

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 请文件的第二轮审核问询函的回复

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）

二〇二三年十二月

上海证券交易所：

贵所 2023 年 11 月 23 日下发的《关于北京海天瑞声科技股份有限公司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）235 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. 关于募投项目	3
2. 关于经营情况.....	39
3. 关于数据合规.....	57

1. 关于募投项目

根据申报材料，1) 本次募投项目，发行人本次发行拟募集资金 66,600 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”，其中场地购置费分别为 1.82 亿元、0.2346 亿元；2) 本次募投项目“AI 大模型训练数据集建设项目”经测算，税后内部收益率为 16.82%，税后投资回收期（含三年建设期）为 5.89 年；3) 数据生产垂直大模型项目总投资 40,651.64 万元，其中 2.64 亿用于购买 220 台 GPU 服务器；4) 报告期末发行人交易性金融资产余额为 38,027.96 万元，均为公司理财所购买的结构性存款产品；5) 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。

请发行人进一步说明：（1）用简单通俗易懂的语言说明本次募投项目的具体内容，补充说明数据生产垂直大模型项目与现有的 AI 训练数据开发的相关技术差异，是否属于发行人现有主营业务，与公司现有主营业务关系；（2）2022 年购置大额房产后，本次募集资金进一步购置大额房产，说明 2022 年及本次募投项目购置的房产性质、单价情况，相关单价或单价测算的公允合理性；结合发行人自身定位、未来业务发展规划、收入及盈利水平较低、上市时为轻资产企业等情况，说明频繁大额购置房产行为是否偏离公司主业，是否与公司科技企业定位相符，是否与发行人收入规模、业务经营情况相匹配。结合前述情况，进一步说明本次募集资金较大金额购置房产的必要性、合理性；（3）结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况，说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确；结合前述情况及效益测算中收入、成本费用等关键指标的具体测算过程及依据等，说明“AI 大模型训练数据集建设项目”效益测算较高的合理性，与发行人现有水平及同行业可比公司相比，相关测算是否谨慎、合理；“数据生产垂直大模型研发项目”募投项目每年折旧摊销、相关费用等情况，补充说明每年成本费用对发行人经营业绩的影响，结合盈利影响情况，说明在发行人收入及盈利水平较低的情况下，以较大金额投入不产生收益项目的必要性、合理性；结合前述情况及公司收入大幅下滑、前募效益较低等情况，说明公司本次募投项目建设规模是否与发行人的收入规模相匹配，募集资金的合理性及必要性；（4）

补充说明手机、电脑等个人设备的具体应用及作为本次募投项目的合理性、必要性；220 台 GPU 服务器的具体应用，是否会受到近期美国扩大芯片出口管制，尤其是对高算力 AI 芯片更加严格管控的影响，如是，项目实施有无替代措施，是否存在障碍；（5）前募资金使用情况报告中补充说明前募项目历年效益情况，说明前募效益测算的依据，效益较低的合理性；前募资金资本化支出比例是否在此前的申报文件中披露，将大额资金变更用于房产购置导致补流比例低于变更前、以变更后的补流比例未超变更前作为符合要求的依据是否充分。

请保荐机构、发行人律师及申报会计师核查并发表明确意见。

回复：

一、用简单通俗易懂的语言说明本次募投项目的具体内容，补充说明数据生产垂直大模型项目与现有的 AI 训练数据开发的相关技术差异，是否属于发行人现有主营业务，与公司现有主营业务关系

（一）本次募投项目的具体内容

1、人工智能产业链中的核心要素及在大模型时代下的变迁

在 AI 产业链中，算力、算法和数据共同构成人工智能发展的三大核心要素。若将人工智能比作人脑，算力对应脑内神经元和突触的数量和类型，算法对应神经元和突触之间的连接和信息传递机制，数据则对应人脑学习的教材或外部信息。以算力为基础支撑，算法模型从技术理论到应用实践的落地过程需要依赖大量的训练数据，训练数据规模越大、类型越丰富、内容越完整、质量越高，模型推断的结论越可靠。就如同人脑需要通过吸收越来越多、越来越丰富的完整高质量教材或外部信息，不断完善神经元和突触之间的连接和信息传递机制。



过去十年，人工智能产业的发展以算法为核心，深度学习算法的持续创新是人工智能产业得以快速发展的基础。未来，随着算法趋于开源、算力大幅提升以及人工智能模型应用实践的落地，数据将在人工智能产业中发挥更重要的作用。

近年来，随着 ChatGPT 成为全球范围内的现象级应用，人工智能迎来了新的发展机遇，大模型技术已经成为人工智能新的发展范式。相比于传统深度学习模型，大模型具备如下差异化特征：

(1) 更大的运算参数规模：传统深度学习模型的运算参数量通常在数百万到数千万量级之间，而大模型的参数量则高达数百亿、甚至数千亿量级；

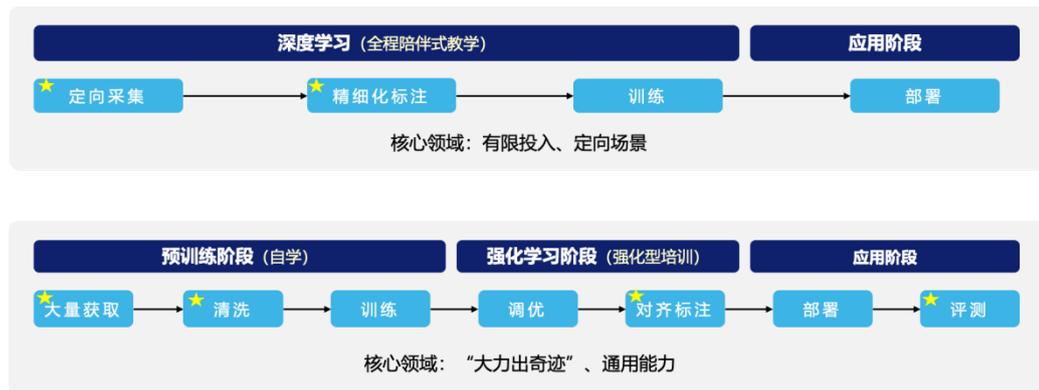
(2) 更强的学习与更广泛的应用方向：模型参数量的几何级增长导致模型拥有更高的复杂度和算力要求，大模型能够更充分地捕捉数据之间的复杂关联，不断提高学习能力，类比于人脑内的神经元和突触之间连接和信息传递机制变得更为活跃，使得人的脑力/智商骤增，具备学习和理解复杂及海量的大千世界数据内容的的能力，应用方向也更为广泛。

差异点	深度学习	大模型
模型参数量	数百万到数千万量级	数百亿到数千亿量级
模型算力	通常个人计算机配套服务器即可满足，一般无需使用大量 GPU 服务器	大量 GPU 服务器，例如数万至数百万个 GPU 服务器训练时长
模型应用方向	<ul style="list-style-type: none"> 主要应用于对事物的判断 专注于特定任务的判断，如日常生活事件的对话交互 	<ul style="list-style-type: none"> 应用更为广泛，除判断外，还能够自己生成内容，如创作一篇文章 能够判断的内容更为丰富、涵盖面更广泛，如各种载体的信息（文本、

差异点	深度学习	大模型
		音视频、图片等），各种专业领域的知识（金融、教育、医疗、政务等）

2、大模型和深度学习模型在学习和训练方式上的差异

在上述背景下，大模型和深度学习模型在学习和训练方式（即如何教导和训练机器）上产生了显著区别，具体如下：



其中，深度学习类比全程陪伴式教学，需要高浓度的定向知识采样以及精细化的标注，以追求最高效的成绩。该类学习模型损耗较低，不太会出现偏离预期的情况，输入多少知识即学多少知识，但也很难产生泛化（即举一反三）和涌现（即自己能够领悟知识点并应用）的能力。大模型则类比自学（预训练阶段）+ 培训（强化学习）的方式，在自学阶段模型首先需要博览群书完成基础教育阶段的学习，形成通识能力，形成自己的价值观雏形；其后在培训阶段，通过专业指导反馈和实践，不断进行纠正学习（类似错题本机制），持续优化在各个垂直应用领域的决策能力，不断提升人机交互效果。

3、大模型训练数据与深度学习训练数据的差异

大模型和深度学习模型在学习和训练方式的区别，导致大模型范式下的训练数据与传统深度学习训练数据亦显著不同，具体如下：

（1）数据规模和数据内容方面，大模型数据规模非常大、覆盖领域非常全，以满足自学阶段博览知识的需要；

（2）生产方式方面，必须设置清洗、筛选、标注等数据处理过程，以满足自学阶段博览“有用”知识及强化培训的训练机制，较传统深度学习训练需要的

定向采集数据和精细化标注过程具有一定的差异性和复杂性；

(3)数据处理技术方面,传统深度学习训练数据的人工标注方式无法长期、持续满足大模型的训练需求,需要提高数据处理的智能化、自动化水平,降低人工参与的比重。只有高效的数据处理技术才能实现快速的大规模、高质量数据供给,形成数据与模型之间的飞轮效应。

差异点	深度学习训练数据集	大模型训练数据集
数据规模	数据规模通常较小,具体而言: <ul style="list-style-type: none"> ■ 文本:数十万至数千万句的文本数据 ■ 音频/视频:几万至十几万小时的音频或视频数据 ■ 图片:数万至数百万张图片 	数据规模较大,通常包含数百万至数十亿的数据点,具体而言: <ul style="list-style-type: none"> ■ 文本:数十亿至数万亿句的文本数据 ■ 音频/视频:几十万至几百万小时的音频或视频数据; ■ 图片:数百万至数亿张图片
数据内容	<ul style="list-style-type: none"> ■ 数据集内容更专注于特定知识 ■ 以文学为例,为训练模型能够对某类型的文章做出摘要,数据集需要包括部分具有典型性的文章,并在文章内将核心意义词做提取和标注,形成深度学习文本数据集 	<ul style="list-style-type: none"> ■ 数据集内容需要包罗通识类的知识 ■ 以文学为例,为训练模型能够对几乎所有类型的文章都能做出有深度的总结而非只是简单的摘要,数据集需要获取短篇、中篇、长篇小说、散文、诗集等,经过高效清洗和筛选,去除无用数据,并对文本的类型、作品风格、特点、摘要总结、读者评价等进行标注,才能形成大模型文本数据集
生产方式	<ul style="list-style-type: none"> ■ 定向采集,清洗和预处理要求低 ■ 数据精细化标注 	<ul style="list-style-type: none"> ■ 原始数据规模大,清洗和预处理要求高,减少无用数据(缺失值、异常值、错误数据等) ■ 根据不同目的进行数据标注、人工反馈,以保证训练模型的准确性和稳健性
数据处理技术	<ul style="list-style-type: none"> ■ 人工标注为主,智能化、自动化为辅 	<ul style="list-style-type: none"> ■ 人工参与比重较低,智能化、自动化水平较高;只有高效的数据处理技术才能实现快速的高质量数据供给,形成数据与模型之间的飞轮效应

4、本次募投项目的主要内容

综合上述技术变迁与新型模型训练需求,与之相对应,为抓住大模型时代下数据行业的发展机遇,满足大模型厂商的训练数据需求,发行人作为从事人工智能训练数据研发设计、生产及销售的专业第三方数据集厂商,拟通过实施本次募投项目实现两个核心目的:

(1)生产一批数据规模较大、覆盖面较全、具有广泛市场需求的大模型数据集,既完整覆盖文本、语音、视频、图像等多种数据模态,又覆盖金融、教育、医疗、政务、数字人等多种下游大模型应用场景。

多样性的训练数据集可以帮助大模型在各种任务和领域中实现更好的性能和应用价值。以政务领域为例，大模型数据集的使用可以帮助政府部门、法律专业人士和广大人民群众等解决政府事务和公共服务方面的需求，提高政务工作的效率。例如，大模型对北京市、上海市等 12345 市长热线、12320 卫生热线中的常见问题进行在线智能解答，高效地支持了公共服务，解决民生问题。又如，在数字人领域，大模型数据集的使用可以提高虚拟数字人的逼真程度，使其更具吸引力和可用性，助力虚拟电商主播、虚拟教育助手、虚拟导游等在相关行业中的应用和发展。

(2) 通过研发数据生产垂直大模型，提升数据生产处理工序的智能化、自动化和工程化，进而提升数据生产效率，提升公司持续生产数据的能力。即发行人利用大模型的方法，智能化生产数据，更好的满足下游各领域人工智能技术与应用的训练需求。

为更好适应大模型数据处理规模较大、数据处理内容及类型多样、数据处理规则复杂、质量要求更高等特点，大模型数据生产将向更加智能化、自动化、工程化方向发展，发行人计划设计、建设更加智能化、自动化的平台体系，即“数据生产垂直大模型”。该平台架构在基础研究、数据生产智能化算法模型以及生产平台工具等技术方面与现有体系均有一定差异，这些差异是发行人跟随技术变迁驱动自身的技术迭代所产生的，具体差异参见本题回复之“一、用简单通俗易懂的语言说明本次募投项目的具体内容，补充说明数据生产垂直大模型项目与现有的 AI 训练数据开发的相关技术差异，是否属于发行人现有主营业务，与公司现有主营业务关系”之“（二）数据生产垂直大模型项目与现有的 AI 训练数据开发的相关技术差异”。

