



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
审核问询函的回复

保荐机构（主承销商）



（深圳市前海深港合作区南山街道桂湾五路 128 号前海深港基金小镇 B7 栋 401）

上海证券交易所：

根据贵所于 2020 年 7 月 17 日出具的上证科审（审核）〔2020〕469 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申请文件的审核问询函》（以下简称“问询函”）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，就问询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

一、关于发行人股权结构、董监高等基本情况	4
1.关于实际控制人.....	4
2.关于置换出资.....	19
3.关于整体变更.....	21
4.关于最近一年新增股东.....	24
5.关于香港海天瑞声.....	29
6.关于行业地位.....	32
7.关于数据安全及其合规性.....	45
二、关于发行人业务	63
8.关于业务资质.....	63
9.关于销售模式和主要客户.....	64
9.1 关于主要服务和产品.....	64
9.2 关于主要客户和境内外销售.....	75
9.3 关于销售收入核查.....	110
10.关于采购模式和主要供应商.....	128
10.1 关于数据服务采购.....	128
10.2 关于采购情况.....	128
三、关于发行人核心技术	222
11.关于技术先进性.....	222
四、关于公司治理与独立性	246
12.关于关联方.....	246
五、关于财务会计信息与管理层分析	251
13.关于营业收入.....	251

13.1 关于收入确认政策.....	251
13.2 关于营业收入波动.....	270
14.关于营业成本.....	274
15.关于毛利率.....	286
16.关于期间费用.....	290
16.1 关于销售费用.....	290
16.2 关于管理费用.....	296
16.3 关于研发费用.....	300
17.关于税务事项.....	312
18.关于货币资金.....	312
19.关于应收账款.....	338
20.关于存货.....	346
21.关于预收账款.....	357
22.关于现金流量表.....	357
六、关于其他事项	366
23.关于前次撤回和本次申报.....	366
24.关于募集资金运用.....	381
25.关于风险因素.....	385
26.关于业务数据来源.....	386
27.关于媒体质疑.....	391
七、保荐机构在充分核查基础上的总体意见	392

一、关于发行人股权结构、董监高等基本情况

1. 关于实际控制人

招股说明书披露，贺琳合计控制公司 42.44%的股份，为发行人控股股东、实际控制人。贺琳于 1987 年至 2017 年 7 月任职于中国科学研究院声学研究所，并于 2005 年即创办海天瑞声有限并担任公司高管。

请发行人说明：（1）贺琳在任职中国科学研究院声学研究所期间创办海天瑞声并担任相关职务，是否符合相关法律法规及中国科学研究院的相关规定，是否存在利用任职单位的资源或便利、相关技术成果创办海天瑞声的情形，是否违反中科院的规章制度或与中科院的劳动合同；（2）发行人的相关专利发明及计算机软件著作权中（包括在申请过程中的）是否属于贺琳在中国科学研究院的职务发明创造或职务作品，如属于，是否与中国科学研究院约定了相关的成果归属，是否存在纠纷或潜在纠纷；未将实际控制人贺琳认定为核心技术人员的原因；（3）贺琳与唐涤飞以及其他股东之间是否存在签署一致行动协议的情形，发行人是否由二人共同控制；（4）发行人股东中是否存在为中科院及声学所的相关员工代持股份的情形。

请保荐机构、发行人律师对上述事项进行核查，说明核查手段、核查方式，并发表明确意见。

答复：

发行人说明事项

（1）贺琳在任职中国科学研究院声学研究所期间创办海天瑞声并担任相关职务，是否符合相关法律法规及中国科学研究院的相关规定，是否存在利用任职单位的资源或便利、相关技术成果创办海天瑞声的情形，是否违反中科院的规章制度或与中科院的劳动合同

一、贺琳在任职中国科学研究院声学研究所期间创办海天瑞声并担任相关职务，是否符合相关法律法规及中国科学研究院的相关规定

（一）相关法律法规及中国科学研究院的相关规定

发行人实际控制人、董事长贺琳于 1987 年至 2017 年 7 月任职于中国科学院声学研究所（以下简称“中科院声学所”），系中科院声学所的普通科研人员。贺琳在中科院声

学所任职期间，2005年5月创办海天瑞声有限并任执行董事、经理职务，2013年11月起至今担任执行董事、董事长职务，贺琳一直为发行人实际控制人。自贺琳创办海天瑞声有限至贺琳从中科院声学所离职的期间内，相关法律法规及中国科学院的相关规定具体如下：

具体规定名称	相关的主要内容
中共中央组织部、人事部、科学技术部《关于深化科研事业单位人事制度改革的实施意见》（人发〔2000〕30号）	鼓励科研人员创办高新技术企业或在完成本职工作的前提下兼职从事研究开发和成果转化。
《国务院关于进一步做好新形势下就业创业工作的意见》（国发〔2015〕23号）	调动科研人员创业积极性。探索高校、科研院所等事业单位专业技术人员在职创业、离岗创业有关政策。
《国务院关于印发实施〈中华人民共和国促进科技成果转化法〉若干规定的通知》（国发〔2016〕16号）	国家设立的研究开发机构、高等院校科技人员在履行岗位职责、完成本职工作的前提下，经征得单位同意，可以兼职到企业等从事科技成果转化活动，或者离岗创业，在原则上不超过3年时间内保留人事关系，从事科技成果转化活动。研究开发机构、高等院校应当建立制度规定或者与科技人员约定兼职、离岗从事科技成果转化活动期间和期满后的权利和义务。
《人力资源社会保障部关于支持和鼓励事业单位专业技术人员创新创业的指导意见》（人社部规〔2017〕4号）	支持和鼓励事业单位专业技术人员到与本单位业务领域相近企业、科研机构、高校、社会组织等兼职，或者利用与本人从事专业相关的创业项目在职创办企业，是鼓励事业单位专业技术人员合理利用时间，挖掘创新潜力的重要举措，有助于推动科技成果加快向现实生产力转化。事业单位专业技术人员兼职或者在职创办企业，应该同时保证履行本单位岗位职责、完成本职工作。专业技术人员应当提出书面申请，并经单位同意；单位应当将专业技术人员兼职和在职创办企业情况在单位内部进行公示。事业单位应当与专业技术人员约定兼职期限、保密、知识产权保护等事项。创业项目涉及事业单位知识产权、科研成果的，事业单位、专业技术人员、相关企业可以订立协议，明确权益分配等内容。
《中国科学院关于科技人员兼职的若干意见》（科发人教字〔2003〕164号）	六、我院科技人员在兼职活动中，应做到：1.正确处理本职工作和兼职的关系，首先保证完成岗位的工作任务和要求。2.个人不得以单位的名义与合作单位签定协议、接受任务或承诺属单位职权的事项。3.遵纪守法，讲究职业道德。在兼职过程中，不得侵犯研究所、合作单位及他人的经济技术权益。4.兼职所得的相应报酬，应如实报研究所人事部门备案，并按国家有关规定，缴纳个人所得税。
《中国科学院关于科技人员兼职的若干规定》（科发人教字〔2006〕172号）	二、科技人员从事兼职活动需报请单位同意，由本单位和兼职单位与个人三方签订兼职协议，对兼职的时间安排、取酬数额及方式、知识产权归属以及福利和保险等问题要有明确的规定。……九、2.……科技人员违反兼职管理规定，严重影响本职工作的，其所在单位应当批评教育；侵害本单位技术权益的，单位有权要求其赔偿损失。必要时，可以责令其停止兼职活动，直至给予行政处分，并保留追究其相应法律责任的权利。……4.对符合下列情况的人员，本单位可根据与之签订的聘任合同有关条款，解除聘用关系，并限期将人事关

	系调离本单位：……（2）长期在外兼职或兼职单位过多，不能胜任本职工作的人员；（3）严重违反本规定或者研究所有关兼职管理规定的人员。
《中国科学院关于领导干部兼职收入的若干规定》（科发人教字〔2006〕67号）	四、兼职取酬原则及审批程序 …… 6、领导干部在企业兼职 …… （4）领导干部不准接受兼职企业转让、赠与的股份及其红利。领导干部依照有关规定，由个人出资合法持有的股份，其红利可不视为个人兼职收入。
《中国科学院工作人员兼职管理规定》（科发人教字〔2013〕22号）	第七条 工作人员从事兼职活动的，须经本单位审批，其中，领导干部从事除本规定第六条所述兼职活动之外的，须报院审批。…… 第八条 经审批同意的兼职活动，由本单位和兼职单位与个人三方签订兼职协议，对兼职的工作内容、时间安排、绩效考核、兼职收入、福利保险、保密义务以及知识产权归属等事项应有明确的约定。…… 第十二条 兼职收入是指工作人员兼职所获得的报酬，包括工资、奖金或其他报酬以及由兼职单位转让、赠予的股权及红利等。……4.从事兼职活动的工作人员，对兼职所得的相应报酬，须如实报本单位人事部门备案，并按国家有关规定申报和缴纳个人所得税；…… 第十六条 工作人员如违反兼职管理规定，严重影响本职工作的，其所在单位应当批评教育，责令其停止兼职活动；侵害本单位经济技术权益的，单位有权要求其赔偿损失；必要时，可以给予处分，并追究其相应法律责任。对下列情况，本单位可依法与之解除聘用关系：1.未经单位同意在外兼职，责令其停止兼职活动后，拒不改正的；2.长期在外兼职或兼职过多，不能胜任本职工作的；3.严重损害本单位利益或给本单位造成重大损失的。
《中国科学院关于新时期加快促进科技成果转化指导意见》（科发促字〔2016〕97号）	为促进科技要素合理流动，院属单位应按照相关政策制定本单位的规章制度，允许科技人员在适当条件下兼职从事科技成果转化，并在兼职中取得合理报酬。各单位应书面约定兼职人员的权利义务，兼职人员须如实将兼职收入报单位备案，按规定缴纳个人所得税。

（二）中科院及中科院声学所出具的相关书面意见

根据中科院相关规定，从事兼职活动的科技人员，对兼职所得的相应报酬，应如实报单位人事部门备案，并按国家有关规定申报和缴纳个人所得税。贺琳 2017 年 7 月中科院声学所离职之前未在发行人处领取薪酬，无需按上述规定备案；从中科院声学所离职后全职在发行人工作并领取薪酬，已按规定由发行人依法代扣代缴个人所得税。

参考上述《中国科学院关于领导干部兼职收入的若干规定》（科发人教字〔2006〕67号）第四条第6项，“……由个人出资合法持有的股份，其红利可不视为个人兼职收入。”，并根据上述《中国科学院工作人员兼职管理规定》（科发人教字〔2013〕22号）第十二条的规定，需如实报本单位人事部门备案的兼职收入是指工作人员兼职所获得的报酬，包括工资、奖金或其他报酬以及由兼职单位转让、赠予的股权及红利等；而贺琳系以自有资金创办海天瑞声并进行后续投资，系基于其个人出资所获取的股权及分

红，并非发行人向贺琳转让、赠予，因此贺琳在中科院声学所任职期间从发行人获取的分红不属于上述规定的兼职收入，无需按上述规定进行备案；此外，贺琳从发行人获取的分红已依法由发行人代扣代缴个人所得税。

基于以上分析，贺琳在中科院声学所任职期间未从发行人处领取薪酬，无需进行备案程序，贺琳从声学所离职后全职在发行人处工作并领取薪酬，已按规定由发行人依法代扣代缴个人所得税；贺琳持有的发行人股权系个人出资行为，并非因其兼职由发行人向其转让或赠予，其获取的分红不属于兼职收入无需进行备案，并且发行人已依法就贺琳所获取的分红代扣代缴个人所得税。

贺琳在中科院声学所任职期间创办海天瑞声有限并任职，该行为发生在国家和中国科学院支持科研人员创业的政策背景下，对比上述相关规定，除就兼职事项未与中科院声学所签署协议外，贺琳在任职中科院声学所期间创办海天瑞声并担任相关职务符合相关法律法规及中国科学院的相关规定。

针对贺琳创办公司并兼职事项，中国科学院、中科院声学所分别出具书面意见如下：

1、中科院出具的证明文件

根据上述《中国科学院工作人员兼职管理规定》第十五条，“领导干部兼职统一归口人事教育局管理，其他工作人员兼职统一归口各单位人事部门管理。……”。

中科院人事局已于 2019 年 5 月 22 日出具书面证明：“中国科学院声学研究所（以下简称“声学所”）是中国科学院所属事业单位法人。按照干部管理权限，贺琳同志在声学所工作期间系该研究所自行管理的干部，有关该同志情况以声学所提供说明为准。”

2、中科院声学所出具的书面确认意见

中科院声学所于 2017 年 7 月 28 日出具书面确认：“贺琳于 1987 年至 2017 年在中国科学院声学研究所（以下简称“本所”）任职，已于 2017 年 7 月离职。贺琳在本所任职期间未曾担任领导职务。本所与贺琳之间未曾签署过保密协议，未签署过竞业禁止协议，也没有竞业禁止安排。贺琳在本所任职期间未违反本单位相关制度规定。本所与贺琳之间未发生过保密纠纷、竞业禁止纠纷或劳动纠纷。”

中科院声学所于 2019 年 5 月 15 日出具书面确认：“声学所知悉贺琳创办海天瑞声公司，并在其中兼职。声学所同意贺琳在海天瑞声公司兼职。贺琳在职期间兼职，未影

响其在声学所履行声学所岗位职责。贺琳在职期间在海天瑞声公司兼职及之后的离职符合声学所相关规定，未损害声学所权益，无遗留问题。”

根据以上证明，贺琳在中科院声学所任职期间创办海天瑞声并任职，虽未就兼职事项与中科院声学所签署协议，但中科院声学所知悉贺琳创办公司并同意其兼职，确认贺琳兼职未影响其履行中科院声学所岗位职责，未损害中科院声学所权益。

综上，在国家和中国科学院支持科研人员创业的政策背景下，贺琳在中科院声学所任职期间创办海天瑞声有限并任职；贺琳在中科院声学所任职期间未从发行人处领取薪酬，所获取的分红不属于兼职收入，根据中科院相关规定无需就兼职收入事宜进行备案；除就兼职事项未与中科院声学所签署协议外，符合相关法律法规及中国科学院的相关规定。贺琳与中科院声学所虽未就兼职事项签署协议，但中科院出具书面证明明确中科院声学所对该事项具有管理权限，中科院声学所出具书面确认同意其兼职，确认贺琳兼职未影响其履行中科院声学所岗位职责，未损害中科院声学所权益，因此不会对发行人本次发行上市构成障碍。

二、是否存在利用任职单位的资源或便利、相关技术成果创办海天瑞声的情形

（一）贺琳在中科院声学所的任职和创立海天瑞声有限的背景

贺琳创办海天瑞声有限时在中科院声学所应用工程部（后隶属于声学所噪声与振动重点实验室）工作，该研究室主要从事声信号处理、语音等领域的基础性科学问题研究，不涉及发行人从事的人工智能基础数据服务业务。贺琳根据对语音行业发展趋势的理解，以及对人工智能行业未来对训练数据的市场需求的判断，决定创办海天瑞声有限，为人工智能行业内大型知名企业提供训练数据产品及服务。

