
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
首次公开发行股票并在科创板上市的
补充法律意见书（六）



北京市天元律师事务所

北京市西城区丰盛胡同 28 号

太平洋保险大厦 10 层

邮编：100032

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
首次公开发行股票并在科创板上市的
补充法律意见书（六）

京天股字（2020）第 355-28 号

致：北京海天瑞声科技股份有限公司

北京市天元律师事务所（以下简称“本所”）根据与北京海天瑞声科技股份有限公司（以下简称“发行人”）签订的《委托协议》，接受发行人委托，担任发行人首次公开发行股票并在科创板上市（以下简称“本次发行上市”）的专项法律顾问。本所已为发行人本次发行上市出具了京天股字（2020）第355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首次公开发行股票并在科创板上市的法律意见书》（以下简称“《法律意见书》”）、京天股字（2020）第355-1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首次公开发行股票并在科创板上市的律师工作报告》（以下简称“《律师工作报告》”）、京天股字（2020）第355-5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首次公开发行股票并在科创板上市的补充法律意见书（一）》、京天股字（2020）第355-9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首次公开发行股票并在科创板上市的补充法律意见书（二）》、京天股字（2020）第355-16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首次公开发行股票并在科创板上市的补充法律意见书（三）》、京天股字（2020）第355-17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首次公开发行股票并在科创板上市的补充法律意见书（四）》、京天股字（2020）第355-23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首次公开发行股票并在科创板上市的补充法律意见书（五）》等法律文件，并已作为法定文件随发行人本次发行上市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”）。

针对发行人本次发行上市申请文件，上交所于2021年2月25日转发了《发行注册环节反馈意见落实函》（以下简称“《注册落实函》”）。本所律师现就《注册落实函》要求发行人律师补充说明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上市之目的使用，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上市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：

问题 2：“2、申请材料显示，发行人主要从事训练数据的研发设计、生产及销售业务，核心技术先进性主要体现在算法与数据并用、工具和平台共建、在语言语音学基础研究方面有深厚积累。发行人所在细分领域为人工智能基础数据服务领域，基础数据服务行业的业务环节分为为训练数据设计、采集、加工和质检。

请发行人用通俗易懂的语言具体说明自身核心技术在主要业务环节的应用情况、方式和效果；结合主营业务及同行业可比公司说明训练数据加工及处理、提供具体服务内容等方面是否存在技术优势与核心竞争力；说明发行人在人工智能细分领域提供基础数据服务所产生的增值服务内容；结合上述分析说明发行人所处行业及细分领域是否符合科创板定位，科创属性行业分类是否准确。请保荐机构和律师核查并发表明确意见。”

（一）请发行人用通俗易懂的语言具体说明自身核心技术在主要业务环节的应用情况、方式和效果

发行人在基础研究、平台工具及训练数据生产3个层次共积累下12项核心技术，并将前述技术运用在训练数据生产环节之中。各项核心技术在发行人主要业

务环节中的应用情况汇总如下：

3个核心技术层次	12项核心技术	核心技术项下的细分技术示例	核心技术在训练数据生产中的应用环节			
			设计	采集	加工	质检
基础研究	语音识别算法	语音数据库质量预估技术	√	√	√	√
	语音合成算法	语音合成数据库评估技术、说话人自适应语音合成技术、语音合成系统评测技术				√
	计算机视觉算法	人脸检测和识别技术、物体识别技术、光学字符识别技术、场景分割技术、行人检测技术、运动轨迹跟踪技术			√	√
	★语音语言学基础研究	基于词典与模型的发音预测技术	√			
	训练数据集设计技术	N元语言模型训练与优化技术、文本正则化技术、基于语言模型的文本易读性评测技术	√			
平台工具	一体化数据处理平台	基于C/S架构的大规模语音处理平台，基于C/S架构的音素标注技术，多语言分布式文本处理技术，基于C/S架构的文本词性标注、文本属性标注技术，基于C/S架构和B/S架构的图像标注技术，大规模数据采集及标注平台	√	√	√	√
训练数据生产	★多语种多模态训练数据设计技术	多语种语料清洗技术、多语种语音库设计技术、多语种音素均衡语料设计技术、混合语言语料设计技术、AudioBook自动切分与文本对齐技术	√			
	多模态多通道数据采集技术	多通道录音技术、移动设备上的语音采集技术、分布式图片内容分析技术、移动设备上的图片采集技术、多语言手写体数据采集技术、多模态数据采集技术		√		
	★数据同步技	音频对齐技术			√	

	术					
	实时采集质控技术	语音信号质量检测技术		√		√
	★大数据驱动的高效数据处理技术	多语种拼写检查技术、视频流中特定帧定位技术、音素边界自动预测技术、基于语音特征的韵律预测技术、基于 CRF 的韵律预测技术、基于 HMM/CRF 的词性预测技术、基于 SVM 的文本分类技术、基于 CRF 的命名实体识别技术、基于前后向最大匹配的中文分词技术			√	
	★分布式高性能自动校验技术	音素标注正确率校验技术、相似说话人自动筛查技术、音字一致校对技术				√

注：上述12项核心技术中，5项核心技术具备较高技术壁垒，在上表中以★表示。

12项核心技术在发行人主要业务环节中的具体应用情况、方式、效果具体如下：

1、核心技术 1：语音识别算法

语音识别（Automatic Speech Recognition，ASR）是让机器能够“听懂”人类语音的技术，它能使机器自动将语音信号转换为对应的文本信息。在语音识别算法领域，发行人拥有基于多种语音识别模型的多语言语音识别技术及相关深度学习技术，已掌握语音数据库质量预估技术等细分技术，取得了 3 项发明专利授权并正在申请 1 项发明专利。

语音识别算法运用于设计、采集、加工、质检 4 个主要业务环节，主要应用在提升训练数据的生产效率和保障训练数据的质量上，应用情况、方式、效果具体如下：

（1）应用情况、方式

①设计环节应用——运用语音识别算法反向检验语音识别训练数据的训练效果、为训练数据设计生产提供指导：发行人用其生产完成的训练数据集，对自有的语音识别算法模型进行训练，测试检验训练数据集的最终训练效果。

②采集、加工及质检环节应用——执行自动化、程序化的数据加工、质检工作，可进行程序化操作，并与人工检查配合：发行人将运用语音识别算法集成在自主开发的数据采集、加工工具之中，一方面可在采集过程中及时校验采集数据情况，向被采集对象提供实时反馈，提高采集的效率和准确率；另一方面运用在语音数据加工环节中，如运用语音识别算法对智能语音数据进行预标注，与人工检查环节结合，提升单位时间加工语音数据的能力，提高生产效率和质量；也可通过语音识别算法对智能语音训练数据执行程序化质检，反向检验人工数据加工结果。

③其他应用——应用在语音识别训练数据训练效果的展示之中：发行人可将语音识别算法模型运用在训练数据的模拟训练和测试之中，如使用特定训练数据集中的部分数据对发行人的语音识别算法进行训练，后续将所需转换的语音输入到算法模型中，为客户直观展示该算法模型经训练后将语音转换为文字的能力、准确度情况，直观反映训练数据的模型训练、测试效果。例如，发行人运用不同语种的训练数据产品对同一语音识别算法模型进行了训练，通过算法模型训练直观展示发行人的训练数据产品具备较优的测试效果，各语种训练数据训练后的语音算法模型的字符错误率/词错误率¹均低于 5%。此外，发行人还运用语音识别算法技术为客户提供部分算法模型拓展服务，协助客户完成模型的拓展和优化。