（二）数据生产垂直大模型与现有的 AI 训练数据开发的相关技术差异

本次募投项目“数据生产垂直大模型研发项目”与公司现有的 AI 训练数据生产平台体系相比，两者实质上均为公司内部数据生产体系，但所服务的下游客户的模型开发需求不同：现有体系服务的是“传统深度学习”技术路线下的模型训练，本次募投项目以服务“大模型”训练数据需求为主要目标，兼顾提升“传统深度学习”的数据服务能力，因此两者体系在设计上存在一定差异，具体差异如下：



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体系（其中灰色部分为公司现有体系，全部蓝色部分为募投项目新增/改造内容）

1、海天瑞声一体化技术支撑平台：数据生产工具平台（上层）

生产平台体系中最上层为数据生产工具平台，主要为数据生产中涉及的各类环节提供相关数据处理工具。本次募投项目是在现有平台工具基础上进行拓展和升级，以使其适配于大模型数据规模大、数据类型多、处理难度高等特点，具备更为高阶的数据处理能力。

例如，相比于公司现有清洗工具，本次募投项目拟开发的数据清洗工具在对大规模数据的清洗效率上具备显著优势：由于传统深度学习框架下，原始语料清洗需要的文本量级通常在 KB 至 MB 级别，所以清洗工具基本以单机单线程的方式实现，实现难度和架构复杂度均较低；但大模型所需数据规模显著提高，通常大语言模型所需的文本量级都是 TB 至 PB 级别，需以服务器集群的方式支持高并发、大吞吐量的清洗工具，以通过多机多线程的方式进行数据清洗。

再例如，相比公司现有标注工具，本次募投项目中对标注工具进行改造，实

现更为复杂的功能。以光学字符识别（OCR）为例，传统深度学习领域的 OCR，其数据处理工具仅需支持图片筛选、区域标框和文字输入等功能即可；但大模型所需数据标注工具除了需要具备上述常规标注能力之外，还需要对于图文关联性（例如图片中多段文字间的互相关系）、图片中非文字部分（例如数学教材中的几何图、科学文献中的实验图表等）、文字部分的语义部分（例如有毒文字的筛选等）等更多复杂功能提供支持。

此外，对于大模型范式下新增的数据集类型，指令数据集的处理需要相应研发新的工具。以指令数据的生成与调优工具为例，需要能够针对不同类型的指令数据（如逻辑推理、生成创作、风格迁移、情感分析、跨语言/跨模态理解等）提供不同的生成、筛选与调优功能，同时适配多种生产界面、交互逻辑与数据处理环节，进而具备一体化能力。

2、海天瑞声数据生产垂直大模型：数据生产智能化算法模型（中层）

数据生产智能化算法模型是支撑数据清洗方案设计、指令数据泛化及迁移、标注数据生成等各数据生产环节对应功能而新增的一层，其作用主要是利用大模型的方法改进数据生产方式，即构建自身的大模型技术底座，并利用该技术底座支撑上层的各类现有平台工具和新增平台工具，促进上层数据生产工具向智能化、自动化方向不断迭代。具体对比情况如下：

对应数据生产环节		是否具备智能化能力	具体功能	对比
设计环节	现有平台	×	-	不具备智能化设计能力，相关方案设计均依赖人工技术团队设计完成。
	本次募投项目平台	✓	下载方案设计、清洗方案设计、指令方案设计、多模态数据方案	引入大模型作为技术底座，通过输入客户个性化数据需求的方式，获得由大模型生成的相应个性化设计方案，仅需花费较少人工完成方案的评审及优化工作即可，大幅提升了设计阶段的处理效率。 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，提升研发效率。
采集环节	现有平台	×	-	现有平台采集的数据均来自于真实场景定向采集，并借助基础的智能化工具，实现采集数据的初步校验（如声音采集过程中的噪音检测、视频采集过程中的亮度检测等），可一定程度提升原始采集数据的基本质量（比如噪音不能高于预设分贝、环境亮度符合预设方案）。

对应数据生产环节		是否具备智能化能力	具体功能	对比
	本次募投项目平台	✓	原始数据生成	<p>引入大模型作为技术底座，在支持真实数据采集的同时，具备数据生成能力，可完成文本、图片、语音等各类数据生成，后续数据处理人员可结合算法需求对各项数据进行筛选、混合、调整后，即可获得高质量原始数据，使原始数据获取不再局限真实场景采集，大幅提升数据获取效率。</p> <p>例如，若想训练提升大模型的逻辑推理能力，大模型需要学习海量高质量数学、物理等题库数据，但目前相关学科电子题库资源较为稀缺，若在现有模式下，需要公司先通过相关学科专家按要求编制题目等方式获得原始题库数据资源。而在本次募投项目建设完成后，可通过输入相关题目要求自动生成题库，而后经过相关专家少量时间复核修订即可完成题库制作，相比现有模式，显著提升数据获取效率。</p>
加工环节	现有平台	✓ (基础能力)	预标注/辅助标注（例如，语音识别、物体检测等）	以传统深度学习框架为基础，结合特定应用场景进行的定向模型开发。例如：针对不同语种开发的多语种语音识别模型，针对智能驾驶场景下不同类型物体（车辆、行人、锥桶等）的识别检测模型等。以多语种语音识别模型为例，目前现有平台具有包括中文、英语、西班牙语等多类主流语种的语音识别模型，借助该类模型先于人工进行预先转写预计可降低约 20%-40%的数据处理成本。
	本次募投项目平台	✓ (高阶能力)	标注结果生成、指令泛化与迁移	<p>(1) 预标注能力增强，降低数据处理/校对成本 例如，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。</p> <p>(2) 将各类预标注功能集中在统一大模型中，提升预标注效率</p>

对应数据生产环节		是否具备智能化能力	具体功能	对比
				例如，多语种语音识别的预标注场景下，鉴于语音大模型所需数据也已不再区分具体语种，往往呈现多语种语音数据混杂的现象，现有单一语种小模型已无法有效支撑该类数据预标注的需求；而在引入本募平台后，无论是否为多国语言，仅需要调用同一个模型即可，预计能够提升 50%以上的预标注效率。
质检环节	现有平台	✓ (基础能力)	基于规则或传统模型的辅助质检能力	仅具有适用于传统深度学习的智能化质检功能，例如同一批次数据的一致性检查（相似语音被错误地转换为不同的文本、相似图片被错误地打上不同的标签等）。
	本次募投资项目平台	✓ (高阶能力)	指令数据验证	与加工环节的预标注功能类似，将各类质检功能集中在统一大模型中，提升质检效率。例如，大模型的训练数据通常需要过滤掉有误导性或价值观偏离的数据，传统模型很难达到可用的准确性，必须依赖大模型的语义理解能力提供智能质检。

3、基础研究（底层）

生产体系中最底层为基础研究，主要为支撑中层的数据生产智能化算法模型的研发提供前期研究和理论基础支持。公司现有体系的基础研究主要针对传统深度学习路线，本次募投项目则侧重于通过开展以大模型范式为主线的基础研究，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。由于两者使用的技术类型存在差异，因此所需进行的底层基础研究也相应产生变化：（1）传统深度学习算法的网络结构相对简单、层数和参数相对较少，而大模型范式下，通常需要理解复杂的网络结构，参数规模达到百亿以上级别，且具备大模型范式下才有的微调技术和基于人类反馈的强化学习技术；（2）由于大模型训练所需的 GPU 服务器等硬件相比传统深度学习在量级上明显提升，进而在模型训练相关的底层工程技术上新增较多研究课题，例如多级多卡并行训练技术等。

（三）本次募投项目是否属于公司现有主营业务，与公司现有主营业务关系

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、人员、生产模式等基础层面共通、产品层面互补、应用领域及客户类型趋同。公司实施本次募投项目旨在跟随技术的发展、更高效的开发训练数据，持续在人工智能产业链上服务下游客户的算法训练需求，与公司自成立之日起至今的产业链位置与主营业务定位一致。本次募投项目是围绕公司现有主营业务开展，属于公司现有业务的延伸，与公司现有业务具有较强的协同关系。

二、2022 年购置大额房产后,本次募集资金进一步购置大额房产,说明 2022 年及本次募投项目购置的房产性质、单价情况,相关单价或单价测算的公允合理性;结合发行人自身定位、未来业务发展规划、收入及盈利水平较低、上市时为轻资产企业等情况,说明频繁大额购置房产行为是否偏离公司主业,是否与公司科技企业定位相符,是否与发行人收入规模、业务经营情况相匹配。结合前述情况,进一步说明本次募集资金较大金额购置房产的必要性、合理性

(一) 2022 年及本次募投项目购置的房产性质、单价情况,相关单价或单价测算的公允合理性

发行人 2022 年购置的房产,其不动产权证证载用途为“办公”,该房产非住宅类或商业类房地产,不涉及住宅类或商业类房地产开发与经营业务,采购单价为 3.9 万元/平米;本次募投项目拟购置的房产性质亦为办公,不涉及住宅类或商业类房地产开发与经营业务,预计采购单价为 4.8 万元/平米,略高于发行人 2022 年购置房产的采购单价,主要原因为具体房产所处地域、房龄、成交市场等个体因素存在差异所致,本次拟购置房产位于北京经济技术开发区”北京数据基础制度先行区”区域内,该区域的房产大多为楼龄较新的信息化产业园区办公楼,而前次 2022 年购置的房产位于北京市东城区,楼龄较长,且卖方因其资金回流需求,成交价亦略低于当时市场价。

本次采购单价参考了拟购买区域北京经济技术开发区及周边同类型场地的销售均价,测算具体情况如下:

序号	标的位置	单价(万元/平米)
1	科创十二街-天通泰科技金融谷	5.36
2	荣华南路-国锐金嶺	4.58
3	天华北街-富兴国际	4.00
平均单价		4.64
拟购置房产单价		4.80



经网络查询，对比本次拟购买标的周边其他同类型场地的单价售价，价格区间处于 4.00 万元/平方米到 5.36 万元/平方米之间，本次拟购买房产标的单位售价略高于平均单价及公司 2022 年采购单价，但属于合理区间，符合当地市场价格，相关单价测算具有公允性、合理性。

(二) 结合发行人自身定位、未来业务发展规划、收入及盈利水平较低、上市时为轻资产企业等情况，说明频繁大额购置房产行为是否偏离公司主业，是否与公司科技企业定位相符，是否与发行人收入规模、业务经营情况相匹配。结合前述情况，进一步说明本次募集资金较大金额购置房产的必要性、合理性

公司本次使用募集资金购置房产主要是基于相关政策对合规运营数据资源的相关要求、满足募投项目的业务和人员需求、保障募投项目的顺利实施以及购置房产相较于租赁房产更具经济性等多方面因素考量，募投项目紧密围绕发行人主营业务、自身定位和未来发展规划进行，未偏离公司主业，与发行人科技企业定位相符，与发行人收入规模、业务经营情况相匹配，具体情况如下：

1、顺应“数据基础制度先行先试”区域政策规划，在区域内购置房产是本次募投项目合规运营数据资源的必要条件

本次募投项目拟购置房地点拟位于北京经济技术开发区“北京数据基础制

度先行区”区域内，主要背景原因为根据中共北京市委、北京市人民政府 2023 年 7 月发布的《关于更好发挥数据要素作用进一步加快发展数字经济的实施意见》，指定北京经济技术开发区等开展数据基础制度先行先试，推动包括公司在内的第三方专业数据服务商依托在该区域内先行先试的基础制度，合规运营各类大模型数据集所需要的部分原始数据（如公共数据）。根据先行区“原始数据不出域”的相关要求，需在先行区内完成原始数据获取和加工，因此公司需在该区域内开展本次募投项目相关业务。此外，政府部门将利用财政资金支持数据服务产业发展，通过资金奖励、产业发展基金等形式，加大对先行区内数据要素型企业的投入力度，公司在本区域落地、发展也将获得相关资金、政策支持，享受先行区政策红利，进一步保障募投项目的顺利实施。

2、购置办公场地是满足本次募投项目新增人员办公及业务需求的必要举措

公司现有办公场所以租赁为主（注），用于满足现有业务办公场地需求（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场所的需求扩张等），无法满足本次募投项目新增人员办公及业务需求。本次募投项目人均办公面积以人均 10 平米测算假设，略高于公司目前人均办公面积，且两者均低于同行业公司，具有合理性。

单位：平米/人

公司现有场地人均面积	公司募投项目人均办公面积	公司募投项目人均办公面积（含展厅和多模态采集室）	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
8.35	10.00	13.44	16.24

注：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租用的主要办公场地，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租用的主要办公场地后，公司预计将节省相关房产租赁支出费用，且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。

3、购置房产可以满足研发活动的保密性要求，有助于优化研发环境，促进研发能力提升

由于本次募投项目建设会涉及到部分公共数据和敏感信息，对办公场地间存在一定的保密性要求，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、建设达到相关安全可控要求，自有场所从事研发活动也更能有效保护商业秘密不致泄露，以保证募投项目实施的稳定性。发行人购置房产能够为本次募投项目相关人员提供长期、稳定的办公场所，为产品和技术的开发提供专业的研发环境，促进研发能力提升。

