



通讯地址：北京市东城区永定门西滨河路8号院7号楼中海地产广场西塔9层
Postal Address: 9/F, West Tower of China Overseas Property Plaza, Building 7, NO.8, Yongdingmen Xibinhe Road, Dongcheng District, Beijing
邮政编码 (Post Code) : 100077
电话 (Tel) : +86(10)88095588 传真 (Fax) : +86(10)88091199

《关于北京海天瑞声科技股份有限公司首次公开发行股票 并在科创板上市申请文件的第四轮审核问询函》之专项核查意见

瑞华专函字【2019】02280017号

上海证券交易所：

根据贵所于2019年6月25日出具的《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申请文件的第四轮审核问询函》上证科审（审核）【2019】316号（以下简称“《第四轮审核问询函》”）相关问题的要求，瑞华会计师事务所（特殊普通合伙）（以下简称“本所”或“申报会计师”）对《第四轮审核问询函》就本所涉及到的有关问题，履行了审慎核查义务。现对《第四轮审核问询函》回复如下：

问题3、关于采购业务

根据专项核查报告，2016年及2017年发行人未保留业务数据，2018年发行人仅保留了部分业务数据，中介机构对采购框架合同、采购订单资料、经双方盖章确认的验收结算单、收到的发票、支付采购结算款项的银行回单进行了核查并对劳务公司进行了函证。

请发行人补充披露各类业务经营过程中采购和销售业务的定价方式和过程。

请发行人说明：（1）结合与客户的保密条款约定，说明发行人未保留业务数据的原因及合理性，发行人是否保存了与经营业务相关的日志数据，如何保证生产经营过程可追溯，进一步说明发行人与财务报表相关的内部控制是否健全并有效执行；（2）发行人对劳务公司提供服务的人数和工作量如何进行复核和确认，定量分析实际服务人数和工作量与结算数据的对比情况、结算调整及差异原

因，验收结算确认单中相关结算信息的依据和来源，如何保证实际采购业务与财务信息一致。

请保荐机构和申报会计师对上述事项进行核查并发表明确意见，说明：（1）在没有保存业务数据的情况下如何对发行人数据服务费支出进行核查，如何保证核查结论的合理和准确；（2）对劳务公司提供服务的真实性进行核查的过程、程序和范围，包括但不限于核查个人的身份登记记录、个人提供服务后确认的签字凭证、劳务公司向个人付款的相关凭证和完税凭据等。

答复：

第一部分：发行人补充披露

报告期内，发行人主要的业务类型为数据资源定制服务和数据库产品，其销售和采购业务的定价方式和过程已在招股说明书“第六节 业务和技术/（三）主要经营模式/3、采购模式 及4、销售模式”补充披露：

1、销售业务的定价方式和过程

（1）数据资源定制服务

定价方式：销售价格以定制项目的预估成本为基础，结合项目技术难度、复杂程度、时限要求等与客户协商确定。例如，采集环境和场景、采集对象年龄结构、语种等因素，都会在定价中予以考虑。

定价过程：确定需求→评估成本→评估技术难度、复杂程度、时限要求等因素→确定报价→与客户协商价格→确定价格、签署合同。

（2）数据库产品

定价方式：综合考虑数据库产品的开发支出、市场需求程度、未来估计重复销售频率等因素，向客户报价，并与客户协商确定销售价格。数据库产品通常以单个数据库为单位进行定价，定价比较灵活。

定价过程：确定需求→参考产品开发成本→评估市场需求程度、未来估计重复销售频率等等因素→确定价格区间→与客户协商价格→确定价格、签署合同。

2、采购业务的定价方式和过程

数据资源定制服务、数据库产品的数据服务采购模式无差异。按照采集、标注两个环节，采购业务的定价方式和过程如下：

(1) 采集

定价方式：采集环节采购的是音频、图像、文本等未经处理的生数据的采集服务。发行人向供应商采购采集服务的采购总价=结算数据量×采集单价。采集单价根据市场上类似服务的一般价格，并参考采集对象资源的稀缺性来确定。例如，外语种、儿童或老人、行车环境等要求较为特殊的采集项目，采集资源有一定稀缺性，则采集单价高于一般项目。结算数据量根据生数据采集服务的来源和内容，单位也有所不同，常见单位包括人、字/词、张等。

定价过程：确定采集方案→向供应商发出采集需求→综合考虑市场价格和该供应商的过往采购价格，与供应商协商价格→确定采集价格。

(2) 标注

定价方式：标注环节采购的是给生数据做人工标记的服务。发行人向供应商采购标记服务的采购总价=有效工时×标记单价。标记单价根据市场上类似服务价格，并参考标记任务的难度来确定。有效工时根据发行人对标产比的测速结果，与标记人员达成一致后确定。有效工时=交付数据量×标产比。

定价过程：确定标注方案→向供应商发出标记需求→综合考虑市场价格和该供应商的过往采购价格，与供应商协商价格；确定项目标产比→确定标记价格。

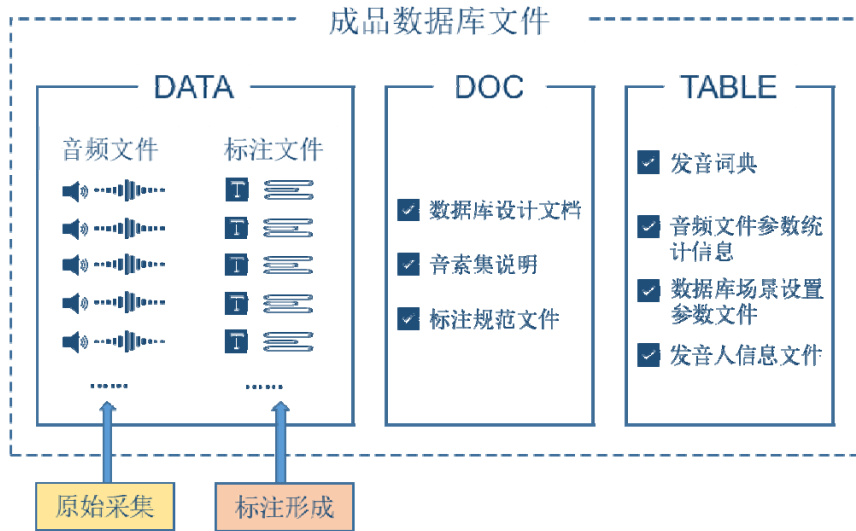
第二部分：发行人对采购模式的进一步说明

一、发行人数据库的特点

无论是数据资源定制服务还是数据库产品，发行人最终向客户交付的均为数据库文件。发行人完整的生产环节包括：数据库设计→采集→标注→质检，从而形成最终可交付的成品数据库。

以智能语音数据库为例，发行人成品数据库由三部分组成：一是“DATA”

文件，为采集的原始音频数据文件和对应的标注文件；二是“DOC”文件，为数据库规范文件；三是“TABLE”文件，为数据库特征信息文件。数据库的具体结构如下图所示：



由上图可以看出，发行人的数据库具有如下特点：

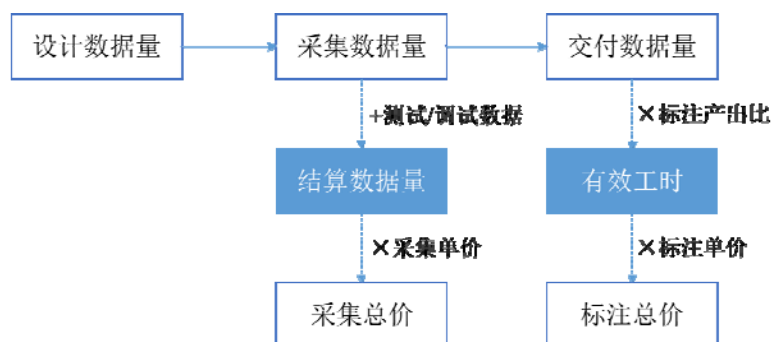
- 1、数据库结构简单透明，由原始采集形成的音频文件、与音频文件对应的带有时间戳的标注文件，在数据库中均独立存储；
- 2、交付数据量决定采购数据量，可从成品数据库向原始采集数据追溯；
- 3、包含了项目绝大部分业务数据信息，包括数据库设计、采集、标注等各个环节。

二、发行人数据服务采购量的确定机制

发行人的数据服务采购涉及采集、标注两个业务环节。其中采集环节采购的是音频、图像、文本等未经处理的原始数据；标注环节采购的是给生数据做人工标记的服务。对于数据采集，发行人以结算数据量为单位与供应商结算；对于标记服务，发行人以有效工时为单位与供应商结算。

发行人数据资源定制服务和数据库产品¹采购各环节的数据量和有效工时之间的关系，以及与数据采购支出的关系如下图所示：

¹发行人主要产品和服务分为数据资源定制服务、数据库产品、数据资源相关的应用服务三类，但数据资源相关的应用服务规模较小，报告期各年收入、毛利占比均不超过 2%，对主营业务不构成重要影响。