（2）应用效果

①使得发行人可结合反向检验测试结果，分析不同结构的训练数据集的训练效果差异，进而在前端训练数据的结构开发、内容设计等环节中应用前述测试反馈和经验，调整训练数据集结构设计，从而生产出具备更佳训练效果的训练数据。

②实现自动化、程序化的数据加工、质检，提升单位时间加工、质检数据的能力、提高了生产效率和质量；并可在采集过程中及时校验采集数据情况，向被采集对象提供实时反馈，提高采集的效率和准确率。

③实现语音识别训练数据训练效果的直观展示，方便客户衡量训练数据质量水平，从而选购与其预期训练效果更为匹配的语音识别训练数据。

¹ 数据来源：发行人测试结果。字符错误率、词错误率为语音识别算法模型识别正确率的反向衡量指标，错误率越高则模型识别准确率越低、识别效果越差。

④协助下游客户完成算法模型的语言拓展、特定算法模块拓展、垂直领域应用拓展等，为客户定制针对特定行业和口音的专属算法模型。

2、核心技术 2：语音合成算法

语音合成（Text to Speech, TTS）是让机器能够“说出”人类语音的技术，它使机器能将文字信息转化为流畅的语音“朗读”出来，相当于给机器安上了人工嘴巴。在语音合成算法方面，发行人拥有参数合成技术、基于神经网络的端到端语音合成技术等细分技术，并运用语音合成技术，基于其生产的语音合成训练数据做了大量的算法模型训练、语音合成效果验证和测试，并通过持续的算法迭代和优化逐步提升语音合成算法技术的研究水平。发行人现已拥有语音合成的 AI 算法模型，在语音合成技术领域具备一定的技术积累，已取得 2 项发明专利授权，正在申请 1 项发明专利。

发行人将语音合成算法技术运用在质检环节，提升了生产效率，具体情况、方式、效果如下：

（1）应用情况、方式

①质检环节应用——检验语音合成训练数据的训练效果：发行人主要将算法模型用于检验语音合成训练数据的合成效果。具体而言，在其语音合成训练数据集生产完成中，发行人将使用该训练数据集对自己储备的语音合成算法模型进行训练，检验该算法经训练后将文字转换为语音的合成效果。

②其他应用——直观展示训练数据的合成效果：发行人使用上述语音合成算法模型，基于不同的训练数据集进行了语音合成效果测试，为客户直观展示训练数据的合成样音，帮助客户选择合意的训练数据。

（2）应用效果

①检验训练数据质量、反哺指导训练数据生产：通过运用语音合成算法技术，发行人充分检验了语音识别训练数据的训练效果，并结合前述训练、测试结果和经验，反向调整训练数据的结构设计，完善采集、加工和质检环节设置，有针对性地提升训练数据质量。

②使得客户可以在实际开展算法模型训练和测试工作之前初步了解特定训

练数据集的语音合成效果，从而选择风格、音色等更符合需求的训练数据，最终取得理想的语音合成效果。

3、核心技术 3：计算机视觉算法

计算机视觉（Computer Vision，CV）是使机器具备“看”的功能的技术，它使得智能家居、手机、安防设备等机器、程序能够代替人眼对目标进行识别、跟踪和测量等。发行人积极在该领域积累算法技术能力，以更好地理解下游客户需求，相应生产、提供高质量的训练数据集，同时将计算机视觉算法技术嵌入到自主研发的一体化数据处理平台中，执行程序化数据加工，提升生产效率。在计算机视觉算法领域，发行人已取得 4 项发明专利授权，另有 1 项发明专利正在申请中。

发行人主要将计算机视觉算法技术运用在加工、质检环节，实现训练数据生产过程的降本增效，具体应用情况、方式、效果如下：

（1）应用情况、方式

加工、质检环节运用——嵌入到自主研发的分布式标注平台之中，运用计算机视觉算法程序执行加工和质检工作。

（2）应用效果

通过计算机视觉助力自动化、程序化训练数据加工，降低了人工加工工作量和成本，提升了数据加工的准确性。

以发行人的 1,000 人手机人像短视频训练数据集（King-AV-025）为例，该训练数据集是发行人针对手机短视频领域人像识别应用所开发的训练数据集，数据集中包括 1,000 个黄种人共计 1,000 段短视频数据，视频时间长度合计达到 25 小时。发行人采集完构建该训练数据集所需的短视频原料数据后，需要对上述短视频数据进行检查和进一步的加工。为保证训练数据集的质量，发行人需要检查、确认上述视频原料数据不存在重复。在上述“去重”的数据检验排查过程中，发行人借助其在计算机视觉算法领域积累的“视频筛选方法、装置及存储介质”技术，针对各个短视频批量自动截取关键帧并执行相似度计算，对于算法检查出的、相似度较高的短视频原料数据，再做人工排查。该技术降低了人工比对成本，提

供了数据加工过程的准确性。

4、核心技术 4：★语音语言学基础研究

语音语言学领域的专业知识是构建高质量语音识别算法和语音合成算法的关键要素。以语音合成为例，语音合成系统可以实现从文本到发音的语音合成过程，在该过程中，发音词典必不可少：发音词典提供了从单词到音素之间的映射关系，可将语言模型建模单位解构为声学模型的建模单元，为后续合成发音奠定基础。具体而言，语音合成系统在接收到文本信息后，首先运用发音词典对文本信息进行语言、韵律处理，将文本（单词、字符等）转换并拆分、解构成一系列对应的发音符号（类似于国际音标）；随后，系统中的语音合成器接收到前述发音符号，运用语音库合成转换为语音对外输出，完成文本到语音的语音合成过程（参见下图）。发音词典质量将直接影响文本转换为语音的发音准确性，进而影响合成效果，在语音合成系统中具备重要作用。



图：语音合成系统框架示意

发行人将语音语言学基础研究成果运用在设计环节，实现训练数据生产过程的降本增效，具体应用情况、方式如下：

（1）应用情况、方式

通过语音语言学基础研究，发行人构建了成熟的发音词典构建流程、构建技术，同时对基于规则和统计的发音预测算法进行了多年研究，结合长期经营获得

的大量语音数据和经验积累，研发并掌握了基于词典与模型的发音预测技术等。基于前述研究成果和技术积累，在面对已有发音词典积累的语种/方言类别时，发行人可较快速地复用此前积累的发音词典，设计、制作相应的训练数据，助力语音识别、合成算法模型在大词汇量的连续语音交互中正确、合理地运用相关的语言模型、语法和词法模型；在面对此前未积累有发音词典的语种、方言，发行人也运用其计算语言学等基础技术，较快地构建新语种/方言训练数据集所需的发音词典，加快训练数据集设计、制作过程。

(2) 应用效果

发行人通过语音语言学领域的基础研究，积累了计算语言学、发音规则、发音词典等基础知识和研究技术成果，截至目前已积累了超过 100 个语种/方言的发音词典，累计词条数超 1,000 万条，稀有语言覆盖数量达 26 种，并将其应用于高质量智能语音训练数据的构建过程之中。