4、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

从经济性角度，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，中长期看，公司购置房产相较于租赁房产更具有经济性。

项目	每年新增费用（万元）
购置房产产生的折旧费	424.51
租赁房产产生的租金	856.29

5、本次购置房产主要服务于募投项目建设，募投项目紧密围绕发行人主营业务、自身定位和未来发展规划进行，未偏离公司主业，与公司科技企业定位相符，与发行人收入规模、业务经营情况相匹配，本次募集资金购置房产具有必要性、合理性

根据上述分析，本次购置的房产主要服务于本次募投项目“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”的建设。发行人作为从事人工

智能训练数据研发设计、生产及销售的专业第三方数据集厂商，拟通过实施本次募投项目，生产一批数据规模较大、覆盖面较全、具有广泛市场需求的大模型数据集，并通过研发数据生产垂直大模型，提升数据生产处理工序的智能化、自动化和工程化，进而提升数据生产效率，提升公司持续生产数据的能力，符合公司“夯实传统业务，探索新型业务”的战略目标。公司将基于过往的数据服务经验，结合行业前沿需求，积极拓展大模型训练数据服务领域，力争将大模型训练数据等创新业务打造成为具有潜在高增长价值的新型业务板块。因此本次募投项目属于公司现有主营业务的拓展及延伸，符合公司自身定位和未来发展规划，未偏离公司主业，与公司科技企业定位相符。

报告期内，发行人的收入及盈利水平有所波动，该波动主要与下游客户 AI 应用行业的整体发展密切相关。近年来智能语音、智能驾驶以及大模型等下游 AI 应用行业面临不同的发展态势，下游客户根据市场发展变化进行需求和预算调整，导致发行人 2023 年收入出现下滑，但 AI 行业长期向好趋势未发生改变。此外，发行人作为专业数据服务商已构建护城河优势，未来收入具备复苏潜力，同时已启动各项费用管控举措，以提升公司盈利能力。具体情况参见本回复之“2. 关于经营情况”之“一、结合发行人 2023 年季报、在手订单及订单交付情况，说明发行人 2023 年是否面临收入大幅下滑及亏损的风险，发行人的持续经营能力与问询回复披露是否一致，是否发生重大变化，公司经营状况和外部环境是否发生重大变化，相关风险是否已经充分提示”之“（二）发行人的持续经营能力与问询回复披露是否一致，是否发生重大变化，公司经营状况和外部环境是否发生重大变化，相关风险是否已经充分提示”。本次购置房产是募投项目顺利实施的必要保障，有助于公司探索寻求新的业绩增长点，与发行人收入规模、业务经营情况相匹配。

此外，根据公开信息检索，与发行人同属软件和信息技术服务业的科技型轻资产上市公司，亦存在较多通过募投项目进行房产购置，以解决募投项目中对运营与办公场地需求的相关案例，具体情况如下：

公司名称	首次申报时间	类型	购置房产的募投项目	购置面积 (平方米)	募集资金 购房金额 (A, 万 元)	募集资金 总额 (B, 万元)	购房金 额占比 (C=A/ B)
赢时胜 (300377. SZ)	2023年 4月	再融资	资管 AMS 平台信创项目	1,940	10,398.40	100,895.00	24.44%
			托管 ACS 平台信创项目	1,710.00	9,165.60		
			金融科技中台创新项目	950.00	5,092.00		
安博通 (688168. SH)	2022年 9月	再融资	数据安全防护与溯源分析平台研发及产业化项目	未披露	3,800.00	13,528.48	28.09%
	2019年 4月	首次公开发行	深度网络安全嵌入系统升级与其虚拟资源池化项目	5,874.11	7,500.00	29,774.00	25.19%
			安全可视化与态势感知平台研发及产业化项目				
安全应用研发中心与攻防实验室建设项目							
东方通 (300379. SZ)	2022年 6月	再融资	Tong 系列中间件产品卓越能力提升项目	11,250.00	56,250.00	220,000.00	25.57%
发行人 (688787. SH)	2023 年 8 月	再融资	AI 大模型训练数据集建设项目	3,450.00	16,560.00	66,600.00	28.18%
			数据生产垂直大模型研发项目	460.00	2,208.00		

综上所述，发行人虽为“轻资产”属性，但公司的业务模式下，募投项目的实施涉及了项目经理、研发工程师及参与数据标注及清洗环节的稳定加工人员等多种人员类型，存在对办公场地的合理需求。鉴于本次使用募集资金购置房产主要是基于相关政策对合规运营数据资源的相关要求、满足募投项目的业务和人员需求、保障募投项目的顺利实施以及购置房产相较于租赁房产更具经济性等多方面因素考量，募投项目紧密围绕发行人主营业务、自身定位和未来发展规划进行，未偏离公司主业，与发行人科技企业定位相符，与发行人收入规模、业务经营情况相匹配，公司本次拟使用募集资金购置房产具有必要性、合理性。

三、结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况，说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确；结合前述情况及效益测算中收入、成本费用等关键指标的具体测算过程及依据等，说明“AI 大模型训练数据集建设项目”效益测算较高的合理性，与发行人现有水平及同行业可比公司相比，相关测算是否谨慎、合理；“数据生产垂直大模型研发项目”募投项目每年折旧摊销、相关费用等情况，补充说明每年成本费用对发行人经营业绩的影响，结合盈利影响情况，说明在发行人收入及盈利水平较低的情况下，以较大金额投入不产生收益项目的必要性、合理性；结合前述情况及公司收入大幅下滑、前募效益较低等情况，说明公司本次募投项目建设规模是否与发行人的收入规模相匹配，募集资金的合理性及必要性

（一）结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况，说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确

1、发行人大模型相关在手订单情况

目前发行人的较多客户已纷纷开展大模型领域的研发和应用，且其中一些客户的大模型产品已经正式推出上线，大模型业务的主要客户及营销渠道与公司现有业务重合度较高，有助于公司打开市场。以美团、科大讯飞、阿里巴巴、腾讯、百度等公司现有业务的主要客户为例，这些客户亦为大模型领域占据主要市场地位的企业，并与公司签署了大模型领域的业务合同或意向性订单。公司已与相关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，预计截至 2023 年末在手合同及意向订单金额合计约 2,900 万元。此外，公司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。

目前大模型领域的业务需求尚未完全释放，主要系大模型客户自身的大模型业务也尚处于发展早期阶段，且由于算力规模直接决定对话模型的反应速度和用户体验，是大模型客户初代产品上线的基本条件，因此大模型客户前期的投入主要集中在算力，训练数据前期主要以其自有和开源数据为主。随着大模型应用的

渗透率逐步提升，为提供更有价值的服务，获取和保持用户持续付费意愿，大模型能力与具体应用结合成为亟待解决的问题，数据将发挥决定性的作用，大规模、高质量的增量通用训练数据、覆盖更多垂直领域的场景训练数据等对于大模型优化和投入应用的效果愈加凸显，大模型客户外购训练数据的需求亦会明显增长。

2、大模型的市场需求情况

据艾瑞咨询预测，2023 年中国 AIGC（人工智能生产内容）产业规模约为 143 亿元，随后进入大模型生态培育期，持续打造与完善底层算力基建、大模型商店平台等新型基础设施，以此孕育成熟技术与产品形态的对外输出。2028 年，中国 AIGC 产业规模预计将达到 7,202 亿元，中国 AIGC 产业生态日益稳固，完成重点领域、关键场景的技术价值兑现，逐步建立完善模型即服务产业生态。

2022-2030年中国AIGC产业规模



数据来源：艾瑞咨询《2023 年中国 AIGC 产业全景报告》

2020 年，随着国内首批语言大模型发布，行业正式进入萌芽期，2022 年中国大模型的人工智能市场规模达到 891.2 亿元。根据沙利文预测，2027 年中国 AI 大模型市场规模为 2,777.5 亿元，2022-2027 年年复合增长率为 25.53%，行业呈高速发展态势。



数据来源：沙利文《2023年中国AI技术变革企业服务白皮书》

在大模型快速发展的背景下，数据已成为决定各类通用及垂直大模型应用及表现的关键因素。

根据沙利文《2023年中国AI技术变革企业服务白皮书》，大语言模型的底层数据质量是重要的竞争壁垒构建之一：以ChatGPT为例，该模型在编程代码撰写方面表现卓越，部分归因于其对Github底层数据源的访问（作为全球领先的编程代码社交平台Github积累了大量的优秀代码样本）。根据《北京市人工智能行业大模型创新应用白皮书》，“高质量数据已成为严重掣肘大模型行业应用的关键。行业大模型对于特定行业数据提出了更高的要求，高质量的行业数据集成为企业之间的竞争壁垒，对行业知识和业务逻辑的深入理解是行业大模型落地应用的重要竞争优势……从技术供给方和需求端看，目前市面上缺乏高质量公开的数据集，大模型厂商无法迭代准确面向行业落地场景训练大模型；多数行业数据丰富的企业用户又缺乏自研大模型的能力，严重掣肘行业模型研发和应用落地。例如，政务、医疗、智慧城市等领域数据获取难度大，隐私性要求高阻碍了大模型应用落地的进程。”

3、截止2023年三季度末收入大幅下滑情况

报告期内，发行人的收入及盈利水平有所波动，该波动主要与下游客户AI应用行业的整体发展密切相关。近年来智能语音、智能驾驶以及大模型等下游AI应用行业面临不同的发展态势，下游客户根据市场发展变化进行需求和预算

调整，导致发行人 2023 年收入出现下滑。虽然发行人最近一期下滑幅度较大，但 AI 下游应用的蓬勃发展和新型应用的不断涌现将带领行业进入下一个阶段，公司依托积累的客户资源和技术优势将有望在此背景下迎来二次增长，长期来看收入增长具备可持续性，发行人预计截至 2023 年底在手订单金额达到 1.39 亿元，同比增长 106%，具体情况参见本回复之“2.关于经营情况”之“二、补充说明发行人在手订单和意向性订单的具体情况，结合 2023 年三季度及四季度订单执行情况，说明问询回复中关于 2023 年全年收入不存在大幅下滑风险的依据是否充分”。随着大模型技术不断向纵深发展，将会带来自然语言处理领域的的数据需求，提升发行人自然语言处理业务收入。此外，随着大模型能力的演进，未来将从以文本交互为主要方式拓展到多模态交互，亦为智能语音、计算机视觉等多模态数据领域带来新的增量需求空间。发行人本次募投项目“AI 大模型训练数据集建设项目”即为发行人延伸大模型数据业务的有力抓手，有助于发行人抓住大模型发展机遇，寻求新的业绩增长点。

综上所述，发行人经营的外部环境未发生重大不利变化、相关业务有较大市场增长潜力的预测依据准确。

（二）结合前述情况及效益测算中收入、成本费用等关键指标的具体测算过程及依据等，说明“AI 大模型训练数据集建设项目”效益测算较高的合理性，与发行人现有水平及同行业可比公司相比，相关测算是否谨慎、合理

结合前述情况，将本次募投项目效益测算中收入、成本费用等关键参数的选取依据与发行人最近 5 年及最近 3 年的现有水平进行对比分析，相关参数选取具有谨慎性、合理性。具体说明如下：

1、营业收入

（1）产品开发投入

募投开发投入根据拟投资建设的数据集规模及数据集单位投入金额进行测算，不同类别数据集单位投入金额与公司最近 5 年（2018-2022 年）/最近 3 年（2020-2022 年）的现有同类数据单位投入金额比较情况如下，不存在明显差异。

数据集类型	数据单位	本次募投产品单位投入金额（元）	公司现有同类数据单位投入金额（元）-最近五年	公司现有同类数据单位投入金额（元）-最近三年	
通用及特定垂直领域的大语言模型训练数据集	万句	3.78	4.19	-（注1）	
多模态大模型训练数据集（注2）	全场景自动驾驶数据集（注3）	视频小时	203.90	144.37	144.37
	多语种语音识别大模型数据集	音频小时	480.00	464.93	517.33
	多语言图文识别大模型数据集	张	18.60	18.92	17.66
	多语种语音生成大模型数据集	音频小时	6,742.86	6,119.67	6,025.96
	多语言图文生成大模型数据集	张	0.84	0.89	-（注4）
	数字人多模态数据集	视频小时	5,920.00	6,050.66（注5）	6,050.66（注5）
大模型评测数据集（注6）	条	8.75	-	-	

注1：公司最近三年同类数据集仅以中文语料为主，未包含其他语种，不具有可比性；

注2：多模态大模型训练数据集因其本身多模态性质，不同细分数据集的数据单位及单位投入金额具有较大差异，因此按照数据集明细进行列示；

注3：全场景自动驾驶数据集的单位投入金额高于现有同类产品单位投入金额，主要原因在于公司现有同类数据集主要以单传感器的视频/图片为主（“单一场景”），而全场景自动驾驶数据集拟包含多个传感器（多角度摄像头、毫米波雷达、激光雷达等）数据，同时多传感器之间的数据要严格同步，因此相比之下单位投入金额较高；