①采集数据量：指发行人通过采集环节获取的、可进入标注环节的合格数据的数量。

②交付数据量：指经过标注、质检环节，可作为成品交付客户的数据的数量。

③结算数据量：指发行人与供应商确认并进行结算的数据的数量。

④标注产出比（标产比）：指标注环节，每产出1小时的交付数据量，需要花费的标记工作的时间。

⑤有效工时：指标注环节，标记人员的有效工作时长，有效工时=交付数据量×标产比。

三、发行人采购模式的特点

发行人成品数据库的以上特点，使得可以通过数据处理进度即可对采集和标记服务的采购进行有效的监控与核算。

（一）采集

成品数据库中的音频文件即为原始采集的合格的音频文件；后续数据处理并非编辑音频文件，而是向数据集中增加标注文本、数据规范文件、数据库特征信息文件，故理论上原始采集音频文件的数量与成品数据库文件中的音频文件数量是相同的。实际生产过程中，总采集数据量通常会留出少量冗余，从而略大于最终要交付的数据量，以备替换偶然出现的不合格录音数据；而发行人通过技术手段、商务约定和严格的项目管理，将冗余量控制在较低范围内。例如，某“150人中文纯净人声录音棚录音项目”，客户订单要求录音人数为150人，实际采集人数为152人。

因此，数据库需要交付多少数据量，采集多少数据量并留出少量冗余即可，不存在额外采集更多数据量的经济必要性。从上述业务逻辑出发，发行人通过采集数据量与交付给客户并验收的数据量、与供应商结算数据量之间的一致性，来

保证采购采集服务的完整性。

（二）标注

原始音频文件采集完毕后，进入标注环节。该环节采购的服务为标记服务，其工作成果体现为标注文件。采集形成的音频文件可以对应到发音人的人数以及录音时长，但标注文件本身无法直接反映出标记的工作量，因此需以某指标为基准进行标准工作量的转换。实际经营中，发行人以交付数据库的数据量为基础，以标产比为转换单位，两者相乘得出标记工作的有效工时。例如，前述“150人中文纯净人声录音棚录音项目”的标记工作，经试标注测速后，标产比确定为2.7（即标注完成1小时音频数据，需要花费2.7个小时的工作量）；假设某标记人员完成了10小时的音频数据标记，则其有效工时核定为 $10 \times 2.7 = 27$ 小时。

四、发行人数据服务采购量控制和核算的具体模式

以数据资源定制服务为例，以下对发行人数据服务采购量的控制和核算的具体模式进行说明。数据库产品与数据资源定制服务相比，对外采购所涉及的采集和标注环节的业务模式无差异。

（一）发行人采购生数据采集服务的具体模式

1、客户合同/订单明确约定交付数据量，发行人根据约定交付数据量采购采集和标记服务

发行人与客户签订的业务合同/订单中会对交付数量进行明确的约定。例如有采集的单一化标注语音业务合同中约定发音人人数和总时长，多样化标注语音合同中约定字/词数或句数。发行人向客户交付数据时，需要满足约定的交付数量并由客户验收通过。

为确保数据质量，发行人采集环节的采集数据量可能略多于交付数据量，以便在剔除不符合要求的数据后仍满足客户交付数量的要求；同时由于发行人启动采集项目前可能进行少量的采集测试、调试采集设备和软件等原因，可能造成结算数据量略大于采集数据量。发行人最终与供应商的结算数据量与向客户的交付数据量之间的差额称为采集损耗，其占总供应商结算数据量的比例称为采集损耗率。发行人通过技术手段、商务约定和严格的项目管理，将采集损耗率控制在较低水平，所需冗余采集量小，不存在额外补充采集较大规模数据的情形。

2、发行人采集损耗率较低，较小的冗余采集量即可满足业务需要

在上述生产模式决定下，当一个项目的交付数据量确定时，发行人的采集损耗率是反映采集数据量的合理性和完整性的主要指标。

(1) 发行人的采集模式自身决定了较低的损耗率

①对于有采集的单一化标注语音，只需被采集人按照给定文本或预设场景，根据既定的录音流程进行录音即可，难度较低，原始采集的音频文件的初始合格率本身即比较高，因此不存在原始采集数据大比例不合格的情况。

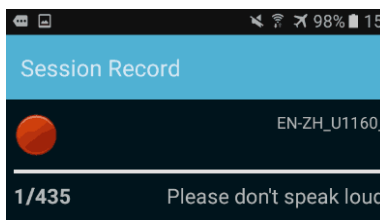
②对于多样化标注语音，由于设计阶段的语料设计和采集阶段的发音人录音要求均较高，且发行人与供应商在合同中对发音人录音事项明确了质量要求，只有录音合格后才会与供应商结算发音人费用，因此后续环节基本不存在录音不合格但仍需与供应商结算该不合格部分的发音人费用的情况。

(2) 发行人控制损耗率的措施

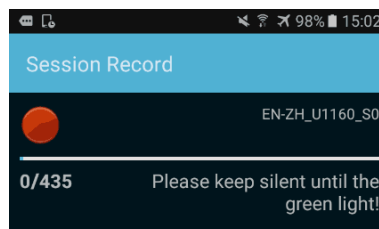
发行人以不超过5%的内部标准，对采集损耗率进行严格的控制。具体措施如下：

①发行人通过技术手段降低采集损耗率

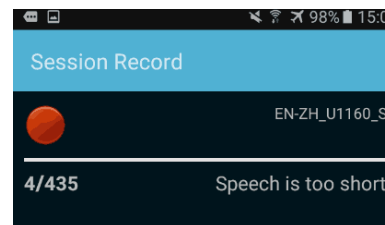
发行人采用了技术手段严格控制采集阶段的不合格数据总量。以单一化标注语音项目的采集为例，发行人在自行开发的各录音软件中均内嵌了质量控制模块，可以通过对语音信号的一些特征（如音量、背景噪声水平等）进行判断，在声音录制的过程中给出反馈，例如音量过大或者过小等，从而可以实时提示发音人对不合格的声音进行重新录制，合格后才能进入下一条语音的录制。软件界面举例如下图所示：



图一：音量过大



图二：未保留足够静音



图三：语音段太短

因此，发行人在采集环节所取得的声音文件在语音信号质量方面已基本满足

后续处理要求，从而能够有效降低后续处理环节中初始采集数量的损耗比例。

②发行人通过商务约定和严格的项目管理来降低采集损耗率

同时，发行人也通过商务约定和严格的项目管理对数据质量进行控制。例如，对于有采集的单一化标注项目，在采集阶段开始前，发行人向供应商提出明确的被采集人特征要求，例如发音人的口音地区和轻重、年龄、性别要求等，在采集环节会对数据进行检查。

③小结

通过以上两种方式，发行人在采集环节中对于所取得数据的数量和质量均有严格把控，确保采集数据量一般仅略高于最终产成品的交付数据量，后续环节的损耗率较低，不存对较多冗余数据量的需求。

（二）发行人采购标记服务的具体情况

对于标注环节，决定标注有效时长的主要因素为标产比。发行人自测标产比来核定标记人员的有效工时，并按该有效工时与供应商进行结算。

1、发行人按标产比折算有效工时，并以其作为采购单位依据的合理性

（1）采用标产比的方式，可以避免标记人员工时统计不准确、工作时长虚增，导致支出核算不准确的问题。此外，采取标产比的方法，有利于督促供应商加强标记人员的管理、提高标记效率。

（2）发行人单位产出的数据服务费支出为量和价相乘的结果。数据服务费的差异主要体现为所消耗有效工时数量的差异（如生数据类型不同、标注环节多少不同），采取标产比的方法即体现了上述差异，因此可以更为准确地反映项目特征与业务实质。

2、发行人确定标产比的方式

发行人通过自行测量的方式确定标产比，并按照自测标产比折算有效工时，依此与供应商进行结算。具体方法为：

（1）发行人将待标注数据拆分为若干任务后，选取少量任务由标记人员进行试标记，测量在连续工作状态下的标产比。

(2) 确定速率后与标记人员达成一致，按照标记人员的实际产出数据量在自测速率下折算为有效工时，与供应商进行结算。

除上述自测标产比情形外，还存在少量的客户约定标产比的情形。该情形下，发行人与客户书面约定总交付数据量、标注工时每小时单价和标产比。

发行人单个项目以固定的标产比进行折算。但由于发行人业务定制性较强，各项目所需处理的数据资源类型差异较大，因此各项目的标产比存在一定范围的差异。报告期内，发行人主要业务的标产比保持稳定，变动原因合理，不存在异常波动，具体情况请参见第二轮审核问询函回复问题7（8）。

综上所述，发行人采集数据量、有效工时与交付数据量直接相关，以交付量确定采购量。采购数据与成品数据具有一一对应关系，采集损耗率控制在较低水平。标记服务的工作量按照标产比折算为有效工时，标产比在报告期内保持稳定，不存在异常波动。

第三部分：发行人说明

(1) 结合与客户的保密条款约定，说明发行人未保留业务数据的原因及合理性，发行人是否保存了与经营业务相关的日志数据，如何保证生产经营过程可追溯，进一步说明发行人与财务报表相关的内部控制是否健全并有效执行；