5、核心技术 5：训练数据集设计技术

训练数据的设计环节有很多学问，以语音识别相关算法、设备为例（如智能音箱），发行人需要结合前述设备的使用场景，以及使用者方言的分布、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，使得最终形成的训练数据集具备更合理、全面的分布结构，助力语音识别算法、设备实现良好的识别效果和体验感。

经过多年经营积累，发行人已具备 150 余个语种/方言的训练数据设计开发能力，并掌握了 N 元语言模型训练与优化技术、文本正则化技术²、基于语言模型的文本易读性评测技术等算法技术，可结合音素均衡、语料主题、覆盖领域、发音人属性、设备特征等因素综合设计满足算法模型开发、训练、拓展需要的训练数据集，并运用前述技术提升训练数据集结构的合理性、科学性。在训练数据集设计技术领域，发行人已取得 1 项发明专利授权。

训练数据集设计技术主要应用在设计环节，发行人应用前述技术科学、合理

² 文本正则化是语音合成系统的主要模块之一，主要作用是将有多种读音方案的文本根据其所处的语境、上下文确定准确发音。以数字文本为例，同一数字在表示具体年份时（如“2019 年”读作“二零一九年”）和表示计数时（如“2019 个苹果”读作“两千零一十九个苹果”）的读音是不同的。文本正则化技术的目的是：尽可能准确地消除文本在读音层面的歧义，以确保后续合成的语音读音准确。

高效地设计智能语音、计算机视觉和自然语言处理等训练数据集结构，为后续生产高质量训练数据奠定基础。该技术的应用情况、方式、效果具体如下：

(1) 应用情况、方式

以语音识别、语音合成领域的训练数据集为例，在原料数据的采集环节，发音人（被采集对象）需要朗读发行人提供的基础语料，并用指定的录音设备录制形成原料音频数据。训练数据集设计环节的工作即包含语料设计工作。发行人应用训练数据设计技术，结合此前训练数据生产过程中的经验及技术诀窍积累，充分考虑如何设计基础语料，使得容量有限的训练数据集能够覆盖尽可能多的自然语言现象（如覆盖更多的语音、语气、语调情况，兼顾多种句子类型（陈述句、疑问句、感叹句等）和句长分布），结合不同的发音习惯、语言特点，实现更好的语料覆盖效果。在训练数据设计技术领域，文本正则化技术、基于语言模型的文本易读性评测技术等细分技术的应用，使得发行人可以设计出更为易读、发音更为准确的语料。

(2) 应用效果

实现高效、科学、合理的训练数据集结构设计，使发行人拥有广泛的文本覆盖能力，例如通用语音识别训练数据产品文本类型覆盖数量可达 1-18 类不等、车载语音识别训练数据产品文本类型总数高达 55 类，使得算法模型获得更好的训练结果，助力客户的语音识别模型能够“听懂”更多实际场景中的语音数据，语音合成模型能够“说出”更为自然的语音效果。

6、核心技术 6：一体化数据处理平台

一体化数据处理平台嵌入了训练数据生产过程中所需的各类工具、软件模块，由发行人自主研发，整合贯通了训练数据生产各环节（设计、采集、加工、质检）及数据安全的需求，整合多维度、多创新点的工具群，构建了流程化、规范化和体系化的训练数据生产体系，提高了产能及效率，降低训练数据生产成本、缩短了训练数据生产处理时间。

在与一体化数据处理平台相关的技术领域，发行人已取得 2 项发明专利授权，正在申请 2 项发明专利。

一体化数据处理平台是发行人的生产系统，应用在发行人全部的主要业务环节（设计、采集、加工、质检）之中，应用情况、方式、效果具体如下：

（1）应用情况、方式

一体化数据处理平台应用于训练数据生产全过程，为发行人业务项目的具体执行、实施提供对应的工具和流程，便于生产过程的数据处理、质量把控，并对生产过程的数据进行记录和保存。例如，在采集环节，一体化数据处理平台整合了多通道桌面录音工具、手机通道录音工具、双目 8 路视频采集工具、深度摄像头采集工具、3D-人脸采集工具等；在加工环节，一体化数据处理平台整合了语音识别、语音合成、3D 点云、通用视频、OCR 等标注工具等。

一体化数据处理平台现由数据处理工具集、终端人管理系统和日志管理系统三大功能模块构成，除可满足训练数据的生产需要外，还可记录采购的数据服务对应的劳务工作量、产出量/数据量等信息，可在验收、决算阶段与发行人业务管理平台生成的财务相关信息进行交叉核对，进一步确保财务信息与生产、业务信息一致、准确。

（2）应用效果

通过平台上各类数据处理工具的开发和整合，实现训练数据的高效高质生产，确保财务信息与生产、业务信息一致准确，实现终端劳务人员等人员的统一化信息管理，留存生产过程日志信息，提升生产和管理效率。



图：一体化数据处理平台模块、功能构成图示

7、核心技术 7：★多语种多模态训练数据设计技术

运用多语种多模态训练数据设计技术可使发行人结合具体场景、应用领域特性，设计与之相匹配的、可用于算法模型开发、训练、拓展及优化的多模态训练数据。如通过多种终端设备同时获取人发出的对话语音信息、唇部动作信息、声音来源方向信息等多维度信息，随后综合运用计算机视觉技术、智能语音技术等算法技术实现视觉、听觉等融合的多维度交互。在多语种多模态训练数据设计技术领域，发行人已取得 4 项发明专利授权。

多语种多模态训练数据设计技术主要应用在设计环节，具体应用情况、方式和效果如下：

(1) 应用情况、方式

发行人的多语种多模态数据设计技术包含多语种音素均衡语料设计技术、混合语言语料设计技术等细分技术，解决了多语种、多模态的训练数据设计环节的具体痛点，具体说明如下：

①多语种训练数据设计

在该领域，发行人掌握的多语种多模态训练数据设计技术主要包含多语种音素均衡语料设计技术、混合语言语料设计技术等子技术，可解决多语种智能语音训练数据集设计时的音素覆盖、句长分布、领域内容分布问题，高效生产混合语料（如中英文混合等）训练数据。对于上述主要细分技术应用情况和方式介绍如下：

多语种音素均衡语料设计技术：该技术可根据训练数据集设计时的原始句长分布需求，从原始语料中选取句数、句长要求符合需求、且与所述原始句长分布相匹配的语料，作为初始句长分布模型（以下简称“初始模型”），形成与原始句长分布一致或较为接近的初始模型；再通过技术化修正处理初始模型得到满足总字数、句数、句长的最终模型，使得最终模型更接近原始语料的句长分布情况，实现语料设计中对于句长分布的要求。

混合语言语料设计技术：该技术运用于训练数据集设计阶段，可基于原始中

英混语料的特征提取，通过转化纯中文语料高效生成中文汉字、英文单词及标点符号的中英混合语料，并兼顾易读性和常见性，形成具备良好中英文混合效果的训练数据。发行人已在该细分技术领域取得了 1 项发明专利授权。

②多模态训练数据设计

在该领域，发行人掌握的多语种多模态训练数据设计技术使得发行人可结合具体场景、应用领域特性，设计与之相匹配的、可用于算法模型开发、训练、拓展及优化的多模态训练数据。如设计科学合理的采集方案，助力后续通过多种终端设备同时采集发音人发出的对话语音信息、唇部动作信息、声音来源方向信息等多维度信息，便于综合运用计算机视觉技术、智能语音技术等算法技术实现视觉、听觉等融合的多维度交互。