注4：公司与多语言图文生成大模型数据集可比的同类数据，主要于2020年之前完成制作；

注5：鉴于公司历史上与数字人多模态数据集同类的的数据规模较小，此处补充将2023年1-6月的同类数据纳入对比范围；

注6：大模型评测数据集在本次募投项目拟建设数据的整体规模中占比较少，主要应用于各类大模型的能力、任务、指标等方面的评测；公司最近5年（2018-2022年）不存在与大模型评测数据集完全可比的同类数据。

（2）加成率

公司根据历史期间训练数据集产品的开发投入及首次销售收入计算本次募投项目的加权平均加成率，最近5年（2018-2022年）为25%，最近3年（2020-2022年）为29%，不存在明显差异，公司选取25%作为本次募投项目产品的加成率更为谨慎。

（3）当期销售率、期后复卖率

公司根据2017-2022年各年度新制成训练数据集产品的销售明细预估本次募投项目的当期销售率、期后复卖率，平均当期销售率为92%，期后平均复卖率区

间为 44%-86%；仅考虑最近 3 年（2020-2022 年），平均当期销售率为 94%，期后平均复卖率区间为 51%-86%，均高于 2017-2022 年的统计结果/区间。当前募投项目选取较低者作为效益测算参数更为谨慎。

2、销售费用

本次募投项目的销售费用率 5%系参考公司和同行业水平进行预测，公司最近五年（2018-2022 年）平均销售费用率为 5.47%，最近三年（2020-2022）平均销售费用率为 6.26%，不存在明显差异。

3、管理费用

本次募投项目的管理费用率 12.14%系参考公司和同行业水平进行预测，公司最近五年（2018-2022 年）平均管理费用率为 15.50%，最近三年（2020-2022 年）平均管理费用率为 17.40%，参考期间平均管理费用率略高，主要是由于 2021 年之后上市财经公关及宣传费、房产购置中介服务费等因素产生的中介服务费支出所致。发行人拟通过进一步加强成本及费用管控，严格落实相关措施，保证募投项目的高效推进。

4、研发费用

由于建设期和纯运营期的研发费用率存在较大差异，本次募投项目的研发费用率系区分建设期和纯运营期并参考与公司可比期间口径的研发费用率进行预测，具体情况如下：

单位：万元

序号	项目	建设期			纯运营期				
		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	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
3	研发费用率	97%	78%	72%	29%	25%	20%	16%	13%
平均研发费用率		77%			21%				
可比期间研发费用率（注）		71%			21%				

注：（1）关于建设期，可比期间选取公司前次 IPO 募集资金建设期的整体研发费用率作为参考；（2）关于纯运营期，鉴于公司自 2021 年开始大规模投入 IPO 募投项目建设，因此 2021 年及之后的研发费用构成既包括前期产品的运营期投入，也包括 IPO 募投产品的

建设期投入，且后者金额较大，不具备可比性，故选取公司 2018-2020 年（IPO 募投项目建设期之前）的平均产品研发费用率作为对比。

此外，本次项目效益测算关键指标与同行业同类项目相比，亦具有谨慎性和合理性，具体如下：

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，公司数据集产品为可进行多次授权销售的标准化产品，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%，同行业同类项目未单独披露类似标准化产品的毛利率，项目整体毛利率均处于较高水平（大于 80%）。

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

综上所述，“AI 大模型训练数据集建设项目”的效益测算与发行人现有水平及同行业可比公司相比，具有谨慎性和合理性。

(三)“数据生产垂直大模型研发项目”募投项目每年折旧摊销、相关费用等情况,补充说明每年成本费用对发行人经营业绩的影响,结合盈利影响情况,说明在发行人收入及盈利水平较低的情况下,以较大金额投入不产生收益项目的必要性、合理性

1、“数据生产垂直大模型研发项目”募投项目每年折旧摊销、相关费用情况,每年成本费用对发行人经营业绩的影响

“数据生产垂直大模型研发项目”募投项目每年产生的折旧摊销及相关费用情况具体如下:

单位:万元

项目	建设期			纯运营期				
	第一年	第二年	第三年	第四年	第五年	第六年	第七年	第八年
技术人员工资	1,050.00	1,667.50	2,185.00	2,310.00	2,310.00	2,310.00	2,310.00	2,310.00
折旧	1,583.35	3,300.74	4,747.02	5,080.74	5,068.27	3,529.68	1,824.43	394.13
摊销	34.07	54.27	65.82	65.82	65.82	61.41	59.20	57.36
托管费	555.42	983.49	1,344.34	1,344.34	1,344.34	1,344.34	1,344.34	1,344.34
合计	3,222.84	6,005.99	8,342.18	8,800.90	8,788.43	7,245.42	5,537.96	4,105.83

上述各年成本费用将导致公司净利润的下降,随着研发项目的推进,影响金额将在第五年后开始降低。

2、发行人收入及盈利水平较低情况下,以较大金额投入不产生收益项目的必要性、合理性

发行人报告期内的收入及盈利水平有所波动,但长期来看收入增长具备可持续性,具体参见本题回复之“三、结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况,说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确.....”之“(一)结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况,说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确”之“3、截止 2023 年三季度末收入大幅下滑情况”。公司投入“数据生产垂直大模型研发项目”的必要性及合理性如下:

(1) 本项目建设是巩固公司的核心技术壁垒，构建长期技术实力支撑的必然手段

数据集生产能力和一体化技术支撑平台是公司核心技术的重要体现。本项目的建设将基于公司在深度学习阶段数据集生产所积累的 know-how，自主研发数据生产垂直大模型，构建大模型数据处理技术通用化解决方案能力，实现完整、可持续迭代的大模型数据技术框架和数据策略，进一步提高公司在人工智能基础数据服务领域的智能化水平，巩固公司的核心技术壁垒，形成长期技术实力支撑。在人工智能技术迅速发展的当下，作为我国人工智能产业链中数据行业的主要从业企业，发行人切实需要提前进行前沿技术储备和建设，为公司的长期发展打下更为坚实的基础，避免技术掉队的情况发生。

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

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

(3) 本项目建设是公司落实国家科技创新发展战略的重要举措

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

(4) 本项目建设有助于长期提升发行人整体盈利水平

虽然“数据生产垂直大模型研发项目”项目不直接产生收益，相关固定资产折旧、摊销及费用增加将导致公司未来期间利润下滑，但作为巩固公司的核心技术

壁垒，构建长期技术实力支撑的必然手段，优化公司数据集生产流程，提升数据服务综合竞争力的有力途径，本项目将有助于长期提升发行人的整体盈利水平。

综上所述，“数据生产垂直大模型研发项目”具有必要性、合理性。

（四）结合前述情况及公司收入大幅下滑、前募效益较低等情况，说明公司本次募投项目建设规模是否与发行人的收入规模相匹配，募集资金的合理性及必要性

公司最近三年一期的收入规模分别为 23,337.40 万元（2020 年）、20,647.65 万元（2021 年）、26,288.79 万元（2022 年）及 10,321.46 万元（2023 年 1-9 月）。除 2023 年 1-9 月因上述相关原因收入下滑外，公司最近三年营业收入稳定在 2 亿元以上。公司最近一期收入大幅下滑的具体原因参见本题回复之“三、结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况，说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确……”之“（一）结合发行人的在手订单、本次募投项目的市场需求及截止 2023 年三季度末收入大幅下滑情况，说明发行人经营的外部环境是否发生变化、相关业务有较大市场增长潜力的预测依据是否准确”之“3、截止 2023 年三季度末收入大幅下滑情况”。

公司前次募集资金投资项目已于 2023 年 5 月结项，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建设期内实际产生的收入为 10,784.83 万元，利润总额为 99.60 万元。前募效益绝对金额较低，主要原因系其为建设期产生的效益。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。

随着 AI 下游应用的蓬勃发展和新型应用的不断涌现将带领行业进入下一个阶段，公司依托积累的客户资源和技术优势将有望在此背景下迎来二次增长，长期来看收入增长具备可持续性，未来随着公司经营规模持续扩大以及募投项目投产并逐步释放效益，公司收入规模预计可合理支撑本次融资规模。发行人在手订单和意向性订单的具体情况参见本回复之“2.关于经营情况”之“二、补充说明发行人在手订单和意向性订单的具体情况，结合 2023 年三季度报及四季度订单执行情

况，说明问询回复中关于 2023 年全年收入不存在大幅下滑风险的依据是否充分”。

综上所述，本次募投项目具有合理性及必要性，其中“AI 大模型训练数据集建设项目”为公司响应国家建立数据基础制度，落实北京建设“北京数据基础制度先行区”的必然选择，践行国家规范生成式人工智能产品要求的重要举措，支撑大模型训练，提升大模型输出能力的有效方式，符合公司“夯实传统业务，探索新型业务”的战略目标；“数据生产垂直大模型研发项目”为巩固公司的核心技术壁垒，构建长期技术实力支撑的必然手段，优化公司数据集生产流程，提升数据服务综合竞争力的有力途径及落实国家科技创新发展战略的重要举措，有助于长期提升发行人整体盈利水平。

综上所述，本次募投项目建设规模与发行人的收入规模基本匹配，募集资金具有合理性及必要性。

四、补充说明手机、电脑等个人设备的具体应用及作为本次募投项目的合理性、必要性；220 台 GPU 服务器的具体应用，是否会受到近期美国扩大芯片出口管制，尤其是对高算力 AI 芯片更加严格管控的影响，如是，项目实施有无替代措施，是否存在障碍

（一）手机、电脑等个人设备的具体应用及作为本次募投项目的合理性、必要性

本次募投项目拟购入 291 台台式机电脑、291 台笔记本电脑、200 部手机、40 台平板电脑。其中，台式机及笔记本电脑为项目人员的个人办公设备，与公司目前人员设备配置情况一致，主要原因为公司研发人员的数据处理工作量较大，对于办公设备的性能需求较高，因此每人通常需配备两台设备协同操作；手机及平板电脑并非个人设备，而是作为语音及多模态原始数据的采集获取工具使用，是大语言模型及多模态大模型数据集产品必需的生产设备。因此，公司将手机、电脑等设备作为本次募投项目的投入设备具有合理性及必要性。

(二) 220 台 GPU 服务器的具体应用，是否会受到近期美国扩大芯片出口管制，尤其是对高算力 AI 芯片更加严格管控的影响，如是，项目实施有无替代措施，是否存在障碍。

1、GPU 服务器的具体应用

GPU 服务器主要用于“数据生产垂直大模型研发项目”中的大模型训练、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（如下表所示）。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、是否会受到近期美国扩大芯片出口管制的影响及相关替代措施

2023 年 10 月，美国公布新的先进计算芯片出口管制规则，扩大限制我国对高端 GPU 的购买。鉴于目前国内人工智能领域所使用的 GPU 服务器仍主要向境外供应商采购，上述芯片出口管制政策出台一定程度上将增加本次募投项目采购相关品牌服务器的难度。

但与此同时，近年来，在政策和市场的双轮驱动下，国产算力已逐步成为国产 AI 大模型的可靠选择。我国目前已拥有华为、新华三、中兴、浪潮等多家信息技术综合服务厂商，在服务器及交换机等网络设备领域具有丰富的产品类型和较好的服务保障能力，且其中部分厂商已推出对标境外品牌的 GPU 服务器。在上述背景下，本次募投项目可购置国产品牌 GPU 服务器，因此近期美国扩大对高算力 AI 芯片出口管制不会对公司募投项目的开展具有重大不利影响，且项目实施存在替代措施，不存在实质性障碍。

五、前募资金使用情况报告中补充说明前募项目历年效益情况，说明前募效益测算的依据，效益较低的合理性；前募资金资本化支出比例是否在此前的申报文件中披露，将大额资金变更用于房产购置导致补流比例低于变更前、以变更后的补流比例未超变更前作为符合要求的依据是否充分。

(一) 前募资金使用情况报告中补充说明前募项目历年效益情况，说明前募效益测算的依据，效益较低的合理性

公司前次募集资金投资项目已于 2023 年 5 月 31 日结项，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。

截至 2023 年 6 月，前次募集资金投资项目“自主研发数据产品扩建项目”实际产生的收入为 10,784.83 万元，利润总额为 99.60 万元。发行人补充说明前募项目历年效益情况及前募效益测算的依据如下：

单位：万元

项目	2020 年下半年	2021 年度	2022 年度	2023 年上半年	合计
项目收入	712.82	3,189.70	5,502.43	1,379.88	10,784.83
利润总额	(74.64)	(101.01)	329.18	(53.93)	99.60

前募效益的测算系依据“自主研发数据产品扩建项目”所形成的新建训练数据集产品口径归集产生的效益情况。其中，前次募集资金项目的项目收入为前募新建训练数据集产品销售产生的收入，前次募集资金项目利润总额为前募新建训练数据集产品产生的收入扣除研发训练数据产品发生的研发费用支出以及相关的销售及管理费用等支出后的金额。

如上表所示，前次募集资金投资项目“自主研发数据产品扩建项目”实际产生的收入为 10,784.83 万元，利润总额为 99.60 万元。前次募集资金投资项目实现的利润总额较小是由于前次募集资金投资项目于今年 5 月 31 日结项，对应的募投项目效益主要处于项目建设阶段所致。

(二) 前募资金资本化支出比例是否在此前的申报文件中披露, 将大额资金变更用于房产购置导致补流比例低于变更前、以变更后的补流比例未超变更前作为符合要求的依据是否充分