一、发行人未保留业务数据的原因及合理性

发行人的业务数据可分为两类：一是发行人生产的数据库本身，包含了采集数据量、交付数据量、标注文本数量以及其他数据库特征信息，其中数据资源定制服务在数据库交付验收后，不再长期保留成品数据库文件，数据库产品业务的成品数据库予以保留；二是生产过程数据，主要以项目采集信息表、项目周报、客户沟通邮件、客户验收报告等线下文档的形式记录，并予以保留。

(一) 发行人不保留定制服务数据库的原因及合理性

对于数据资源定制服务，发行人在数据库交付验收后，不再长期保留成品数据库文件。其原因及合理性分析如下：

1、数据资源定制服务完成后，数据库所有权归属客户，发行人不应保留

发行人数据资源定制服务所形成的数据库所有权在销售时已经转移至客户，公司不再拥有该等数据资源的所有权，如果仍然长时间保留备份，则存在相关信息失密和侵权的风险。因此，发行人对定制服务数据不予保留。

2、重要客户对于服务完成后不得留存数据有明确要求

发行人的客户包括大型科技公司、人工智能企业和科研机构等，高度重视数据安全和保密工作，并对其供应商的数据安全和保密作出严格规定，对于发行人定制服务涉及的数据均有相应的保密要求；部分重要客户通过合同或者专项文件、供应商管理制度等形式对服务完成后的数据留存做出了明确要求，具体如下：

序号	客户名称	相关文件名称	相关条款内容
1	微软	SSPA (Supplier Security and Privacy Assurance, 供应商安全和隐私保证)	<p>“#13 确保 Microsoft 个人数据和机密数据的保留时间不超过执行相关服务所需的时间，法律要求继续保留 Microsoft 个人数据和/或机密数据的情况除外。</p> <p>#14 确保 Microsoft 自行决定 Microsoft 个人数据或机密数据在供应商的处理或控制下归还给 Microsoft，或应 Microsoft 的请求在服务执行后进行销毁。</p> <p>必须在应用程序内制定相关流程以确保从应用程序中安全删除数据（无论是由用户明确删除，还是根据数据期限之类的其他诱因删除）。</p> <p>如果需要销毁 Microsoft 个人数据或机密数据，供应商必须烧毁、粉碎或撕毁包含 Microsoft 个人数据和/或机密数据的实物资产，以便无法读取或重建相关信息。”</p>
2	阿里巴巴集团	阿里巴巴集团外包安全管理和人员行为规范细则	<p>“1.5.1 乙方人员不得将文档（包括但不限于打印件、复印件、各种甲方内部资料等）带出驻场或供应商场地，如涉及纸质文档必须用碎纸机销毁。</p> <p>1.5.2 外包项目结束后，乙方必须将所有交付物（包括但不限于代码、文档、图片等）移交给甲方，不得私自保留拷贝。”</p>
3	百度时代网络技术(北京)有限公司	语音数据标注服务框架协议(第二期)	<p>“4.1 本合同项下的一切数据资源（包含原始数据和标注数据资源，下同）的所有权利（包括但不限于知识产权）归甲方享有。乙方不得对数据资源以及数据资源享有的知识产权进行任何形式的使用，包括但不限于转让、销售、公开或许可第三方使用或用于本合同以外的任何目的。乙方英语数据提交验收合格后对数据的语音文件及文本文件进行删除，不得备份。”</p>
4	腾讯科技(深圳)有限公司	AI 语音素材采购框架协议	<p>“2.所有包含保密信息的或与之有关或相关的材料, 无论是否由披露方所提供, 包括但不限于废弃图纸、文件碎片、照相底片或计算机输入或输出信息, 并且包括任何种类的</p>

序号	客户名称	相关文件名称	相关条款内容
			全部复印件和复制件，乙方应在收到甲方书面要求或相关订单履行完成（以较早者为准）之日起七（7）天内，在合理可行的范围内立即归还或销毁。自此，接收方不得再为可目的使用保密信息。”
5	我国某大型科技公司	EAI 语义引擎项目采购说明书	“接受方应立即返还或销毁所有根据‘本协议’所接受的披露方的‘保密信息’，包括但不限于以任何形式存在的保密信息的原件、复印件、复制品及对“保密信息”的概述摘要，并向披露方提供已经返还或销毁“保密信息”的书面确认。”
6	某国际消费电子产品厂商	CONFIDENTIALITY AGREEMENT (保密协议)	“7. Within ten (10) business days of receipt of Discloser's written request, and at Discloser's option, Recipient will either return to Discloser all tangible Confidential Information, including but not limited to all electronic files, documentation, notes, plans, drawings, and copies thereof, or will provide Discloser with written certification that all such tangible Confidential Information of Discloser has been destroyed.” (在收到披露方的书面请求后十（10）个工作日内，并且根据披露方的选择，接收方应将任何形式的机密信息返回给披露者，包括但不限于所有电子文件、文档、注释、计划、图纸及其副本，或向披露方提供所有任何形式的机密信息已被销毁的书面证明。)

3、发行人数据资源定制服务数据库包含了客户的个性化保密信息

发行人定制化数据库的结构参见本问题回复之“（2）/一、发行人产品与生产模式的特点”。以智能语音数据库为例，定制化数据库中记录的音频文件、标注文件、数据库设计信息、被采集对象信息、音频/图像参数和数据库参数等信息，均属于客户的个性化信息，为客户的保密信息。

综上所述，发行人定制化数据库交付验收后所有权转移给客户，重要客户对于服务完成后不得留存数据有明确要求，而定制化数据库包含了客户的个性化保密信息，因此发行人不留存定制化数据库是合理的。

（二）对于数据库产品，发行人对成品数据库予以保留，并不删除

数据库产品为发行人根据对人工智能算法模型应用领域、行业发展趋势、市场需求等的评估和研判设计开发，开发完成后授权给客户使用，销售完成后，数据库产品的所有权不发生转移，仍归发行人所有。因此，数据库产品在销售后发行人予以保留，并不删除。

二、发行人是否保存了与经营业务相关的日志数据，如何保证生产经营过程可追溯

（一）发行人未在系统中保存与经营相关的日志数据，但不影响发行人对采购业务量的监控和核算

发行人的“一体化人工智能数据处理技术支撑平台”为一系列采集和标注软件及工具的集合，业务数据随着生数据采集和数据处理的过程记录和保存于成品数据库中，未以日志数据的形式存在，未在信息系统中保留经营业务相关的日志数据。

发行人依据采集数据量和有效工时与供应商结算数据服务采购费用，两者均与交付给客户的数据量直接相关，在项目执行过程中，已完成的采集数据量可通过音频文件个数和长度、图片文件个数、文本文件的条数等直观看出，已完成的有效工时可通过已处理的数据量乘以标产比的计算过程得出，故发行人通过数据处理进度即可对采集和标注服务的采购进行有效的监控。例如“云南500人1500小时重口音手机录音项目”，在项目执行过程中的某时点，服务器上已形成300人/900小时的音频文件及其对应标注文件，则发行人按300人/900小时为依据进行采集数据量和有效工时的核算，并不需要依赖于信息系统中记录的日志数据。

因此，发行人尽管未保留与经营业务相关的日志数据，但不影响对采购业务量的监控和核算。

（二）发行人通过线下文档的形式记录和保留生产过程数据，生产过程可追溯

除了线上数据，发行人的生产过程数据还以项目采集信息表、项目周报、客户沟通邮件、客户验收确认文件等线下文档形式记录和保留。

项目采集信息表：为发行人进行音频采集时，对每个发音人信息的具体记录（如性别、年龄、国籍、来自省份或城市、语言或口音等）。

项目周报：为项目执行当中，各业务负责人对项目进展进行定期汇总并向总经理、主管业务的副总经理汇报，相关信息包括项目编号、内容、客户、项目负责人、交付时限、当前状态、已完工比例等。所发送周报均保存于发行人163企

业邮箱系统内，为第三方系统，不存在篡改的可能性。

客户沟通邮件：对于数据资源定制服务，在项目执行过程中，发行人与客户针对文本设计、标产比、项目进度等，进行交流与确认。客户沟通邮件通过163企业邮件系统完成，不存在篡改的可能性。

客户验收确认文件：成品数据库完成质检向客户提交；客户验收通过后，向发行人发送验收确认文件，或在客户系统中完成验收。客户验收确认通过企业邮件系统或客户系统完成，不存在篡改的可能性。

此外，对于数据库产品，开发完成后，纳入发行人的产品列表，包括产品说明文档、产品文件等主要信息全部归档备份管理。

上述文件覆盖了发行人数据资源定制服务和数据库开发的生产过程，保证发行人的生产过程可追溯。

（三）进一步说明发行人与财务报表相关的内部控制是否健全并有效执行

1、发行人与财务报表相关的内部控制健全，并得到有效执行

综上所述，发行人的采集、标注数据服务采购与交付给客户的数据量紧密挂钩。发行人采集损耗率保持较低水平，保证了采集服务的采购结算量可通过对比拟交付数据量得到有效控制；发行人采取以交付数据量为基数，用标产比来折算有效工时的方式来核算标记工作量，而标产比为实际测试得到或由外部客户约定，保证了标记服务的采购结算量可通过对比交付数据量并控制标产比的合理性而得到有效控制。