（二）发行人创办过程中的技术发展和市场开拓情况

贺琳创办海天瑞声有限未利用中科院声学所的资源或便利、相关技术成果，未侵犯中科院声学所的技术权益：

1、技术方面：海天瑞声有限 2005 年设立初期，海天瑞声有限从事语音采集、标注业务，业务量较小，主要从事语音采集及利用业内公知、基础的开源算法技术对采集的语音进行加工处理。随着人工智能行业迅猛发展，对训练数据开始产生较大的市场需求，同时行业技术也在快速更新升级，市场需求迫使发行人提升技术研发水平，以实现训练数据大量、快速处理的需求。在此背景下，2008 年末以来，核心技术人员郝玉峰、

李科、唐涤飞先后加入发行人，开始带领研发、业务人员逐步进行训练数据生产的平台工具、基础研究、生产技术等方面的技术研发，以便为客户提供大量训练数据的高质量、高效率的服务和产品。贺琳作为执行董事、董事长，主要负责战略、融资、外部关系等重大经营管理事项，未参与具体技术研发。

2、市场方面：贺琳 2005 年设立海天瑞声有限时，市场上专门从事训练数据服务和产品的企业较少。基于贺琳多年积累的人脉基础，海天瑞声有限开始为微软等企业提供语音训练数据服务。随着人工智能行业迅猛发展，依赖前期业绩和客户口碑，以及发行人销售人员对于市场的深度开拓，发行人在 2009 年至 2012 年期间形成与阿里巴巴、三星、百度、腾讯等人工智能领域大型科技公司良好稳定的合作关系。上述大型科技公司与发行人的良好合作关系延续至今。

（三）中科院声学所的书面确认

中科院声学所于 2019 年 4 月 3 日出具书面确认：“本所与北京海天瑞声科技有限公司之间不存在知识产权争议或纠纷。”

中科院声学所于 2019 年 5 月 15 日出具书面确认：“声学所知悉贺琳创办海天瑞声公司，并在其中兼职。声学所同意贺琳在海天瑞声公司兼职。贺琳在职期间兼职，未影响其在声学所履行声学所岗位职责。贺琳在职期间在海天瑞声公司兼职及之后的离职符合声学所相关规定，未损害声学所权益，无遗留问题。”

根据 2020 年 8 月 12 日中介机构对中科院声学所主管人员的访谈，上述 2019 年 4 月 3 日中科院声学所书面确认后发行人新取得或申请的专利及计算机软件著作权和中科院声学所没有冲突，没有争议和纠纷。

综上，贺琳不存在利用任职单位的资源或便利、相关技术成果创办海天瑞声的情形。

三、是否违反中科院的规章制度或与中科院的劳动合同

如前所述，贺琳在中科院声学所任职期间创办海天瑞声有限并担任相关职务，除就兼职事项未与中科院声学所签署协议外，符合中国科学院的相关制度规定。

贺琳具体与中科院声学所签署聘用合同。根据中科院声学所的聘用合同，其中并未就员工创办企业或兼职做出限制性规定。

中科院声学所于 2017 年 7 月 28 日出具书面确认：“贺琳，于 1987 年至 2017 年在中国科学院声学研究所任职，已于 2017 年 7 月离职。贺琳在本所任职期间未曾担任领导职务。本所与贺琳之间未曾签署过保密协议、未签署过竞业禁止协议，也没有竞业禁止安排。贺琳在本所任职期间未违反本单位相关制度规定。本所与贺琳之间未发生保密纠纷、竞业禁止纠纷或劳动纠纷。”

中科院声学所于 2019 年 5 月 15 日出具书面确认：“声学所知悉贺琳创办海天瑞声公司，并在其中兼职。声学所同意贺琳在海天瑞声公司兼职。贺琳在职期间兼职，未影响其在声学所履行声学所岗位职责。贺琳在职期间在海天瑞声公司兼职及之后的离职符合声学所相关规定，未损害声学所权益，无遗留问题。”

综上，贺琳在中科院声学所任职期间创办海天瑞声有限并担任相关职务，除就兼职事项未与中科院声学所签署协议外符合中国科学院的相关制度规定，未违反与中科院声学所的聘用合同。贺琳与中科院声学所虽未就兼职事项签署协议，但中科院声学所出具书面确认同意其兼职，确认贺琳兼职未影响其履行中科院声学所岗位职责，未损害中科院声学所权益，因此不会对发行人本次发行上市构成障碍。

(2) 发行人的相关专利发明及计算机软件著作权中（包括在申请过程中的）是否属于贺琳在中国科学研究院的职务发明创造或职务作品，如属于，是否与中国科学研究院约定了相关的成果归属，是否存在纠纷或潜在纠纷；未将实际控制人贺琳认定为核心技术人员的原因

一、发行人的相关专利发明及计算机软件著作权中（包括在申请过程中的）是否属于贺琳在中国科学研究院的职务发明创造或职务作品，如属于，是否与中国科学研究院约定了相关的成果归属，是否存在纠纷或潜在纠纷

(一) 贺琳不是发行人专利的发明人

截至目前，发行人及其子公司已取得的专利及其具体发明人情况如下所示，不涉及贺琳为发明人的情形：

序号	专利名称	所有权人	专利类型	专利号	专利申请日	授权公告日	发明人
1	语音对齐方法及装置	发行人	发明	ZL 2018 1 0449585.3	2018年5月11日	2020年6月23日	邵志明、郝玉峰
2	音素误标注的检	发行人	发明	ZL 2018 1	2018年7月	2020年1月	孟君、廖晓玲、

	测方法和装置			0827500.0	25日	31日	郝玉峰
3	语音检测方法和装置	发行人	发明	ZL 2018 1 0883930.4	2018年8月6日	2020年6月23日	邵志明、曹琼、宋琼、郝玉峰
4	韵律标注方法、装置和设备	发行人	发明	ZL 2018 1 0988973.9	2018年8月28日	2020年1月7日	孟君、曹琼、廖晓玲、郝玉峰
5	中英混合语料的生成方法、装置、设备及存储介质	发行人	发明	ZL 2019 1 0022453.7	2019年1月10日	2020年3月24日	杨福星、曹琼、郝玉峰
6	音频质量评估方法、装置、电子设备及存储介质	发行人	发明	ZL 2019 1 0073390.8	2019年1月25日	2019年10月29日	刘永辉、曹琼、郝玉峰
7	语料选取处理方法、装置、设备及计算机可读存储介质	发行人	发明	ZL 2019 1 0072150.6	2019年1月25日	2020年1月14日	杨福星、曹琼、郝玉峰
8	视频中音频聚类的处理方法和装置	发行人	发明	ZL 2019 1 1289077.4	2019年12月16日	2020年5月5日	闫启伟、黄宇凯、郝玉峰、曹琼、李科
9	个性语料获取方法及个性语料获取装置	发行人	发明	ZL 2019 1 1314319.0	2019年12月19日	2020年5月5日	杨福星、郝玉峰、黄宇凯、邵志明、曹琼、李科
10	语音库的语音获取方法、装置、电子设备及存储介质	发行人	发明	ZL 2019 1 1314504.X	2019年12月19日	2020年5月5日	孟君、黄宇凯、郝玉峰、曹琼、李科、宋琼
11	短语语料获取方法及短语语料获取装置	发行人	发明	ZL 2019 1 1352915.8	2019年12月25日	2020年5月5日	杨萌萌、郝玉峰、黄宇凯、邵志明、曹琼、李科
12	端到端的语音合成方法、装置及存储介质	发行人	发明	ZL 2020 1 0007974.8	2020年1月6日	2020年5月19日	邵志明、黄宇凯、郝玉峰、曹琼、李科、宋琼
13	视频抽选方法、装置及存储介质	发行人	发明	ZL 2020 1 0100697.5	2020年2月19日	2020年6月2日	刘杰辰、曹琼、郝玉峰、黄宇凯、李科
14	视频筛选方法、装置及存储介质	发行人	发明	ZL 2020 1 0096575.3	2020年2月18日	2020年6月26日	刘杰辰、曹琼、黄宇凯、郝玉峰、李科
15	图像标注方法、图像标注装置及计算机存储介质	发行人	发明	ZL 2020 1 0096586.1	2020年2月18日	2020年8月4日	陈佃文、曹琼、黄宇凯、郝玉峰、李科、张哲玮
16	跨语言非标准词识别方法及装置	发行人	发明	ZL 2020 1 0122519.2	2020年2月27日	2020年7月14日	闫启伟、郝玉峰、黄宇凯、曹琼、李科、宋琼

17	语音数据库的质量评估方法、装置及计算机存储介质	发行人、清华大学	发明	ZL 2020 1 0164556.X	2020年3月11日	2020年6月19日	张卫强、李科、黄宇凯、郝玉峰、曹琼
18	用于度量语音数据库覆盖性的无监督模型训练方法及装置	发行人、清华大学	发明	ZL 2020 1 0309303.7	2020年4月20日	2020年7月28日	李科、张卫强、黄宇凯、郝玉峰、宋琼
19	语音转写方法、语音转写装置及计算机存储介质	发行人	发明	ZL 2020 1 0283135.9	2020年4月13日	2020年7月28日	孟君、黄宇凯、郝玉峰、曹琼、李科、宋琼
20	基于麦克风的音频处理方法和装置	中瑞智	发明	ZL 2017 1 0539306.8	2017年7月4日	2020年3月31日	黄大伟、李虹辰
21	语音采集设备	发行人	实用新型	ZL 2019 2 2276799.8	2019年12月18日	2020年3月31日	王涛、黄宇凯、郝玉峰、曹琼、李科
22	智能语音采集设备	中瑞智	外观设计	ZL 2020 3 0062563.X	2020年2月27日	2020年8月18日	王昆

截至目前，发行人及其子公司正在申请过程中的专利及其发明人的情况如下，不涉及贺琳为发明人的情形：

序号	专利名称	申请人	专利类型	申请号	专利申请日	发明人
1	一种基于语音识别的中文语音校对方法和装置	发行人	发明	201910797082.X	2019年8月27日	孟君、郝玉峰、曹琼
2	端到端的语音合成方法、装置及存储介质	发行人	发明	202010007126.7	2020年1月3日	邵志明、郝玉峰、黄宇凯、曹琼、李科、宋琼
3	图像标注方法、图像标注装置及计算机存储介质	发行人	发明	202010078873.X	2020年2月3日	刘杰辰、陈佃文、曹琼、郝玉峰、黄宇凯、李科
4	图像分割方法、装置、电子设备及计算机存储介质	发行人	发明	202010089863.6	2020年2月13日	陈佃文、曹琼、郝玉峰、黄宇凯、李科
5	基于线性预测残差负熵的语音音质度量评价方法及装置	发行人、清华大学	发明	202010659644.7	2020年7月10日	李科、张卫强、黄宇凯、郝玉峰、宋琼、廖晓玲
6	基于识别模型的识别方法、模型训练方法及装置	发行人、清华大学	发明	202010659647.0	2020年7月10日	赵泽宇、李科、黄宇凯、郝玉峰、邵志明、张卫强

（二）贺琳不是发行人计算机软件著作权的创作人员

根据发行人确认，发行人及其子公司已取得及正在申请中的计算机软件著作权的主要创作人为唐涤飞、李科、黄宇凯、郝玉峰、曹琼、邵志明、郝志峰、王涛、陈兆全、

李虹辰、董立、王越、司尧、闫启伟、孙文涛、张哲玮、孟君、杨福星等，不涉及贺琳作为创作人。

发行人及其子公司目前拥有 117 项计算机软件著作权，具体见招股说明书“第六节 业务与技术/五、发行人的主要固定资产和无形资产/（二）主要无形资产”部分所述。

发行人及其子公司目前正在申请的计算机软件著作权情况如下：

序号	计算机软件著作权名称	申请人	首次发表日期	填报申请日期
1	中文文本易读性打分软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
2	海天瑞声中考作文打分标注软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
3	海天瑞声移动版调查问卷软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
4	海天瑞声线上 MOS 评测软件 V1.3.1	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
5	海天瑞声图形图像语义分割校对软件 V0.1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
6	海天瑞声双语对齐检测软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
7	海天瑞声手机录音实时检测后端 ASR 服务软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 8 日
8	海天瑞声日语手写体采集标注软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日
9	海天瑞声多音字标注软件 V1.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日
10	海天瑞声安卓版 ASR 检测版在线录音软件 V3.0.0.3	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日
11	海天瑞声 IOS 版 ASR 检测版在线录音软件 V1.0.5	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日
12	海天瑞声安卓版 OCR 图片采集软件 V1.0.9	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日
13	海天瑞声语音合成在线校对平台 V4.0.0.0	发行人	2020 年 4 月 30 日	2020 年 5 月 6 日

（三）发行人的相关专利发明及计算机软件著作权（包括在申请过程中的）为发行人自主研发形成

1、发行人相关专利发明及计算机软件著作权系发行人的研发团队自主研发形成

发行人自创立以来一直重视技术人才的引进，并持续推进技术研发。随着核心技术人员郝玉峰、李科、唐涤飞加入发行人，发行人开始逐步在训练数据生产的平台工具、基础研究、生产技术等方面进行深度的技术研发。发行人核心技术人员带领研发和业务人员利用发行人的物质和技术条件，结合自身专业能力及在发行人工作期间积累的实践

经验，按照业务经营中对训练数据生产相关的具体需求持续进行自主研发，不断申请计算机软件著作权和专利，逐步形成发行人核心技术体系，发行人享有相关研发成果的知识产权。

2、发行人发明及创作人员具备相关技术研发能力

发行人专利发明及计算机软件著作权（包括在申请中的）的发明及创作人员具备相关发明创作的专业知识、技术储备和从业经验，其中主要人员相关情况如下：

序号	姓名	职务	教育背景	行业相关经验	工作背景情况
1	唐涤飞	董事、总经理	中国科学技术大学通信与电子系统专业工学硕士	20 余年人工智能及软件行业从业经验	曾经任职于联想集团技术中心、微软中国研发中心、阿里巴巴云计算公司，长期在语音语义、数据应用和移动平台等技术、产品、服务方面进行研发。
2	李科	董事、副总经理	清华大学信息与信号处理专业硕士	10 余年人工智能及软件行业从业经验	曾任职于 IBM 中国技术开发中心从事软件技术开发。
3	黄宇凯	董事、副总经理、技术总监	复旦大学计算机科学与技术专业硕士	10 余年人工智能及软件行业从业经验	曾任职于微软（中国）有限公司从事软件技术开发。
4	郝玉峰	技术总监、首席科学家	东南大学生物医学工程专业博士	10 余年人工智能及软件行业从业经验	多年从事语音技术相关的研究工作。现任中国中文信息学会语音信息专业委员会委员、中国人民大学信息学院工程硕士导师。
5	曹琼	计算机视觉业务部经理	北京航空航天大学检测技术与自动化装置专业博士	10 余年人工智能及软件行业从业经验	曾任职于富士通研究开发中心有限公司，从事计算机视觉数据算法研究相关工作。
6	邵志明	高级研发工程师	北京邮电大学模式识别专业硕士	5 年以上人工智能及软件行业从业经验	曾任职于百度公司，从事语音识别研究相关工作。
7	宋琼	语音识别采集业务部总监	北京交通大学信号与信息处理专业硕士	10 余年人工智能及软件行业从业经验	曾任职于北京中软国际信息技术有限公司，从事语音识别算法研究、语音识别数据库开发等。
8	廖晓玲	语音合成业务部总监	天津大学计算机专业硕士	10 余年人工智能及软件行业从业经验	曾任职于阿里巴巴云计算事业部，从事语音合成算法研究、语音合成数据库开发等。
9	孟君	高级研发工程师	北京信息科技大学计算机应用技术专业硕士	5 年以上人工智能及软件行业从业经验	从事语音信号处理、语音识别等技术研究。