前募资金资本化支出比例已在此前报送的《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“2. 关于前次募投项目披露”回复中披露如下:

1、前次 IPO 募集资金到账后使用计划 (调整前)

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
	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%

2、前次 IPO 募集资金实际使用情况 (调整后)

根据发行人 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, 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。上述事项已经公司董事会及监事会审议通过，独立董事发表了明确同意独立意见，该事项履行了必要的审批程序。

前次 IPO 募集资金投资项目调整内部投资结构后，前次 IPO 募集资金实际使用情况如下：

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%。

公司前次 IPO 募投项目实际用于非资本性支出的金额占前次募集资金净额的比例为 64.54%，未超过募集资金到账后使用计划中非资本性支出的比例 68.98%。因此，发行人本次融资的募集资金规模未进行调整。

六、核查程序及核查意见

（一）核查程序

保荐机构、发行人律师及申报会计师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、发行人关于本次募投项目内容的说明文件，了解本次募投项目的具体内容、与公司现有技术的差异、与主营业务的关系等；

2、查阅了发行人 2022 年的房产购置及相关协议文件、房产不动产权证书，检索了本次募投项目拟购买房产区域及周边同类型场地的销售均价，了解相关房产性质及单价情况；查阅了本次募投项目的可行性研究报告及测算依据、关于“数据基础制度先行先试”区域政策规划的相关文件，检索了同行业上市公司通过募投项目进行房产购置的相关案例，了解本次募集资金购置房产的必要性、合理性；

3、查阅了本次募投相关的行业研究报告、发行人大模型领域的业务合同及意向订单，了解相关业务的市场增长潜力；查阅了本次募投项目的可行性研究报告及测算依据，了解募投项目效益测算、折旧摊销及相关费用情况、对发行人经营业绩的具体影响以及募集资金的合理性、必要性；查阅发行人前次募集资金投资项目的可研报告等资料，了解前次募集资金使用情况及效益情况；查阅公司定期报告，了解公司收入变动情况及近期收入下滑原因；

4、查阅了本次募投项目的可行性研究报告，了解手机、电脑、GPU 服务器等设备的具体应用；查阅了近期美国扩大芯片出口管制的相关政策及国产厂商的发展情况，并向发行人了解了相关政策对募投项目实施的实际影响；

5、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效益情况、前募效益测算的依据、效益较低的合理性；

6、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、发行人律师及申报会计师认为：

1、本次募投项目围绕公司现有主营业务开展，属于公司现有业务的延伸，与公司现有业务具有较强的协同关系；

2、本次募投项目拟购置房产相关单价测算具有公允性、合理性；募投项目围绕发行人主营业务、自身定位和未来发展规划进行，未偏离公司主业，与发行人科技企业定位相符，与发行人收入规模、业务经营情况相匹配；公司本次使用募集资金购置房产主要是基于相关政策对数据资源获取的相关要求、满足募投项目的业务和人员需求、保障募投项目的顺利实施以及购置房产相较于租赁房产更具经济性等多方面因素考量，具有必要性、合理性；

3、发行人经营的外部环境未发生重大不利变化、相关业务有较大市场增长潜力的预测依据准确；本次募投项目效益测算中收入、成本费用等关键参数的选取依据与发行人现有水平及同行业同类项目相比，具有谨慎性、合理性；“数据生产垂直大模型研发项目”每年产生的成本费用将导致公司净利润的下降，发行人报告期内的收入及盈利水平有所波动，但长期来看收入增长具备可持续性，公司投入“数据生产垂直大模型研发项目”具有必要性、合理性；本次募投项目建设规模与发行人的收入规模基本匹配，募集资金具有合理性及必要性；

4、公司将手机、电脑等设备作为本次募投项目的投入设备具有合理性及必要性，近期美国扩大对高算力 AI 芯片出口管制不会对公司募投项目的开展具有重大不利影响，且项目实施存在替代措施，不存在实质性障碍；

5、前次募集资金投资的“自主研发数据产品扩建项目”的效益测算系依据所形成的新建训练数据集产品口径归集产生的效益情况。截至 2023 年 6 月，前次募集资金投资项目“自主研发数据产品扩建项目”实际产生的收入为 10,784.83 万元，利润总额为 99.60 万元。前次募集资金投资项目实现的利润总额较小是由于前次募集资金投资项目于今年 5 月 31 日结项，对应的募投项目效益主要处于项目建设阶段所致。

6、前募资金资本化支出比例已在此前报送的《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

“2. 关于前次募投项目披露”回复中披露。公司前次 IPO 募投项目实际用于非资本性支出的金额占前次募集资金净额的比例为 64.54%，未超过募集资金到账后使用计划中非资本性支出的比例 68.98%。因此，发行人本次融资的募集资金规模未进行调整。

2. 关于经营情况

根据申报材料 1) 发行人 2023 年三季报, 公司主营收入 1.03 亿元, 同比下降 38.25%; 归母净利润-3,211.99 万元, 同比下降 272.4%; 扣非净利润-4,252.09 万元, 同比下降 1,095.65%; 其中 2023 年第三季度, 公司单季度主营收入 2,875.37 万元, 同比下降 45.1%; 单季度归母净利润-1,487.85 万元, 同比下降 1,772.52%; 单季度扣非净利润-1,941.63 万元, 同比下降 389.18%; 2) 根据问询回复, 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, 预计全年收入在 1.8 亿元至 2.1 亿元之间, 下滑幅度预计在 20%至 30%之间, 因此 2023 年全年收入不存在大幅下滑的风险; 3) 报告期, 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)。

请发行人进一步说明: (1) 结合发行人 2023 年三季报、在手订单及订单交付情况, 说明发行人 2023 年是否面临收入大幅下滑及亏损的风险, 发行人的持续经营能力与问询回复披露是否一致, 是否发生重大变化, 公司经营状况和外部环境是否发生重大变化, 相关风险是否已经充分提示; (2) 补充说明发行人在手订单和意向性订单的具体情况, 结合 2023 年三季报及四季度订单执行情况, 说明问询回复中关于 2023 年全年收入不存在大幅下滑风险的依据是否充分; (3) 部分大客户账龄较长的原因及合理性, 是否有纠纷, 收入确认是否真实准确; 结合发行人 2023 年三季报及期后的应收账款回款、坏账、客户回款意愿等情况, 说明发行人应收账款回款情况是否已经改善, 相关不利因素是否已经消除或改善。

请保荐机构及申报会计师核查并发表明确意见。

回复:

一、结合发行人 2023 年三季报、在手订单及订单交付情况，说明发行人 2023 年是否面临收入大幅下滑及亏损的风险，发行人的持续经营能力与问询回复披露是否一致，是否发生重大变化，公司经营情况和外部环境是否发生重大变化，相关风险是否已经充分提示

(一) 结合发行人 2023 年三季报、在手订单及订单交付情况，说明发行人 2023 年是否面临收入大幅下滑及亏损的风险

2023 年 1-9 月，公司营业收入为 10,321.46 万元，同比下降 38.25%，扣非前归属于母公司所有者净利润为-3,211.99 万元，扣非后归属于母公司所有者净利润为-4,252.09 万元。发行人结合截至三季度末在手订单及意向订单情况，对 2023 年全年业绩情况进行估算如下：

项目	金额（万元）
2023 年 1-9 月确认收入（A）	10,321.46
在手订单（截至 2023 年 9 月 30 日）（B）	7,973.14
B 预计可在 2023 年交付并确认收入（C）	2,937.96
2023 年 10-11 月新增在手订单及预计 2023 年底前意向订单可转化为的在手订单的金额(D)	13,016.63
D 预计可在 2023 年确认收入金额（E）	3,616.29
2023 年度营业收入估算（F=A+C+E）	16,875.71
2022 年度营业收入（E）	26,288.79
下降幅度（E-D）/E	35.81%
考虑估算差异及偶发因素影响，估算 2023 年收入同比下降幅度区间	约 30.00%~40.00%
考虑估算差异及偶发因素影响，估算 2023 年收入区间	约 15,773~18,402

注：以上数据未经审计或审阅，上述对全年营收的估算不构成业绩预测或承诺

结合上表分析，发行人 2023 年收入存在大幅下滑（较 2022 年下滑幅度超过 30%）的风险，同时考虑到发行人毛利率及前三季度平均单季度的费用率水平，公司 2023 年全年存在亏损风险。

(二) 发行人的持续经营能力与问询回复披露是否一致，是否发生重大变化，公司经营情况和外部环境是否发生重大变化，相关风险是否已经充分提示

首轮问询回复中关于发行人持续经营能力的主要披露如下：

“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。”

虽然发行人 2023 年面临收入大幅下滑及亏损风险，但发行人持续经营能力未发生重大变化，与首轮问询回复披露一致，公司经营状况和外部环境亦未发生重大变化。

1、外部环境方面，虽然受 AI 行业应用变迁、客户研发投入波动等因素使得发行人 2023 年收入出现下滑，但 AI 行业长期向好趋势未发生改变

发行人的收入波动与下游客户 AI 应用行业的整体发展密切相关，近年来智能语音、智能驾驶以及大模型等下游 AI 应用行业面临不同的发展态势，下游客户根据市场发展变化进行需求和预算调整，影响到发行人报告期内的智能语音收入、计算机视觉收入、自然语言处理收入。具体而言：

(1) AI 应用变迁对发行人收入波动的影响

- 从全行业来看，AI 技术经历了前期智能语音投入以及日常交互应用化的蓬勃发展，2020 年至 2022 年行业发展进入平台期，发行人报告期内**智能语音收入**规模有所下滑；
- 发行人积极开拓智能驾驶业务作为第二增长曲线，2022 年以来发行人与智能驾驶相关的**计算机视觉业务**实现较快增长，但依然处于相对早期阶段；
- 2022 年底，随着大模型范式的推出引发 AI 应用的质变突破，发行人与之相关的**自然语言处理收入**有望迎来高速增长，但目前规模不大，主要是由于大模型客户自身的大模型业务也尚处于发展早期阶段、且现阶段主要将采购预算投入算力，数据采集需求尚在逐步释放中；
- 此外，大模型技术对机器理解能力的提升将带动语音交互应用场景不断扩展，与之相关，发行人的**智能语音收入**规模有望恢复二次增长。

同行业公司 Appen、数据堂虽然与公司收入结构存在一定差异，但从公开披

露的财务数据来看，2023 年 1-6 月的营业收入均出现了不同程度的同比下滑。

1) 报告期内智能语音收入有所下滑，主要由于智能语音应用在 2020 年进入技术瓶颈期，下游客户投入有所放缓，大模型促使智能语音应用有望步入 2.0 时代

■ **智能语音 1.0 时代**：2016 年至 2022 年，智能语音应用从高速增长发展到稳定

2016 年-2019 年是 AI 技术的第三次浪潮，本次浪潮主要体现在 AI 技术从实验室走向产品应用，在语音方面的应用上体现最为显著，比如天猫精灵、小爱同学（小米）、小度等智能音箱，苹果 Siri、三星 Bixby、腾讯微信语音交互等手机终端语音助手等产品纷纷推出市场。发行人主营智能语音数据起家，对应期间（2016 年-2019 年）的业绩获得高速增长。

2020 年开始，智能语音应用进入稳定期，主要由于智能语音助手产品的部分技术和用户体验有待提升。一个完整的智能语音交互链条一般需经历语音信号输入并识别、转写为文本——NLP（自然语言处理，即文本语义理解）——机器输出理解后的文本——语音合成输出语音信号等步骤。一方面，NLP 领域在当时尚未获得技术突破，导致智能语音助手对话的自然度和流畅度都有待提高，且经常出现人机交互到三四个轮次就出现机器“答非所问”的情况等，用户体验很难显著提升；另一方面语音交互主要的场景是泛生活化场景，用户付费意愿较低，商业价值遇到了瓶颈。以上原因造成客户在智能语音上的研发投入自 2020 年以后趋平，由此导致发行人报告期内智能语音业务收入有所下降。

■ **智能语音 2.0 时代**：2023 年至未来 3-5 年，智能语音应用有望迎来二次快速增长

2022 年底 ChatGPT 的出现，掀起又一波人工智能发展热潮，大模型技术路线的落地为人工智能开启发展新范式，机器的学习能力和理解能力随着算法参数的几何式增长及驾驭算力能力的显著扩展，得到了爆发式增强，原本限制智能语音应用的 NLP 技术瓶颈得到了突破，进而推动语音交互的应用场景从泛生活化扩展到更为专业的领域，如机器人的语音交互、高阶的万物互联的语音交互、一些垂直场景如政府服务、医疗服务等。下游客户纷纷在智能语音方面增加研发投入

入以不断提升产品体验，获取增量客户，带来增量收入。如 2023 年下半年，Amazon 重启语音助手 Alexa 重大升级项目，三星重启语言拓展计划以提升语音助手 Bixby 的交互能力等。

2023 年下半年发行人的语音数据产品询盘量较上半年明显增长，大模型的不断迭代有望促使智能语音应用步入 2.0 时代，即语音交互应用更加深入到垂直场景且体验升级的时代，发行人的智能语音业务收入也有望恢复二次增长。

