

---

北京市天元律师事务所  
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 股股票的  
补充法律意见书（四）

---



北京市天元律师事务所  
北京市西城区金融大街 35 号  
国际企业大厦 A 座 509 单元  
邮编：100033

**北京市天元律师事务所**

**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 股股票的**

**补充法律意见书（四）**

京天股字（2023）第424-13号

致：北京海天瑞声科技股份有限公司（以下简称“发行人”或“公司”）：

北京市天元律师事务所（以下简称“本所”）接受发行人委托，担任发行人2023年度向特定对象发行A股股票（以下简称“本次发行”）的专项法律顾问，本所已为发行人本次发行出具了京天股字（2023）第424号《北京市天元律师事务所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股股票的法律意见书》（以下简称“《法律意见书》”）、京天股字（2023）第424-1号《北京市天元律师事务所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股股票的律师工作报告》（以下简称“《律师工作报告》”）、京天股字（2023）第424-4号《北京市天元律师事务所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股股票的补充法律意见书（一）》、京天股字（2023）第424-7号《北京市天元律师事务所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股股票的补充法律意见书（二）》、京天股字（2023）第424-10号《北京市天元律师事务所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股股票的补充法律意见书（三）》等法律文件，并已作为法定文件随发行人本次发行的其他申请材料一起上报至上海证券交易所（以下简称“上交所”）。

上交所针对发行人本次发行申请文件出具了上证科审（再融资）[2023]235号《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》（以下简称“《第二轮问询函》”）。本所律师现就《第二轮问询函》要求发行人律师补充核查的有关问题出具本补充法律意见书。本所律师在《法律意见书》中所作的声明、承诺同样适用于本补充法律意见书。如无特殊说明，本补充法律意见书使用的简称与

《律师工作报告》使用的简称含义相同。本补充法律意见书仅供发行人为本次发行之目的使用，不得被任何人用于其他任何目的。本所在此同意，发行人可以将本补充法律意见书作为本次发行申请所必备的法定文件，随其他申请材料一起上报上交所。

基于上述，本所律师依据相关法律法规规定，按照律师行业公认的业务标准、道德规范和勤勉尽责精神，出具如下补充法律意见：

## 一、第 1 项问题：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。”

（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

### 1、本次募投项目的具体内容

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，本次募投项目包括“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”。其具体内容如下：

（1）“AI 大模型训练数据集建设项目”：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，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。

（2）“数据生产垂直大模型研发项目”：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”。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，提升发行人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。

### 2、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”与发行人现有的 AI 训练数据生产平台体系相比，两者实质上均为发行人内部数据生产体系，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：

(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。

相比于发行人现有清洗工具，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。相比发行人现有标注工具，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。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	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。

对应数据生产环节		是否具备智能化能力	具体功能	对比
环节	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	<p>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。</p> <p>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。</p>
采集环节	现有平台	×	-	<p>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）。</p>
	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）	<p>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。</p>

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	✓ (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 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己不再区分具体语种，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己无法有效支撑该类数据预标注的需求；而在引入本次募投项目平台后，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。</p>
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。发行人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。

### **3、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**

本次募投项目为发行人现有产品矩阵的拓展和现有技术体系的迭代升级，与发行人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。发行人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，与发行人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发行人现有主营业务开展，属于发行人现有业务的延伸，与发行人现有业务具有较强的协同关系。

综上，本所律师认为，本次募投项目围绕发行人现有主营业务开展，属于发行人现有业务的延伸，与发行人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**

#### **1、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，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-国锐金崮	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

经核查，本次拟购买房产单价略高于上述本次拟购买标的周边其他同类型场地的平均单价及发行人 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。

综上，本所律师认为，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。

2、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

发行人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，未偏离发行人主业，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产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。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，推动包括发行人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（包括公共数据）。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，因此发行人需在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，发行人在本区域落地、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

发行人现有办公场所以租赁为主，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，略高于发行人目前人均办公面积，且两者均低于同行业公司，具有合理性。

单位：平方米/人

发行人现有场地人均面积	发行人募投项目人均办公面积	发行人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：发行人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，用于满足发行人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 平方米，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 平方米。

(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，使用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		<b>6.0</b>

经测算，从经济性角度，发行人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，发行人购置房产相较于租赁房产更具有经济性。

(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，未偏离发行人主业，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，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，提升发行人持续生产数据的能力，符合发行人“夯实传统业务，探索新型业务”的战略目标。发行人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发行人现有主营业务的拓展及延伸，符合发行人自身定位和未来发展规划，未偏离发行人主业，与发行人科技企业定位相符。

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，以提升发行人盈利能力。本次购置房产是募投项目顺利实施的必要保障，有助于发行人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%

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)
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		

因此，发行人虽为“轻资产”属性，但发行人的业务模式下，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。

综上，本所律师认为，本次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，未偏离发行人主业，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，发行人本次使用募集资金购房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。

(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

1、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发行人现有业务重合度较高，有助于发行人打开市场。以美团、科大讯飞、阿里巴巴、腾讯、百度等发行人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，并与发行人签署了大模型领域的业务合同或意向性订单。发行人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，发行人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