(2) 应用效果

①解决多语种智能语音训练数据集设计时的音素覆盖、句长分布、领域内容分布问题，高效生产混合语料（如中英文混合等）训练数据，有效扩充中英混合语料库，为中英混合语音识别的模型训练提供充足的训练数据，提高算训练数据的质量、训练效果。

②使得发行人设计的多模态训练数据集机构科学合理，并配合设计科学合理的采集方案，为后续采集多维度信息、形成高质量多模态训练数据奠定基础。

8、核心技术 8：多模态多通道数据采集技术

多模态多通道采集过程是指通过部署多台设备实现多通道多模态原料数据采集，具体工作方式举例如下：针对同一发音人，同时部署多台录音设备、多路视频采集设备，从而同时采集发音人的语音、视频、唇形等多维度不同形态的数据。

在多模态多通道数据采集过程中，各录音设备、视频采集设备等需接入同一服务器，可能出现多路设备冲突导致丢失数据包、视频帧、语音视频不一致的情形。发行人针对上述现象，自主研发了多模态多通道数据采集技术，并开发完善了相应的采集软件、软硬件一体系统，能够实现多维度原料数据的采集和有机整合。

针对上述问题和生产痛点，发行人自主研发了多模态多通道数据采集技术，并在该领域取得了 2 项发明专利授权及 1 项实用新型专利授权。

多模态多通道数据采集技术主要应用在采集环节，将该技术的具体应用情况、方式和效果举例说明如下：

(1) 应用情况、方式

该技术的研发成果主要体现在发行人自主研发的采集软件、软硬件一体系统等采集工具之中，在发行人 8 类采集设备、6 类采集通道上均可应用，解决了多路采集设备、软件录入数据同时录入、自动对齐整合的采集需求。

(2) 应用效果：实现多路原料数据的同时采集、自动对齐

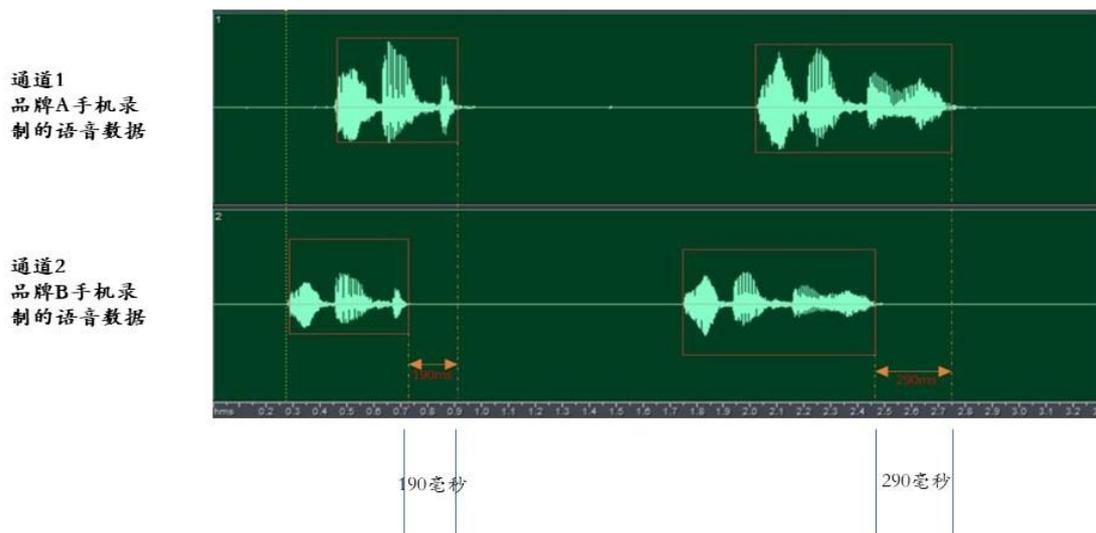
以发行人的 200 人中文普通话语音识别数据库（King-ASR-010）和 250 人唇语语音视频数据库（King-AV-018）生产过程为例：

King-ASR-010 是发行人生产的中文普通话语音识别训练数据集，该数据集覆盖了 4 个通道（在同一桌面不同位置布设的 4 个录音设备）的音频数据，每个通道覆盖 200 人共 6,000 句语音数据。发行人自主研发的多通道采集工具很好地支持了 4 路语音数据同时录入、自动对齐整合的采集需求。

King-AV-018 是发行人针对虚拟主播需求开发的视频训练数据集，该数据集中含有 250 个中国人共计 610 句的唇语视频信息。发行人在常见的语音数据采集设备上叠加了针对唇语信息的视频采集设备，运用自主研发的多模态数据采集软件同时录入不同硬件设备的数据，并保持多维度数据的对齐。

9、核心技术 9：★数据同步技术

数据同步性是训练数据集质量的重要评价标准。而在数据采集过程中，往往会出现多通道（多个设备）同时采集的需求，如多个收音设备同时记录声音、多个拍摄设备同时记录影像等。因设备的物理限制，受设备采集开始时间不同步、设备本身时钟不同步、结束时间不同步等因素影响，多个设备采集的原始数据很容易发生不同步的情况（参见下图示例），使得后续数据加工难度大大增加。



图：不同通道录制的语音数据不对齐情况示例——针对同一音源，不同品牌手机同时录制的语音数据的时间起点不一致，且不同语音片段的终点也不一致，且起点至终点之间并不是完全线性平移的关系，例如第一段语音的终点差距为 190 毫秒、第二段语音的终点差距则达到了 290 毫秒

针对上述问题，发行人自主研发了数据同步技术，并在该技术领域取得了 1 项发明专利授权。

该技术主要运用在加工环节，具体应用情况、方式、效果如下：

（1）应用情况、方式

发行人应用数据同步技术对已采集的、存在不同步问题的语音类原料数据进行加工，执行自动对齐的加工处理。具体而言，该技术通过获取不同录音设备采集的同一语音内容对应的多个语音数据，并从任一语音数据选取任一语音片段作为样本，通过确定所选语音样本的帧数提取其语音特征参数，在各其他语音数据中确定与上述样本相似度最高的目标语音片段，进而实现多个语音数据的时间轴对齐处理工作。

（2）应用效果：自动化、程序化实现多通道采集原料数据的事后精准对齐，提升了训练数据生产效率和产品质量

以在发行人生产的训练数据集山东重口音普通话语音数据库（手机）（King-ASR-384-12）的生产过程为例：King-ASR-384-12 是发行人针对山东普通话手机端语音识别应用开发的典型训练数据集。该数据集的原料音频数据共涉及三个通道，分别由三个不同的手机同时采集，每个通道采集约 542 个小时的语音，

对应句子数达 500,866 句。在 King-ASR-384-12 的制作过程中，需要对录制的 50 多万组、每组 3 句话的语音数据进行比对和对齐，工作量巨大，采用纯人工操作难以较快完成。发行人通过运用核心技术数据同步技术项下的音频对齐技术，对上述 50 多万组语音数据进行了并行处理及对齐，高效地完成了前述音频文件的同步需求。