在上述业务特点的基础上，发行人对数据服务采购的核算建立了如下内部控制流程。发行人与财务报表相关的采购核算内部控制健全，并得到有效执行。

业务管理系统流程

财务系统流程

注：标*为关键控制点

1. 根据合同确认履约订单、一一对应建立项目并录入系统；对于数据库产品，根据产品开发决策立项、项目信息录入系统

2. 履约订单与项目一一对应，确保可按项目对收入、支出进行独立归集和核算；每个项目进行项目预算申请及审核，通过后立项成功

3. 项目负责人监控项目执行，并记录采集数据量、有效工时等关键业务信息

4. 依据项目过程记录的关键信息，将结算数据量、有效工时录入业务管理系统，进行支出报销申请；复核人员复核关键业务信息，确保系统中数据服务采购量的完整、准确

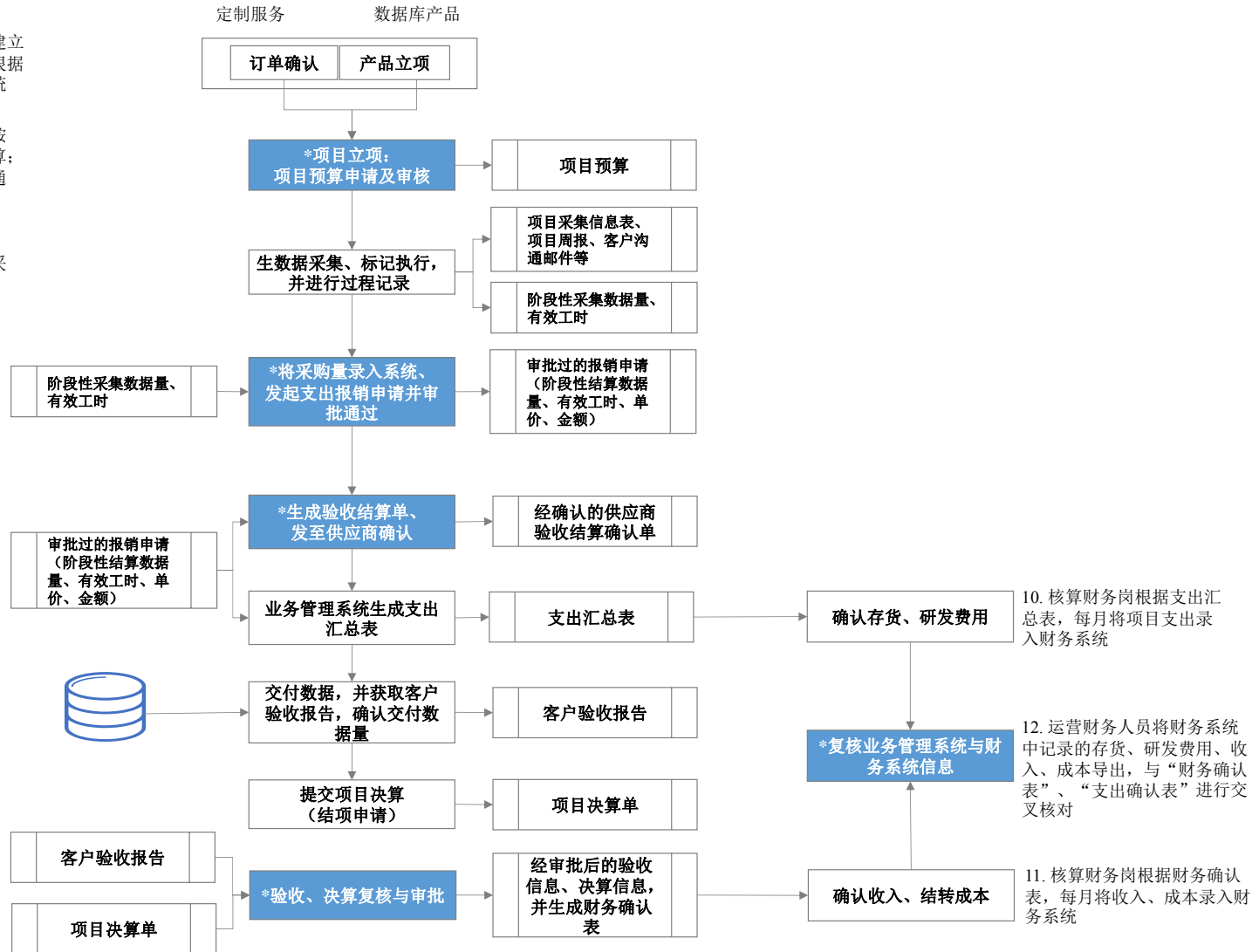
5. 根据支出报销审批信息生成验收结算单，并发送给供应商，由其确认后形成双方结算依据

6. 业务管理系统根据系统记录的经审批的支出报销信息，自动生成支出汇总表，按项目、任务记录每笔支出

7. 各阶段数据开发完毕后，形成数据库并交付给客户，客户确认交付数据量与数据质量无误后，出具验收报告

8. 项目完成且所有支出报销完毕后，申请结项

9. 复核人员复核验收、决算信息，并审批通过



针对上述内部控制流程中的关键控制点所实施的控制措施如下：

(1) 流程2——项目立项：项目预算申请及审核

项目预算申请及审核，是发行人保证成本完整性、准确性的重要控制点，一般项目为三级审批制，重大项目为五级审批制。主要复核环节如下：

①将预计采集数据量与履约订单约定的交付数据量之间进行核对，保证预算采购量的真实、完整、准确；②对比同类项目间的采集产出比、标注产出比，保证预计采集数据量和有效工时的合理、完整；③综合考虑市场价格和该供应商的过往采购价格，对单价进行复核，保证预算中采集、标记单价的合理性；④根据履约订单金额、预算支出金额，测算预算毛利率，复核项目盈利能力的合理性。

(2) 流程4——将采购量录入系统、发起支出报销申请并审批通过

针对生数据采集环节，复核人员针对采集数据量、结算数据量、交付数据量进行交叉比对，如存在重大差异，需由项目负责人员解释原因；如与预算对比出现重大差异，需经业务分管负责人、财务负责人审批：

①设计数据量对比采集数据量，是否总体基本一致；②音频文件个数对比采集数据量，是否一致；③采集数据量对比结算数据量，是否存在异常结算情况；④结算数据量对比预算采购量，是否存在较大偏离；⑤复核实际采集产出比，是否与商务约定或预算方案一致；⑥采集单价对比预算单价。

针对标注环节，复核人员针对标注产出比、有效工时进行合理性复核；如有效工时与预算对比出现重大差异，需经业务分管负责人、财务负责人审批：

①对比采集数据量，复核标注工作覆盖的数据量是否完整；②标记产出比对比预算，是否出现重大偏离；③有效工时的计算是否准确；④标记单价对比预算单价。

(3) 流程5——生成验收结算单、发至供应商确认

①根据支出报销审批信息生成验收结算单，并发送给供应商；②由供应商确认后盖章，形成双方结算依据。

(4) 流程9——验收、决算复核与审批

复核验收、决算信息，保证验收金额的真实、准确，项目支出的准确、完整。复核内容主要包括：

①交付数据量对比验收数据量；②验收量对比履约订单中的销售量；③验收单价对比履约订单中销售单价；④对比预算，复核项目决算中覆盖的环节、各环节的采购数量是否完整；⑤决算支出对比预算支出，出现重大偏离需提交项目决算分析表，解释原因。

（5）流程12——复核业务管理系统与财务系统信息

①项目采购支出、收入、成本等关键信息通过业务管理系统自动汇总并生成支出汇总表、财务确认表；②财务系统经过人工录入后，生成存货、研发费用、收入、成本等信息；③由专门人员将两个系统的记录导出后进行复核比对，如有差异，调查原因，并进行相应修订。

2、发行人进一步完善相关内部控制手段的措施

发行人尽管建立了较为完善的线下内部控制以保证数据服务采购的真实、准确、完整，但未能充分利用信息系统实现全部生产过程信息的记录和留存。为了进一步提升精细化管理水平、完善相关内部控制手段、提升内部控制的有效性，发行人将加强业务信息系统建设，实现对生产过程数据的自动化记录和保存。拟采取如下改进措施：

①进一步完善人员信息管理模块，包括管理人员、公司研发人员、供应商管理人员、被采集人、标记人员等。保证所有人员的个人信息在系统内的完整、真实、可追溯。

②进一步完善日志管理模块，涵盖数据资源开发的全流程，对于与经营业务相关的日志数据，做到集中处理、冗余备份、实时监控。

③新增数据分析模块，包括对业务数据、日志数据、用户行为数据等多维度数据做全面定时定量分析，提供对业务情况、财务情况的交叉验证能力。

④加强支出报销申请与项目决算申请中的复核机制。除目前执行的复核措施外，由项目负责人提交上述系统记载信息（包括但不限于采集数据量、测试标产比等）与实际支出报销信息的匹配说明，由主管部门负责人及运营财务部审批。