序号	姓名	职务	教育背景	行业相关经验	工作背景情况
10	陈兆全	高级研发主管	大连大学计算机科学与技术专业学士	10年以上软件行业从业经验	曾任职于北京神州泰岳科技有限公司从事软件开发。

（四）中科院声学所的书面确认

中科院声学所于 2019 年 4 月 3 日出具书面确认：“本所与北京海天瑞声科技有限公司之间不存在知识产权争议或纠纷。”

中科院声学所于 2019 年 5 月 15 日出具书面确认：“声学所知悉贺琳创办海天瑞声公司，并在其中兼职。贺琳在职期间在海天瑞声公司兼职及之后的离职符合声学所相关规定，未损害声学所权益，无遗留问题。”

根据 2020 年 8 月 12 日中介机构对中科院声学所主管人员的访谈，上述 2019 年 4 月 3 日中科院声学所书面确认后发行人新取得或申请的专利及计算机软件著作权和中科院声学所没有冲突，没有争议和纠纷。

根据发行人确认、中国裁判文书网 (<http://wenshu.court.gov.cn/>)、人民法院公告网 (<http://rmfygg.court.gov.cn/>)、中国执行信息公开网 (<http://zxgk.court.gov.cn/>) 等网络公开查询检索及走访发行人住所地人民法院，发行人不存在与核心技术相关的争议或纠纷。

根据贺琳承诺，发行人相关专利发明及计算机软件著作权（包括在申请过程中的）不属于贺琳在中科院声学所的职务发明创造或职务作品，如发行人因相关专利发明及计算机软件著作权与中科院声学所发生争议，贺琳将承担因此而给发行人造成的全部相关损失。

综上，发行人目前的专利发明及计算机软件著作权（包括在申请过程中的）均为相关专利发明人或著作权主要创作人在发行人工作过程中自主研发的职务成果，不属于贺琳在中科院声学所的职务发明创造或职务作品，不存在纠纷或潜在纠纷。

二、未将实际控制人贺琳认定为核心技术人员的原因

贺琳作为公司实际控制人、董事长，主要行使董事长职责并具体负责战略规划制订、重大决策研究、重要外部关系管理、融资管理等事项，不参与具体技术研发，并非发行人知识产权的发明人或创作人，因此未将贺琳认定为核心技术人员。

发行人核心技术人员为唐涤飞、李科、黄宇凯、郝玉峰、曹琼、邵志明，上述人员在发行人核心技术自主研发的过程中起到重要作用。上述核心技术人员对发行人研发的具体贡献请见招股说明书“第六节 业务与技术/六、发行人的研发和技术/（九）研发人员和核心技术人员情况”所述。

（3）贺琳与唐涤飞以及其他股东之间是否存在签署一致行动协议的情形，发行人是否由二人共同控制

一、根据贺琳、唐涤飞以及其他股东签署的《确认函》或《股东调查表》，贺琳与唐涤飞以及其他股东之间不存在签署一致行动协议的情形。

二、发行人的实际控制人为贺琳，非由贺琳、唐涤飞二人共同控制，具体原因如下：

（一）贺琳实际控制的发行人股份数量远高于唐涤飞，所享有的表决权足以对股东大会的决议、人员的任免、重大事项提案的提出和审议通过产生重大影响

贺琳目前直接持有及通过中瑞安合计控制发行人 13,623,853 股股份，占发行人股份总数的 42.44%，并通过中瑞立间接持有发行人 110,092 股股份，占发行人股份总数的 0.34%，担任发行人董事长。

唐涤飞目前直接持有发行人 3,577,982 股股份，占发行人股份总数的 11.15%，唐涤飞持股 50%的创慧科瑞通过中瑞立间接持有发行人 550,459 股股份，占发行人股份总数的 1.71%，并担任发行人总经理。

贺琳直接及通过中瑞安间接控制发行人的股份比例在 40% 以上，实际控制的发行人股份数量远高于唐涤飞，所享有的表决权足以对股东大会的决议、人员的任免、重大事项提案的提出和审议通过产生重大影响。

（二）贺琳为发行人创始股东，在发行人公司创立和发展中起到重大作用

2005 年 5 月贺琳创立海天瑞声有限，贺琳持有 90% 股权并担任执行董事、经理，是公司创始股东；2005 年设立至 2008 年期间，发行人业务规模总体较小，贺琳全面负责具体经营管理事务。2009 年至 2012 年期间，由于郝玉峰、李科先后加入，贺琳作为

执行董事、经理主要负责重大经营管理事项，李科、郝玉峰等人具体负责产品开发、技术研发、业务开拓等日常管理事项。

2013年3月唐涤飞加入海天瑞声有限，贺琳考虑到海天瑞声有限经营管理事务增多且其精力有限，2013年11月任命唐涤飞代替贺琳担任公司经理，并在此之后唐涤飞一直担任公司总经理，主要负责公司日常经营管理事项，包括日常业务运营、项目管理、人员管理等；贺琳一直担任公司执行董事、董事长，具体负责公司战略规划制订、重大决策研究、重要外部关系管理、融资管理、公司治理机制完善等重大事项，并对高级管理人员的履职情况进行监督，对重大经营管理事项提出意见。

贺琳作为实际控制人、董事长，对发行人重大经营管理事项决策和重要人员提名起到主导作用。报告期内发行人重大经营管理事项，包括2017年加大计算机视觉业务发展力度、2019年引进投资者等，均由实际控制人、董事长贺琳主导决策。股份公司董事中，股份公司设立时七名董事由海天瑞声有限董事会提名，具体人选由贺琳提出，股份公司设立后新增两名董事由控股股东贺琳提名；有限公司董事贺琳、唐涤飞、志鹏由贺琳提名。高级管理人员中的总经理、董事会秘书由贺琳提名。

报告期内发行人管理团队，除了总经理唐涤飞之外，发行人副总经理李科主要分管业务开拓和产品开发，副总经理及技术总监黄宇凯主要分管技术研发，技术总监及首席科学家郝玉峰主要分管语言拓展和数据质检，副总经理、财务负责人、董事会秘书吕思遥主要分管财务管理、公共关系管理及证券事务。上述高级管理人员分工合作，共同管理公司日常经营管理事务。

（三）公司实际经营管理中，贺琳、唐涤飞依据各自意愿行使表决权，不存在一致行动协议或共同控制的其他安排

根据发行人的《公司章程》等内部治理制度，发行人对股东大会、董事会审议事项不存在表决权的特殊安排，关于发行人实际控制权不存在特殊约定。发行人股东之间包括贺琳、唐涤飞之间不存在一致行动协议或共同控制的其他安排。

最近2年内，发行人公司治理结构健全且运行良好，贺琳、唐涤飞在公司董事会、股东大会决策过程中依据各自意愿进行表决，不存在一致行动的情形。贺琳以其所享有的表决权对股东大会的决议、人员的任免、重大事项提案的提出和审议通过产生了重大影响。

综上，发行人并非由贺琳、唐涤飞二人共同控制，发行人实际控制人为贺琳。

(4) 发行人股东中是否存在为中科院及声学所的相关员工代持股份的情形

根据发行人全体股东确认，其系发行人的股份的真实持有人，其及其直接或间接出资人不存在受托代其他任何自然人、公司、企业或其他组织持有发行人的股份的情况，其所持发行人股份不存在重大权属纠纷。

综上，发行人股东中不存在为中科院及中科院声学所的相关员工代持股份的情形。

中介机构核查事项

(1) 核查程序

保荐机构、发行人律师主要履行了如下核查程序：

- 1、查询相关法律法规及中国科学院的相关规定；
- 2、取得了中国科学院的证明以及中国科学院声学研究所的确认函；
- 3、访谈了中国科学院声学所的主管工作人员；
- 4、查阅发行人向贺琳发放报酬、分红及缴纳个人所得税的流水记录和凭证；
- 5、网络公开查询检索及走访发行人住所地人民法院，查询其技术成果争议纠纷情况；
- 6、查阅发行人及其子公司的专利证书、软件著作权证书，通过专利查询系统平台查询发行人及其子公司的专利（包括申请中的专利），取得发行人申请中的软件著作权文件及创作人名单，查阅发明人及创作人简历；
- 7、访谈发行人部分核心技术人员和软件著作权主要创作人员，了解其在发行人进行技术研发的相关情况；
- 8、访谈发行人实际控制人贺琳，了解贺琳创办海天瑞声有限的背景和公司发展历程；
- 9、查阅发行人股东填写签署的《股东调查表》、《确认函》，查阅发行人公司章程、工商变更档案；
- 10、查阅发行人公司章程及内部治理制度，自设立以来的董事会（包括各专门委员会）、股东大会文件，董事、高级管理人员提名函，股东临时提案等公司治理文件。

(2) 核查结论

经核查，保荐机构、发行人律师认为：

1、基于国家和中国科学院支持科研人员创业的政策背景，贺琳在中科院声学所任职期间创办海天瑞声有限并任职。贺琳在中科院声学所任职期间未从发行人处领取薪酬，所获取的分红不属于兼职收入，根据中科院相关规定无需就兼职收入事宜进行备案。除就兼职事项未与中科院声学所签署协议外，符合相关法律法规及中国科学院的相关规定。贺琳与中科院声学所虽未就兼职事项签署协议，但中科院出具书面证明明确中科院声学所对该事项具有管理权限，中科院声学所出具书面确认同意其兼职，确认贺琳兼职未影响其履行中科院声学所岗位职责，未损害中科院声学所权益，因此不会对发行人本次发行上市构成障碍。贺琳不存在利用任职单位的资源或便利、相关技术成果创办海天瑞声的情形，未违反中科院的规章制度，未违反其与中科院声学所的聘用合同。

2、发行人的相关专利发明及计算机软件著作权中（包括在申请过程中的）均为发行人自主研发形成，不属于贺琳在中国科学研究所的职务发明创造或职务作品，不存在纠纷或潜在纠纷；贺琳作为公司实际控制人、董事长，主要负责战略规划制订、重大决策研究、重要外部关系管理、融资管理、公司治理机制完善等事项，不参与具体技术研发，并非发行人知识产权的发明人或创作人，所以未将贺琳认定为核心技术人员。

3、贺琳与唐涤飞以及其他股东之间不存在签署一致行动协议的情形，发行人并非由二人共同控制，贺琳为发行人的实际控制人。

4、发行人股东中不存在为中科院及声学所的相关员工代持股份的情形。

2. 关于置换出资

根据申请材料，2014年3月贺琳及其配偶蔡惠智以非专利技术“基于互联网的数据处理平台技术”增资400万元，2015年4月以货币出资400万元置换了专利技术出资400万元。

请发行人说明变更出资方式的原因，是否涉及非专利技术纠纷、是否会影响发行人的正常经营。

请保荐机构及发行人律师对上述事项进行核查，并发表明确意见。

答复：

发行人说明事项

(1) 变更出资方式的原因

基于根据业务需求提高注册资本的需要，2014年3月海天瑞声有限股东贺琳、蔡惠智将非专利技术“基于互联网的数据处理平台技术”作为出资向海天瑞声有限增资，贺琳、蔡惠智以该非专利技术的出资金额分别为360万元、40万元。2015年1月，贺琳、蔡惠智各自将其持有的海天瑞声有限知识产权出资40万元、40万元转让给唐涤飞。

2014年3月上述非专利技术增资后，发行人将该非专利技术应用于公司数据加工业务中。由于该领域技术更新迭代速度较快，出于谨慎角度考虑，为保证出资价值，并避免存在出资不实的风​​险，2015年4月贺琳、唐涤飞分别向海天瑞声有限新增投入货币出资320万元、80万元，将无形资产出资方式变更为货币出资。

北京东审会计师事务所有限公司于2015年5月19日出具东审字[2015] 04-239号《验资报告》，确认截至2015年4月10日，海天瑞声有限已收到贺琳置换的注册资本320万元，唐涤飞置换的注册资本80万元。

(2) 是否涉及非专利技术纠纷

根据发行人及贺琳、蔡惠智、唐涤飞的确认，上述变更出资方式不涉及非专利技术纠纷。

根据中国裁判文书网（<http://wenshu.court.gov.cn/>）、人民法院公告网（<http://rmfygg.court.gov.cn/>）、中国执行信息公开网（<http://zxgk.court.gov.cn/>）等网络公开查询检索及走访发行人住所地人民法院，发行人不存在与上述非专利技术相关的争议或纠纷。

(3) 是否会影响发行人的正常运营

2015年4月发行人变更出资方式后，贺琳、唐涤飞同意该非专利技术仍归发行人所有由发行人继续使用。发行人后续基于该非专利技术进行研发创新，将其应用于发行人数据生产流程中。因此，变更出资方式并不影响发行人的正常运营。

综上，由于非专利技术所属领域技术更新迭代速度较快，出于谨慎角度考虑，为保证出资价值，并避免存在出资不实的风​​险，发行人股东决定变更出资方式，该事项不涉及非专利技术纠纷，不会影响发行人的正常经营。

中介机构核查事项

(1) 核查程序

保荐机构、发行人律师主要履行了如下核查程序：

- 1、查阅海天瑞声有限 2014 年 3 月非专利技术出资至 2015 年 4 月货币置换非专利技术出资期间的工商变更档案文件、验资报告等；
- 2、就该非专利技术出资及置换情况，与贺琳、唐涤飞进行访谈；
- 3、取得发行人及贺琳、蔡惠智、唐涤飞的确认函；
- 4、通过网络公开查询检索以及走访发行人所在地人民法院，查询发行人是否存在非专利技术纠纷。

(2) 核查结论

经核查，保荐机构、发行人律师认为：

由于非专利技术所属领域技术更新迭代速度较快，出于谨慎角度考虑，为保证出资价值，并避免存在出资不实的风​​险，发行人股东决定变更出资方式，该事项不涉及非专利技术纠纷，不会影响发行人的正常经营。