10、核心技术 10：实时采集质控技术

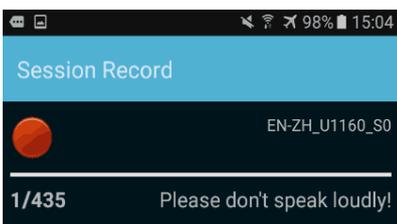
在原料数据采集过程中，受采集设备的物理限制、被采集人的主客观情况或者采集环境变化等因素影响，无法完全保证采集到的每条原料数据都满足后期加工的质量要求。例如，在手机录音采集任务中，当发音人说话声音太响，在音频采集时会出现截幅现象；但说话声音太小时，又会造成所采集到的语音片段难以听清内容，不符合后期加工要求；此外，音频录制时，若周围噪音过大，也会导致音频数据的信噪比过小，不符合质量要求。

针对上述原料数据质量问题，发行人总结多年行业经验自主研发了实时采集质控技术，并在该技术领域取得了 1 项发明专利授权、1 项实用新型专利授权。

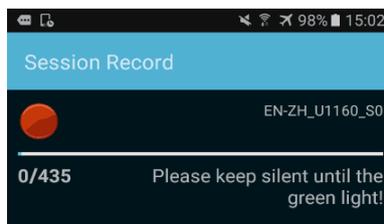
实时采集质控技术主要应用在采集、质控环节，应用情况、方式、效果具体如下：

（1）应用情况、方式

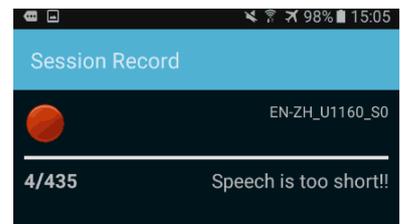
应用该技术，发行人将声音能量检测、音频截幅检测、静音检测、信噪比检测等数字信号处理技术整合到了实际数据采集设备中，在采集的同时进行质量检查，实时发现采集数据的质量问题并及时提供现场反馈（参见下图示例），提高了采集数据的有效性、降低了无效数据占比，提升了最终产品的质量。



图一：实时反馈——音量过大



图二：实时反馈——未保留足够静音



图三：实时反馈——语音段太短

图：原料数据采集过程中实时质检反馈示意

以发行人生产的西班牙语手机对话类语音数据库（King-ASR-321）的生产

过程为例：King-ASR-321 是发行人针对西班牙语自由对话语音识别应用开发的训练数据集。该训练数据集涉及普通环境下的呼叫中心+手机双通道，每个通道采集了约 213 个小时的语音数据，对话内容覆盖 21 个主题。在 King-ASR-321 原料对话数据的采集过程中，应用了发行人实时采集质控技术的录音设备能够实时检测信噪比、语音能量和静音时长等信息，一旦出现录制声音过大或过小、信噪比过低、首尾静音时长等不符合要求的语句，会即时提示要求发音人重录；同时，该设备将依据长音频中切分的各段语音长度和对应文本字数计算发音人的语速分布，后续可执行统计分析、找到异常情况并校正，有效提高了训练数据产品质量。

(2) 应用效果

在采集同时通过实时质检，即时提示被采集对象按要求规范采集，实现前端原料数据的高质量采集，使发行人采集损耗率总体控制在 5% 以内，并为后续生产高质量训练数据奠定基础。

11、核心技术 11：★大数据驱动的高效数据处理技术

发行人在训练数据研发、生产领域深耕多年，积累了深厚的开发经验，并在训练数据的设计、采集和处理环节积累了大量数据，具体包括：发行人自行采集的大量原料数据、发行人拥有自主知识产权的训练数据产品；以及训练数据生产过程中产生的大量日志数据等。

发行人充分利用前述数据及经验积累，通过分析训练数据生产过程中的数据分布规律和行为数据模型设计开发出了高效率的数据处理平台与工具集，形成了大数据驱动的高效数据处理技术。在该技术领域，发行人已取得 4 项发明专利授权。

大数据驱动的高效数据处理技术主要应用在加工环节，具体应用情况、方式、效果如下：

(1) 应用情况、方式

发行人对积累的原料数据及加工后数据等大数据的异同点进行了分析，运用统计模型归纳出一般性规律，将之应用在采集、标注过程所运用的辅助工具与算

法的开发和优化，以及加工任务类型的划分和流程管控的科学设计之中，提升了数据加工的工作效率。例如：发行人基于多年积累的训练数据生产经验开发了词性预测工具，针对不同语种/方言反复调整参数和算法，不断提升预测的准确率，提升数据加工效率、降低人力需求；此外，发行人分析了标注人员的专业领域、属性特征，与其从事数据加工的效率情况和工作质量情况，从而指导自身针对具体数据加工需求更科学地进行数据处理任务分配、提升数据加工效率；发行人也对数据加工过程中的人工行为的模式及原因加以分析，用以指导数据加工流程的优化和工具界面调整设计，最终达到提升数据开发效率的目的。

（2）应用效果

发行人现已积累了 25 个运用于标注过程的预测算法或工具，利用上述工具提升了数据加工过程的效率及准确率，保证了发行人的产能及数据质量。从产能角度，在语音识别、语音合成、计算机视觉、自然语言领域，报告期内每月处理数据量可分别高达 2 万小时语音数据、15 万句语音数据、500 万个目标对象、2 万工作小时；从效率角度，以词性预测算法工具为例，其将词性预测算法的准确率由 86% 提升至 96%³。

在大数据驱动的高效数据处理技术领域，发行人还积累有音素边界自动预测技术、基于语音特征的韵律预测技术等细分技术，前述技术均使得发行人实现了加工效率的提升、有效减少了人工工作量。具体说明如下：

①音素边界自动预测技术：音素是根据语音的自然属性划分出来的最小语音单位，依据音节里的发音动作来分析，一个动作构成一个音素。音素分为元音与辅音两大类。如汉语音节啊（ā）只有一个音素，爱（ài）有两个音素，代（dài）有三个音素等。音素及音素边界的正确性是衡量语音合成训练数据质量的重要指标。发行人通过大量数据训练声学模型自主研发掌握了音素边界自动预测技术，可通过语音识别算法的运用得到音素边界的预测值，标记人员可基于音素边界自动预测技术的标注结果执行人工复核和少量精确人工调整，减少了人工操作的工作量和准确度。

②基于语音特征的韵律预测技术：韵律反映的是语音语句中各字词之间的停

³ 上述数据来自发行人实验室模拟运算分析。

顿时长情况，加工环节中的韵律标注操作即需标出句子中不同字词的具体停顿长度，如以“#1”代表语法词的停顿长度、“#2”代表≤90 毫秒的停顿长度等。由于语音韵律具有较强的差异性和个人特色，因此韵律标注通常由标注人员通过语音听辨人工完成，标注人员根据韵律标注规范，判断每个语法词的韵律边界类型。发行人掌握的基于语音特征的韵律预测技术，可以利用语音识别模型，对音素进行自动切分并预测韵律边界，利用语音识别模型预测结果，修正基于文本得到韵律预测结果；结合标注人员的后续人工听辨、修改，提升了韵律预测与实际音频数据的贴合度，提高了韵律预标的准确率，有效减少了标注过程中约 25% 的人工标注工作量。

12、核心技术 12：★分布式高性能自动校验技术

质检环节是训练数据生产过程的关键步骤，是发行人保证训练数据集质量的重要手段。发行人生产的训练数据集将直接运用于客户算法模型的训练过程中，对训练数据集的质量、准确率均有较高的要求。

