



关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的
科创板上市委会议意见落实函的回复

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）

上海证券交易所：

根据贵所于 2020 年 11 月 27 日印发的上证科审（审核）〔2020〕970 号《关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的科创板上市委员会意见落实函》（以下简称“落实函”）的要求，华泰联合证券有限责任公司（以下简称“华泰联合”或“保荐机构”）作为北京海天瑞声科技股份有限公司（以下简称“海天瑞声”、“发行人”或“公司”）首次公开发行股票并在科创板上市的保荐机构（主承销商），会同发行人及发行人律师北京市天元律师事务所（以下简称“天元律师”）和申报会计师毕马威华振会计师事务所（特殊普通合伙）（以下简称“毕马威会计师”）等相关各方，本着勤勉尽责、诚实守信的原则，就落实函所提问题逐项进行认真讨论、核查与落实，并逐项进行了回复说明。具体回复内容附后。

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首次公开发行股票并在科创板上市招股说明书（申报稿）》（以下简称“招股说明书”）一致。涉及招股说明书补充披露或修改的内容已在招股说明书中以**楷体加粗**方式列示。

2、本回复中若出现总计数尾与所列值和不符的情况，均为四舍五入所致。

3、本回复中涉及的我国、我国经济以及行业的事实、预测和统计，包括本公司的市场份额等信息，来源于一般认为可靠的各种公开信息渠道。本公司从上述来源转载或摘录信息时，已保持了合理的谨慎，但是由于编制方法可能存在潜在偏差，或市场管理存在差异，或基于其它原因，此等信息可能与国内或国外所编制的其他资料不一致。

目 录

问题1.....	3
问题2.....	6

问题 1

1. 请发行人在招股说明书中用通俗、可理解的语言说明其产品服务内容、作业模式和业务实质。

答复：

发行人对招股说明书进行了修订完善，并以通俗、可理解的语言对产品服务内容、作业模式和业务实质进行了补充披露：

1、发行人在“重大事项提示/一、发行人业务定位特别提示”中，对发行人业务定位进行了补充完善，并提示投资者阅读关于产品服务内容、作业模式、业务实质的通俗化说明，具体如下：

“发行人主要从事训练数据研发设计、生产及销售业务，提供用于人工智能算法模型开发训练所需的专业数据集，报告期内前述业务¹收入占比超过 98%。发行人的下游客户主要包括大型科技公司、人工智能企业及科研机构等，下游客户使用发行人提供的训练数据对算法模型进行开发、拓展和提升。当前发行人提供的训练数据主要是可满足多种场景应用需求的通用型训练数据，正在逐步探索更多垂直领域特征的训练数据覆盖。

发行人从属于软件和信息技术服务业，通过设计数据集结构、组织数据采集、对取得的原料数据进行加工及质检，形成训练数据集。发行人的主要产品、服务均以训练数据集为核心。成品训练数据集主要由数据文档、说明文档、技术文档三部分构成，并以软件形式向客户交付。发行人并非 AI 产业链上的算法模型研发商或提供商，提请广大投资者认真阅读本招股说明书关于发行人主营业务的具体说明。其中，对于发行人产品服务内容、作业模式、业务实质的通俗化说明，请参见本招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（一）主营业务概述/2、产品服务内容、作业模式和业务实质介绍”。

2、发行人在“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（一）主营业务概述/2、产品服务内容、作业模式和业务实质介绍”中以通俗、可理解的语言对产品服务内容、作业模式、业务实质进行了补充披露，内容如下：

¹即发行人的训练数据定制服务、训练数据产品业务。

“2、产品服务内容、作业模式和业务实质介绍

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

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

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

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

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

(2) 作业模式

总体而言，发行人的业务过程主要是下述几方面工作：①为了实现更好的算法模型训练效果，发行人需要对训练数据集的结构和内容进行设计；②需要采集大量数据，比如声音、文本、图像、视频；③需要对采集到的数据进行加工，把需要让算法学会

的信息，通过标注的方式告诉算法；④而在采集、加工的过程中，发行人则要对质量进行把控，持续进行质检。因此形成了训练数据生产的四大环节：设计、采集、加工、质检。

（3）业务实质

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

3、发行人对招股说明书“第六节 业务和技术/一、发行人主营业务、主要产品或服务的情况/（二）主要产品及服务/1、主要产品及服务按业务类型分类”进行了修改完善，具体如下：