3. 关于整体变更

招股说明书披露，发行人系由海天瑞声有限公司于 2017 年 11 月整体变更设立的股份有限公司。

请发行人说明整体变更时相关股东个人所得税的缴纳情况，是否符合税收法律法规规定。

请发行人律师对上述事项进行核查，并发表明确意见。

答复：

发行人说明事项

(1) 整体变更时相关股东个人所得税的缴纳情况，是否符合税收法律法规规定

发行人整体变更时全部发起人股东及其涉及的个人所得税缴纳情况如下：

序	股东姓名/名	个人所得税缴纳依据及相关缴纳情况
---	--------	------------------

号	称	
1	贺琳	根据《财政部、国家税务总局关于将国家自主创新示范区有关税收试点政策推广到全国范围实施的通知》（财税[2015]116号）：“自2016年1月1日起，全国范围内的中小高新技术企业以未分配利润、盈余公积、资本公积向个人股东转增股本时，个人股东一次缴纳个人所得税确有困难的，可根据实际情况自行制定分期缴税计划，在不超过5个公历年度内（含）分期缴纳，并将有关资料报主管税务机关备案。” 发行人整体变更时属于上述通知规定的中小高新技术企业，符合个人股东分期缴纳个人所得税的条件。
2	唐涤飞	2017年12月9日，贺琳、唐涤飞已取得于北京市海淀区地方税务局第四税务所备案的《<个人所得税分期缴纳备案表（转增股本）>及其填报说明》，贺琳、唐涤飞应于2021年度内分别缴纳其因发行人整体变更而应缴纳的个人所得税1,355,945.27元、559,596.07元，此前年度内无需缴纳。
3	中瑞安	合伙人中的自然人尚未缴纳个人所得税。
4	中瑞立	合伙人中的自然人尚未缴纳个人所得税。
5	清德投资	根据《财政部、国家税务总局关于合伙企业合伙人所得税问题的通知》（财税[2008]159号），“合伙企业以每一个合伙人为纳税义务人。合伙企业合伙人是自然人的，缴纳个人所得税。” （1）清德投资2019年12月转让的1,255,454股股份对应的自然人合伙人因整体变更应缴纳的个人所得税已缴纳完毕 2019年12月，清德投资向中移投资转让1,255,454股股份，股份转让价款为76,059,618元，清德投资就上述股份转让已为自然人合伙人缴纳个人所得税，以上述股份转让收入76,059,618元扣减转让股份对应的整体变更前原始投资成本9,578,785元后的金额作为股份转让收益计算自然人合伙人应缴纳的个人所得税，缴纳金额一并包含了该部分股份对应的自然人合伙人因整体变更应缴纳的个人所得税。 （2）清德投资目前所持2,545,463股股份对应的自然人合伙人因整体变更应缴纳的个人所得税尚未缴纳
6	上海丰琬	（1）上海丰琬2019年12月转让的927,424股股份对应的自然人合伙人因整体变更应缴纳的个人所得税已缴纳完毕 2019年12月，上海丰琬向中移投资转让927,424股股份，股份转让价款为56,186,453元，上海丰琬就上述股份转让已为自然人合伙人缴纳个人所得税，以上述股份转让收入56,186,453元扣减转让股份对应的整体变更前原始投资成本7,877,412元后的金额作为股份转让收益计算自然人合伙人应缴纳的个人所得税，缴纳金额一并包含了该部分股份对应的自然人合伙人因整体变更应缴纳的个人所得税。 （2）上海丰琬目前所持1,880,374股股份对应的自然人合伙人因整体变更应缴纳的个人所得税尚未缴纳
7	上海兴富	（1）上海兴富2019年12月转让的652,576股股份对应的自然人合伙人因整体变更应缴纳的个人所得税已缴纳完毕 2019年12月，上海兴富向中移投资转让652,576股股份，股份转让价款为39,535,216元，上海兴富就上述股份转让已为自然人合伙人缴纳个人所得税，以上述股份转让收入39,535,216元扣减转让股份对应的整体变更前原始投资成本6,262,196元后的金额作为股份转让收益计算自然人合伙人应缴纳的个人所得税，缴纳金额一并包含了该部分股份对应的自然人合伙人因整体变更应缴纳的个人所得税。 （2）上海兴富目前所持1,323,112股股份对应的自然人合伙人因整体变更应缴纳的个人所得税尚未缴纳

8	杭州银杏数	<p>(1) 杭州银杏数 2019 年 12 月转让的 262,576 股股份对应的自然人合伙人因整体变更应缴纳的个人所得税已缴纳完毕</p> <p>2019 年 12 月, 杭州银杏数向中移投资、芜湖青和、芜湖博信分别转让 7,576、165,062、89,938 股股份, 股份转让价款分别为 45.8963 万元、1,000 万元、544.875 万元, 杭州银杏数就上述股份转让已为自然人合伙人缴纳个人所得税, 以上述股份转让收入 1,544.875 万元扣减转让股份对应的整体变更前原始投资成本 4,293,944 元后的金额作为上述股份转让收益, 并汇总其他投资收益后综合计算 2019 年度自然人合伙人应缴纳的个人所得税, 缴纳金额一并包含了该部分股份对应的自然人合伙人因整体变更应缴纳的个人所得税。</p> <p>(2) 杭州银杏数目前所持 532,378 股股份对应的自然人合伙人因整体变更应缴纳的个人所得税尚未缴纳</p>
9	天津金星	为有限责任公司, 不涉及缴纳个人所得税
10	杭州士兰	

根据中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数的税务合规证明, 上述合伙企业未受到税务部门的行政处罚。

针对上述中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数的自然人合伙人未就发行人整体变更缴纳完毕个人所得税的情形, 中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数已出具承诺: “若未来因北京海天瑞声科技股份有限公司(以下简称“海天瑞声”)整体变更纳税事宜本企业收到税务主管部门催缴或代扣代缴通知的, 本企业将根据税务主管部门要求进行纳税或代扣代缴, 保证不因该等纳税事宜影响本企业作为海天瑞声股东的资格, 确保不对海天瑞声本次申请首次公开发行股票并在科创板上市构成不良影响或审核障碍。”

综上, 发行人整体变更时, 自然人股东贺琳、唐涤飞已按规定办理了个人所得税分期缴纳备案, 符合相关法律法规的规定; 中瑞安、中瑞立的自然人合伙人尚未缴纳发行人整体变更的个人所得税; 清德投资、上海丰琬、上海兴富、杭州银杏数自然人合伙人已缴纳 2019 年 12 月转让股份对应的发行人整体变更的个人所得税, 尚未缴纳目前所持股份对应的发行人整体变更的个人所得税; 中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数未受到税务部门的行政处罚; 中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数已出具承诺, 将根据税务主管部门要求进行纳税或代扣代缴, 保证不对发行人本次发行上市构成不良影响或审核障碍; 因此上述个人所得税缴纳情况不会对发行人本次发行上市构成障碍。

中介机构核查事项

（1）核查程序

- 1、查阅发行人整体变更时的工商登记档案资料；
- 2、查询有关整体变更个人所得税的法律法规；
- 3、查阅经北京市海淀区地方税务局第四税务所备案的《〈个人所得税分期缴纳备案表（转增股本）〉及其填报说明》，并核查发行人是否符合中小高新技术企业的条件；
- 4、与清德投资、上海丰琬、上海兴富、杭州银杏数工作人员了解其为自然人合伙人缴纳 2019 年 12 月股份转让涉及个人所得税的计算过程，查阅其税款缴纳凭证；
- 5、取得中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数的税务合规证明及承诺函。

（2）核查结论

经核查，发行人律师认为：

发行人整体变更时，自然人股东贺琳、唐涤飞已按规定办理了个人所得税分期缓缴备案，符合相关法律法规的规定；中瑞安、中瑞立的自然人合伙人尚未缴纳发行人整体变更的个人所得税；清德投资、上海丰琬、上海兴富、杭州银杏数自然人合伙人已缴纳 2019 年 12 月转让股份对应的发行人整体变更的个人所得税，尚未缴纳目前所持股份对应的发行人整体变更的个人所得税；中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数未受到税务部门的行政处罚；中瑞安、中瑞立、清德投资、上海丰琬、上海兴富、杭州银杏数已出具承诺，将根据税务主管部门要求进行纳税或代扣代缴，保证不对发行人本次发行上市构成不良影响或审核障碍；因此上述个人所得税缴纳情况不会对发行人本次发行上市构成障碍。

4. 关于最近一年新增股东

招股说明书披露，本次申报前一年内公司新增中移投资、中网投、芜湖青和与芜湖博信四家股东。中移投资作为发行人股东，受国有独资企业中国移动通信集团有限公司控制，尚未取得国有资产监督管理机构关于国有股权管理的相关批复。

请发行人披露：（1）上述国有股权管理批复的取得进展情况，预计取得时间，是否存在法律障碍；（2）除中移投资外，是否存在其他需要取得国有股权批复的国有股东。

请保荐机构、发行人律师就上述事项以及新股东与发行人其他股东、董事、监事、高级管理人员、本次发行中介机构负责人及其签字人员是否存在亲属关系、关联关系、委托持股、信托持股或其他利益输送安排，新股东与发行人或相关股东是否签署相关对赌协议等进行核查，并发表明确意见。

答复：

发行人披露事项

(1) 上述国有股权管理批复的取得进展情况，预计取得时间，是否存在法律障碍

公司已于招股说明书“第五节 发行人基本情况/八、发行人股本情况/（二）本次发行前的前十名股东情况”及“第五节 发行人基本情况/七、持有发行人 5%以上股份的主要股东及实际控制人情况/（二）其他持有本公司 5%以上股份的主要股东的基本情况/2、中移投资”补充披露：

2020年7月15日，发行人股东中移投资已取得国务院国有资产监督管理委员会核发的《关于北京海天瑞声科技股份有限公司国有股东标识有关事项的批复》（国资产权[2020]324号），“如海天瑞声发行股票并上市，中移投资在证券登记结算公司设立的证券账户应标注“CS”标识。”

(2) 除中移投资外，是否存在其他需要取得国有股权批复的国有股东

根据《关于进一步明确非上市股份有限公司国有股权管理有关事项的通知》（国资厅产权〔2018〕760号）规定，“三、非上市股份有限公司拟首次在证券交易所申请发行股票并上市，其股东符合《上市公司国有股权监督管理办法》（国资委 财政部 证监会令第36号）第三条和第七十四条所规定情形的，由国有资产监督管理机构进行标识管理。”

根据《上市公司国有股权监督管理办法》规定：“第三条 本办法所称国有股东是指符合以下情形之一的企业和单位，其证券账户标注“SS”：（一）政府部门、机构、事业单位、境内国有独资或全资企业；（二）第一款中所述单位或企业独家持股比例超过 50%，或合计持股比例超过 50%，且其中之一为第一大股东的境内企业；（三）第二款中所述企业直接或间接持股的各级境内独资或全资企业。”，“第七十四条 不符合本办法规定的国有股东标准，但政府部门、机构、事业单位和国有独资或全资企业通过投资关系、协议或者其他安排，能够实际支配其行为的境内外企业，证券账户标注为

“CS”，所持上市公司股权变动行为参照本办法管理。”，“第七十八条 国有出资的有限合伙企业不作国有股东认定，其所持上市公司股份的监督管理另行规定。”

发行人目前共有 14 名股东，除中移投资外，还有自然人股东 2 名，合伙企业股东 9 名，法人股东 2 名。根据上述《上市公司国有股权监督管理办法》第七十八条规定，国有股东认定不涉及有限合伙企业；法人股东 2 名中，天津金星系香港联合交易所上市公司小米集团（股票代码：HK01810）控制的公司，实际控制人为雷军；杭州士兰最终权益持有人均为自然人。

因此，除中移投资外，发行人不存在其他需要取得国有股权批复的国有股东。公司已于招股说明书“第五节 发行人基本情况/八、发行人股本情况/（二）本次发行前的前十名股东情况”补充披露：“除中移投资外，不存在其他需要取得国有股权批复的国有股东。”

保荐机构、发行人律师核查事项说明

（1）新股东与发行人其他股东、董事、监事、高级管理人员、本次发行中介机构负责人及其签字人员是否存在亲属关系、关联关系、委托持股、信托持股或其他利益输送

经保荐机构、发行人律师核查，并根据四家新股东中移投资、中网投、芜湖青和、芜湖博信签署的《股东调查表》、《确认函》以及发行人其他股东、董事、监事、高级管理人员签署的《调查表》，新股东与发行人其他股东、董事、监事、高级管理人员、本次发行中介机构负责人及其签字人员不存在亲属关系、关联关系、委托持股、信托持股或其他利益输送。

（2）新股东与发行人或相关股东是否签署相关对赌协议

2019 年 12 月 26 日，新股东中移投资、中网投与发行人及其他股东签署《股东协议》，其中约定中移投资、中网投享有回购权、估值调整的对赌相关权利，具体内容如下：

权利条款	主要内容
------	------

回购权	<p>如果(i) 公司未能在成交日后三年内递交上市申请材料；(ii) 创始股东违反法律法规或发生重大诚信问题，对公司合格上市构成实质障碍；(iii) 创始股东或公司严重违反本协议的约定（“实质性违约”，系指包括但不限于：(u)公司或创始股东在本协议中所做的陈述和保证在重大方面不真实、不完整或不准确，或存在重大误导；(v)创始股东从事严重违反本协议中的承诺和义务的行为，包括但不限于直接或间接开展任何竞争性业务、在法律法规规定的法定退休年龄之前未经投资人同意擅自从公司离职等；(w) 创始股东通过关联交易方式转移公司的收入或利润，向其关联方输送利益；(x)因创始股东或公司的不作为，公司连续两年未能召开董事会或股东大会；(y)因目标集团拥有或使用的核心知识产权产生权属纠纷或涉嫌侵权等原因对公司主营业务的开展造成实质障碍；以及(z)创始股东的其他违约行为给目标集团或竞争性业务造成重大损失，且未能在投资人合理要求的时间内纠正违约行为并充分赔偿投资人)；(iv) 公司任何其他股东根据其与公司或/或创始股东的协议约定要求行使回购权或类似权利；以及(v) 如果公司因数据合规方面的合法性出现重大问题或受到重大处罚，从而导致公司无法持续经营，或严重影响公司实现合格上市的，投资人有权于其后的任何时间不时以书面通知要求公司和/或创始股东在投资人发出行使回购权的书面通知后九十日内尽快完成回购或购买，回购价款应当为下列二者孰高者：(x) 届时投资人所持公司股份经独立资产评估确认的公允市场价值；和(y) 投资人实际支付的投资款（受限于本协议项下包括但不限于估值调整条款项下的各项调整），加上投资款根据每年百分之八(8%)的单利计算得出自相关协议项下所定义的成交日（对于投资人在相关交易项下取得的公司股份而言）至回购价款支付完毕之日期间的收益。</p> <p>公司和创始股东同意：公司就上述全部回购事件承担回购义务并作为回购义务人，创始股东仅就前述(i)、(ii)、(iii)、(iv)和(v)项中因创始股东自身原因导致的回购事件和数据合规事件与公司承担连带回购义务并作为回购义务人，且创始股东就回购价款的支付义务以其届时持有的公司股份价值为限。</p>
估值调整	<p>公司和创始股东特此承诺，公司 2019 年度经审计税后净利润不低于人民币七千万（RMB70,000,000）元。</p> <p>各方理解并确认，如公司未完成全部业绩指标，投资人有权选择要求公司对投资人进行现金补偿，其数额为：认购价款 * (1 - (2019 年实际净利润 ÷ 业绩指标))。</p>

除了上述对赌权利条款以外，《股东协议》约定中移投资、中网投还享有随售权、优先认购权、更优惠的条件、优先清算权等特殊权利，并对创始股东及管理层股东股权转让作出限制。上述《股东协议》同时约定，在发行人为合格上市正式提交上市申报文件前的一个工作日或各方协商一致同意的其他日期前，上述约定条款以及其他根据发行人申请合格上市需要而应终止的条款应予终止并停止执行。

中移投资、中网投与发行人、贺琳已于 2020 年 5 月 13 日签署《股东特殊权利终止协议》，约定各方同意上述回购权条款自该协议签署生效之日起自动终止，上述其他条款在发行人为合格上市正式提交申报文件前的一个工作日自动终止，投资人不在任何时间以任何事由向公司、创始股东或其他方主张行使上述条款所规定的特殊权利。因此，上述约定的对赌等特殊权利条款已终止。