2) 计算机视觉业务作为发行人第二增长曲线，2022 年以来收入增长显著，随着高阶自动驾驶逐步普及，发行人计算机视觉业务保持快速增长

考虑到智能语音业务发展进入技术瓶颈期，发行人自 2021 年底开始加大力度布局第二增长曲线，即计算机视觉业务，并锁定智能驾驶场景作为核心方向持续进行研发投入，计算机视觉业务也自 2022 年起贡献较多收入增长。

目前智能驾驶正处于由低阶自动驾驶（L1、L2 级）向高阶自动驾驶（L3 及以上）逐步过渡的阶段，随着高阶自动驾驶进入量产时间表，车载传感器特别是车载雷达的数量不断提升，数据处理的维度逐步从 2D 图像向 3D、4D 图像演进，数据处理的难度、效率以及精准度要求都将明显提高，发行人经过前期持续投入，已经完成了自动驾驶智能化标注平台的研发、合规体系搭建以及优质客户资源的储备，未来发行人的计算机视觉业务将保持快速增长。

3) 报告期内自然语言处理业务占比较低，在大模型时代有望高速增长

报告期内发行人自然语言处理业务收入规模较小，主要原因在于此前一段时间机器学习能力没有得到显著提升，NLP 技术瓶颈尚未突破，技术应用范围较为有限，训练数据投入的边际效益并不明显。自然语言处理（NLP）技术旨在实现机器能够像人一样理解语言意图，所需的原料数据主要为对话文本等。2023 年以来，随着语言大模型初代产品的陆续推出，下游客户持续投入预算进行大模型的训练升级，由于自有和开源数据无法满足大模型训练的持续和长期需求，下游客户将逐步转向外购数据，向发行人这类专业数据服务商进行采购，以获取更多垂直应用场景下的增量数据资源。

2023 年前三季度，发行人该项业务收入未见明显增长主要系大模型客户自身的大模型业务也尚处于发展早期阶段，且由于算力规模直接决定对话模型的反

应速度和用户体验，是大模型客户初代产品上线的基本条件，因此大模型客户前期的投入主要集中在算力，训练数据前期主要以其自有和开源数据为主。随着大模型应用的渗透率逐步提升，为提供有价值的服务，获取和保持用户持续付费意愿，大模型能力与具体应用结合成为必须解决的问题，而数据将在此阶段发挥决定性的作用，大规模、高质量的增量通用训练数据、覆盖更多垂直领域的场景训练数据等对于大模型优化和投入应用的效果愈加凸显，大模型客户外购训练数据的需求亦会明显增长。发行人拟通过实施本次募投项目“AI 大模型训练数据集建设项目”，抓住大模型时代的发展机遇。

(2) 下游客户研发投入的变动对发行人收入的影响

购买训练数据是客户在人工智能领域研发预算的一个必要组成部分，通常数据上的采购投入会占到客户整体 AI 技术投入的 10%-15%左右。影响客户整体 AI 技术投入的因素比较多，核心是其产品方向和整体的预算安排。除上述 (1) 中提及的 AI 应用变迁对客户在不同技术方向上的投入有较大影响外，客户整体财务表现也是影响研发预算的核心要素。

发行人的客户多为大型科技企业，研发预算与宏观经济景气度息息相关，紧随市场发展变化进行预算调整，是下游客户的一个显著特点。在宏观经济增速放缓阶段，客户自身收入增速也在放缓，但投资者反而会更关注业绩表现和确定性，因此自然引发裁员、预算收缩、组织结构调整等各种动作优化开支，甚至直接削减一些不能短期产生收入的研发类项目，在这样的背景下，发行人的收入也会相应受到影响，如发行人境外主要客户 Microsoft、Amazon、Meta 等在 2022 年底至 2023 年初期间均公布了裁员计划，由此直接导致发行人 2023 年的境外收入受到一定负面影响。

2、公司经营状况方面，发行人作为专业数据服务商已构建护城河优势，未来收入具备复苏潜力，同时费用增长势头已有所趋缓，盈利能力有望复苏

(1) 2023 年发行人单季度收入下滑幅度有所收窄，各项业务具备复苏潜力

自 2023 年二季度以来，发行人收入下滑趋势呈现收窄态势，2023 年一季度，发行人营业收入较过去两年同期平均值（即 2021 年一季度、2022 年一季度收入

平均值) 下滑 37.43%; 2023 年二季度, 发行人单季度营收相较过去两年同期平均值(即 2021 年二季度、2022 年二季度收入平均值) 下滑 29.05%, 2023 年三季度, 发行人单季度营收相较过去两年同期平均值(即 2021 年三季度、2022 年三季度收入平均值) 下滑 25.46%, 下滑幅度有所收窄。此外, 发行人 2023 年前三季度单季度毛利率分别为 47.79%、65.83%、60.45%, 自二季度起, 发行人毛利率已有所回升。公司预计截至 2023 年底, 在手订单预计可达到 1.39 亿元, 较去年同期增长 106%。

关于收入的复苏潜力, 分业务进一步分析如下:

1) 智能语音领域, 发行人已在数据集数量、行业理解、下游客户储备等方面构筑护城河优势, 可支撑在智能语音 2.0 时代中获取更大的商业机会

发行人以智能语音数据训练产品起家, 凭借多年在该领域的持续深耕, 在标准数据集储备、语音数据处理工具能力、垂直行业理解、下游客户资源等方面都处于行业领先地位, 可支撑在智能语音 2.0 时代中获取更大的商业机会。2023 年下半年以来, 语音数据的需求量已有明显回暖趋势, 截至目前, 发行人已收到来自若干国际知名客户关于智能语音产品库的询盘采购需求, 整体金额超过 7,000 万元, 部分预计可在年底前转化为在手订单。

2) 计算机视觉领域, 发行人已完成平台工具及客户资源的前期储备, 有能力适配未来不断变化的智能驾驶行业技术和应用发展新趋势

计算机视觉领域, 智能驾驶行业蓬勃发展带动数据需求高速增长, 由此将带来发行人计算机视觉业务收入保持快速增长。发行人已研发完成的智能驾驶平台工具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, 将助力公司适配智能驾驶行业技术和应用发展的趋势。目前发行人智能驾驶领域客户已超过 70 家, 覆盖传统车企、新势力车企、头部智能驾驶技术公司等, 最近一年内, 发行人通过新增定点或签署框架协议等方式拓展了较多国内领先自动驾驶领域客户。

3) 自然语言处理领域, 现有优质的客户群资源可以助力公司抓住大模型时代发展机遇

自然语言处理领域, 随着大模型技术不断向纵深发展, 将会带来自然语言处

理领域的客户需求，提升发行人自然语言处理业务收入。此外，随着大模型能力的演进，未来将以文本交互为主要方式拓展到多模态交互，亦为智能语音、计算机视觉等多模态数据领域带来新的增量需求空间。发行人本次募投项目“AI 大模型训练数据集建设项目”即为发行人延伸大模型数据业务的有力抓手，有助于发行人抓住大模型发展机遇，寻求新的业绩增长点。目前发行人的较多客户已纷纷开展大模型领域的研发和应用，且其中一些客户的大模型产品已经正式推出上线，大模型业务的主要客户及营销渠道与公司现有业务重合度较高，有助于公司打开市场。以美团、科大讯飞、阿里巴巴、腾讯、百度等公司现有业务的主要客户为例，这些客户亦为大模型领域占据主要市场地位的企业，并与公司签署了大模型领域的业务合同或意向性订单，2023 年全年合计金额约 2,900 万元。

（2）公司持续推动各运营环节降本增效，费用增长势头趋缓

2020 年至 2023 年 1-9 月，发行人为配合整体战略发展及业务拓展目标，持续加大人员储备、升级管理架构及激励体系、大力拓展自动驾驶数据处理平台研发等，导致员工人数及人均薪酬均呈现上涨态势，其中人员数量增长更为显著，2020 年底至 2023 年三季度末，发行人员工总人数分别为 151 人、245 人、273 人、226 人，2023 年三季度末员工数量较上年末有所收缩，费用持续上涨趋势已有所改善。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。

发行人 2023 年下半年以来持续调整优化管理模式，减少不必要的支出，推动各运营环节降本增效。从单季度费用来看，发行人三季度销售费用、管理费用、研发费用的单季度规模均较二季度有所降低，销售费用及管理费用的同比增幅也有收窄趋势，费用增长势头趋缓。2023 年研发费用较去年同期下降幅度较大主要系 2022 年公司加速进行 IPO 募集资金项目以及智能驾驶等主要数据处理平台的投入建设所致。

单位：万元

期间	项目	销售费用	管理费用	研发费用
2023 年一季度	金额	480.04	1,560.22	1,693.31
	同比增幅	51.01%	62.70%	-36.80%

2023 年二季度	金额	733.29	1,784.85	1,528.05
	同比增幅	43.60%	35.87%	-38.25%
2023 年三季度	金额	617.27	1,535.45	1,355.96
	同比增幅	26.36%	18.53%	-5.58%

3、发行人已在募集说明书中进行充分风险提示

发行人已在募集说明书及其他申报材料中对于业绩下滑进行了充分风险提示如下：

“（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、补充说明发行人在手订单和意向性订单的具体情况，结合 2023 年三季度及四季度订单执行情况，说明问询回复中关于 2023 年全年收入不存在大幅下滑风险的依据是否充分

首轮问询回复中，发行人结合截至 2023 年 8 月底的在手订单及意向性订单情况对 2023 年全年收入进行估算，彼时预计全年收入在 1.8 亿元至 2.1 亿元之间，下滑幅度预计在 20%至 30%之间。上述在手订单和意向订单截至目前的执行情况如下：

（一）在手订单具体情况及执行情况

首轮问询回复时点，发行人在手订单 7,320.81 万元，预计 2023 年可实现收

入 4,799.42 万元，截至本轮回复出具日，发行人结合上述在手订单的执行情况重新进行测算，上述在手订单预计可在 2023 年实现收入 3,899.02 万元，差异金额为 900.39 万元，形成差异的主要原因可归纳为两类，一类是项目执行进度不及预期，譬如因项目难度较高，跟客户协商后适当减缓进度，或者数据已交付客户，但客户反馈内部质检没有排期，预计 2023 年无法对数据进行验收等因素；另一类是部分客户的需求出现调整，譬如因其研发进度调减需求，或客户侧人员变动或优先级调整导致需求下发进度放缓等。

单位：万元

业务分类	首轮问询时点 在手订单金额	首轮问询时点预计 2023 年确认收入金额	最新预计 2023 年 确认收入金额	差异金额
智能语音	4,822.81	3,152.11	2,676.37	475.74
自然语言	1,094.59	722.55	606.60	115.95
计算机视觉	1,346.05	888.31	600.99	287.32
应用服务	57.37	36.45	15.06	21.39
合计	7,320.81	4,799.42	3,899.02	900.39

（二）意向订单具体情况及执行情况

首轮问询回复时点，发行人结合客户沟通邮件、询盘记录等进行梳理，意向订单金额约为 17,170.32 万元，预计 2023 年可实现收入约为 6,258.58 万元，截至本轮回复出具日，发行人结合上述意向订单的具体转化和执行情况重新测算，上述意向订单预计可在 2023 年实现收入 2,914.75 万元，意向订单中剩余大部分预计可转化为 2023 年底在手订单，预计 2023 年底转化为在手订单的金额约 11,120.37 万元。同时截至本轮回复出具日，发行人新增意向订单 17,400.32 万元，该部分意向订单预计可转化为 2023 年收入的金额为 2,615.84 万元。综上，意向订单转化为 2023 年收入的差异为 727.98 万元，差异的主要金额集中在智能语音业务线。

智能语音业务线形成差异的主要原因为部分海外大客户的意向订单转化进度有所延期，以某国际消费电子产品厂商为例，首轮问询回复时点预计金额 5,950 万元的意向订单因客户内部需求确定进度延缓导致无法在今年确认收入（首轮问询时点预计 2023 年可确认收入 1,112 万元），最新预计该意向可转化为年底在手订单（预计 5,000 万元）；以某海外科技头部公司客户为例，首轮问询回复时点预计金额 1,330 万元的意向订单因客户内部采购预算下单周期调整

的原因改为预计在 2024 年一季度签单（预计订单金额调整为 2,300 万元），因此无法在今年确认收入(首轮问询时点预计 2023 年可确认收入 870 万元)。

单位：万元

业务分类	首轮问询时点情况		最新预计情况				差异金额 (B-C-E)
	意向订单 金额 (A)	预计 2023 年确认收 入金额 (B)	预计 A 可 转化为 2023 年收 入金额(C)	预计 A 可 转化为 2023 年底 在手订单 金额	首轮问询 时点至今 新增意向 订单金额 (D)	D 预计转化 为 2023 年 收入金额 (E)	
智能语音	13,029.11	4,599.53	1,738.79	9,163.57	11,245.29	1,991.48	869.26
自然语言	394.63	178.30	30.50	237.26	1,545.73	200.91	-53.11
计算机视觉	3,703.07	1,456.78	1,145.46	1,719.54	4,419.64	418.94	-107.63
应用服务	43.50	23.97	-	-	189.66	4.51	19.46
合计	17,170.32	6,258.58	2,914.75	11,120.37	17,400.32	2,615.84	727.98