“（1）训练数据定制服务

……

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

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

……

（2）训练数据产品

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

问题 2

2、.请保荐人、发行人律师对发行人向境外销售提供数据产品或定制服务是否符合国家数据安全、个人信息保护及境外出口限制等相关法规发表明确核查意见。

答复：

一、发行人向境外销售提供数据产品或定制服务符合国家数据安全、个人信息保护等相关法规；

关于发行人向境外销售提供数据产品或定制服务涉及的个人数据信息跨境提供，目前关于数据安全及个人信息保护的相关法规或规范中对此的专门规定均为原则性要求，具体的细则及相关的配套规定尚未出台，现有相关专门规定如下：

序号	法规名称	相关内容
1	《互联网个人信息安全保护指南》	6.2 保存 个人信息的保存行为应满足以下要求： a) 在境内运营中收集和产生的个人信息应在境内存储，如需出境应遵循国家相关规定；
2	《信息安全技术 个人信息安全规范》	9.8 个人信息跨境传输 在中华人民共和国境内运营中收集和产生的个人信息向境外提供的，个人信息控制者应遵循国家相关规定和相关标准的要求。

目前有关部门正在加紧推进数据安全和个人信息保护包含跨境提供等方面的立法工作，已公布了《中华人民共和国个人信息保护法（草案）》（2020年10月21日经第十三届全国人民代表大会常务委员会第二十二次会议审议发布公开征求意见）和《中华人民共和国数据安全法（草案）》（2020年7月3日经第十三届全国人民代表大会常务委员会第二十次会议审议发布公开征求意见）。其中《中华人民共和国个人信息保护法（草案）》主要针对向境外提供个人信息的个人信息处理者需要具备的条件、向境外提供个人信息需要事先告知被采集人并征求同意、处理个人信息达到规定数量的个人信息处理者向境外提供个人信息进行安全评估等事项作出规定；《中华人民共和国数据安全法（草案）》则规定对与履行国际义务和维护国家安全相关的属于管制物项的数据依法实施出口管制。上述法规草案均处于征求意见稿阶段，尚未正式出台，也不存在具体实施细则。

关于向境外销售提供数据产品或定制服务，发行人会告知被采集人并取得其授权同意；发行人与境外客户订立合同前，会对境外客户的数据安全和个人信息保护能力进行评估；发行人与境外客户订立合同，会约定双方就数据安全和个人信息保护的权利和义务，包括约定双方的个人信息处理活动遵守数据安全保护相关法规等内容。此外，发行人仅将经技术处理后的训练数据提供给客户，并不涉及提供被采集对象个人身份信息，客户后续使用上述业务数据也无需个人身份信息。同时，发行人研发使用一体化数据处理平台，通过采用技术措施减少数据接触范围，对于境内运营中收集和产生的数据和个人信息均在境内存储，并制定和公开了《个人信息保护政策》，定期开展个人信息安全影响评估，包含评估境外客户履约情况等，从而能够在产品或定制服务的境外销售过程中保障数据及个人信息安全。因此，发行人符合现行有效的国家数据安全、个人信息保护方面相关法规要求。

综上，发行人满足国家数据安全、个人信息保护方面的相关法规要求，并已经建立并不断强化完善了相关的数据安全和个人信息保护等合规性的措施，报告期内亦不存在被监管处罚的情形。伴随未来该领域配套监管政策出台，在目前规范化的基础上，发行人将积极采取措施满足新法规要求，不会对发行人未来生产经营造成重大不利影响。

二、发行人向境外销售提供数据产品或定制服务符合境外出口限制等相关法规；

（一）发行人向境外销售提供数据产品或定制服务涉及的境外出口限制等相关法规情况

目前对向境外销售数据信息进行限制约束的法律法规具体包括《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》、《中华人民共和国网络安全法》及特定行业（例如医疗健康行业）的法律规章。

逐项对照上述法律规章中关于数据信息出口的相关要求，对发行人报告期内向境外销售提供数据产品或定制服务涉及法规具体说明如下：

序号	法规名称	相关内容	与发行人实际情况的对照说明
1	《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》	《中华人民共和国技术进出口管理条例》： 第三十一条 属于限制出口的技术，实行许可证管理，未经许可，不得出口。 第三十二条 出口属于限制出口的技术，应当向国务	1、发行人向境外提供的是训练数据产品或服务，不涉及对境外客户提供语