根据艾瑞咨询《2023 年中国 AIGC 产业全景报告》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。

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《2023 年中国 AI 技术变革企业服务白皮书》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。

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，发行人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%。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，本所律师认为，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。

2、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发行人最近5年（2018-2022年）/最近3年（2020-2022年）的现有同类数据集单位投入金额不存在明显差异，比较情况如下：

数据集类型		数据单位	本次募投产品单位投入金额（元）	发行人现有同类数据集单位投入金额（元）-最近五年	发行人现有同类数据集单位投入金额（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：发行人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，主要原因在于发行人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：发行人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：鉴于发行人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；发行人最近 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

发行人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，发行人选取 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

发行人根据 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%系参考发行人和同行业水平进行预测，发行人最近五年（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%系参考发行人和同行业水平进行预测，发行人最近五年（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发行人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，可比期间选取发行人前次 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，鉴于发行人自 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，故选取发行人 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，将发行人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，发行人与同行业同类项目相比差异较小；内部收益率方面，发行人内部收益率为 16.82%，处于可比区间；毛利率方面，发行人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，本所律师认为，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。

3、“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	<b>3,222.84</b>	<b>6,005.99</b>	<b>8,342.18</b>	<b>8,800.90</b>	<b>8,788.43</b>	<b>7,245.42</b>	<b>5,537.96</b>	<b>4,105.83</b>

上述各年成本费用将导致发行人净利润的下降，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，发行人投入“数据生产垂直大模型研发项目”的必要性及合理性如下：

①本项目建设是巩固发行人的核心技术壁垒，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发行人核心技术的重要体现。本项目的建设将基于发行人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，进一步提高发行人在人工智能基础数据服务领域的智能化水平，巩固发行人的核心技术壁垒，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，为发行人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本项目建设是优化发行人数据集生产流程，提升数据服务综合竞争力的有力途径

大模型训练数据集的生产流程包括设计、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发行

人的数据处理技术，促进数据资源处理经验的进一步沉淀，长期来看，可以大幅提高发行人的数据处理能力、效率，提升服务范围 and 水平，适应人工智能发展的新阶段，获得有效长期的发展动力，进一步巩固和提升发行人在数据服务领域的竞争力。

③本项目建设是发行人落实国家科技创新发展战略的重要举措

发行人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，升级发行人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。本项目的建设是发行人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发行人未来期间利润下滑，但作为巩固发行人的核心技术壁垒，构建长期技术实力支撑的必然手段，优化发行人数据集生产流程，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，本所律师认为，发行人“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，投入“数据生产垂直大模型研发项目”具有必要性、合理性。

**4、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**

发行人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 月因相关原因收入下滑外，发行人最近三年营业收入稳定在 2 亿元以上。

发行人前次募集资金投资项目已于 2023 年 5 月结项，发行人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。与发行人规划的前次募投项

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，发行人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，未来随着发行人经营规模持续扩大以及募投项目投产并逐步释放效益，发行人收入规模预计可合理支撑本次融资规模。

如前所述，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”为发行人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，符合发行人“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”为巩固发行人的核心技术壁垒，构建长期技术实力支撑的必然手段，优化发行人数据集生产流程，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，本所律师认为，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**

#### **1、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，与发行人目前人员设备配置情况一致，主要原因为发行人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，发行人将手机、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。

**2、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。发行人预计投入 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。

**(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发行人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。

综上，本所律师认为，发行人将手机、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发行人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

### 1、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

发行人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，提高了发行人训练数据产品的规模和种类，丰富完善了发行人的产品体系，满足了下游市场对于训练数据产品延伸开发的更多需求，增强了发行人的竞争力，达到了上述预期效果，促进了发行人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	<b>10,784.83</b>
利润总额	(74.64)	(101.01)	329.18	(53.93)	<b>99.60</b>

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。

### 2、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

### （1）前次 IPO 募集资金到账后使用计划（调整前）

发行人首次公开发行实际募集资金净额为 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%。

### （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，发行人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，根据发行人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，发行人增加“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。上述事项已经发行人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。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%。

发行人前次 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。

综上，本所律师认为，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，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 2023 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；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

本所律师会同保荐机构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 服务器等设备的具體应用；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、第 3 项问题：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据数量进行区分适用。

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。”

（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

### 1、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。

## 2、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。因此，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。

综上，本所律师认为，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。

(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：

**1、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。

**2、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，本所律师认为，《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，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。

### （三）核查程序

本所律师会同保荐机构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。

（本页以下无正文）

(本页无正文，为《北京市天元律师事务所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 股股票的补充法律意见书（四）》的签字盖章页)

北京市天元律师事务所 (盖章)



负责人: 朱小辉

朱小辉

经办律师 (签字): 周世君

周世君

王韶华

王韶华

顾鼎鼎

顾鼎鼎

2023 年 12 月 5 日

本所地址: 北京市西城区金融大街 35 号  
国际企业大厦 A 座 509 单元, 邮编: 100033