发行人在首轮问询回复时点，综合在手订单以及根据客户沟通邮件、询盘记录等梳理的意向订单金额，对收入下滑幅度进行了合理预测，预计 2023 年全年收入不存在大幅下滑风险，但由于后续订单执行及转化过程中，因客户需求变化、部分项目执行进度放缓等因素，导致最新预计的收入下滑幅度有所扩大，最新的全年收入估算情况参考本题回复“（一）结合发行人 2023 年三季度、在手订单及订单交付情况，说明发行人 2023 年是否面临收入大幅下滑及亏损的风险”部分，发行人首轮问询回复中关于 2023 年全年收入不存在大幅下滑风险的依据充分。

三、部分大客户账龄较长的原因及合理性，是否有纠纷，收入确认是否真实准确；结合发行人 2023 年三季度及期后的应收账款回款、坏账、客户回款意愿等情况，说明发行人应收账款回款情况是否已经改善，相关不利因素是否已经消除或改善。

最近三年两期（2023 年 6 月 30 日、2023 年 9 月 30 日）末，发行人账龄在 1 年以上的应收账款前五名客户的情况如下：

单位：万元

2023 年 9 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	期后 回款	期后回 款比例

科大讯飞股份有限公司	科大讯飞	228.01	104.93	52.73	-	-
北京中核久安科技有限公司	中核久安	214.56	214.56	107.28	-	-
普强时代（珠海横琴）信息技术有限公司	普强	165.41	165.41	82.71	-	-
阿里云计算有限公司	阿里巴巴	147.35	143.85	124.46	54.62	37.06%
LumenVox Corporation	LumenVox	122.06	120.70	58.20	-	-
合计		877.39	749.45	425.37	54.62	6.22%
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	期后回款	期后回款比例
普强时代（珠海横琴）信息技术有限公司	普强	225.41	225.41	112.71	60.00	26.62%
北京中核久安科技有限公司	中核久安	214.56	214.56	107.28	-	-
阿里云计算有限公司	阿里巴巴	155.35	143.85	124.69	64.01	41.20%
LumenVox Corporation	LumenVox	122.84	114.09	14.20	-	-
科大讯飞股份有限公司	科大讯飞	157.30	94.42	43.06	3.80	2.42%
合计		875.46	792.33	401.94	127.81	14.60%
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	期后回款	期后回款比例
普强时代（珠海横琴）信息技术有限公司	普强	235.41	235.27	28.25	70.00	29.73%
北京中核久安科技有限公司	中核久安	214.56	214.56	25.75	-	-
阿里云计算有限公司	阿里巴巴	198.24	143.55	75.98	106.90	53.92%
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%
北京京东世纪贸易有限公司	京东	84.09	84.09	42.04	-	-
合计		832.30	777.47	272.02	276.90	33.27%
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	期后回款	期后回款比例
阿里云计算有限公司	阿里巴巴	311.58	200.69	28.60	267.34	85.80%
北京小米智能科技有限公司	小米	314.08	174.67	17.69	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%
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	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	1.50	2.64%
合计		894.55	596.21	216.63	794.91	88.86%
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	期后回款	期后回款比例
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%
Verint Systems Ltd.	Verint	60.96	60.96	30.48	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	1.50	2.64%
合计		217.86	217.86	137.38	162.09	74.40%

注：期后回款数据截至2023年11月30日

（一）部分大客户账龄较长的原因及合理性，是否有纠纷，收入确认是否真实准确

前述表格中部分大客户账龄较长的原因主要包括客户内部付款流程较长、回款受到客户下游回款影响较大、客户经营出现问题及其他外部因素等，发行人已积极进行催收跟进回款。经访谈发行人及部分客户、公开查阅“中国裁判文书网”、等信息，截至本回复出具日，除深圳市天道日新科技有限公司外，发行人与相关客户就应收账款不存在纠纷。针对深圳市天道日新科技有限公司拖欠款项，公司已通过诉讼方式维护自身合法权益，截至目前已全部收回该客户的应收款项。关于收入真实性，经查阅发行人与客户的业务合同及交付验收单据，并访谈部分客户，相关收入确认真实准确。主要长账龄客户的具体情况如下：

客户名称	账龄较长的原因及合理性	是否有纠纷	收入确认是否真实准确
客户9	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	目前无纠纷，但客户回款意愿不强，公司不排除后续通过司法手段追回款项	是
客户10	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	否	是
客户11	客户内部付款流程较长	否，已100%回款	是

客户名称	账龄较长的原因及合理性	是否有纠纷	收入确认是否真实准确
客户 12	客户内部付款流程较长	否, 已 100%回款	是
客户 2	客户内部产品线调整, 项目结算较慢	否	是
客户 13	客户部门调整, 业务对接人更换	否	是
客户 14	客户研发周期较长, 项目结算受项目整体进度影响, 公司已发送律师函催收	否	是
客户 15	受 2020 年宏观环境波动影响, 回款周期拉长	否, 已基本 100%回款	是
境外客户 J	该公司提供了已完成付款的证据, 但因对方将款项付款至欺诈账户, 导致发行人并未收到该笔回款, 目前该笔款项尚在追回中, 双方正积极进行商讨解决方案	目前无纠纷, 但客户回款意愿不强, 公司不排除后续通过司法手段追回款项	是
深圳市天道日新科技有限公司	客户拖欠款项	是, 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)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	是
上海未来伙伴机器人有限公司	客户破产重整, 发行人已全额计提了坏账准备。2021 年 7 月 12 日, 上海市徐汇区人民法院裁定批准重整计划, 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	否	是

(二) 结合发行人 2023 年三季报及期后的应收账款回款、坏账、客户回款意愿等情况，说明发行人应收账款回款情况是否已经改善，相关不利因素是否已经消除或改善

1、应收账款整体规模及回款情况有一定改善

单位：万元

项目	2023 年 9 月 30 日	2023 年 6 月 30 日	2022 年 12 月 31 日	2021 年 12 月 31 日	2020 年 12 月 31 日
应收账款余额	10,123.94	10,154.08	11,065.82	9,726.78	6,700.63
回款金额（截至 2023 年 11 月 30 日）	2,956.57	4,489.62	8,118.03	8,660.21	6,422.72
回款比例（截至 2023 年 11 月 30 日）	29.20%	44.21%	73.36%	89.03%	95.85%
回款金额（截至 2023 年 8 月 31 日）	-	2,221.70	7,423.60	8,429.37	6,370.22
回款比例（截至 2023 年 8 月 31 日）	-	21.88%	67.09%	86.66%	95.07%

如上表所示，相较 2022 年年末，2023 年二季度末及三季度末，发行人应收账款余额整体规模有所下降。截至 2023 年 11 月 30 日，各期末应收账款的期后回款比例分别达到 95.85%、89.03%、73.36%、44.21%、29.20%，回款情况整体良好。最近一期末期后回款比例较低的原因主要是由于期后时间较短，仅为 2 个月，而客户信用期多为 60 天或 90 天所致。

截至 2023 年 11 月 30 日，2022 年末应收账款的期后回款金额相较截至 2023 年 8 月 31 日增加 694.43 万元，回款比例上涨 6.27%，整体回款情况有所改善；此外，2023 年 9 月 30 日的应收账款余额截至 2023 年 11 月 30 日的回款比例为 29.20%，2023 年 6 月 30 日的应收账款余额截至 2023 年 8 月 31 日的回款比例为 21.88%，同样期后 2 个月的时间内，回款比例有所提升。

因此，基于上述回款指标分析，发行人应收账款期后回款情况有一定改善。

2、报告期内坏账计提金额及计提比例有所提升，发行人正积极通过加强对销售人员回款考核、加强前端客户信用风险把控等方式逐步消除不利影响

报告期内，应收账款坏账计提金额及计提比例持续提升，主要系部分客户的逾期时间有所拉长所致，发行人已加强对销售人员的回款考核，督促业务人员采用各种方式加强催收，同时持续加强对客户信用情况的前端把控及持续跟踪，以

降低应收账款回款风险。

单位：万元

项目	2023年9月30日	2023年6月30日	2022年12月31日	2021年12月31日	2020年12月31日
应收账款余额	10,123.94	10,154.08	11,065.82	9,726.78	6,700.63
坏账计提金额	1,460.64	908.88	887.79	707.13	422.34
坏账计提比例	14.43%	8.95%	8.02%	7.27%	6.30%

3、发行人长账龄客户较为稳定且大部分具备回款意愿，少数客户回款意愿不强但对发行人不利影响有限，发行人计提坏账准备较为充分

发行人报告期内1年以上账龄的前五大客户较为稳定，经查阅发行人与客户沟通记录、客户回款凭证、项目组邮件询问客户回款意愿，最近三年两期（2023年6月30日、2023年9月30日）末1年以上的前五大客户中，回款比例较低（未达到或接近100%）的客户回款意愿及坏账计提情况如下表所示，除中核久安、Lumenvox等少数客户回款意愿不强外，其他客户回款意愿正常，上述回款意愿不强的客户主要系由于个体原因，相关客户未回款对发行人财务报表影响有限。发行人一方面积极通过内外部渠道加强催收，另一方面发行人已针对长账龄客户计提较高比例的坏账准备，会计处理较为谨慎。

客户名称	回款意愿	坏账计提情况
客户10	回款意愿正常，2023年6月30日应收款项余额为155.35万元，1年以上账龄金额为143.85万元，截至2023年8月31日，期后回款比例为7.41%，截至2023年11月30日，期后回款比例为41.20%，回款有所改善。	截至2023年9月30日，坏账计提比例为84.46%
客户2	回款意愿正常，目前该客户已通知发行人开票191.52万元，预计将陆续回款。	截至2023年9月30日，坏账计提比例为23.13%（1年以上账龄占比相对较低，因此整体坏账比例偏低）
客户14	回款意愿正常，已发送律师函催收，回款进度较慢，2023年6月30日应收款项余额为225.41万元，全部为1年以上账龄，截至2023年8月31日，期后回款比例为4.44%，截至2023年11月30日，期后回款比例为26.62%，回款有所改善。	截至2023年9月30日，坏账计提比例为50.00%
客户13	回款意愿正常，发行人与该客户正在沟通协商中。	截至2023年9月30日

		日，坏账计提比例为100%
客户9	回款意愿不强，原因为该公司提供了已完成付款的证据，但因对方将款项付款至欺诈账户，导致发行人并未收到该笔回款，目前该笔款项尚在追回中，双方正积极进行商讨解决方案。	截至2023年9月30日，坏账计提比例为47.68%
北京中核久安科技有限公司	回款意愿不强，因其自身现金流紧张，发行人多次催收无果。	截至2023年9月30日，坏账计提比例为50.00%
上海未来伙伴机器人有限公司	回款意愿不强，因执行破产重整程序，2023年7月，已回款1.5万元，后续无回款计划。	已全额计提坏账准备

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、获取发行人2023年三季报，了解业绩下滑的原因，审阅发行人提供的在手订单及执行情况，分析2023年收入是否面临大幅下滑及亏损风险。

2、访谈发行人管理层，了解公司经营状况和外部环境是否发生重大变化，持续经营能力是否存在重大不确定性风险。

3、查阅并分析发行人报告期内应收账款及坏账变动情况，重点关注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，是否存在纠纷，公司对加强应收账款回款采取的管理措施等。获取长账龄客户的销售合同、验收单据、发票等支持性证据以验证其收入的真实性和准确性。

4、查阅发行人报告期各期末应收账款整体的期后回款情况，针对账龄较长、回款比例较低的客户，获取发行人与客户沟通记录、期后客户回款凭证、并通过邮件与客户沟通回款意愿及纠纷情况，查询中国裁判文书网公开信息，查看发行人与客户的诉讼信息等。

（二）核查意见

经核查，保荐机构、申报会计师认为：

1、发行人2023年存在收入大幅下滑（较2022年下滑幅度超过30%）及亏损的风险，但发行人持续经营能力未发生重大变化，与首轮问询回复披露一致，公司经营状况和外部环境亦未发生重大变化，相关风险已经充分提示。

2、首轮问询回复时点，发行人综合在手订单以及根据客户沟通邮件、询盘记录等梳理的意向订单金额，对收入下滑幅度进行了合理预测，预计 2023 年全年收入不存在大幅下滑风险，预测依据充分。但由于后续订单执行及转化过程中，因客户需求变化、部分项目执行进度放缓等因素，导致最新预计的收入下滑幅度有所扩大，具备合理性。

3、部分大客户账龄较长的原因具备合理性，除深圳市天道日新科技有限公司存在诉讼纠纷外，其他客户目前不存在纠纷，深圳市天道日新科技有限公司应收款项已全额收回，发行人对相关客户的收入确认真实准确。基于正文部分的回款指标分析，发行人应收账款回款情况已有一定改善，发行人正在采取加强销售人员回款考核、积极推进内外部催收、前端审核评价客户信用风险等方式消除应收账款逾期的不利影响，少数客户由于自身原因回款意愿不强，但对发行人影响有限，发行人已计提较高比例坏账准备。

3. 关于数据合规

根据申报材料，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》，发行人已申报的数据出境安全评估已获通过。2023年9月28日国家互联网信息办公室发布《规范和促进数据跨境流动规定（征求意见稿）》，进一步对《评估办法》规定的出境数据的安全评估的适用范围进行界定，且按照预计一年内向境外提供个人信息的数量进行区分适用。

