借资金

自本次发行相关董事会决议日前六个月起至本回复出具日，公司不存在拆借资金的情形。

6、委托贷款

自本次发行相关董事会决议日前六个月起至本回复出具日，公司不存在委托贷款的情形。

7、购买收益波动大且风险较高的金融产品

自本次发行相关董事会决议日前六个月起至本回复出具日，公司不存在购买收益波动大且风险较高的金融产品。

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

二、最近一期末公司是否存在持有金额较大、期限较长的财务性投资（包括类金融业务）情形

截至 2023 年 6 月 30 日，发行人未持有较大金额的财务性投资，发行人与投资相关的会计科目列示如下：

单位：万元

序号	科目	截至 2023 年 6 月 30 日账面价值	财务性投资金额	财务性投资占归属母公司所有者净资产比例
1	交易性金融资产	38,027.96	-	-
2	预付款项	1,237.29	-	-
3	其他应收款	204.68	-	-
4	其他流动资产	308.37	-	-
5	长期待摊费用	37.10	-	-
	合计	39,815.40	-	-

注：2023 年 6 月 30 日数据未经审计。

1、交易性金融资产

截至 2023 年 6 月 30 日，公司交易性金融资产明细如下：

单位：万元

科目	账面余额
以公允价值计量且其变动计入当期损益的金融资产	38,027.96
其中：理财产品	38,027.96
合计	38,027.96

截至 2023 年 6 月末，发行人理财产品的明细如下：

单位：万元

产品类型	发行方	产品名称	金额	起始日	到期日
保本	宁波银行股份有限公司	单位结构性存款 237363	500.00	2023/6/6	2023/9/6
保本	中信银行股份有限公司	共赢智信汇率挂钩人民币结构性存款 14640 期	600.00	2023/4/17	2023/7/18
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36288 期	1,300.00	2023/5/29	2023/8/29
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36294 期	880.00	2023/6/1	2023/8/30
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36298 期	600.00	2023/6/5	2023/7/7
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36299 期	4,000.00	2023/6/5	2023/9/5

产品类型	发行方	产品名称	金额	起始日	到期日
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36303 期	2,540.00	2023/6/12	2023/9/12
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36319 期	18,120.00	2023/6/19	2023/9/19
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36330 期	500.00	2023/6/22	2023/7/24
保本	中信银行股份有限公司	共赢智信利率挂钩人民币结构性存款 36331 期	4,940.00	2023/6/22	2023/9/22
保本	招商银行股份有限公司	招商银行点金系列看涨两层区间 92 天结构性存款	4,000.00	2023/6/5	2023/9/5
合计			37,980.00	-	-

注：上述投资金额合计数与 2023 年 6 月末交易性金融资产科目余额差异为理财产品利息导致的公允价值变动。截至本回复出具日，上述产品部分已到期，公司已赎回并根据资金使用计划进行滚动投资。

截至 2023 年 6 月 30 日，公司购买的上述结构性存款系为提高资金使用效率而使用闲置自有资金进行现金管理，上述结构性存款为风险较低、流动性较强的理财产品，不属于“收益波动大且风险较高的金融产品”，不属于财务性投资。