以语音合成训练数据的音素信息为例，对音素信息标注的准确率往往需要达到 99% 以上。对大规模数据进行质检并避免重复人工检查、提升最终的训练数据准确率是质检环节面临的挑战。基于传统的信号处理技术已无法很好地满足要求，发行人自主研发了分布式高性能自动校验技术，运用成熟的人工智能算法来辅助完成训练数据质检。在分布式高性能自动校验技术领域，发行人已取得 4 项发明专利授权。

分布式高性能自动校验技术主要应用在质检环节，具体应用情况、方式、效果如下：

(1) 应用情况、方式

发行人在其部署的分布式高性能集群系统上运行自主研发的平台架构、工具及全自动校验技术相关算法，实现质检环节各计算节点间的数据共享、安全传输、任务动态分配及消息队列自动管理等功能，同时支持自动添加、删除和更新节点服务器，弹性地支撑实际业务需求。

在分布式高性能自动校验技术领域，发行人拥有相似说话人自动筛查技术及

音素标注正确率校验技术等细分质检技术，具体应用情况如下：

①相似说话人自动筛查技术：为覆盖不同说话人的声学特性，一个大规模语音识别训练数据集往往包含上千个说话人录制的语音数据。为避免同一说话人录制多份数据影响训练数据的质量，发行人自主研发了相似说话人自动筛查技术。该技术首先建立说话人的语音模型，然后对说话人进行相似度打分，筛选出疑似相同说话人，指导后续人工校验过程。在该技术领域，发行人已取得 1 项发明专利授权。

②音素标注正确率校验技术：音素及音素边界的正确性是衡量语音合成训练数据质量的重要指标。发行人利用基于三音子模型的音素边界自动预测技术，通过训练声学模型，对音素进行识别，基于一定数学分布的假设先行筛选出错误可能性较高的音素，交由人工检查进一步修正。在该技术领域，发行人已取得 1 项发明专利授权。

(2) 应用效果

①分布式高性能自动校验技术使得发行人整体实现了质检、比对、处理等计算任务的高效运行，可在约 24 小时内完成约 660G 的数据质检工作。

②细分技术相似说话人自动筛查技术将相似说话人筛查等质检校验的人工工作量减少了 98%以上，同时提高了工作的准确度；音素标注正确率校验技术则同样通过程序先行、人工检查修正的方式提升了音素标注质量和效率。

综上，本所律师认为，发行人在基础研究、平台工具、训练数据生产三个层次积累了 12 项核心技术，广泛运用于训练数据生产的各个环节之中。

(二) 结合主营业务及同行业可比公司说明训练数据加工及处理、提供具体服务内容及方式等方面是否存在技术优势与核心竞争力

1、同行业可比公司对比

发行人同行业可比公司、竞争对手主要为 Appen Limited（以下简称“Appen”）、北京慧听科技有限公司（以下简称“慧听科技”）及标贝（北京）科技有限公司（以下简称“标贝科技”）。结合前述竞争对手的公开披露信息及主营业务，对训练数据加工处理、服务内容方式等各方面对比如下：

(1) 主营业务情况

主营业务情况				
项目	海天瑞声	Appen	慧听科技	标贝科技
主营业务范围概述	训练数据+训练数据相关的应用服务	训练数据	训练数据	训练数据+ AI算法模型解决方案和技术产品
主营业务	训练数据的研发设计、生产及销售业务。	主要面向机器学习、人工智能开发提供高质量人工标注数据集。	业务包括语音识别、语音合成、语音评测、语言文本类、多媒体类的多领域数据制作，以及语音合成、语音识别、输入法系统研发等；涵盖语音训练数据、音乐数据、语音质量评测等。	主要提供语音交互相关服务，包括语音合成整体解决方案，以及语音合成、语音识别、图像识别等数据服务。

在主营业务方面，发行人与其主要的同行业可比公司、竞争对手主营业务均覆盖了训练数据生产销售、研发相关业务。相比同行业可比公司，发行人业务还包含训练数据相关的应用服务，除提供训练数据外还可基于训练数据提供算法模型相关的训练服务，助力下游客户完成算法模型的语言拓展、算法模块拓展、垂直应用领域拓展，也可为客户定制针对特定场景应用的专属算法模型。同行业可比公司则在训练数据以外也提供AI算法模型解决方案和技术产品，涉足下游AI算法技术领域。

以下围绕训练数据加工处理、提供具体产品服务的内容及方式展开具体对比。

(2) 训练数据加工处理技术指标对比

训练数据加工处理情况及技术指标				
项目	海天瑞声	Appen	慧听科技	标贝科技
技术实力情况	海天瑞声自主开发了一体化数据处理支撑平台，在基础研究、平台工具、训练数据生产三个维度下均积累核心技术，将多项具体核心技术整合为发行人特有的核心技术体系。	Appen 拥有人工智能辅助数据注释平台，在全球 130 多个国家与 100 多万名专业承包商合作，训练数据涵盖科技、汽车、金融服务、零售、医疗健康和政府等	采用全程质量控制流程，执行完善的标注流程，配合保密管理手段，提供质量上乘的数据服务。	拥有语音合成模型和算法，通过算法+专业的人工数据处理方式，为客户提供优质的语音合成服务。拥有 TOBI 标注体系，通过自主研发的 TTS

		各个领域。		评测系统，为客户提供高质量的数据服务。
数据处理能力/效率指标	①语音识别领域：报告期内平均每月采集近 4,000 人语音数据，每月最高可处理 2 万小时语音数据； ②语音合成领域：报告期内平均每月采集近 150 万字/词语音数据，每月最高可处理 15 万句语音数据； ③计算机视觉领域：每月最高可在图片数据中标注超过 500 万个目标对象； ④自然语言：每月标注自然语言可达 2 万工作小时	未公开披露	未公开披露	未公开披露
行业机构评价	“北京市级企业科技研究开发机构”、“新一代人工智能产业创新重点任务入围揭榜单位”、“2019 中国信息技术人工智能行业优秀产品”、“2019 中国信息技术人工智能数据服务领军企业”、2020 年国家专精特新“小巨人”公示名单入选企业等	2017-2019 年德勤亚太区“高科技高增长 500TM”、2012-2019 德勤澳洲“高科技高增长 50TM”、2019 美国常识顾问(CSA Research)第八大语言服务供应商	“慧听中文重口音语音数据库”曾获中国语音产业联盟“2015 中国语音创新产品”	中国企业联合会“2019 全球人工智能 TOP 50”、融资中国“新经济领域最具成长性企业 TOP30”、中国语音产业联盟会员单位
语种/方言覆盖能力	150 余个	180 余个	19 个	9 个
应用领域覆盖能力	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言、音乐	智能语音、计算机视觉、自然语言、音乐
专利储备（已取得专利授权）	24 项（22 项发明专利、1 项实用新型专利及 1 项外观设计专利）	2 项	无	1 项
计算机	133 项	未公开披露	13 项	25 项

软件著作权数量				
成品训练数据集储备	699 个 (截至 2020 年 9 月 30 日)	276 个	23 个	94 个

数据来源及说明:

①Appen、慧听科技、标贝科技数据:截至 2020 年 12 月,前述公司官网及公开披露信息;国家知识产权局中国及多国专利审查信息查询平台(<http://cpquery.sipo.gov.cn/>)、中国版权保护中心 CPCC 微平台等公开信息查询渠道及第三方机构查询信息。