⑤实现业务管理系统信息向财务系统自动传输，生成记账凭证，避免信息导入财务系统过程中人为操作失误的风险。

(2) 发行人对劳务公司提供服务的人数和工作量如何进行复核和确认，定量分析实际服务人数和工作量与结算数据的对比情况、结算调整及差异原因，验收结算确认单中相关结算信息的依据和来源，如何保证实际采购业务与财务信息一致。

一、发行人对劳务公司提供服务人数和工作量进行复核和确认的方式

发行人对劳务公司提供服务的人数和工作量进行复核和确认的方式如下：

1、项目负责人员根据已完成的采集数据量/已发生的有效工时，在业务管理系统中填报支出申请；

2、部门经理对支出申请进行审核，交叉比对已完成数据量及其他业务信息；如果支出与预算出现较大偏差，需经业务分管负责人审批；

3、财务部门对支出申请进行审核，交叉比对业务信息；如果支出与预算出现较大偏差，需经财务负责人审批；

4、支出申请审批通过，业务管理系统生成供应商验收结算确认单；验收结算单发送给供应商，供应商确认后盖章并发回。

上述复核和确认的具体过程和方式，请参见本问题回复“第三部分/（2）/（三）进一步说明发行人与财务报表相关的内部控制是否健全并有效执行”中发行人内部控制的相关内容。

二、实际采集人数和标注工作量与结算数据量的对比情况、结算调整及差异原因

1、实际采集人数与结算数据量的对比情况

发行人有采集的单一化标注项目类型涉及的采集人数较多。与实际采集人数相比，结算人数略多于实际采集人数，主要是由于：（1）正式采集前，发行人会根据项目情况采集少量数据进行测试，相应的测试人数不作为有效的实际采集人数，但该部分数据量需要与供应商结算；（2）由于发行人的采集设备和软件

问题导致所采集的数据不符合要求，则不计入实际采集人数，但该部分数据量需要与供应商结算。

根据对发行人采集环节的核查情况，按报告期内各年数据资源定制服务和当期首次销售数据库产品收入合计占比50%以上的标准抽取项目，其合计结算人数分别为19,946人、21,360人、31,910人，超过实际采集人数的数量分别为334人、352人、812人，占实际结算人数的比例分别为1.67%、1.65%、2.54%。结算人数不低于实际采集人数，且相差较小。

2、实际标注工作量与结算数据量的对比情况

发行人按交付数据量×标产比来确定标记人员的有效工时，并按此与供应商结算相关数据服务费用，因此实际标注工作量与结算数据量之间不存在差异。

三、供应商验收结算单相关信息的来源

供应商验收结算确认单中的相关结算信息主要为单价、数量、服务费、管理服务费和结算总金额。其中， $服务费=单价 \times 数量$ ， $管理服务费=服务费 \times 管理费率$ ， $结算总金额=服务费+管理服务费$ 。

单价：请参见前文本问题“请发行人补充披露各类业务经营过程中采购和销售业务的定价方式和过程”之回复。

数量：①对于采集数量信息，发行人项目负责人员根据已完成采集的数据量对应的采集服务支出填报，经过部门经理、主管领导和财务部门复核审批；供应商核对确认后盖章。②对于标记工作量信息，项目负责人员根据已完成标注的数据量×标产比得到的有效工时数量填报，经过部门经理、主管领导和财务部门复核审批；供应商核对确认后盖章。

四、保证实际采购业务和财务信息一致的方法

（一）业务管理系统中记录的实际采购业务信息完整、准确

如前所述，发行人采集数据量、有效工时均与交付数据量直接挂钩，保证了业务管理系统记录的实际采购业务信息是完整、准确的。

发行人在内部管理中，通过预算管理、支出管理和项目决算管理，从项目计划到实施完毕，始终通过交叉复核的方式监控采集、结算、交付数据量的一致性；

通过测速管理、同类项目标产比比较分析等措施保证标产比的合理性；通过交付数据量的核对，保证有效工时计算的准确性。从而保证业务管理系统中的采购业务信息是完整、准确的。

（二）财务系统中的采购相关信息与业务管理系统中的采购业务信息一致

项目采购支出信息通过业务管理系统自动汇总并生成支出汇总表，人工录入财务系统后，生成存货、研发费用等信息。由专门人员将两个系统的记录导出后，在表格内进行总额的逐一比对，如有差异，调查原因，并进行相应修订。从而保证采购相关财务信息与业务信息的一致性。

保证采购业务与财务信息一致性的具体内控措施请参见本问题回复“第三部分/（2）/（三）进一步说明发行人与财务报表相关的内部控制是否健全并有效执行”的相关内容。

第四部分：申报会计师核查情况

经核查，申报会计师认为：发行人未保留数据资源定制服务数据库的原因合理；发行人未保存与经营业务相关的日志文件，主要通过线下控制手段保证生产经营过程的可追溯；发行人与财务报表相关的业务采购内部控制健全并得到有效执行，并制定了进一步完善内部控制的有效措施；发行人实际采购业务与财务信息一致。

申报会计师核查过程：

（1）在没有保存业务数据的情况下如何对发行人数据服务费支出进行核查，如何保证核查结论的合理和准确；

如前所述，发行人定制化数据在交付之后不予保留，但数据库产品仍然保留；生产过程数据通过项目采集信息表、项目周报、客户沟通邮件、客户验收确认文件等线下形式予以保留。申报会计师结合对发行人生产模式和产品特点的理解，通过留存数据对生产模式进行了验证，并通过抽样的方式对数据采集的完整性、标产比的准确性和稳定性等方面进行了详细复核比对；对终端被采集人和标记人员进行了电话抽样访问，并抽查了被采集人签字记录、供应商向被采集人和标记人员的付款记录等，核查发行人采购数据服务的真实性。

此外，申报会计师通过供应商函证、走访，确认发行人采购数据服务费真实、准确；查阅了发行人实际控制人、董监高等关联方银行流水，获取了供应商的相关说明和承诺，确认发行人不存在其他方代垫成本费用的情形，确认发行人采购数据服务费的完整性；对发行人报告期内的毛利率、采集产出比、标注产出比进行了详细的分析性复核，核查发行人采购数据服务费的完整性。

具体核查方式和核查过程如下：

一、关于发行人生产模式的有效性验证以及相关核查

申报会计师通过核查发行人的生产模式和采购数量确定机制的合理性，以及复核交付验收数据量、采集数据量、有效工时等主要指标的一致性，来验证数据服务采购的完整性。由于智能语音业务为报告期内发行人最主要的业务领域，报告期内各年收入占比均超过 80%，故本次核查以智能语音业务为主。同时对于数据库产品，由于数据库音频文件在发行人服务器上予以保留，申报会计师对数据库产品发音人音频文件本身的真实、有效性也进行了核查。

（一）对发行人采集和标注环节业务数据的核查

申报会计师对数据资源定制服务项目和数据库产品的抽查比例情况如下：

单位：万元

	2018 年	2017 年	2016 年
①抽查项目收入	5,726.43	3,511.87	2,950.80
②发行人总营业收入	19,265.77	11,907.09	8,422.86
占比（①/②）	29.72%	29.49%	35.03%
③发行人不含重复销售的数据库产品的营业收入	13,877.05	8,312.39	6,649.69
占比（①/③）	41.27%	42.25%	44.37%
④抽查项目成本（定制服务）+产品研发费用（数据库产品）	2,936.03	2,031.21	1,916.56
⑤发行人营业成本+产品研发费用	8,051.38	4,758.51	4,227.56
占比（④/⑤）	36.47%	42.69%	45.33%

注：（1）由于重复销售的数据库无成本或研发费用发生，故上表除列示抽查项目收入对总营业收入的占比外，还列示了对不含重复销售的数据库产品的营业收入的占比。（2）由于数据库产品的开发支出计入研发费用，故将数据库产品的研发费用与定制服务的成本合并计算。

1、采集环节的核查

（1）对有采集的单一化标注语音项目采集人员数量完整性执行的核查程序

核查方式：

①抽查发行人报告期内项目的客户合同或订单、客户验收文件、项目采集信息表、供应商结算单等业务和财务资料，统计核对客户验收发音人总人数、供应商采集总人数、供应商结算总人数等数据的匹配关系；

②统计分析供应商结算总人数与客户验收发音人总人数总人数的差额占前者的比例（“采集损耗率”），统计分析采集损耗率在报告期内的稳定性、合理性，验证分析采集支出的准确性、完整性；

③随机抽取发行人服务器上存储的数据库产品音频文件进行试听，验证音频文件有效性，同时对抽取的数据库产品发音人数量与该数据库产品的采集数据量对比分析。

有采集的单一化标注语音项目采集环节的核查情况：

单位：人

项目年份	抽查项目数量（个）	客户验收发音人数	供应商采集总人数	供应商结算总人数	采集损耗率
2018年	86	30,285	31,038	31,581	4.10%
2017年	55	20,826	21,350	21,730	4.16%
2016年	28	19,355	19,837	20,198	4.17%