		<p>院经贸主管部门提出申请。</p> <p>《中国禁止出口限制出口技术目录》： （限制出口部分） （十五）计算机服务业 编号：056101X 技术名称：信息处理技术 控制要点： 9. 汉字、语音识别技术 10. 汉语或少数民族语音合成技术 11. 汉字压缩、还原技术 12. 印刷体汉字识别技术、程序结构、主要算法和源程序 13. Videotex[可视图文]系统的汉字处理技术及网间控制技术 14. 具有交互和自学习功能的脱机手写汉字识别系统及方法 15. 用于计算机汉字输入识别方法中的手写体样张、印刷体样张以及汉语语料库 16. 汉字识别的特征抽取方法和实现文本切分技术的源程序 17. 语音合成技术（包括语料库设计、录制和标注技术，语音信号特征分析和提取技术，文本特征分析和预测技术，语音特征概率统计模型构建技术等） 18. 人工智能交互界面技术（包括语音识别技术，麦克风阵列技术，语音唤醒技术，交互理解技术等） 19. 语音评测技术（包括朗读自动评分技术，口语表达自动评分技术，发音检错技术等）</p>	<p>音识别、语音合成等相关的技术或技术服务； 2、发行人生产的计算机视觉、自然语言类训练数据中涉及15项“用于计算机汉字输入识别方法中的手写体样张、印刷体样张以及汉语语料库”，但报告期内发行人严格遵守法规要求，均不出口销售该部分训练数据。</p>
2	《健康医疗大数据标准、安全和服务管理办法（试行）》	<p>第四条 本办法所称健康医疗大数据，是指在人们疾病防治、健康管理等过程中产生的与健康医疗相关的数据。</p> <p>第三十条 责任单位应当具备符合国家有关规定要求的数据存储、容灾备份和安全管理条件，加强对健康医疗大数据的存储管理。健康医疗大数据应当存储在境内安全可信的服务器上，因业务需要确需向境外提供的，应当按照相关法律法规及有关要求进行安全评估审核。</p>	<p>发行人向境外提供的数据信息不含健康医疗大数据。</p>
3	《人类遗传资源管理条例》	<p>第二条 本条例所称人类遗传资源包括人类遗传资源材料和人类遗传资源信息。</p> <p>人类遗传资源材料是指含有人体基因组、基因等遗传物质的器官、组织、细胞等遗传材料。</p> <p>人类遗传资源信息是指利用人类遗传资源材料产生的数据等信息资料。</p> <p>第七条 外国组织、个人及其设立或者实际控制的机构不得在我国境内采集、保藏我国人类遗传资源，不得向境外提供我国人类遗传资源。</p> <p>第九条 采集、保藏、利用、对外提供我国人类遗传资源，应当符合伦理原则，并按照国家有关规定进行</p>	<p>发行人向境外提供的数据信息不含人类遗传资源信息。</p>

		伦理审查。	
4	《中华人民共和国网络安全法》	第三十七条 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个人信息和重要数据应当在境内存储。因业务需要，确需向境外提供的，应当按照国家网信部门会同国务院有关部门制定的办法进行安全评估；法律、行政法规另有规定的，依照其规定。 第三十一条 国家对公共通信和信息服务、能源、交通、水利、金融、公共服务、电子政务等重要行业和领域，以及其他一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的关键信息基础设施，在网络安全等级保护制度的基础上，实行重点保护。关键信息基础设施的具体范围和安全保护办法由国务院制定。	发行人不属于关键信息基础设施的运营者，不适用该规定。

（二）发行人符合《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》等相关法规规定

关于上文所述的《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》的相关内容，发行人符合其相关规定，具体分析论述如下：1、关于《中国禁止出口限制出口技术目录》（限制出口部分）“（十五）计算机服务业/编号:056101X /技术名称：信息处理技术”中第15项“用于计算机汉字输入识别方法中的手写体样张、印刷体样张以及汉语语料库”：手写体样张、印刷体样张属于公司计算机视觉业务，汉语语料库属于公司自然语言业务。报告期内发行人严格遵守法规要求，不存在从中国境内出口该部分训练数据的情形。