2、预付款项

截至 2023 年 6 月 30 日，公司预付款项账面余额为 1,237.29 万元，主要为新办公场地装修设计费用及数据服务费等，不属于财务性投资。

3、其他应收款

截至 2023 年 6 月 30 日，公司其他应收款账面余额按性质列示如下：

单位：万元

项目	账面余额
房租押金	168.35
保证金	18.21
员工备用金	27.72
其他	1.12
合计	215.40

注：合计数与前表账面价值数存在差异系由于此处以余额口径列示明细。

截至 2023 年 6 月 30 日，公司其他应收款账面余额为 215.40 万元，主要为房租押金、保证金、员工备用金，不属于财务性投资。

4、其他流动资产

截至 2023 年 6 月 30 日，公司其他流动资产明细情况如下：

单位：万元

项目	账面余额
待抵扣及待认证进项税	235.70
预缴所得税	3.96
待摊房屋租赁费	68.70
合计	308.37

截至 2023 年 6 月 30 日，公司其他流动资产账面余额为 308.37 万元，主要为待抵扣及待认证进项税、预缴所得税和待摊房屋租赁费，不属于财务性投资。

5、长期待摊费用

截至 2023 年 6 月 30 日，公司长期待摊费用明细情况如下：

单位：万元

项目	账面余额
办公场所装修及其他	37.10
合计	37.10

截至 2023 年 6 月 30 日，公司长期待摊费用账面余额为 37.10 万元，主要为办公场所装修等，不属于财务性投资。

综上所述，最近一期末，发行人不存在持有金额较大、期限较长的财务性投资（包括类金融业务）情形。

三、核查程序及核查意见

（一）核查程序

针对上述事项，保荐机构和申报会计师履行了如下核查程序：

1、查阅《证券期货法律适用意见第 18 号》《监管规则适用指引——发行类第 7 号》等中国证监会关于财务性投资及类金融业务的有关规定，了解财务性投资及类金融业务的认定标准。

2、核查发行人报告期内的财务报表及附注，取得发行人交易性金融资产、其他应收款、其他流动资产等报表科目的明细表，并与发行人管理层访谈，了解

自本次董事会决议日前六个月至今发行人是否存在实施或拟实施财务性投资的情况：

3、核查发行人与数安易签署的投资协议，出资款支付凭证、数安易的主要合同，并与发行人管理层访谈了解收购数安易的背景，判断是否属于财务性投资；

4、取得最近一期末的理财合同及台账，判断是否属于财务性投资。

（二）核查意见

经核查，保荐机构、申报会计师认为：

本次发行董事会决议前六个月至本回复出具日，发行人不存在实施或拟实施财务性投资及类金融业务的情形，最近一期末发行人不存在持有金额较大、期限较长的财务性投资（包括类金融业务）情形。发行人符合《证券期货法律适用意见第 18 号》第 1 条、《监管规则适用指引一一发行类第 7 号》第 1 条的相关规定。

8. 关于其他

根据申报材料，发行人控股子公司安徽瑞天数智科技有限公司业务中包括组织文化艺术交流活动、网络文化经营，互联网信息服务；控股子公司山西瑞天数智科技有限公司业务中包括组织文化艺术交流活动，非居住房地产租赁。

请发行人说明：公司是否存在文化传媒业务、互联网平台业务和房地产业务，若是，请说明相关业务的具体内容、经营模式收入利润占比等情况，以及后续业务开展的规划安排。请保荐机构和发行人律师核查并发表意见。

回复：

一、公司不存在文化传媒业务、互联网平台业务和房地产业务

（一）公司主要从事 AI 训练数据的研发设计、生产及销售业务，不存在文化传媒业务、互联网平台业务和房地产业务

公司主要从事 AI 训练数据的研发设计、生产及销售业务，不存在文化传媒业务、互联网平台业务和房地产业务。根据《审计报告》、半年度报告及发行人确认，报告期内，公司营业收入全部为主营业务收入，具体情况如下：

单位：万元

项目	2023年1-6月		2022年		2021年		2020年	
	金额	比例	金额	比例	金额	比例	金额	比例
主营业务收入	7,446.09	100.00%	26,288.79	100.00%	20,647.65	100.00%	23,337.40	100.00%
其他业务收入	-	-	-	-	-	-	-	-
合计	7,446.09	100.00%	26,288.79	100.00%	20,647.65	100.00%	23,337.40	100.00%

（二）公司子公司安徽瑞天数智、山西瑞天数智的主营业务为训练数据产品及服务，不涉及开展文化传媒业务、互联网平台业务和房地产业务，发行人子公司已变更经营范围

公司子公司安徽瑞天数智、山西瑞天数智的原经营范围和主营业务情况如下表所示：

公司名称	原经营范围	主营业务
安徽瑞天数智	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；组织文化艺术交	训练数据产品及服务