注：采集损耗率=（供应商结算总人数-客户验收发音人总人数）/供应商结算总人数。

2016年至2018年，按当期数据资源定制服务和当期首次销售数据库产品收入合计占比50%以上并尽量覆盖各种语种、场景和采集设备²的标准，分别抽查了28个、55个、86个有采集的单一化标注语音项目。供应商采集总人数指采集数据量，与项目采集信息表（包括采集人员姓名、性别、年龄、地区、联系方式、采集文本编号、采集场景、采集日期等信息）总人数一致；抽查项目的采集损耗率区间为4.10%至4.17%；供应商结算总人数为发行人与供应商之间的结算数据量，样本结算总人数略大于供应商采集总人数。经核查，上述关键数据量指标基本一致，采集损耗率稳定且保持较低水平。

2016年至2018年，在各年度随机抽取了5个存储在服务器上的数据库产品

² 语种：选取中文、英语、德语、韩语、法语、西班牙语等主要语种/方言；场景：覆盖车载、桌面、录音棚等常见场景；采集设备：覆盖手机、音箱、麦克风等录音通道。

音频文件进行试听，经核查，服务器上存储的数据库产品音频文件真实、有效，抽样数据库产品发音人数量与该数据库产品的供应商结算数据量基本一致。

核查结论：

有采集的单一化标注语音项目的供应商结算总人数（结算数据量）不低于供应商采集总人数（采集数据量），供应商采集总人数不低于客户验收总人数（交付数据量），符合实际情况，不存在已采集未结算的数据量；同时采集损耗率在报告期内基本保持稳定，且维持较低水平，所需冗余采集量较小；发行人服务器存储的数据库产品音频文件真实、有效，抽样数据库产品的发音人数量与该数据库产品的供应商结算人数基本一致。

（2）对多样化标注语音项目采集字词、句数量完整性进行的核查程序

核查方式：

①抽查报告期内项目向供应商采集的字词、句数量，与发行人设计文本字词数量，以及交付客户并验收字词数量进行核对，计算供应商结算字词数对客户验收字词数的损耗率；

②抽样听取了部分多样化标注语音数据库产品的音频文件，并与其对应的语料文本内容核对一致性，验证多样化标注语音项目采集支出的完整性、真实性。

多样化标注语音项目采集环节的核查情况：

项目年份	抽查项目数量（个）	客户验收字词数	公司设计文本字词数	供应商结算字词数	采集损耗率	抽样音频文件是否与字词文本一致
2018年	18	4,411,026	4,418,198	4,427,664	0.38%	是
2017年	17	3,952,738	3,955,516	4,000,342	1.19%	是
2016年	10	2,436,716	2,446,166	2,466,766	1.22%	是

注：采集损耗率=（供应商结算字词数-客户验收字词数）/供应商结算字词数

2016年至2018年，按当期数据资源定制服务和当期首次销售的数据库产品

合计收入占比 50%以上并尽量覆盖各种语种、发音人类型、细分标注环节³的标准，分别抽查了 10 个、17 个、18 个多样化标注语音项目。样本项目的公司设计文本字词数和供应商采购字词数比客户验收字词数略多，主要原因是公司为保证客户验收合格率和客户满意度，会额外设计并采集少量的备用字词录音。样本项目中，供应商采购字词数对客户验收字词数的损耗率为 0.38%至 1.22%，报告期内基本稳定。

由于部分多样化标注语音项目在启动前期，发行人会有发音人遴选的过程，即安排多个发音人按相同的设计文本录制小样，并最终确定一个发音人。参与遴选但未被选中的发音人，发行人仍需与供应商结算相应的发音人费用。由于上述原因，使得供应商结算的采购字词数比设计文本字词数略多。

核查结论：

对于多样化标注语音数据库，公司设计文本字词数大于客户验收的字词数，公司向供应商采集录音的字词数大于客户验收的字词数，符合实际情形，不存在已采集未结算的数据量；公司设计文本字词数对客户验收字词数覆盖率、供应商采购字词数对客户验收字词数覆盖率在报告期内基本保持稳定，且维持较低水平，所需冗余采集量较小；数据库音频文件真实、有效；多样化标注语音项目采集字词、句数量准确、完整。

2、标注环节的核查

(1) 对自测标产比执行的核查程序

核查方式：

抽查智能语音业务部分项目，统计供应商完成标注、质检等业务环节的有效工作时长，计算标产比，现场观摩同类型项目标产比，验证标产比的合理性。

①有采集的单一化标注语音的核查情况：

项目年份	抽查项目数量 (个)	客户验收小时数 (小时)	供应商完成工时 数 (小时)	标注产出比
2018 年	86	42,700	129,080	3.02

³ 语种：选取中文、英文、西班牙语、葡萄牙语等主要语种/方言；发音人方面：覆盖成年人、儿童及男性、女性；标注环节：覆盖了韵律、词性、音素边界标注、校对等标注环节。

项目年份	抽查项目数量 (个)	客户验收小时数 (小时)	供应商完成工时 数(小时)	标注产出比
2017年	55	19,234	103,389	5.38
2016年	28	21,121	59,690	2.83

注：标注产出比=供应商完成工时数/客户验收小时数。

2016年至2018年，仍选取采集环节抽取的样本，样本平均标注产出比分别是2.83、5.38和3.02。

2017年抽样项目的平均标注产出比较高，主要因为当期抽取的55个项目中有8个自由对话项目，其中还包括泰米尔语、古吉拉特语、泰卢固语等标注难度较高的印度小语种自由对话项目。自由对话类项目在采集环节无预设文本，由发音人自由发挥，公司需要先进行人工文本生成，再进行文本与声音一致性检查、声音事件分类标注等标注环节，因此相对有预设文本类项目而言标注复杂度更高，标注用时更长，因此提升了2017年抽样项目的平均标注产出比。

申报会计师现场抽取了5个有采集的单一化标注语音中文项目，涵盖了中文和外语种，对标记工作独立进行了现场监督和测速，测得上述项目有预设文本类的项目标注产出比区间为2.89~3.01，自由对话类项目标注产出比区间为16.23~19.30，与抽查项目的有采集的单一化标注语音标注产出比基本一致，并且自由对话类项目标注难度较大，标注产出比较高。

核查结论：

申报会计师抽取了有采集的单一化标注语音项目，对其标注环节独立进行了监督测速，与将测速结果与样本的标注产出比进行比对，两者基本一致，同时自由对话类项目标注难度较大，标注产出比较高，符合公司业务实际情况。

②无采集的单一化标注语音标注环节的核查情况

项目年份	抽查项目数量 (个)	客户验收小时数 (小时)	供应商完成工时 数(小时)	标注产出比
2018年	50	54,287	680,140	12.53
2017年	50	24,378	295,423	12.12
2016年	49	30,152	353,136	11.71

注：标注产出比=供应商完成工时数/客户验收小时数。

公司报告期内无采集的单一化标注语音项目没有数据库产品，只有数据资源

定制服务收入，2016年至2018年，按当期数据资源定制服务收入占比50%以上并尽量覆盖各种语种、音频长度、主题⁴的标准分别抽查了49个、50个、50个无采集的单一化标注语音项目，标注产出比区间为11.71~12.53，报告期内基本稳定，呈小幅上升趋势的原因是由于长句音频项目逐年增多，长句通常语速较快且变化更为复杂，人物角色和领域较为复杂，多为客服等领域，因而难度较高，增加了语音标注的难度。

申报会计师现场抽取了无采集的单一化标注语音的3个短音频项目、3个长音频项目，按标注、质检等不同环节分别进行测速，汇总各环节标注时长后，除以加工的音频文件时长，得出该项目的标注产出比。分别测得上述短音频项目的标注产出比区间为10.29~11.00，长音频项目的标注产出比区间为14.42~15.00，与抽查项目的无采集的单一化标注语音标注产出比区间基本一致。

核查结论：

申报会计师抽取了无采集的单一化标注语音项目，对其标注环节独立进行了监督测速，与将测速结果与样本的标注产出比进行比对，两者基本一致，不存在较大差异。

③多样化标注语音标注环节的核查情况

项目年份	抽查项目数量 (个)	客户验收小时数(小时)	供应商有效标注 工时数(小时)	标注产出比
2018年	18	504	32,803	65.10
2017年	17	412	35,018	85.20
2016年	10	292	19,415	66.57

2016年至2018年，仍选取采集环节抽取的样本，样本的标产比分别为66.57、85.20和65.10。

2017年多样化标注语音项目标注产出比较高的原因是：抽样项目中有2个项目分别为西班牙语和法语项目，由于文本中夹杂着大量的英语单词，需要先就西班牙语或法语进行标注，然后对英语进行标注，相当于标注工作需要重复两遍进行；另外抽样项目中有1个为马来西亚语项目，由于马来西亚语难度较高，该