②发行人数据:除特别标注外,均为截至本补充法律意见书签署日数据。

结合上述对比可以看出,相比同行业上市公司,发行人的技术优势主要体现在:经过多年深耕,以长期的业务实践和创新积累为基础,整合了多项核心技术,构建起围绕基础研究、平台工具和训练数据生产三个层次的完整核心技术体系。与同行业竞争对手相比,发行人的语种/方言覆盖能力处于领先水平,积累的自主知识产权的训练数据产品数量和规模远大于境内竞争对手,具备较强的技术优势、核心竞争力。发行人在专利储备方面也具备显著优势,截至本补充法律意见书签署日,发行人已取得 24 项专利授权(包含 22 项发明专利授权、1 项实用新型专利授权及 1 项外观设计专利授权),显著优于同行业竞争对手,充分反映了发行人的技术竞争优势和核心竞争力。

(3) 提供具体产品服务内容和方式对比

产品/服务提供方式				
项目	海天瑞声	Appen	慧听科技	标贝科技
训练数据相关的产品服务形式	训练数据定制服务(采集加工、纯加工)、训练数据产品	定制开发+训练数据产品	定制开发+训练数据产品	定制开发+训练数据产品

在训练数据业务领域,发行人与同行业可比竞争对手提供产品服务的方式基本相同,均可提供训练数据的定制化开发服务,也可提供成品训练数据集。

基于公开渠道可查询信息,以下主要基于发行人和同行业可比公司储备的成品训练数据集(即训练数据产品)情况进行产品服务具体内容对比:

①训练数据产品整体情况对比

发行人与同行业可比公司的训练数据产品情况整体如下：

项目	海天瑞声	Appen	慧听科技	标贝科技
智能语音类训练数据产品				
数量	534	119	19	83
覆盖语种/方言数量	100	82	11	10
录音场景类别	9	15	4	2
采集设备种类	8	2	5	2
采集通道覆盖数	6	5	7	1
自然语言类训练数据产品				
数量	95	152	2	11
覆盖语种/方言数量	43	100	2	5
计算机视觉训练数据产品				
数量	70	5	2	0

②各应用领域训练数据产品具体对比⁴

A. 智能语音训练数据产品

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

选取较有代表性的几类智能语音训练数据产品，与同行业主要竞争对手对比如下：

a.通用语音识别训练数据产品

训练数据集内容结构属性	海天瑞声	Appen	慧听科技	标贝科技

⁴ 注：本部分对比中，Appen、慧听科技、标贝科技的训练数据产品数量等相关信息来源为其官方网站或公开信息披露所列举的信息，数据截至2020年12月；发行人的训练数据产品信息为截至2020年9月30日之情况。

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

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

b. 车载语音识别训练数据产品

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

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

理。

c.特色语音识别训练数据产品

训练数据集 内容结构属性	海天瑞声	Appen	慧听科技	标贝科技
稀有语言覆盖数量	26	38	0	0

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

与同行业主要竞争对手公开披露、可在训练数据产品结构维度加以比较的智能语音类训练数据产品相对比，发行人的训练数据产品的优势和竞争力主要体现在：

覆盖的噪声类型、录音通道数量、录音文本内容类型更为丰富。以通用语音识别训练数据产品、车载语音识别训练数据产品、特色语音识别训练数据产品这几类典型的智能语音类训练数据产品为例，发行人的训练数据产品整体而言可覆盖更丰富的录音文本内容类型、噪音环境类型、录音通道数量，在内容上相比境内竞争对手可覆盖更多稀有语言，具备广泛性、强适用性，具备竞争优势。

B. 计算机视觉训练数据产品

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

训练数据产品 内容结构 属性	海天瑞声	Appen	慧听科技	标贝科技
人像识别检测相关	包含了不同人脸数据，覆盖黑、白、黄、棕等不同肤色。覆盖闸机口安检库、手机拍照库、人物表情库、等应用场景。	包含数千名中国人、黑人在不同角度和光线下的人脸图片；以及 1,000 名中国人身体动作视频等	包含 2 万张人物轮廓图片，覆盖不同年龄段男女老少、各类着装打扮，各类行动姿势（站立、行走、蹲坐等）	未公开披露

OCR (光学字符识别)	覆盖 10 余个语种，覆盖菜单、路牌、收据、书籍等 17 类不同场景	包含中文、泰语、芬兰语 3 个语种，覆盖书籍、报刊、门票、路牌、菜单等 16 个不同场景	未公开披露	未公开披露
-----------------	------------------------------------	--	-------	-------

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

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

C. 自然语言训练数据产品

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

③产品服务内容对比小结

结合上述对比情况可以看出，在产品服务内容方面，与竞争对手在公开信息渠道列示披露的训练数据产品相比，发行人的核心竞争力和优势主要体现在下述方面：

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

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

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

文本、环境覆盖范围更广：整体而言，发行人的智能语音识别训练数据产品覆盖的噪声类型、录音通道数量、录音文本内容类型更为丰富。以通用语音识别训练数据产品、车载语音识别训练数据产品、特色语音识别训练数据产品这几类典型的智能语音类训练数据产品为例，发行人的训练数据产品整体而言可覆盖更丰富的录音文本内容类型、噪音环境类型、录音通道数量，在内容上相比境内竞争对手可覆盖更多稀有语言，具备广泛性、强适用性，具备核心竞争力。

整体而言，与同行业可比公司相比，发行人可提供覆盖多语种/方言、多场景、多领域的，采集方案更为复杂的，文本、环境覆盖更广的训练数据产品，在产品服务内容方面具备全面性、专业性优势。发行人在产品服务内容方面具备核心竞争力。

2、技术优势与核心竞争力小结

结合上述同行业可比公司对比情况可以得出：

在训练数据加工处理及相关技术方面：发行人构建了围绕基础研究、平台工

具和训练数据生产3个层次的完整核心技术体系，具备较强的技术优势和核心竞争力，体现在语种方言覆盖能力、自主知识产权训练数据产品资源积累、专利技术储备等方面。

在具体产品服务内容和方式方面：发行人的产品服务提供形式与同行业可比公司基本相同；但着眼到各应用领域的具体产品服务内容方面，则发行人整体呈现出训练数据产品数量更多、产品覆盖语种/方言更多、采集方案复杂度更高、文本及环境覆盖更广的特点，在产品服务内容方面具备广泛性、全面性、专业性特征，具备核心竞争力。

综上，本所律师认为，发行人在训练数据加工及处理、提供的具体服务内容和方式等方面存在技术优势和核心竞争力。

（三）说明发行人在人工智能细分领域提供基础数据服务所产生的增值服务内容

发行人在人工智能细分领域提供基础数据服务所产生的增值服务内容主要体现在以下两个方面：

1、实现从原料数据到训练数据的生产过程，助力人工智能算法模型发展

（1）训练数据是人工智能算法发展和演进的“燃料”，从自然数据源简单收集取得的原料数据不能直接用于算法训练

①算法、算力、数据是人工技术发展的三大要素，其中训练数据是算法发展和演进的“燃料”

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。AI技术企业需要使用训练数据对算法模型进行训练、优化，才能使得算法模型符合实际场景中的应用需求。在当前技术发展进程中，深度学习算法是推动人工智能技术取得突破性发展的关键技术理论，而大量训练数据的训练支撑则是深度学习算法实现的基础。深度学习分为“训练”和“推断”两个环节：训练需要海量数据输入，训练出一个复杂的深度神经网络模型；推断指利用训练好的模型，去“推断”现实场景中的待判断数据，并得出各种结论。训练数据越多、越完整、质量越高，模型推断的结论越可靠。