综上，新股东中移投资、中网投和发行人及相关股东曾经签署相关对赌协议，中移投资、中网投享有的对赌等特殊权利条款目前已全部终止。

中介机构核查事项

(1) 核查程序

1、查询国有股权管理的相关法律法规，取得《关于北京海天瑞声科技股份有限公司国有股东标识有关事项的批复》（国资产权[2020]324号）；

2、收集、查阅天津金星的公司章程及小米集团2019年度报告，收集、查阅杭州士兰及其股东的公司章程，在国家企业信用信息公示系统查询天津金星、杭州士兰及其股东的股权结构情况；

3、查阅发行人股东签署的《股东调查表》、《确认函》，查阅发行人董事、监事、高级管理人员签署的《调查表》，在国家企业信用信息公示系统查询发行人股东的股权结构情况；

4、查阅发行人《公司章程》，查阅新股东与发行人及相关方签署的《投资协议》、《股份认购协议》、《股份转让协议》、《股东协议》、《股东特殊权利终止协议》等文件。

(2) 核查结论

经核查，保荐机构、发行人律师认为：

1、中移投资已取得国有股权管理批复，如发行人发行股票并上市，中移投资的证券账户应标注“CS”标识；除中移投资外，发行人不存在其他需要取得国有股权批复的国有股东；

2、新股东与发行人其他股东、董事、监事、高级管理人员、本次发行中介机构负责人及其签字人员不存在亲属关系、关联关系、委托持股、信托持股或其他利益输送安排；

3、新股东中移投资、中网投和发行人及相关股东曾经签署相关对赌协议，中移投资、中网投享有的对赌等特殊权利条款目前已全部终止。

5. 关于香港海天瑞声

招股说明书披露，发行人设立子公司香港海天瑞声时未办理境外投资项目备案手续。

请发行人说明：（1）香港海天瑞声设立及增资时履行外汇手续的情况；如未履行，相关法律后果及对本次发行上市的影响；（2）设立时未办理境外投资项目备案手续可能面临的行政处罚及对公司境外业务的影响，是否会影响持续经营。

请保荐机构、发行人律师结合相关法律法规以及《上海证券交易所科创板股票发行上市审核问答》第3条的规定，对上述事项进行核查，并发表明确意见。

答复：

发行人说明事项

（1）香港海天瑞声设立及增资时履行外汇手续的情况；如未履行，相关法律后果及对本次发行上市的影响；

根据《国家外汇管理局关于进一步简化和改进直接投资外汇管理政策的通知》（汇发[2015]13号），取消境内直接投资项下外汇登记核准和境外直接投资项下外汇登记核准两项行政审批事项，改由银行直接审核办理境内直接投资项下外汇登记和境外直接投资项下外汇登记，国家外汇管理局及其分支机构通过银行对直接投资外汇登记实施间接监管。根据《直接投资外汇业务操作指引》，境内机构在以境内外合法资产或权益（包括但不限于货币、有价证券、知识产权或技术、股权、债权等）向境外出资前，应到注册地银行申请办理境外直接投资外汇登记；银行通过外汇局资本项目信息系统为境内机构办理境外直接投资外汇登记手续后，境内机构凭业务登记凭证直接到银行办理后续资金购付汇手续。

香港海天瑞声设立时，海天瑞声有限公司于2016年2月6日取得招商银行股份有限公司北京清华园支行的《业务登记凭证》，业务类型为ODI中方股东对外义务出资。2016年2月22日，海天瑞声有限公司凭上述凭证向香港海天瑞声缴纳出资款50万美元。

2018年7月，发行人对香港海天瑞声增资30万美元。发行人于2018年7月24日取得招商银行股份有限公司北京清华园支行的《业务登记凭证》，业务类型为ODI中

方股东对外义务出资。2018年7月24日，发行人凭上述凭证向香港海天瑞声缴纳增资款30万美元。

发行人已在设立、增资香港海天瑞声时到注册地银行办理境外直接投资外汇登记，凭银行出具的《业务登记凭证》向香港海天瑞声缴纳出资。

综上，香港海天瑞声设立及增资时履行了必要的外汇手续。

(2) 设立时未办理境外投资项目备案手续可能面临的行政处罚及对公司境外业务的影响，是否会影响持续经营。

一、香港海天瑞声境外投资项目备案手续办理情况

海天瑞声有限2015年11月设立香港海天瑞声，投资总额为50万美元，当时已取得北京市商务委员会2016年1月22日核发的《企业境外投资证书》（境外投资证第N1100201600082号），但未办理发改部门境外投资项目备案手续。

发行人后于2018年5月向香港海天瑞声增资30万美元，已取得北京市商务委员会2018年5月23日核发的《企业境外投资证书》（境外投资证第N1100201800230号）及北京市发展和改革委员会2018年5月25日核发的《项目备案通知书》（京发改[2018]1007号）。

二、相关法律法规对于境外投资项目办理备案手续的规定

根据香港海天瑞声设立时有效的《境外投资项目核准和备案管理办法》第七条、第八条、第二十九条，除涉及敏感国家和地区、敏感行业之外的境外投资项目实行备案管理，其中地方企业实施的中方投资额3亿美元以下境外投资项目，由各省、自治区、直辖市及计划单列市和新疆生产建设兵团等省级政府投资主管部门备案；对于投资主体应申请办理核准或备案但未依法取得核准文件或备案通知书而擅自实施的项目，以及未按照核准文件或备案通知书内容实施的项目，一经发现，国家发展改革委将会同有关部门责令其停止项目实施，并提请或者移交有关机关依法追究有关责任人的法律和行政责任。

三、发行人未因香港海天瑞声设立时未办理发改部门备案而此受到行政处罚

2015年11月，发行人设立香港海天瑞声时未办理发改部门备案手续。但香港海天瑞声已经有效登记并持续运营，发行人2018年5月对香港海天瑞声增资时已在北京市

发展和改革委员会办理完毕增资的备案手续，未受到有关部门责令停止项目实施或追究有关责任人责任等相关的行政处罚。

根据保荐机构和发行人律师对北京市发展和改革委员会主管工作人员的访谈，发行人设立香港海天瑞声时未在发改委办理备案手续，发改委未对发行人进行处罚，未责令停止项目实施；香港海天瑞声后续增资事项已在发改委办理备案，发改委同意发行人向该项目追加投资。

四、香港海天瑞声的运营情况

香港海天瑞声自设立以来承担发行人部分境外业务拓展和服务采购等业务。根据君合律师事务所出具的法律意见，报告期内香港海天瑞声有效登记、存续。

综上，发行人设立香港海天瑞声时未办理发改部门境外投资项目备案手续，但香港海天瑞声已经有效登记并持续运营，且香港海天瑞声后续增资事宜已办理发改部门境外投资备案手续，发行人未受到相关部门责令停止项目实施或追究有关责任人相应责任等行政处罚；目前香港海天瑞声仍有效存续，因此，上述情况不会影响发行人的境外业务，不影响发行人的持续经营。

中介机构核查事项

(1) 核查程序

保荐机构、发行人律师主要履行了如下核查程序：

1、查阅香港海天瑞声设立、增资时银行出具的《业务登记凭证》及出资缴纳的银行凭证；

2、查阅外汇管理的相关法律法规，通过国家外汇管理局北京外汇管理部网站查询发行人行政处罚信息；

3、查阅香港海天瑞声设立、增资时的商委和发改委备案文件，查询境外投资发改委备案的相关法律法规，访谈北京市发展和改革委员会主管工作人员；

4、查阅君合律师事务所出具的关于香港海天瑞声的法律意见。

5、结合相关法律法规及《上海证券交易所科创板股票发行上市审核问答》第3条的规定，对香港海天瑞声设立时未办理境外投资项目备案手续是否属于重大违法行为进行核查。

(2) 核查结论

经核查，保荐机构、发行人律师认为：

1、香港海天瑞声设立及增资时履行了必要的外汇手续；

2、发行人设立香港海天瑞声未办理发改部门境外投资项目备案手续，但香港海天瑞声已经有效登记并持续运营，香港海天瑞声后续增资事宜已办理发改部门境外投资备案手续，发行人未受到被责令停止项目实施或追究有关责任人责任等行政处罚，相关事项已经过访谈确认。目前香港海天瑞声仍有效存续，因此不会影响发行人的境外业务，不影响发行人的持续经营。上述事项不属于《上海证券交易所科创板股票发行上市审核问答》第3条规定的重大违法行为。

6. 关于行业地位

招股说明书披露，海天瑞声是我国领先的 AI 训练数据专业提供商，是行业的头部阵营。根据艾瑞咨询发布的《2020 年中国 AI 基础数据服务行业研究报告》，2019 年，以海天瑞声、百度众包为代表的 AI 基础数据服务行业前五大企业的市场份额为 26.2%。

请发行人披露：（1）2019 年 AI 基础数据服务行业前五大企业的市场份额情况；（2）充分分析发行人的竞争劣势。

请发行人说明：（1）“我国领先”“头部阵营”等表述是否有充分依据，如不充分，请删除相关表述；（2）AI 基础数据服务行业的技术壁垒以及处于低集中度竞争的原因，需求方自建基础数据团队是否为行业发展趋势，发行人与需求方在技术水平、数据资源和人才储备上是否具有竞争优势。

答复：

发行人披露事项

(1) 2019 年 AI 基础数据服务行业前五大企业的市场份额情况

根据艾瑞咨询公开发布的 AI 基础数据服务行业报告，以海天瑞声、百度众包为代表的 AI 基础数据服务行业前五大企业的市场份额为 26.2%，但报告并未披露 AI 基础数据服务行业前五大企业中其他企业情况和前五大企业对应的市场份额情况。

经发行人对该研究报告撰写方的咨询确认，获得以下与前五大企业的市场份额相关信息（出于保密要求，研究报告撰写人未提供对应企业名称），公司已于招股说明书“第

六节 业务与技术/二、发行人所处行业的基本情况和竞争状况/（四）发行人产品或服务的市场地位/1、发行人是国内领先的训练数据提供商”补充披露：

企业	对应市场份额占比 (%)
企业 1	11
海天瑞声	8
企业 3	2-3
企业 4	2-3
企业 5	1-2

上述市场份额情况是艾瑞咨询 AI 基础数据服务行业研究团队在对市场规模调研推算的基础上调查得出的，艾瑞咨询 AI 基础数据服务行业研究团队于 2019 年 12 月通过主要供应商（品牌服务商、中小数据供应商、自建数据团队的需求方）、及主要数据需求方（如互联网公司、智能硬件公司、AI 公司、汽车公司）的背对背调研获取，样本的配比与单体样本的选择会对结果产生一定影响。

（2）充分分析发行人的竞争劣势

公司已于招股说明书“第六节 业务与技术/二、发行人所处行业的基本情况和竞争状况/（四）发行人产品或服务的市场地位/4、竞争优势与劣势”补充披露：

一、相比国际大型 AI 基础数据服务公司，发行人资金实力较弱、规模体量较小、覆盖的语种/方言数量相对较少

虽然发行人在国内 AI 基础数据服务行业拥有领先的行业地位，但与国际大型训练数据公司，如 Appen 相比，发行人在营收规模、净利润规模、员工数量等指标上均小于 Appen，在体量上具备一定劣势。具体来看，Appen 2019 年度取得营业收入 5.36 亿澳元，净利润 4,161 万澳元，发行人同期取得营业收入 2.38 亿人民币，净利润 8,159 万人民币。截至 2019 年末，Appen 员工总数为 781 人，而同期发行人员工仅为 142 人，规模相对较小。

此外，Appen 已在 AI 基础数据服务行业深耕 20 余年，相较于发行人而言开展业务的时间更久，语种或方言的覆盖能力更强。截至目前，根据公开信息可查到的 Appen 覆盖的语种或方言数量已达 180 余个，高于发行人覆盖的 150 余个。

二、海外市场的业务拓展、及海外品牌影响力相对较弱

发行人通过优质的产品与服务以及与国内主要下游客户的长期合作，在国内市场获得了较大的品牌知名度，但在海外市场，发行人的品牌影响力相较于国际大型训练数据公司而言仍相对较弱。例如，Appen 已经在包括其所在国在内的全球主要国家和地区开展业务，覆盖的终端人范围已经超过 130 个国家和地区，相比之下，发行人的海外业务仍有进一步的拓展空间。同时，发行人员工、自营场地均主要分布在中国大陆地区，在海外的营销、业务人员、经营网点布局均相对有限，尚未形成较强的海外品牌影响力，在海外市场的竞争力及对海外客户的吸引力具有一定劣势，仍有较大提升空间。

发行人说明事项

(1) “我国领先”“头部阵营”等表述是否有充分依据，如不充分，请删除相关表述

关于发行人“我国领先”、“头部阵营”等表述主要通过同与同行业可比公司对比、公司下游客户及应用产品、自身技术积累、参与制定行业标准、承担重大项目、所获荣誉、资质和奖项等方面体现，具体依据如下：

一、从与同行业可比公司对比的角度

虽然发行人与国际训练数据供应商 Appen 相比，在经营情况、语言能力等方面存在差距，但与国内竞争对手慧听科技、标贝科技在技术实力、关键业务等方面的对比来看，发行人具有竞争优势，具体情况如下：

项目	海天瑞声	慧听科技	标贝科技
技术实力概述	海天瑞声自主开发了“一体化数据处理平台”，在基础研究、平台工具、训练资源开发三个维度下均积累核心技术，有机搭建起发行人的训练数据生产一站式解决方案，将多项具体核心技术整合为发行人特有的核心技术体系。	采用全程质量监控流程，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 评测系统，为客户提供高质量的数据服务。
拥有的自有知识产权训练数据产品数量	600 余个	23 个	101 个
训练数据产品种类	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
语种/方言覆盖能力	150 余个	19 个	19 个

专利储备情况	已取得 22 项专利授权	2 项专利正在申请	已取得 1 项专利授权
计算机软件著作权数量	117 个	13 个	23 个

注：国内竞争对手中百度众包业务以图像类数据为主，故未列入比较。

技术实力方面：在专利领域，发行人及其子公司已取得的专利数量 22 个，显著高于公开信息渠道可查询的慧听科技、标贝科技的专利数量；在计算机软件著作权领域，发行人及其子公司共拥有计算机软件著作权 117 项，远高于公开信息可查询的慧听科技、标贝科技的计算机软件著作权数量。

关键业务方面：发行人在语种方言覆盖能力、拥有的自有知识产权训练数据产品数量方面均高于公开信息可查询的慧听科技、标贝科技的相关数量，具有明显优势。

二、从公司下游客户及应用产品角度

经过多年发展和沉淀，发行人产品和服务已经过市场检验并得到了国内外知名企业的认可。目前，发行人的产品和服务已获得阿里巴巴、腾讯、百度、微软、三星、亚马逊等大型科技公司，科大讯飞、商汤科技、云知声、海康威视等 AI 企业以及中国科学院、清华大学等科研机构的认可，应用于其研发的个人助手、智能音箱、语音导航、搜索服务、智能驾驶、机器翻译等多种 AI 产品相关算法模型的训练过程中。目前发行人客户累计数量超 400 家，覆盖了主要的大型科技公司、AI 企业及科研机构。下游众多行业龙头客户的认可以及应用产品的广泛覆盖印证了公司在 AI 基础数据服务行业的领先地位。