⁴ 语种：选取中文、英文等主要语种/方言；长度：长句音频、短句音频；主题：覆盖会议、演讲、客服、呼叫中心等常见主题。

项目由马来西亚籍标注员和中国标注员分别校对了一次，合计标注时间较长。综上所述，由于抽样项目中部分外语种项目的标注难度较高、标注时间较长，提升了2017年抽样项目的平均标注产出比。

多样化标注语音项目的细分标注环节较多，包括韵律、词性、音素边界标注、校对，整体标产比为各环节标产比的累加，环节越多，所需标注时长越长。同时，中文和外语种的标注难度有显著差异，中文的标产比明显低于外语种。

申报会计师现场抽取了6个目前正在执行的多样化标注语音项目，分别按韵律标注、词性标注、音素边界标注、校对4个环节进行了测速，测得标产比情况如下：

明细环节	中文项目平均标产比	外语种项目平均标产比
韵律标注	8.12	10.54
词性标注	12.37	15.16
音素边界标注	36.55	52.34
校对	9.89	13.25

前述抽查的多样化标注语音项目按包含的明细标注环节可主要分为下表列示的两大类。将各类被抽项目的标产比与测速项目进行对比，情况如下：

项目包含的明细标注环节	样本均值 2018年	样本均值 2017年	样本均值 2016年	中文项目测速标产比	外语种项目测速标产比
韵律+词性+音素边界+校对	74.21	87.76	88.81	66.92 ^注	91.29 ^注
韵律+音素边界+校对	59.02	70.75	70.74	54.55 ^注	76.14 ^注

注：数值为前表中相应语种的对应环节标产比累加

由上表可以看出，抽查的多样化标注语音项目标产比均介于测速得到的中文和外语种标产比数值之间。

核查结论：

申报会计师抽取了多样化标注语音项目，对其标注环节独立进行了监督测速，与将测速结果与样本的标注产出比进行比对，两者基本一致，不存在较大差异。

(2) 对客户确认标产比情况执行的核查程序

核查方式：

客户确认标产比的情形主要存在于部分计算机视觉、自然语言项目，确认方式为邮件。申报会计师抽查了部分项目，查阅客户的速率确认邮件，统计供应商完成标注、质检等业务环节的有效工时；并用客户确认的标产比×交付数据量，计算得出客户认可的工作小时，与供应商完成的有效工作小时数进行对比分析，验证标注环节支出的合理性、完整性。

核查情况：

项目年份	抽查项目数量 (个)	客户验收工时数 (小时)	供应商有效标注 工时数(小时)	标注产出比
2018年	92	53,879	71,137	1.32
2017年	47	27,402	34,740	1.27
2016年	20	24,944	28,624	1.15

注：客户验收工时数=客户验收数据量×客户确认的工作速率；标注产出比=供应商有效标注工时数/客户验收工时数。

2016年至2018年，有部分计算机视觉和自然语言项目按客户确认的标产比进行结算，按此类结算项目合计数量50%以上的标准各期分别抽查了20个、47个、92个项目，抽查的项目中，标注产出比区间为1.15至1.32，标注产出比高于1的原因主要是由于客户确认的工作量不包含质检，供应商的有效工时包含了质检。报告期内标注产出比总体较为稳定，逐年略有上升的主要原因是随着客户对数据标注要求越来越高，质检的抽查比例提升，使得每小时产出所需总的标注工作小时数量增加所致。

核查结论：

对于客户确认标产比的项目，与供应商结算的有效工时大于按客户确认速率计算的工作小时，主要原因为与供应商结算的有效工时中额外包含了质检工时，不存在未结算的有效工时；抽样项目的标注产出比在报告期内总体较为稳定，标注产出比逐年略有上升原因具有合理性。

（二）项目周报的核查

公司管理人员为了有效监督各业务主要项目工作进度，形成了各业务负责人每周向公司管理人员汇报项目进度的项目周报制度，申报会计师在前述“（一）对发行人采集和标注环节业务数据的核查”抽查的项目中又按项目数量抽取了三年合计30%以上的项目，对抽查项目是否在项目周报中归集并汇报的情况进行了

核查，核查情况如下：

单位：个

业务类型	抽样项目数量	周报覆盖数量	占比
有采集的单一化标注语音项目	62	62	100.00%
无采集的单一化标注语音项目	47	47	100.00%
多样化标注语音项目	17	17	100.00%
计算机视觉和自然语言项目	50	49	98.00%
合计	176	175	99.43%

经核查，发行人通过项目周报的形式对绝大部分项目进行了监控和管理；周报以邮件的形式由业务负责人向发行人总经理、主管业务的副总经理发出，发行人使用 163 企业邮箱系统，不存在人为篡改邮件系统的可能，相关项目周报的控制流程真实、可追溯。

（三）对最终采集对象和标记人员的相关核查

对最终采集对象和标记人员的相关核查情况请参见“问题 3/（2）对劳务公司提供服务的真实性进行核查的过程、程序和范围，包括但不限于核查个人的身份登记记录、个人提供服务后确认的签字凭证、劳务公司向个人付款的相关凭证和完税凭据等。”

二、供应商函证、实地走访

1、对数据服务供应商的函证情况

申报会计师对发行人报告期内主要数据服务供应商进行函证，核查采购金额和往来余额的真实性和准确性，各年发函比例如下，所发函证均取得回函，回函金额不存在差异；

单位：万元

年度	数据服务供应商发函情况			数据服务供应商回函情况	
	采购金额	发函金额	发函比例	回函金额	回函比例
2018 年	7,352.66	6,232.50	84.77%	6,232.50	100.00%
2017 年	4,057.31	3,181.51	78.41%	3,181.51	100.00%
2016 年	3,396.15	2,691.43	79.25%	2,691.43	100.00%

2、对数据服务供应商的实地走访情况

申报会计师对发行人报告期内主要数据服务供应商进行了实地走访，所走访的数据服务供应商的对应各年数据服务采购金额占各年数据服务采购总金额的比例超过 90%，访谈过程中，主要就双方合作时间、合作方式、提供服务类型、各年交易规模、发行人采购占其同类销售占比、验收结算方式、与外部劳务人员签订协议情况等情况进行详细了解，核查采购交易的真实性、合理性，以及采购模式的可持续性。

经核查，发行人对主要数据服务供应商的采购真实、合作商业背景合理。

三、关于关联方是否代垫成本费用的核查

1、申报会计师获取了发行人报告期内主要数据服务供应商的工商资料和报告期内税务合规证明，并穿透获得其股东和其他关键人员、关联公司名单，检查与发行人、发行人关联方及其董事、监事、高级管理人员等关键人员是否存在关联关系。

2、申报会计师获取了发行人报告期内主要数据服务供应商出具的说明和提供的财务数据，核查其与发行人之间采购定价是否公允，采购交易是否符合商业实质并为供应商带来商业收益，确认双方之间及双方股东、董事、监事、高级管理人员、关键经办人员之间是否存在关联关系，是否存在提供劳务服务以外的其他交易和资金往来并为发行人代垫成本费用的行为，核查发行人成本费用的完整性。

3、申报会计师获取了发行人实际控制人及其配偶、发行人总经理控制、担任董事或高级管理人员、具有重大影响的企业（包括注销或转让的）及其董事、监事、高级管理人员出具的说明，确认该等企业及其董事、监事、高级管理人员与发行人供应商之间是否在交易、资金往来或其他利益安排，是否存在为发行人代垫成本费用的行为，核查发行人成本费用的完整性。

经核查，发行人主要数据服务供应商与发行人、发行人关联方及其董事、监事、高级管理人员等关键人员不存在关联关系，发行人主要数据服务供应商与发行人之间采购定价公允，不存在除提供劳务服务以外的其他交易和资金往来，不存在其他方为发行人代垫成本费用的情况，报告期内发行人成本费用完整。

四、对发行人毛利率、采集产出比、标注产出比等经营指标进行分析性复

核

申报会计师对发行人报告期内对营业收入和营业成本的匹配性进行了重点核查，对毛利率、采集产出比、标注产出比、单位价格和单位成本等关键经营指标进行了详细的分析性复核，通过高管访谈、查阅项目资料和财务资料、了解行业情况以判断波动是否符合行业趋势等方式对存在较大波动的情形进行了分析，获得了合理解释。相关分析内容参见第二轮审核问询函回复问题 7、10 和第三轮审核问询函回复问题 5、6 之相关内容。

经核查，发行人营业收入和营业成本具有较高的匹配性，相关指标波动的具体业务背景原因合理，有效验证了数据服务采购的完整性。

(2) 对劳务公司提供服务的真实性进行核查的过程、程序和范围，包括但不限于核查个人的身份登记记录、个人提供服务后确认的签字凭证、劳务公司向个人付款的相关凭证和完税凭据等。