②原料数据需经专业化的采集、加工处理，才能形成工程化数据集供算法模型训练使用

通常，从自然数据源简单收集取得的原料数据并不能直接用于深度学习算法的训练，必须经过专业化的采集、加工处理，形成相应的工程化数据集后才能供深度学习算法等算法、模型训练使用。

以发行人的成品训练数据集为例说明：一个成品训练数据集内部包含数据文档、说明文档和技术文档，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。前述语音录音数据文件需要经过截取等加工操作，还需要结合语音录音数据情况进行标注。通俗而言，标注文件将告知机器和算法与语音录音相关的多种信息，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，仅从发音人等数据源简单录制取得的语音录音数据并不能直接用于算法训练。此外，训练数据集结构设计的合理性、科学性也将影响最终的算法模型训练效果。

(2) 发行人通过其专业化业务流程，实现从原料数据到训练数据的生产过程

发行人通过设计数据集结构、组织数据采集、对取得的原料数据进行加工，最终形成可供AI算法模型训练使用的专业数据集。

为了助力下游AI技术企业使用训练数据更好地训练、教导算法模型，实现更好的训练效果，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，使得最终形成的训练数据集具备更合理、全面的分布结构。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。

发行人的主要业务环节包括设计、采集、加工、质检，业务过程涉及下述几方面工作：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需要根据算法模型的训练需要，搭建/选择合适的采集环

境，选取合适的采集设备，采集大量符合要求的原料数据，比如声音、文本、图像、视频等；③需要对采集到的原料数据进行加工，把想让算法学会的信息，通过标注的方式提供给算法；④而在采集、加工的过程中，发行人则要对数据质量进行把控，伴随着持续进行的质检。

因此，发行人的业务实质是：通过执行设计、采集、加工、质检的工序，实现从原料数据到训练数据的生产过程，最终形成可供下游AI产业链各机构训练其算法模型所使用的训练数据集。

发行人在人工智能细分领域提供基础数据服务所产生的增值服务内容包括：①研发设计合理的训练数据集结构；②采集形成符合要求的原料数据包；③通过加工、质检等操作最终形成可供AI算法模型训练使用的成品训练数据集，助力下游AI产业算法模型及技术的发展优化。

2、运用核心技术开展训练数据生产工作，运用科技力量实现赋能增值

经过多年积累，发行人具备对主流人工智能核心算法的理解能力、前瞻性的专业数据集设计能力、丰富的语言覆盖能力及场景采集能力，以及AI辅助标注和精细标注能力。发行人充分发挥前述创新实力，将核心技术积累运用于训练数据生产过程，运用科技力量为生产过程赋能增值，显著区别于通过简单组织大量劳务人员进行原料数据采集、标注的企业。发行人的技术实力主要体现在下述3个方面：

（1）算法与数据处理技术并用：公司积累了12项核心技术，其中在多语种的语音语言学基础研究和高效数据处理两大领域积累了5项关键核心技术。通过算法与处理技术并用，可以满足算法在不同应用场景下的训练数据需求，可对大规模数据进行高效的加工或质检。截至2020年9月30日，公司积累有接近700个自有知识产权的训练数据产品，可提供超过150余个语种/方言的训练数据。

（2）工具和平台共建：公司自主开发了一体化数据处理平台，融入项目管理、质量把控、数据安全管控，嵌入生产所需的各类工具、软件模块，可实现高效、高质的训练数据生产，提升生产效率及质量控制水平。

（3）在语音语言学基础研究方面有深厚积累：公司建立了成熟的发音词典

构建流程、积累了深厚的语音语言学基础研究成果。截至本补充法律意见书签署日，公司的产品/服务已覆盖150余个语种/方言，公司已积累下超过100个语种/方言的发音词典，累计词条数超过1,000万条，可构建高质量智能语音训练数据。

截至目前，公司已经取得22项发明专利授权、1项实用新型专利授权及1项外观设计专利授权，133项计算机软件著作权，另有4项专利已提交申请。

多年积累的核心技术和专业服务能力，使得公司能够更大规模、更有效率、更加精准地生产训练数据，在提升自身产出效率的同时也有效提高了训练数据对于客户算法模型的改善、优化效果。

综上，本所律师认为，发行人在人工智能细分领域提供基础数据服务所产生的增值服务内容主要体现在：实现从原料数据到训练数据的生产过程，助力人工智能算法模型发展；运用核心技术开展训练数据生产工作，运用科技力量实现赋能增值。

（四）结合上述分析说明发行人所处行业及细分领域是否符合科创板定位，科创属性行业分类是否准确

结合上述分析，发行人主要从事训练数据的研发设计、生产及销售业务，生产过程中需要设计训练数据集结构、采集原料数据、对原料数据进行加工处理和质检，最终形成可供AI算法模型训练使用的训练数据集，与数据加工处理息息相关。

结合国家统计局《战略性新兴产业分类（2018）》，发行人所从事的训练数据生产业务属于“新一代信息技术产业—新兴软件和新型信息技术服务—新型信息技术服务—信息处理和存储支持服务—数据加工处理服务”行业。

因此，发行人属于《科创板企业发行上市申报及推荐暂行规定》中规定的行业领域“新一代信息技术领域”中的战略性新兴产业，符合科创板行业领域要求；此外，发行人连续荣获“国家高新技术企业”及“中关村高新技术企业”称号，并于2018年被评为国家规划布局内重点软件企业，2019年入选工业和信息化部“新一代人工智能产业创新重点任务入围揭榜单位”，2020年入选工业和信息化部

部国家专精特新“小巨人”公示名单，科创属性、技术实力得到行业、主管机构的高度认可。

综上，本所律师认为，发行人所处行业及细分领域属于“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”，符合科创板定位，科创属性行业分类准确。

（五）核查程序

本所律师执行了下述核查程序：

1、访谈发行人主要管理人员，查阅发行人专利证书并通过国家知识产权局中国及多国专利审查信息查询系统查询专利情况，了解核心技术的形成、储备和使用情况；

2、登录发行人信息系统，查看发行人自主研发的一体化数据处理平台及其集成的技术、工具情况、在生产过程中的使用情况；

3、与发行人主要业务人员了解发行人的整体业务流程、产品服务情况；通过公开渠道查询了同行业可比公司的核心技术储备及应用情况、产品服务情况、知识产权情况等；结合行业报告、行业专家访谈，了解发行人的核心技术优势、核心竞争力情况，以及发行人在人工智能基础数据服务领域产生的增值服务内容情况；

4、查阅国家统计局《战略性新兴产业分类（2018）》，结合发行人主营业务情况，核对其所处行业、细分领域，核查其与《科创板企业发行上市申报及推荐暂行规定》中行业领域的匹配情况；查阅发行人获得的证书奖项。

（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首次公开发行股票并在科创板上市的补充法律意见书（六）》的签署页)

北京市天元律师事务所（盖章）



负责人：_____

朱小辉

经办律师：_____

周世君

王韶华

顾鼎鼎

2021年3月2日

本所地址：北京市西城区丰盛胡同28号

太平洋保险大厦10层，邮编：100032