三、从自身技术积累角度

通过多年对训练数据服务领域的深耕，发行人已掌握 12 项核心技术，可以满足 AI 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，发行人的核心技术情况参见招股说明“第六节 业务与技术”之“六、发行人的研发和技术”。截至目前，公司已积累超过 600 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，已积累超过 100 个语种/方言的发音词典，累计词条数超过 1,000 万条。此外，公司已经取得 20 项发明专利、1 项实用新型专利授权及 1 项外观设计专利，117 项计算机软件著作权，另有 6 项专利已提交申请。

四、从参与制定行业标准的角度

发行人现任中国人工智能产业发展联盟理事单位、中国语音产业联盟理事单位，并担任全国信息安全标准化技术委员会成员单位、中国通信标准化协会会员单位、国家人工智能标准化总体组成员单位、中国电子工业标准化技术协会发布的《信息技术-人工智能-面向机器学习的数据标注规程》标准参与起草单位，积极参与并推动各项标准、规范的制定和出台，以专业性和创新性获得了行业的普遍认可。

五、从承担重大项目的角度

2019年，经过多轮多方评议，发行人入围“新一代人工智能产业创新重点任务入围揭榜单位”，开展“人工智能训练与测试数据资源库及服务平台”项目技术攻关工作。为贯彻落实国务院《新一代人工智能发展规划》和工业和信息化部《促进新一代人工智能产业发展三年行动计划（2018-2020）》（以下简称“《三年行动计划》”），工业和信息化部于2018年11月启动“新一代人工智能产业创新重点任务揭榜工作”。围绕《三年行动计划》确定的重点任务方向，揭榜工作主要涵盖智能产品、核心基础、智能制造关键技术装备和支撑体系四大方向，四大方向下囊括多个细分领域，发行人所属的细分领域为支撑体系下的行业训练资源库。

根据工业和信息化部科技司对于《新一代人工智能产业创新重点任务揭榜工作方案》的官方解读文件，人工智能揭榜工作主要特点是优中选优，树标杆育主力，在人工智能主要细分领域，选拔领头羊、先锋队。揭榜工作在每个细分领域择优遴选不超过5家揭榜单位，择优公布揭榜成功单位不超过3家，旨在树立人工智能领域标杆，激发竞争活力、培育行业龙头，推动形成一批具有国际竞争力的人工智能企业。首次评选历时一年，全国参与评选的企业与机构共1,248家，最终137家单位榜上有名，入选率仅为10.98%。除发行人外，同时入选的还有阿里巴巴、腾讯、百度、科大讯飞等知名的人工智能企业与机构。本次入选体现了行业相关部门对公司技术实力和行业地位的高度认可。

六、从所获荣誉、资质和奖项的角度

发行人连续多年荣获“国家高新技术企业”及“中关村高新技术企业”称号，并于2018年起连续被评为国家规划布局内重点软件企业，此外，公司荣膺多项国内外AI领域奖项，充分证明了发行人的行业地位。截至目前，发行人及其产品服务所获奖项、荣誉或认证、称号情况如下：

序号	名称	颁发单位	颁发时间
1	北京市新技术新产品（服务） ¹	北京市科委、北京市发展改革委等六家单位联合颁发	2020年
2	北京市级企业科技研究开发机构	北京市科学技术委员会	2019年
3	新一代人工智能产业创新重点任务入围揭榜单位	工业和信息化部	2019年
4	2019中国信息技术人工智能行业优秀产品	中国信息协会、信息化观察网	2019年
5	2019中国信息技术人工智能数据服务领军企业	中国信息协会、信息化观察网	2019年
6	2019中关村高成长企业TOP100	中关村高新技术企业协会	2019年
7	腾讯2019年服务类供应商“杰出贡献奖”	深圳市腾讯计算机系统有限公司	2019年
8	菜鸟网络“驼峰计划”突出贡献奖	菜鸟网络	2019年
9	GB/T22080-2016/ISO/IEC27001:2013《信息技术 安全技术 信息安全管理体系要求》标准认证	北京赛西认证有限责任公司	2019年
10	中国人工智能产业发展联盟理事单位	中国人工智能产业发展联盟	2018年
11	国家规划布局内重点软件企业	国家发展改革委、工业和信息化部、财政部、商务部、国家税务总局	2018年
12	中关村高新技术企业	中关村科技园区管理委员会	2018年
13	国家高新技术企业	北京科学技术委员会、北京市财政局、北京市国家税务局和北京市地方税务局联合颁布	2018年
14	中关村高新技术企业协会理事单位	中关村高新技术企业协会	2018年
15	信息化影响中国·2018年人工智能行业最佳产品	中国信息协会、信息化和软件服务网	2018年
16	2018“中国声谷”杯智能语音创新大赛行业先锋奖	中国语音产业联盟	2018年
17	2018中关村高成长企业TOP100	中关村高新技术企业协会	2018年
18	中国语音产业联盟理事会员单位	中国语音产业联盟	2017年
19	2017中国信息技术年度人工智能数据服务领域领军企业奖	中国信息协会、信息化观察网	2017年
20	25 Most Recommended Artificial Intelligence Companies 2017（2017年度25家最具推荐价值人工智能企业）	美国 APAC CIO Outlook 杂志	2017年

¹发行人产品/服务：uTrans 在线语音转写标注质检平台服务、King-Lexicon-051 上海话发音词典、King-ASR-384 中文口音语音识别系列数据库获评北京市新技术新产品（服务）称号。

序号	名称	颁发单位	颁发时间
21	Oriental COCODA 2016 ITN Best Paper Award	Oriental COCODA	2016 年

综上所述，发行人关于“我国领先”“头部阵营”等的表述具有充分依据。

(2) AI 基础数据服务行业的技术壁垒以及处于低集中度竞争的原因，需求方自建基础数据团队是否为行业发展趋势，发行人与需求方在技术水平、数据资源和人才储备上是否具有竞争优势

一、AI 基础数据服务行业的技术壁垒

报告期内，发行人主要从事训练数据的研发设计、生产及销售。发行人所处的 AI 基础数据服务行业正处于快速发展阶段，行业技术壁垒正在逐步凸显，主要体现在下述方面：

(一) 对下游 AI 行业发展趋势及需求的洞察力

AI 基础数据服务行业服务的是 AI 产业链下游的各类机构。优质的训练数据提供商需要紧跟业界技术的最新研发成果和技术趋势，充分理解下游客户在前沿算法开发、训练等环节对训练数据的需求，从而设计开发出训练效果更优的训练数据。

优质的训练数据提供商在基础研究等方面通常有所布局，能够更好地把握技术未来的发展方向，从而理解因此衍生的新兴需求，构建了在下游客户专业需求理解能力方面的专业门槛。

(二) AI 基础研究体系中的语言开发能力

训练数据提供商需要建立和完善基础研究体系，从而支撑起从设计到采集、处理、质检的训练数据开发各环节。基础研究中重要的是语言开发能力。AI 应用本质上解决的是语义理解和推理的问题，而语言是语义的主要载体，其携带的信息和价值是通过数据资源开发的形式从基础数据提供商处产生、并传递到下游客户的最终应用场景中去。以智能语音领域为例，不管是语音识别还是语音合成，包括语言模型训练、发音预测等技术的语音语言学的研究都是最关键的技术之一。

优质的训练数据提供商在语言能力积累和新语言开发上具备明显优势。该类提供商对于语言学和 AI 应用有效结合有较为深刻的认识，积累起了不同语言的丰富知识，能

够迅速地整合到开发环节中去。同时，该类提供商也有完整的新语言开发体系，积累了相关工具和算法能力，能够更迅速地响应客户新的语言需求。

（三）训练数据设计能力

训练数据设计环节要求训练数据提供商从应用场景出发，结合算法模型对训练数据的专业要求，设计出高质量的训练数据结构，提供优质的训练数据。

以智能语音训练数据开发为例，如果未能设计较为科学有效的训练数据结构，据此开展采集，那么随机收集的原料数据往往难以覆盖足够多的语言现象，并符合音素均衡等要求。因此，设计开发优质训练数据要求提供商具备丰富的专业储备、较强的语料设计能力，在设计环节能够结合音素均衡、语料主题等各种因素综合进行结构设计，增强语料数据对自然语言现象的覆盖效果，实现以有限数据集覆盖尽可能多的语言现象，达到更好的训练效果。

此外，随着 AI 算法模型和技术应用的发展，其对所需训练数据有了更多专业要求，也要求训练数据产品和服务提供商具备设计及开发提供新品种训练数据的能力。

优质的训练数据提供商通常掌握了较为深厚的专业知识储备，具备覆盖多种领域、场景、语种的训练数据设计能力，更能迅速满足客户的新兴需求，在训练数据设计能力领域构建了技术壁垒和专业门槛。

（四）训练数据开发能力

除前述设计环节外，训练数据开发的其他环节主要包含采集、处理、质检环节，训练数据服务商凭借多年经验，在各环节均建立了壁垒。具体如下：

1、采集环节：需要高效准确地采集多种原料数据，具体涉及采集设备、环境的专业选择以及采集设备的空间布置等；也涉及多通道同时采集、多模态采集、训练数据同步等具体技术。优质的训练数据提供商拥有更丰富的采集渠道和经验积累，有的训练数据提供商自主研发了便于数据采集的软件设备、工具、平台等，并在多通道采集、多模态数据资源采集、数据资源同步技术方面也积累下专业技术，提升了效率和质量，构建了较强的技术壁垒和专业门槛。

2、加工环节：完成采集后，需要对大量原料数据进行校对、标注等处理，最终形成可用于算法模型训练的高质量、结构化训练数据。涉及的具体标注工作根据具体训练

数据对应的领域、类型不同有所差别，标注类型总量可多达几十种。处理环节涉及大量原料数据、数据处理工作和数据处理人员，既要保障数据流转的安全性，又要保证质量和效率。优质的训练数据提供商往往结合自身的专业经验，积累了高效率的数据处理平台、工具集，并将 AI、大数据等专业技术运用于数据处理，具备完成复杂、精确的数据标注处理工作的能力，提升数据处理的自动化程度，能够实现更快、更好的数据处理结果，在训练数据处理方面构建下技术壁垒和专业门槛。

3、质检环节：质检环节是训练数据开发流程中的重要一环。该环节涉及大量的校验和交叉比对计算等工作，对计算性能是一大挑战。优质的训练数据提供商往往设计了科学有效的质检流程及工具平台，并部署了可支撑质检环节计算需求的服务器系统及应用，在该环节构建下技术壁垒和专业门槛。

二、AI 基础数据服务行业处于低集中度竞争的原因

（一）行业处于快速增长期，技术壁垒逐步凸显，中小型数据供应商虽目前广泛存在，行业集中度较低，但未来行业集中度会逐步提升

伴随国内 AI 行业的爆发，大量 AI 企业取得融资，为了不断提高算法精度，数据采标需求也空前增长，催生了 AI 基础数据服务行业的繁荣。早期的 AI 基础数据服务门槛较低，各种类型体量的公司参与其中，使行业集中度较低。未来随着竞争加快，行业技术壁垒的逐渐凸显，AI 公司对训练数据的质量要求也不断提高，有着更强行业需求的洞察力、语言开发能力、训练数据设计和开发能力的公司将逐步占据更多市场份额。并且当产业落地成为主旋律，需求方对垂直场景的定制化数据采标需求成为主流，众多小型 AI 基础数据服务公司在数据质量和采标能力上达不到要求，或被淘汰，或依附大平台，未来行业集中度将会逐渐提高。

（二）AI 基础数据服务行业对于劳动力存在密集性需求，使得中小型数据供应商长期存在，行业集中度较低

AI 基础数据服务行业提供的训练数据本质上是集合了承载人类特征的数据，用于训练 AI 模型，使其对实际场景和事物的判断最大程度接近人类。因此，采集环节是采集人类特征形成原料数据，需要大量的被采集对象；标注环节是将人类对事物的判断赋予在数据上，需要大量的标注人员。这就造成了在采集及标注环节对于劳动力存在密集性需求，所以致力于人力输送的服务形式会持续存在，因此中小型数据供应商仍会占有

一定的市场份额。随着技术壁垒的凸显、客户需求多样化等情况成为常态，下游客户对于训练数据提供商的技术水平要求越来越高，AI 产业链的专业化分工将愈加清晰。作为训练数据提供商，需要对 AI 基础技术，如智能语音、计算机视觉、自然语言处理等算法有更深刻的理解，对行业发展趋势拥有洞察能力，充分发挥在语言研究、数据库设计、数据采标平台及标准开发、质量控制等方面的能力，才能研发出紧跟 AI 技术发展趋势的训练数据。仅能满足密集型劳动力需求的中小型供应商不具备该等核心竞争优势。专业化的品牌训练数据服务商将在 AI 产业链上的扮演更加重要的角色，大部分中小型数据供应商将逐渐远离利润中心，品牌数据服务商阵营将占有市场的主要份额，行业的集中度将逐步提升。

三、需求方自建基础数据团队是否为行业发展趋势

出于对数据安全性、产业链整体布局等的考虑，需求方自建基础数据团队始终存在，但并不会对品牌数据供应商的发展产生挤出效应，品牌数据供应商和需求方自建基础数据团队的共同发展将成为行业发展趋势。

（一）需求方自建基础数据团队增加，但未对行业产生挤出效应

出于对数据安全性、成本和整体布局的考虑，AI 和科技型互联网领域的部分头部公司开始组建自有标注团队，所处理项目从少量较为简单基础的标注需求，逐渐向大量复杂任务发展。但需求方自建基础数据团队并未对市场产生明显挤出效应。其原因如下：

首先，包括大型 AI 和科技型互联网领域的头部公司在内的整个市场训练数据需求总量大，新数据类型、新应用场景的需求，随着下游 AI 行业的蓬勃发展不断涌现，整个 AI 基础数据行业的市场规模迅速增长。

其次，需求方自建基础数据团队主要是为满足 AI 和科技型互联网领域的部分头部公司等需求方自身研发需求而设立，但由于缺乏长期、全面的训练数据设计及加工技术体系、采标资源、数据研发流程管理经验等领域的建设，其自建团队通常只能解决部分自身研发需求，需求方仍然会通过采购品牌数据服务商的服务，满足更全面的自身研发需求。

最后，需求方自建基础数据团队因需求方自身出于数据安全性、成本等考虑而产生，主要服务于需求方自身需求，其自身的数据资源出于数据保密、竞争因素等考虑，通常以自用为主，极少通过分享对外输出，并未能形成较强的应对外部市场多样化需求、对

外提供对应产品和服务的能力。且由于其拥有的数据缺乏分享能力，也就无法通过数据分享提升数据质量形成规模效应，相较于采购品牌数据服务商的服务，相关数据研发成本较高。因此，需求方自建基础数据团队并未广泛出现于大型科技公司和人工智能企业，已经存在的需求方自建基础数据团队很难挤占除需求方自身需求以外的其他市场需求对应的市场空间。

（二）未来品牌数据服务商与需求方自建基础数据团队将共同发展，替代中小型供应商阵营，占有市场的主要份额。同时，品牌数据服务商将更具优势

从 AI 基础数据行业的发展趋势来看，市场竞争日趋激烈，行业处于集中度不断提升的阶段，虽然中小型数据供应商的整体体量仍然可观，但随着行业业务与技术门槛提升、下游数据需求和应用场景的多样化和专业化，更多的中小型数据供应商由于缺乏核心竞争力、人员成本等的压力，将会被淘汰。相比于中小型数据供应商，品牌数据服务商和需求方自建基础数据团队，在能力、管理及技术方面存在竞争优势。品牌数据服务商可以全面应对市场化为需求，需求方自建基础数据团队可以应对部分自身需求。品牌数据服务商和需求方自建基础数据团队将替代中小型供应商，占有市场的主要份额。