公司名称	原经营范围	主营业务
	流活动；信息系统集成服务；软件销售；计算机软硬件及辅助设备零售；人工智能公共数据平台；大数据服务；数据处理和存储支持服务；卫星遥感数据处理；数据处理服务；工业互联网数据服务；会议及展览服务；租赁服务（不含许可类租赁服务）（除许可业务外，可自主依法经营法律法规非禁止或限制的项目）许可项目：网络文化经营；互联网信息服务；测绘服务。	
山西瑞天数智	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批发;数据处理服务;人工智能公共数据平台;信息系统集成服务;组织文化艺术交流活动;会议及展览服务;非居住房地产租赁。	训练数据产品及服务,目前尚未实际开展业务

安徽瑞天数智系发行人设立于安徽蚌埠的数据标注产业基地，其原经营范围中包含“组织文化艺术交流活动、网络文化经营，互联网信息服务”，但并未开展文化传媒业务、互联网平台业务和房地产业务。安徽瑞天数智已办理经营范围变更，删除了上述经营范围中“组织文化艺术交流活动、网络文化经营，互联网信息服务”的内容。

山西瑞天数智系发行人设立于山西大同的数据标注产业基地，目前尚未实际开展业务，其原经营范围中包含“组织文化艺术交流活动，非居住房地产租赁”，但不会开展文化传媒业务、互联网平台业务和房地产业务。山西瑞天数智已办理经营范围变更，删除了“组织文化艺术交流活动，非居住房地产租赁”的内容。

综上，发行人不存在文化传媒业务、互联网平台业务和房地产业务。

二、核查程序及核查意见

（一）核查程序

保荐机构及发行人律师履行了以下核查程序：

1、取得并查验了发行人《审计报告》，年度、半年度报告及说明，核查发行人主营业务及收入情况；

2、取得发行人子公司安徽瑞天数智、山西瑞天数智的《营业执照》、财务报表，核查发行人及其子公司的经营范围和主营业务情况；

3、取得安徽瑞天数智、山西瑞天数智变更经营范围后新取得的《营业执照》，核查其经营范围变更情况。

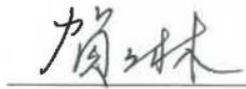
(二) 核查意见

经核查，保荐机构及发行人律师认为：发行人不存在文化传媒业务、互联网平台业务和房地产业务。

发行人董事长声明

本人已认真阅读北京海天瑞声科技股份有限公司本次审核问询函的全部内容，确认回复的内容真实、准确、完整，不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性承担相应法律责任。

发行人董事长：



贺琳

北京海天瑞声科技股份有限公司

2021年9月19日

（本页无正文，为北京海天瑞声科技股份有限公司《关于北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票申请文件的审核问询函的回复》之签章页）

北京海天瑞声科技股份有限公司
2023年9月19日



(本页无正文，为华泰联合证券有限责任公司《关于北京海天瑞声科技股份有限公司 2023 年度向特定对象发行 A 股股票申请文件的审核问询函的回复》之签章页)

保荐代表人：

杨阳

杨阳

张鹏

张鹏

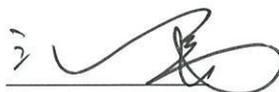
华泰联合证券有限责任公司

2023年9月19日

保荐人法定代表人声明

本人已认真阅读北京海天瑞声科技股份有限公司本次问询意见回复报告的全部内容，了解报告涉及问题的核查过程、本公司的内核和风险控制流程，确认本公司按照勤勉尽责原则履行核查程序，问询意见回复报告不存在虚假记载、误导性陈述或者重大遗漏，并对上述文件的真实性、准确性、完整性、及时性承担相应法律责任。

保荐人法定代表人：



江禹

华泰联合证券有限责任公司

2023年9月19日