一、对劳务供应商提供服务真实性的总体核查情况

1、对发行人采购劳务服务的穿行核查

申报会计师对发行人报告期内向劳务供应商采购劳务服务进行了穿行核查，从账面记录的数据服务费采购出发，获取采购框架合同、采购订单资料、经双方盖章确认的验收结算单、收到的发票、支付采购结算款项的银行回单，检查供应商名称、验收服务的内容、验收时间、结算数量、结算单价、结算金额、支付金额等信息与发行人记录是否一致，核查采购交易的真实性、准确性。2016 年、2017 年和 2018 年，核查金额占发行人当期营业成本与产品研发费用之和的比例分别为 60.46%、67.95%和 74.19%。

2、对劳务供应商进行函证和走访

申报会计师对劳务供应商进行函证和走访的情况请参见前文“二、供应商函证、实地走访”。

3、对终端劳务服务人员的核查

申报会计师对劳务公司提供服务真实性进行了一系列核查，包括对提供劳务的采集和标注人员电话访谈、查阅由采集人员签字的现场采集登记表、核查劳务

公司向个人付款的相关凭证和完税凭据等。具体核查内容、核查范围、核查情况如下。

二、对终端劳务服务人员进行穿透核查的情况

（一）对采集和标注人员电话访谈的劳务真实性核查

为了验证劳务公司提供服务的真实性，申报会计师在前述“（一）对发行人采集和标注环节业务数据的核查”抽查的项目中又按项目数量抽取了三年合计30%以上的项目，合计超过1,100名采集和标注人员进行了电话访谈，请访谈对象对其是否真实参与了某项目的录音/标注工作、工作完成后是否收到报酬、报酬金额或报酬单价的区间进行了确认。

2016年至2018年，对采集和标注人员电话访谈的核查情况如下：

单位：个/人

业务类型	电话访谈项目数量	电话访谈人数	电话访谈有效回复人数	确认比例
有采集的单一化标注语音项目	62	857	648	75.61%
无采集的单一化标注语音项目	47	145	145	100%
多样化标注语音项目	17	51	51	100%
计算机视觉和自然语言项目	50	69	69	100%

其中有采集的单一化标注语音项目的确认比例低于其他类型项目，主要是由于该类型项目主要为临时性、一次性录音，并且有的发音人参与录音的时间距今较久，有部分发音人在电话中不愿配合回答问题。对于未确认的被访谈对象，申报会计师进一步核查了现场采集登记表，其中有190位被访谈对象在登记表中留有签字，占有采集的单一化标注语音项目电话访谈未有效回复人数的90.91%。

（二）劳务公司向个人付款的相关凭证核查

申报会计师对前述“一、对采集和标注人员电话访谈的劳务真实性核查”中抽查的项目除进行电话访谈以外，又向劳务供应商索取了对应项目的付款明细表和银行回执等向终端劳务人员付款的相关凭证，核查是否向劳务人员如实支付相关费用；同时将劳务供应商付款明细表与发行人报销明细表进行核对，查验支付给劳务人员的费用与发行人报销金额是否一致。

核查情况：

单位：万元

业务类型	抽样项目数量	发行人支付给供应商的服务费 ^注	供应商支付给个人的费用	供应商向个人支付费用占比
有采集的单一化标注语音项目	62	267.78	266.58	99.55%
无采集的单一化标注语音项目	47	700.49	696.59	99.44%
多样化标注语音项目	17	262.30	241.36	92.02%
计算机视觉和自然语言项目	50	112.48	111.75	99.35%
合计	176	1,343.05	1,316.28	98.01%

注：“发行人支付给供应商的服务费”不含向供应商支付的管理服务费。

经核查，上述抽样项目劳务公司支付明细表明细加总金额与发行人报销明细加总金额基本一致，抽样项目发行人合计支付给供应商服务费 1,343.05 万元，供应商支付给个人的费用合计 1,316.28 万元，抽样项目的供应商向个人支付费用整体占发行人支付给供应商服务费的比例为 98.01%。

上述抽样项目电话访谈和供应商付款凭证的抽查比例情况如下：

单位：万元

	2018 年	2017 年	2016 年
①抽查项目收入	2,659.81	819.55	655.11
②发行人总营业收入	19,265.77	11,907.09	8,422.86
占比（①/②）	13.81%	6.88%	7.78%
③发行人不含重复销售的数据库产品的营业收入	13,877.05	8,312.39	6,649.69
占比（①/③）	19.17%	9.86%	9.85%
④抽查项目成本（定制服务）+产品研发费用（数据库产品）	1,250.85	489.64	373.96
⑤发行人营业成本+产品研发费用	8,051.38	4,758.51	4,227.56
占比（④/⑤）	15.54%	10.29%	8.85%

注：（1）由于重复销售的数据库无成本或研发费用发生，故上表除列示抽查项目收入对总营业收入的占比外，还列示了对不含重复销售的数据库产品的营业收入的占比。（2）由于数据库产品的开发支出计入研发费用，故将数据库产品的研发费用与定制服务的成本合并计算。

（三）对个人的身份登记记录、个人提供服务后确认的签字凭证的核查

报告期内大部分时间，发行人以项目采集信息表电子表格的形式留存被采集对象的相关信息。2018 年开始，为了进一步加强对采集活动的管理，发行人开始要求被采集对象在现场采集时，需在纸质版的项目采集信息表上签字作为现场

采集登记表，并有专人负责到现场采集登记表进行收集和整理。

申报会计师对前述抽查的2018年86个有采集的单一化标注项目的现场采集登记表留存情况进行了核查，经核查，其中60个项目保留了现场采集登记表，被采集对象在现场采集登记表上留有签字。上述项目对有采集的单一化标注业务的收入占比、对成本和产品研发费用之和的占比分别为37.82%、29.79%。

（四）完税凭据的核查

劳务供应商主要为发行人提供生数据采集、标记服务。

1、采集服务

关于生数据采集业务，劳务供应商未提供相应被采集人的完税凭据。存在以下两种情形：

（1）有采集的单一化标注语音项目、计算机视觉类项目和自然语音类项目

该等项目特点为采集人数众多、人均结算单价较低。报告期内，有采集的单一化标注语音项目人均结算单价分别是183.67元、141.79元和202.28元，因此为发行人提供此类服务的个体单次收入较低，平均而言未达到每次800元的劳务报酬个人所得税起征点。

（2）多样化标注语音项目

该等项目特点为采集人数较少，人均结算单价较高，发音人单次收入平均而言达到了个人所得税缴纳标准。报告期内，发行人为采购此类采集服务向境内劳务供应商支付的数据服务费金额分别为216.86万元、251.34万元、336.45万元。假设发音人均需以全部劳务服务收入为应纳税所得额，按照20%的税率缴纳个人所得税，据此估算，劳务供应商需代扣代缴的个人所得税总额分别不超过43.37万元、50.27万元、67.29万元。

2、标记服务

发行人智能语音类、计算机视觉类、自然语言类业务均采购了劳务供应商的标记服务，根据标记人员与劳务供应商之间的关系可以分为以下两类：

（1）因发行人部分客户有持续订单需求，希望有相对较为稳定、熟练的标

记人员为其提供服务。因此，针对该部分需求，劳务供应商与部分标记人员签署了正式劳动合同，并为其依法代扣代缴了个人所得税。申报会计师取得了目前仍在职的 33 名标记人员在主管税务机关开具的《个人所得税纳税清单》。

(2) 部分标记人员为临时性兼职人员，单次劳务服务收入较低，且未与劳务供应商签署正式劳动合同，劳务供应商未能提供相关完税凭据。申报会计师对前述“一、对采集和标注人员电话访谈的劳务真实性核查”抽查项目中采购标记服务支出的数据服务费总额、标记人员数量进行了统计，测算出报告期内平均单个标记人员的年均劳务收入低于 8,500 元，金额较低，推断标记人员涉及个人所得税纳税义务的可能性较小。

此外，申报会计师取得了报告期内前五大数据服务供应商提供的确认文件、中国境内数据服务供应商主管税务机关出具的证明，根据上述确认文件和证明，报告期内，关于数据服务提供商为发行人提供劳务涉及向劳务人员发放的劳务费报酬，数据服务提供商已依法代扣代缴个人所得税，不存在因此而受到税务机关处罚的情形。

核查意见：

为核查劳务供应商提供服务的真实性，申报会计师对采购劳务服务进行了穿行核查，对主要劳务供应商进行了走访和函证，并对终端劳务服务人员提供服务的情况进行了核查。对于终端劳务人员的核查，申报会计师抽取了一定数量的劳务服务人员进行了电话访谈，对访谈对象为公司提供劳务服务并取得报酬相关事项进行了确认；抽查了部分项目现场采集登记表，对其留存和签字情况进行了核实；抽取了一定数量的项目检查供应商向劳务服务人员的付款情况并进行了逐笔核对和加总复核，核实供应商向终端劳务服务人员付款的真实性和金额的准确性；获得了部分劳务服务人员的完税凭证。综合上述核查情况，申报会计师认为，发行人的劳务供应商向发行人提供的服务是真实的。

【本页系瑞华会计师事务所（特殊普通合伙）《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申请文件的第四轮审核问询函之专项核查意见》之签章页】



中国注册会计师：

郭勇

郭勇



中国注册会计师：

左东强

左东强



2019年7月5日