同时，与需求方自建基础数据团队相比，面对国内外行业的新需求、新趋势，新兴业务拓展及新成立的 AI 创业公司拉动的对于不同应用场景、不同类型高质量训练数据的增量市场，品牌数据服务商在品牌效益、规模效应、团队建设、专业技术体系建设、资质、服务意识、业务能力等方面均有优势。同时，由于需求方自建基础数据团队服务于需求方自身需求的局限性，品牌数据服务商将在未来的竞争中占有更大的主动性。

综上所述，品牌数据供应商和需求方自建基础数据团队的共同发展将成为行业发展趋势，同时，品牌数据供应商在全面应对市场化需求方面具有优势，将占有较多市场的份额。

四、发行人与需求方在技术水平、数据资源和人才储备上是否具有竞争优势

（一）技术水平

在 AI 基础数据领域，技术含量差距较大，需求方自建基础数据团队虽然对于需求方相关的应用场景及下游 AI 技术较为熟悉，但整个 AI 基础数据市场对于训练数据的类型及应用场景存在多样化的需求，且 AI 基础数据市场对于训练数据具有专业性、精

细化、复杂性数据处理要求。在高质量训练数据领域，需求方自建基础数据团队在专业性、流程管理、质量把控等维度难以直接与包括发行人在内的专业训练数据供应商竞争。在除与自身相关应用场景以外的训练数据需求方面，需求方自建基础数据团队无法熟悉各个应用场景训练数据需求痛点，较难与专业训练数据供应商相竞争。

（二）数据资源

1、发行人积累的训练数据类型和数据内容涉及的应用场景的广度具有明显优势，可迅速响应客户需求

需求方自建基础数据团队拥有的数据更多的涉及需求方相关的应用场景和领域，具有一定的专用性，同时需求方掌握的数据以实际场景的原料数据为主，后续需要进行标注等各种处理才能形成训练数据。而发行人是专业的训练数据产品和服务供应商，掌握的数据以训练数据为主，其拥有及开发的自有知识产权的训练数据产品数量较多，产品和服务涉及的训练数据类型、应用场景更为广泛，拥有根据不同应用场景需求及时响应，提供对应种类训练产品和服务的能力。

2、发行人数据具有共享性，从而可形成规模效应，降低数据投入成本

从行业逻辑上看，发行人为所有下游客户提供训练数据，因而可以借鉴不同客户的需求提升数据质量、扩大数据库的范围，是一种正向促进的循环，但需求方自建基础数据团队形成的数据出于数据保密、竞争因素等考虑，通常以自用为主，缺少了通过分享提高数据质量、扩大数据规模的正向促进循环，因而数据的规模效应差，导致仅通过自建团队的方式进行训练数据方面的投入成本会更大。

（三）人才储备

和需求方自建基础数据团队相比，多年来，通过开展训练数据业务，海天瑞声形成了较为强大的技术人才储备，建立起了包含核心技术人员在内的业务能力突出、行业经验丰富的技术团队。

从核心技术团队角度来看，发行人的6名核心技术人员均为计算机、信息工程、自动化等领域的专业人士，毕业于各大国内知名院校，并在人工智能行业从业多年，经验丰富。其教育背景、行业从业经验、相关专利积累及专业背景情况如下：

姓名	职务	教育背景	行业相关经验	专利积累	专业背景情况
----	----	------	--------	------	--------

姓名	职务	教育背景	行业相关经验	专利积累	专业背景情况
唐涤飞	董事、总经理	学士和硕后就读于中国科学技术大学	20余年人工智能及软件行业从业经验	拥有6项美国、中国专利	加入海天瑞声之前，曾经任职于联想集团技术中心、微软中国研发中心、阿里巴巴云计算公司。拥有20余年人工智能和软件行业的研发和管理经验，领导跨国团队在语音语义、数据应用和移动平台等技术、产品、服务取得重要突破。 在海天瑞声，打造了以基础算法研究、数据处理平台，算法技术与工具平台相结合验证数据库效果为核心的公司技术战略布局，确立了公司的竞争优势。
李科	董事、副总经理	学士和硕后就读于清华大学	10余年人工智能及软件行业从业经验	13项已授权专利和5项在申请专利的共同发明人	加入海天瑞声之前，曾任职于IBM中国技术开发中心。 在海天瑞声主导研发了智能化和高效率的一体化数据处理平台，奠定了公司产品和服务的基础；组织、负责公司知识产权相关工作，带领公司研发、技术团队为公司取得多项专利及软件著作权。
黄宇凯	董事、副总经理、技术总监	学士和硕后就读于上海交通大学和复旦大学	10余年人工智能及软件行业从业经验	13项已授权专利和5项在申请专利的共同发明人	加入海天瑞声之前，曾任职于微软（中国）有限公司。 在海天瑞声，重构了一体化数据处理平台的软件架构；组织开发并改进数据处理各环节所需工具以高效支撑智能语音、计算机视觉、自然语言处理等公司业务；遴选公司人工智能基础算法的研究成果并转化为软件系统使之应用于数据开发业务。
郝玉峰	技术总监、首席科学家	学士、硕士、博士先后就读于东南大学	10余年人工智能及软件行业从业经验	20项已授权专利、6项在申请专利的共同发明人	加入海天瑞声之前，已经多年从事语音技术相关的研究工作。现任中国中文信息学会语音信息专业委员会委员、中国人民大学信息学院工程硕士导师。 在海天瑞声，负责通过语音识别、语音合成、自然语言处理、图像处理等AI技术手段，提高训练数据设计生产的工作效率和质量，任职期间设计研发公司相关产品设计流程、算法、信号处理技术。以上算法和技术被广泛应用于公司数据设计、开发和质量控制流程。
曹琼	计算机视觉业务部经理	学士、硕士、博士先后就读于北京航空航天大学	10余年人工智能及软件行业从业经验	17项已授权专利、4项在申请专利的共同发明人	加入海天瑞声之前，于富士通研究开发有限公司任职近十年。 在海天瑞声，设计并实现使用计算机视觉数据自动检测算法与人工标注结合的人机交互数据生产方案，主导开发了计算机视觉数据处理通用平台，其成果被应用于公司计算机视觉业务中，成为公司竞争优势之一。
邵志明	高级研发工程师	学士、硕士先后就读于邮电大学	5年以上人工智能及软件行业从业经验	5项已授权专利、2项在申请专利的共同发明人	加入海天瑞声之前，曾经任职于百度公司，从事语音识别研发相关工作。 在海天瑞声，从事语音信号处理、语音合成、说话人识别、语种识别、自然语言处理等技术研究，其主导研发的语音多通道对齐技术和ASR唤醒词训练数据自动转写系统，显著提升相应业务的效果和效率。

除核心技术人员外，截止2020年3月底，公司员工数量为142人，其中技术团队人员（包含核心技术人员在内）98人，占比69.01%，涉及研发中心、数据中心、语言研究部、质量保证部，形成了一支完备的技术团队。在技术团队（包含核心技术人员在

内)中,骨干成员 29 人,平均训练数据行业相关从业经验在 7 年以上,硕士以上学历人数占比在 65.52%以上,均为信息与信号处理学、生物计算与人工智能、信息与通信工程、语言学等训练数据研发相关专业背景。技术团队的持续建设为公司实现技术创新提供了关键的人才储备。

此外,公司积极鼓励团队开展科研工作,重视科研队伍建设,并与清华大学、北京语言大学等多所重点院校开展产学研合作与人才培养,构建起积极的人才储备战略,已建立完善的人才梯队和行之有效的人才培养和激励机制,在长期的项目实践中着手培养并储备了一大批既精通专业技术又具备行业经验的复合型人才。

综上所述,与需求方自建基础数据团队比,发行人在技术水平、数据资源和人才储备上方面具有更为专业化的竞争优势。

7. 关于数据安全及其合规性

报告期内,发行人主要从事 AI 训练数据的研发设计、生产及销售业务。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。

请发行人说明:(1)发行人及其原料数据采集供应商相关数据的获取方式及其合规性,发行人是否享有数据的所有权或获得相关数据主体的授权许可,相关授权许可是否存在使用范围、主体或期限等方面的限制,发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形,是否存在数据内容侵犯个人隐私或其他合法权益的风险;(2)《个人信息安全规范》《数据安全法》(草案)等法规的出台对发行人业务的具体影响,发行人及其原料数据采集供应商是否存在违反上述规定的情形,请发行人就相关监管政策出台对行业可能产生的影响进行重大事项提示和风险揭示。

请保荐机构及发行人律师对上述事项进行核查,并发表明确意见。

答复:

发行人说明事项

(1) 发行人及其原料数据采集供应商相关数据的获取方式及其合规性，发行人是否享有数据的所有权或获得相关数据主体的授权许可，相关授权许可是否存在使用范围、主体或期限等方面的限制，发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形，是否存在数据内容侵犯个人隐私或其他合法权益的风险；

一、发行人及其原料数据采集供应商相关数据的获取方式及其合规性，发行人是否享有数据的所有权或获得相关数据主体的授权许可

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。发行人业务经营中使用的数据，主要通过发行人及原料数据采集供应商主动收集获取，在为客户提供数据转写、标注等纯加工定制服务时涉及的数据则由客户自行提供。上述两种方式涉及的数据所有权、授权许可及合规性等情况具体如下：

(一) 发行人及原料数据采集供应商主动收集数据

1、发行人及其原料数据采集供应商相关数据的获取方式

发行人主要通过原料数据采集供应商采集数据，原料数据采集供应商按照发行人的采集要求寻找匹配的终端人员，组织其使用发行人研发的技术、平台、工具等，为发行人提供数据采集服务。在个别业务中，由于采购量较小，作为补充性采购，发行人也存在少量直接采集数据的情形。

结合发行人业务经营中的数据应用领域，各领域涉及的主要数据资源形式及其具体获取方式如下：

应用领域	数据资源形式	数据具体获取方式
智能语音	语音	按照发行人要求的语种/方言类别、语言风格、交流形式等，被采集人根据发行人提供的语料文本内容等进行朗读或自由对话，对被采集人的语音进行录制
计算机视觉	个人人像、视频	被采集人按照发行人要求的姿势、动作、场景等配合作出姿势、动作等，对被采集人进行照片拍摄、视频录制
	其他图像、视频	终端人员按照发行人要求的具体场景、采集内容等，对公开场景、公开物品进行录制拍摄，例如拍摄自动驾驶场景视频
自然语言	个人短信、邮件、手写文本	被采集人按照发行人要求的文本内容、字体样式等提供个人短信、邮件、手写文本
	其他自然语言文本	终端人员按照发行人要求的文本内容，收集提供公开文本内容片段

2、发行人是否享有数据的所有权或获得相关数据主体的授权许可

发行人和原料数据采集供应商主动收集数据，被采集对象为个人的，例如采集语音、人像等，已经事前告知被采集人或其监护人采集内容以及采集目的、方式等规则，征得被采集人或其监护人同意后，由被采集人按发行人具体要求主动提供，因此主动收集数据已取得被采集人或其监护人的授权许可。在被采集人或其监护人授权许可范围内，采集的语音、人像等业务数据的知识产权根据协议约定和业务类型（提供训练数据定制服务或训练数据产品）归属发行人或客户，原料数据采集供应商并不享有相关权益，因此发行人获取数据不需要获得供应商的授权许可。

发行人和原料数据采集供应商主动收集数据，被采集对象不是个人的，例如按照发行人具体要求拍摄公开场景、收集公开文本片段等，不涉及侵犯其他第三方权益，因此无需取得第三方的授权许可。采集数据的知识产权根据协议约定和业务类型（提供训练数据定制服务或训练数据产品）归属发行人或客户，原料数据采集供应商并不享有相关权益，因此发行人获取数据不需要获得供应商的授权许可。

3、相关数据获取方式的合规性

（1）个人信息数据获取的合规性

①个人信息数据获取的相关法律规定

根据《中华人民共和国民法总则》第一百一十一条，“自然人的个人信息受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。”

根据《信息安全技术 个人信息安全规范》（GB/T 35273—2017）5.3 条款，“a）收集个人信息前，应向个人信息主体明确告知所提供产品或服务不同业务功能分别收集的个人信息类型，以及收集、使用个人信息的规则（例如收集和使用个人信息的目的、收集方式和频率、存放地域、存储期限、自身的数据安全能力、对外共享、转让、公开披露的有关情况等），并获得个人信息主体的授权同意；b）间接获取个人信息时：1) 应要求个人信息提供方说明个人信息来源，并对其个人信息来源的合法性进行确认；2) 应了解个人信息提供方已获得的个人信息处理的授权同意范围，包括使用目的，个人信息主体是否授权同意转让、共享、公开披露等。如本组织开展业务需进行的个人信息处理活动超出该授权同意范围，应在获取个人信息后的合理期限内或处理个人信息前，征得个人信息主体的明示同意。”

②获取个人信息数据已经事先明确告知并获得授权同意

被采集对象为个人且涉及采集个人语音、人像等个人信息的，发行人和原料数据采集供应商在数据采集前已明确告知被采集人或其监护人采集内容、目的、方式等规则，并经其授权同意，不存在非法获取的情形，符合上述法律规定。

在采集上述个人语音、人像等业务数据之外，特定项目会对被采集人的年龄、口音区等设定具体要求，发行人和原料数据采集供应商会结合被采集人的身份证明文件等信息进行核验，该等项目采集时被采集人会自愿填写姓名、年龄、性别、身份证号、联系方式等个人身份信息，留存数据采集记录以保障被采集人依法行使撤回授权、删除等个人信息权利。根据发行人内部控制的要求，发行人会将上述个人身份信息中的身份证号、联系方式等个人敏感信息进行去标识化处理。因发行人对业务数据进行加工处理是为AI算法模型的训练拓展提供可靠的训练素材，不需要也不涉及识别到特定个人，发行人将经技术处理后的业务数据提供给客户，并不提供个人身份信息，客户后续使用上述业务数据也无需个人身份信息或识别到特定个人。

(2) 非个人信息数据获取的合规性

发行人和原料数据采集供应商主动收集非个人信息数据的，如按照发行人具体要求拍摄公开场景（例如自动驾驶场景）视频时，发行人已将视频中存在的人像进行去标识化处理；公开收集文本内容片段，系收集通用文本内容片段，再将其中字词进行重新组合。上述情形不涉及侵犯其他第三方的合法权益，不存在非法获取的情形。

(3) 发行人对原料数据采集供应商获取数据的合规控制措施

为了对原料数据采集供应商采集数据进行规范管理，发行人主要采取以下措施：

①通过制定《数据安全管理制度》、与相关供应商签署《数据安全和保密协议》以及在采集前对供应商进行培训等方式，明确要求供应商采集个人信息数据进行事先告知、获得授权许可等合规事项，要求供应商获取数据不得侵犯第三方权益，规定保证数据来源合法及保密等相关责任；

②收到采集数据后，发行人核实、留存被采集人的授权声明；发行人自行研发、目前应用的终端人管理系统设置了授权许可管理模块，对被采集人采集前的授权程序进行更为严格的控制，被采集人通过授权后才能开始采集；