2、关于《中国禁止出口限制出口技术目录》（限制出口部分）中“（十五）计算机服务业/编号:056101X /技术名称：信息处理技术”中第9项“汉字、语音识别技术”、第10项“汉语或少数民族语音合成技术”、第17项“语音合成技术（包括语料库设计、录制和标注技术，语音信号特征分析和提取技术，文本特征分析和预测技术，语音特征概率统计模型构建技术等）”及第18项“人工智能交互界面技术（包括语音识别技术，麦克风阵列技术，语音唤醒技术，交互理解技术等）”：发行人向客户提供的训练数据产品或定制服务均以训练数据集为核心，通常以软件形式向客户交付，主要由数据文档、说明文档、技术文档三部分构成。其中，数据文档是交付成果的核心，说明文档和技术文档一般作为附加文件同时提供给客户，作为数据文档相关标准、规范、使用的说明。数据文档根据业务领域不同所交付的内容有所不同：

（1）智能语音业务中数据文档通常包括原始采集形成的音频文件、与音频文件对

应的带有时间戳的标注文件；

(2) 计算机视觉业务中数据文档通常包括图像和视频等数据文件、标注文件；

(3) 自然语言业务中数据文档通常包括对话文本等数据文件、标注文件。

因此，发行人向客户提供的交付成果中并不包含《中国禁止出口限制出口技术目录》（限制出口部分）中“（十五）计算机服务业”中第9、10、17、18项所列示的相关技术。此外，发行人在训练数据集的生产和质检环节中会运用语音识别、语音合成算法技术提高生产效率、检测产品质量，但发行人并不对外出售前述语音识别、语音合成算法技术。

另外，发行人向客户销售提供数据产品或定制服务，提供的数据信息包括声音、文本、图形图像、视频、肖像等数据信息，不涉及提供健康医疗数据、人类遗传资源信息。

综上，发行人向境外客户提供训练数据产品或训练数据定制服务符合《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》、《健康医疗大数据标准、安全和服务管理办法（试行）》、《人类遗传资源管理条例》等境外出口限制的相关法规规定，发行人不存在违反上述相关法规的情形。

中介机构核查事项

一、核查程序

保荐机构及发行人律师执行的主要核查程序如下：

1、查阅对照《互联网个人信息安全保护指南》、《信息安全技术 个人信息安全规范》等现行有效法规中关于国家数据安全及个人信息保护的相关规定，研究查询《中华人民共和国个人信息保护法（草案）》、《中华人民共和国数据安全法（草案）》等法律文件中相关规定和立法进展情况；

2、向发行人数据保护官了解业务经营涉及的向境外提供数据的保护措施，查阅《个人信息保护政策》、《数据安全管理制度》等相关文件；

3、查阅对照《中华人民共和国技术进出口管理条例》及《中国禁止出口限制出口技术目录》等法律法规中关于境外出口限制的相关规定，查阅发行人境外业务相关合

同、与发行人业务主管人员沟通了解确认发行人向境外提供的数据信息及最终交付成果的内容及形式等；

4、检索报告期内发行人是否因数据安全、个人信息保护、境外出口事项而被处罚或发生争议或纠纷。

二、核查结论

经核查，保荐机构及发行人律师认为：发行人向境外销售提供数据产品或定制服务符合现行有效的国家数据安全、个人信息保护及境外出口限制等相关法规。伴随未来数据安全、个人信息保护领域配套监管政策出台，在目前规范化的基础上，发行人将积极采取措施满足新法规要求，不会对发行人未来生产经营造成重大不利影响。

（本页无正文，为《关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的科创板上市委会议意见落实函的回复》之签章页）

发行人董事长（签名）：贺琳

贺琳



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科创板上市委员会意见落实函回复的全部内容，确认科创板上市委员会意见落实函回复中不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性、及时性承担相应法律责任。

发行人董事长（签名）： 贺琳

贺琳

北京海天瑞声科技股份有限公司



（本页无正文，为《关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的科创板上市委会议意见落实函的回复》之签章页）

保荐代表人： 张鹏

张 鹏

葛青

葛 青



保荐机构总经理声明

本人已认真阅读北京海天瑞声科技股份有限公司本次科创板上市委员会意见落实函回复的全部内容，了解科创板上市委员会意见落实函回复涉及问题的核查程序、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，本回复中不存在虚假记载、误导性陈述或重大遗漏，并对其真实性、准确性、完整性、及时性承担相应法律责任。

保荐机构总经理 (签名):



马 骁

华泰联合证券有限责任公司