请发行人进一步说明：数据出境安全评估申报审批的具体流程及对发行人未来境外客户订单的影响，发行人认为《规范和促进数据跨境流动规定（征求意见稿）》未来如出台对发行人开展境外业务的影响将进一步减小的依据是否充分。

请保荐机构及发行人律师核查并发表明确意见。

回复：

一、数据出境安全评估申报审批的具体流程及对发行人未来境外客户订单的影响

（一）数据出境安全评估申报审批的具体流程

根据《评估办法》及同时施行的《数据出境安全评估申报指南（第一版）》，数据出境安全评估申报审批的具体流程如下：

1、就拟进行的应履行安全评估申报程序的数据出境交易，数据处理者向所在地省级网信部门申报数据出境安全评估

数据处理者通过所在地省级网信部门申报数据出境安全评估，向所在地省级网信部门送达书面申报材料，包括数据出境安全评估申报书、数据出境风险自评估报告、数据处理者与境外接收方拟订立的数据出境相关合同或者其他具有法律效力的文件影印件等。

2、省级网信部门收到申报材料后进行完备性查验，通过完备性查验的上报国家网信部门

省级网信部门收到申报材料后，在5个工作日内完成申报材料的完备性查验。通过完备性查验的，省级网信部门将申报材料上报国家网信部门；未通过完备性查验的，数据处理者将收到申报退回通知。

3、国家网信部门受理并书面通知数据处理者

国家网信部门自收到省级网信部门上报申报材料之日起7个工作日内，确定是否受理并书面通知数据处理者。数据处理者如被告知补充或者更正申报材料，应当及时按照要求补充或者更正材料。无正当理由不补充或者更正申报材料的，安全评估将会终止。

4、国家网信部门进行安全评估

国家网信部门受理申报后，根据申报情况组织国务院有关部门、省级网信部门、专门机构等进行安全评估。安全评估过程中，发现数据处理者提交的申报材料不符合要求的，国家网信部门可以要求其补充或者更正。数据处理者无正当理由不补充或者更正的，国家网信部门可以终止安全评估。

国家网信部门应当自向数据处理者发出书面受理通知书之日起45个工作日内完成数据出境安全评估；情况复杂或者需要补充、更正材料的，可以适当延长并告知数据处理者预计延长的时间。

5、国家网信部门评估完成后向数据处理者告知评估结果

国家网信部门评估完成后，数据处理者将收到评估结果通知书。对评估结果无异议的，数据处理者须按照数据出境安全管理相关法律法规和评估结果通知书的有关要求，规范相关数据出境活动；对评估结果有异议的，数据处理者可以在收到评估结果通知书15个工作日内向国家网信部门申请复评，复评结果为最终结论。

通过数据出境安全评估的结果有效期为2年，自评估结果出具之日起计算。有效期届满，需要继续开展数据出境活动的，数据处理者应当在有效期届满60个工作日前重新申报评估。

（二）对发行人未来境外客户订单的影响

1、发行人部分境外业务的开展适用《评估办法》，发行人作为数据处理者应履行数据出境安全评估申报审批程序，该部分境外业务收入占比较小，因此数据出境安全评估申报审批具体流程对发行人整体业务发展的影响较小

根据《评估办法》第二条、第四条，发行人作为数据处理者自上年1月1日起累计向境外提供1万人敏感个人信息，因此发行人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，适用《评估办法》，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。

发行人境外业务主要包括以下三种类型：

(1) 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。

(2) 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。

(3) 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。《评估办法》于 2022 年 9 月 1 日施行后，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。

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，发行人开展的境外业务中，根据《评估办法》规定属于应履行数据出境安全评估申报审批程序的数据出境交易的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，根据《评估办法》属于应进行数据出境安全评估申报审批的境外业务在发行人整体业务中占比较小，数据出境安全评估申报审批流程对发行人整体业务发展的影响较小。

2、《评估办法》自 2022 年 9 月 1 日施行后，数据出境安全评估申报审批流程作为影响因素之一阶段性影响境外客户该类需进行安全评估审批的业务需求进度安排，目前数据出境安全评估申报审核政策已明晰，发行人已熟悉掌握数据出境安全评估申报审批流程及申报文件准备等相关事项，网信部门通过发行人已通过的申报审批事项对发行人的数据合规基本情况予以认可，有利于后续该类业

务交易顺利推进申报审批流程，数据出境安全评估申报审批具体流程对发行人未来该类境外客户订单的影响较小

《评估办法》自 2022 年 9 月 1 日施行后，受数据出境安全评估申报文件准备、申报程序及审批周期的影响，同时出于合规性及审慎性考虑，发行人与存在该类业务需求的境外客户进行了及时沟通，双方均处于观望状态，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阶段性减少。

目前数据出境安全评估申报审批政策已明晰，尤其发行人已就拟进行的一项该类业务交易申报并于 2023 年 9 月 25 日取得国家互联网信息办公室网络数据管理局出具的《评估结果通知书》，经评估，发行人所申报的数据项准予出境。

境外客户逐渐根据业务需求与发行人积极洽谈该类业务合作。后续如未来境外客户增加该类需进行数据出境安全评估审核的订单，发行人将按照相关法律法规的要求增加申报数据出境安全评估。鉴于发行人已熟悉掌握数据出境安全评估申报审批所要求的申报文件、具体流程，申报材料质量进一步完善，申报材料准备时间更为可控，且网信部门通过前述已通过审批的申报对发行人的数据合规基本情况予以认可，将更有利于后续该类业务交易顺利推进申报审批流程，能够从申报程序角度减小对发行人未来境外客户订单的影响。

综上，数据出境安全评估申报审批流程对发行人未来境外客户订单的影响较小。

二、发行人认为《规范和促进数据跨境流动规定（征求意见稿）》未来如出台对发行人开展境外业务的影响将进一步减小的依据是否充分

国家互联网信息办公室于 2023 年 9 月 28 日发布《规范和促进数据跨境流动规定（征求意见稿）》（以下简称《数据跨境流动规定》），为进一步规范和促进数据依法有序自由流动，进一步明确了《评估办法》规定的出境安全评估审批的具体适用范围和标准。《数据跨境流动规定》未来如出台，对发行人开展境外业务的影响将进一步减小，其依据充分，具体如下：

（一）结合发行人以往及目前预计该类业务开展情况，根据《数据跨境流动规定》发行人无需或可以不申报数据出境安全评估

《数据跨境流动规定》明确按照预计一年内向境外提供个人信息数量区分是否进行数据出境安全评估申报。《数据跨境流动规定》第五条、第六条规定：

“五、预计一年内向境外提供不满 1 万人个人信息的，不需要申报数据出境安全评估、订立个人信息出境标准合同、通过个人信息保护认证。但是，基于个人同意向境外提供个人信息的，应当取得个人信息主体同意。

六、预计一年内向境外提供 1 万人以上、不满 100 万人个人信息，与境外接收方订立个人信息出境标准合同并向省级网信部门备案或者通过个人信息保护认证的，可以不申报数据出境安全评估；向境外提供 100 万人以上个人信息的，应当申报数据出境安全评估。但是，基于个人同意向境外提供个人信息的，应当取得个人信息主体同意。”

根据发行人数据出境安全申报文件，因发行人自 2022 年 1 月 1 日起至 2023 年 3 月 31 日累计向境外提供个人信息涉及 17,590 人，因此属于《评估办法》第四条第（三）款规定的“自上年 1 月 1 日起累计向境外提供 1 万人敏感个人信息的数据处理者向境外提供个人信息”这一情形，根据《评估办法》规定应作为数据处理者履行数据出境安全评估申报程序。根据发行人结合目前境外业务开展情况预计，发行人未来一年内向境外提供个人信息的数量预计不会超过 100 万人。

因此根据《数据跨境流动规定》第五条、第六条，如发行人预计一年内向境外提供不满 1 万人个人信息，不需要申报数据出境安全评估、订立个人信息出境标准合同、通过个人信息保护认证；如预计一年内向境外提供 1 万人以上、不满 100 万人个人信息，发行人可以不申报数据出境安全评估，与境外接收方订立个人信息出境标准合同并向省级网信部门备案或者通过个人信息保护认证，避免数据出境安全评估申报审批流程对该类业务的影响。

（二）如预计一年内向境外提供 1 万人以上、不满 100 万人个人信息时，发行人将采取与境外接收方订立个人信息出境标准合同并向省级网信部门备案或者通过个人信息保护认证的方式进行数据出境，将更有助于发行人与境外客户及时签署订单并完成交付

1、与境外接收方订立个人信息出境标准合同并向省级网信部门备案不会影响合同的及时签订和履行

根据《评估办法》规定，数据出境安全评估申报审批流程存在一定的审核周期，且发行人需在审批通过后才能与境外客户签订并履行合同。

而根据国家互联网信息办公室 2023 年 2 月 3 日发布的《个人信息出境标准合同办法》（以下简称“《标准合同办法》”），个人信息处理者应当在标准合同生效之日起 10 个工作日内向所在地省级网信部门备案，备案应当提交标准合同及个人信息保护影响评估报告。因此，如发行人预计向境外提供 1 万人以上、不满 100 万人个人信息时，为尽快完成订单签署及交付，发行人可以与境外客户协商按照个人信息出境标准合同进行签署。这种情况下，发行人仅需将已签署合同提交省级网信部门备案即可，不会影响合同的及时签订和履行。

2、通过个人信息保护认证，发行人可在证书有效期内进行数据出境，无需履行网信部门审批备案程序

根据国家市场监督管理总局、国家互联网信息办公室 2022 年 11 月 4 日发布的《个人信息保护认证实施规则》，认证机构根据认证委托资料、技术验证报告等相关资料信息进行综合评价，作出认证决定，对符合认证要求的，颁发认证证书；认证证书有效期为 3 年，在有效期内，通过认证机构的获证后监督，保持认证证书的有效性。

上述个人信息保护认证是对发行人主体资格的认证，发行人一直遵守相关法律法规和规范性文件开展业务经营，发行人将在《数据跨境流动规定》正式出台后尽快启动办理个人信息保护认证事项。一旦通过个人信息保护认证，在认证有效期内，发行人将正常开展全部境外业务，且不再受网信部门数据出境安全评估申报审批或标准合同备案流程的影响。

综上，发行人认为《数据跨境流动规定》未来如出台对发行人开展境外业务的影响将进一步减小的依据充分。

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、根据《评估办法》《数据出境安全评估申报指南（第一版）》核查数据

出境安全评估申报审批的具体流程；对发行人业务负责人员进行访谈，了解申报审批流程对发行人境外客户订单的影响；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》适用范围内境外业务的收入占比；取得并查阅国家互联网信息办公室网络数据管理局出具的《评估结果通知书》，核查发行人数据出境安全评估的审核情况，分析数据出境安全评估申报审批流程对发行人境外订单的影响。

2、取得并查阅发行人数据出境安全评估申报文件，与发行人业务负责人员了解发行人数据出境中涉及的个人信息人数情况以及后续业务预计；查阅研究《数据跨境流动规定》、《标准合同办法》、《个人信息保护认证实施规则》，对照分析减小数据出境安全评估申报审批流程对发行人境外业务影响的依据，与发行人业务负责人员、数据保护官了解《数据跨境流动规定》出台后签署标准合同、申报个人信息保护认证等事项。

（二）核查意见

经核查，保荐机构及发行人律师认为：

1、数据出境安全评估申报审批具体包括申报、完备性查验、受理、安全评估、结果告知等流程；发行人部分境外业务的开展适用《评估办法》应履行数据出境安全评估申报审批程序，该部分境外业务收入占比较小，因此数据出境安全评估申报审批具体流程对发行人整体业务发展的影响较小；《评估办法》自 2022 年 9 月 1 日施行后，数据出境安全评估申报审批流程作为影响因素之一阶段性影响境外客户该类需进行安全评估审批的业务需求，目前数据出境安全评估申报审核政策已明晰，发行人已熟悉掌握数据出境安全评估申报审批流程及申报文件准备等相关事项，网信部门通过发行人已通过的申报审批事项对发行人的数据合规基本情况予以认可，有利于后续该类业务交易顺利推进申报审批流程，数据出境安全评估申报审批具体流程对发行人未来境外客户订单的影响较小；

2、《数据跨境流动规定》明确按照预计一年内向境外提供个人信息的数量区分是否进行数据出境安全评估申报，结合发行人以往及目前预计该类业务开展情况，发行人无需或可以不申报数据出境安全评估，从而不受数据出境安全评估申报审批流程的影响；如预计向境外提供 1 万人以上、不满 100 万人个人信息时，

发行人将采取与境外接收方订立个人信息出境标准合同并向省级网信部门备案或者通过个人信息保护认证的方式进行数据出境，将更有助于发行人与境外客户及时签署订单并完成交付；因此发行人认为《数据跨境流动规定》未来如出台对发行人开展境外业务的影响将进一步减小的依据充分。

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

2023年12月5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 股股票申请文件的第二轮审核问询函的回复》之签章页）



北京海天瑞声科技股份有限公司

2023 年 12 月 5 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 股股票申请文件的第二轮审核问询函的回复》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12月5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12月5日