③发行人设置被采集人等相关方投诉渠道并在官方网站公开，便于被采集人等相关方及时向发行人反馈原料数据采集供应商的违规侵权事项。

发行人主要原料数据采集供应商已出具《合规经营承诺函》，承诺并保证为发行人提供数据采集服务过程中所收集并向发行人提供的个人信息及其他数据符合相关法律法规及监管规定，数据来源合法，不存在任何信息、数据权属及知识产权争议，不存在与数据获取相关的争议、纠纷或相关负面报道。

通过中国裁判文书网（<http://wenshu.court.gov.cn/>）、人民法院公告网（<http://rmfygg.court.gov.cn/>）、12309中国检察网（<http://www.ajxxgk.jcy.gov.cn/>）、中国执行信息公开网（<http://zxgk.court.gov.cn/>）等网络公开查询检索，发行人及原料数据采集供应商不存在与数据获取相关的争议、纠纷或相关负面报道。

综上，发行人和原料数据采集供应商主动收集数据的获取方式合规，发行人或客户享有数据的知识产权，已获得相关数据主体的授权许可。

（二）客户自行提供数据

1、相关数据的获取方式，发行人是否享有数据的所有权或获得相关数据主体的授权许可

发行人向客户提供数据转写、标注等纯加工定制服务时，是由客户向发行人提供业务数据并委托发行人进行加工。客户仅向发行人提供语音、图像、视频等业务数据，不提供涉及的个人身份信息。客户向发行人提供数据，主要通过加密传输等方式提供，部分情形客户要求发行人委派人员在客户工作场所的工作平台上直接进行加工。

发行人不参与客户数据收集过程。发行人不享有数据的知识产权，仅经客户授权许可对数据进行加工。

2、相关数据获取方式的合规性

根据《信息安全技术 个人信息安全规范》（GB/T 35273—2017）5.3 b）条款，“间接获取个人信息时：1) 应要求个人信息提供方说明个人信息来源，对其个人信息来源的合法性进行确认；2) 应了解个人信息提供方已获得的个人信息处理的授权同意范围，包括使用目的，是否授权同意转让、共享、公开披露等。”

客户向发行人提供包括个人语音、人像等涉及个人的业务数据，不会提供个人信息，发行人加工过程也不需要个人信息，因此发行人无法识别到具体特定个人。发行人会和客户工作人员沟通了解其数据来源和授权许可情况，协议中约定客户提供数据的相关义务。

发行人主要客户为阿里巴巴、三星、腾讯、微软等人工智能产业链上知名大型企业，该企业均制定有隐私政策等相关个人信息保护制度，要求经明确告知并获得授权同意后获取个人数据，重视业务经营中涉及的数据来源和使用的合法性。客户在与发行人合作过程中也通过问卷调查、现场安全审计等方式对发行人的个人信息保护机制和数据安全能力进行评估并严格要求。

通过中国裁判文书网（<http://wenshu.court.gov.cn/>）、人民法院公告网（<http://rmfygg.court.gov.cn/>）、12309中国检察网（<http://www.ajxxgk.jcy.gov.cn/>）、中国执行信息公开网（<http://zxgk.court.gov.cn/>）等网络公开查询检索，发行人不存在与客户提供数据相关的争议、纠纷或相关负面报道。

综上，发行人从客户取得数据的获取方式合规，发行人不享有数据的知识产权，已获得客户的授权许可。

二、相关授权许可是否存在使用范围、主体或期限等方面的限制，发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形

（一）相关授权许可是否存在使用范围、主体或期限等方面的限制

1、发行人及其原料数据采集供应商主动收集数据的相关授权许可

发行人及其原料数据采集供应商主动收集数据，被采集对象为个人的，已取得被采集人或其监护人的授权许可。根据发行人制定、签署的授权声明或协议，相关授权许可存在使用范围、主体或期限等方面的限制，具体限制如下：

（1）被采集人自愿向发行人提供信息或数据，同意发行人使用上述信息和数据中的声音、文本、图形图像、视频、肖像进行处理并用于向第三方提供训练数据定制服务、训练数据产品，上述信息和数据中的姓名、国籍、性别、年龄、联系方式、监护人信息等由发行人用于验证声明人的适格主体身份；

(2) 除经被采集人许可发行人使用数据和信息用于向第三方提供训练数据资定制服务、训练数据产品外，在未得到被采集人的许可之前，发行人不会把被采集人的任何个人数据或信息提供给其他第三方（包括公司或个人），但下列情况除外：①根据有关的法律法规要求；②行政机关依据法律作出的强制行为；③司法机关依据法律作出的决定、裁定或判决。

2、客户提供数据的相关授权许可

客户向发行人提供语音、图像、视频等业务数据，并授权许可发行人进行加工处理。发行人和客户签署的委托加工相关协议中包含许可及保密等条款，相关授权许可存在使用范围、主体、期限等方面的限制，具体限制如下：

发行人应对客户提供的资料负有安全保密责任，任何时候不得向第三方泄露；除对数据进行加工处理以外，不得随意保留、再应用；发行人应按照协议约定时间清除全部资料；未经客户同意，发行人不得将委托业务转包给第三方。

(二) 发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形

1、发行人使用数据不存在超出上述限制的情形

(1) 发行人使用主动收集数据未超出授权许可限制

如上文所述，发行人使用主动收集数据的用途包含在上述授权许可范围内。另外，发行人及其原料数据采集供应商主动收集的业务数据，主要通过发行人研发的采集软件直接上传至发行人一体化数据处理平台进行处理，该平台整合了采集、加工、质检环节所需的软件工具和模块，融入数据安全管控，能够防范个人私自对数据进行处理。发行人不存在向第三方非法提供或公开披露业务数据、个人身份信息的情形，未超出授权许可的限制使用数据。

(2) 发行人使用客户提供数据未超出授权许可限制

发行人将不同客户提供数据分开存储，严格按照协议约定对数据进行加工后再返还给客户。对于部分情形客户要求发行人委派人员在客户工作场所的工作平台上直接进行加工的，发行人及其委派人员无法控制数据。发行人未将客户提供数据用于其他用途，或向第三方非法提供或公开披露，未超出客户授权许可限制使用数据。

(3) 发行人对使用数据的内部控制措施

发行人采取一系列内部控制措施保护数据安全，避免发生超出授权限制使用数据的情形：

①数据脱敏和加密传输：发行人对收集的个人信息中的身份证号、联系方式等个人敏感信息等进行去标识化处理，并将去标识化后的个人信息与业务数据分开存储；采用加密等安全措施存储和传输数据信息；

②设置保存期限及数据删除：发行人根据与客户协议约定和管理需要设置收集数据的保存期限，超出设置的保存期限后进行删除，或者按照客户约定交付客户后进行删除；

③权限、人员设置及内部审批：对被授权访问数据的内部人员设置仅能访问职责所需的数据信息的权限，对安全管理人员、数据操作人员的角色进行分离设置，对数据信息的重要操作设置内部审批流程，如批量修改、拷贝、下载等；

④人员保密协议及培训：发行人与访问、操作数据的员工签署保密协议，进行数据安全的培训。

根据发行人确认及中国裁判文书网 (<http://wenshu.court.gov.cn/>)、人民法院公告网 (<http://rmfygg.court.gov.cn/>)、12309 中国检察网 (<http://www.ajxxgk.jcy.gov.cn/>)、中国执行信息公开网 (<http://zxgk.court.gov.cn/>) 等网络公开查询检索，发行人不存在超出授权许可的限制使用数据的情形，未发生数据信息泄露事件，不存在个人信息保护、数据安全方面的纠纷或诉讼。

2、原料数据采集供应商使用数据不存在超出上述限制的情形

(1) 主要原料数据采集供应商使用数据未超出限制

原料数据采集供应商为发行人提供业务数据采集、标注服务，主要通过发行人研发的采集、标注软件、工具及工作平台进行，采集、标注形成的业务数据通过前述软件、工具、工作平台直接上传至发行人数据处理平台，原料数据采集供应商无法保存、私自使用该类业务数据。因特定项目具体需要，原料数据采集供应商存在少量使用第三方工具/通用工具（即非发行人研发的工具）进行业务数据采集、标注的情形，在该种情形下，发行人要求原料数据采集供应商将业务数据交付给发行人后即予以删除，不再留存备份数据。原料数据采集供应商未向其他第三方提供或公开披露其收集、接触的数据和个人身份信息。

（2）主要原料数据采集供应商对使用数据的内部控制措施

根据主要原料数据采集供应商访谈及主要原料数据采集供应商提供的保密制度等相关文件，主要原料数据采集供应商采取与相关工作人员签署保密协议、对人员进行数据保密相关培训、对数据信息储存设置密码和人员权限等措施，避免发生超出限制使用数据的情形。

（3）发行人对原料数据采集供应商使用数据的合规控制措施

为避免接触数据信息的原料数据采集供应商超出限制使用数据信息，发行人采取以下主要措施：

①通过制定《数据安全管理制度》、与相关供应商签署《数据安全和保密协议》以及对供应商进行培训等方式，明确供应商承担对数据信息保密的义务和责任，不得超出协议约定使用相关数据信息，不得私自泄露或使用，一旦出现数据违规使用将停止合作；

②发行人会对主要原料数据采集供应商进行不定期检查，查看了解原料数据采集供应商与相关人员签署的保密协议、人员培训情况、对数据信息的储存方式及密码、权限管理等情况；

③发行人设置被采集人等相关方投诉渠道并在官方网站公开，便于接受对原料数据采集供应商违规使用的投诉。

发行人主要原料数据采集供应商已出具《合规经营承诺函》，承诺并保证与发行人合作过程中使用个人信息及其他数据符合相关法律法规及监管规定，不存在超出限制使用数据的情形，未发生数据信息泄露事件，不存在与数据使用相关的争议、纠纷或相关负面报道。

根据中国裁判文书网（<http://wenshu.court.gov.cn/>）、人民法院公告网（<http://rmfygg.court.gov.cn/>）、12309中国检察网（<http://www.ajxxgk.jcy.gov.cn/>）、中国执行信息公开网（<http://zxgk.court.gov.cn/>）等网络公开查询检索，发行人主要原料数据采集供应商不存在个人信息保护、数据安全方面的纠纷或诉讼。

综上，相关授权许可存在使用范围、主体或期限等方面的限制，发行人及其原料数据采集供应商不存在超出上述限制使用数据的情形。

三、是否存在数据内容侵犯个人隐私或其他合法权益的风险

关于发行人和原料数据采集供应商主动收集数据信息，不存在数据内容侵犯个人隐私或其他合法权益的风险。关于客户提供数据，发行人不参与数据收集过程。

（一）个人隐私权、肖像权

根据《中华人民共和国民法总则》第一百一十条，“自然人享有生命权、身体权、健康权、姓名权、肖像权、名誉权、荣誉权、隐私权、婚姻自主权等权利。”

根据《中华人民共和国民法通则》第一百条，“公民享有肖像权，未经本人同意，不得以营利为目的使用公民的肖像。”

涉及个人作为采集对象时，发行人和原料数据采集供应商采集其语音、人像等已经向被采集人或其监护人作出明确告知，并经其授权同意、主动提供配合并放弃相关权益，不涉及欺诈、诱骗、强迫被采集人；拍摄的公开场景视频中涉及个人的，发行人已将其中人像进行去标识化处理。因此，不存在数据内容侵犯被采集人隐私权、肖像权等的风险。

（二）个人信息权益

根据《中华人民共和国民法总则》第一百一十一条，“自然人的个人信息受法律保护。任何组织和个人需要获取他人个人信息的，应当依法取得并确保信息安全，不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息。”

根据《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第三条、第四条，“未经被收集者同意，将合法收集的公民个人信息向他人提供的，属于刑法第二百五十三条之一规定的“提供公民个人信息”，但是经过处理无法识别特定个人且不能复原的除外。”，“违反国家有关规定，通过购买、收受、交换等方式获取公民个人信息，或者在履行职责、提供服务过程中收集公民个人信息的，属于刑法第二百五十三条之一第三款规定的“以其他方法非法获取公民个人信息。”

发行人和原料数据采集供应商主动收集涉及个人的业务数据，及在数据收集过程中被采集人自愿提供个人身份信息，发行人和原料数据采集供应商已事先明确告知被采集人或其监护人收集数据信息的内容、目的、方式和范围等，并经被采集人或其监护人授权许可。发行人或原料数据采集供应商仅将数据信息用于由发行人向客户提供训练数据定制服务、训练数据产品，未超出授权许可限制使用个人数据信息，不存在非法收集、

使用、提供或公开他人个人信息的情形。因此，不存在数据内容侵犯被采集人个人信息权益的风险。

（三）知识产权

根据《中华人民共和国民法总则》第一百二十三条，“民事主体依法享有知识产权。知识产权是权利人依法就下列客体享有的专有的权利：（一）作品；……”。

根据《中华人民共和国民法通则》第九十四条，“公民、法人享有著作权（版权），依法有署名、发表、出版、获得报酬等权利。”

根据《中华人民共和国著作权法》第十七条，“受委托创作的作品，著作权的归属由委托人和受托人通过合同约定。合同未作明确约定或者没有订立合同的，著作权属于受托人。”

根据《中华人民共和国著作权法实施条例》第二条，“著作权法所称作品，是指文学、艺术和科学领域内具有独创性并能以某种有形形式复制的智力成果。”

根据相关协议约定，发行人及原料数据采集供应商主动采集数据系按发行人或客户的要求实施完成，相应知识产权由发行人或客户享有，原料数据采集供应商或终端人员放弃相关权益。发行人及原料数据采集供应商公开收集文本内容片段等业务数据，系收集通用文本内容片段，再将字词进行重新组合，不涉及侵犯作品的著作权。因此，不存在数据内容侵犯他人知识产权的风险。

综上，发行人及其原料数据采集供应商获取相关数据不存在数据内容侵犯个人隐私或其他合法权益的风险。

（2）《个人信息安全规范》《数据安全法》（草案）等法规的出台对发行人业务的具体影响，发行人及其原料数据采集供应商是否存在违反上述规定的情形，请发行人就相关监管政策出台对行业可能产生的影响进行重大事项提示和风险揭示。

答复：

一、《个人信息安全规范》《数据安全法》（草案）等法规的出台对发行人业务的具体影响，发行人及其原料数据采集供应商是否存在违反上述规定的情形

（一）发行人及其原料数据采集供应商目前个人信息、数据安全保护情况

1、发行人

在《中华人民共和国民法典》、《个人信息安全规范》、《数据安全法》（草案）等法规出台前，发行人已按照现行有效的《中华人民共和国民法通则》、《中华人民共和国民法总则》、《信息安全技术 个人信息安全规范》（GB/T 35273—2017）等相关法律、法规、规范性文件建立个人信息、数据安全保护机制：

①在数据收集、存储、使用、删除、转让等数据流转全流程确立个人信息、数据安全的规范要求

如前所述，发行人及原料数据采集供应商主动收集数据涉及个人的，已向被采集人或监护人事先明确告知收集的内容、目的、使用范围等规则，并取得其授权同意，收集业务经营必要的的数据信息；发行人对收集的个人信息中的身份证号、联系方式等个人敏感信息等进行去标识化处理，并将去标识化后的个人信息与业务数据分开存储；发行人使用数据未超出授权许可的范围，未向第三方非法提供或公开披露数据；存储的数据按照客户协议约定及管理需要及时删除；发行人向客户提供加工后的业务数据，不提供个人信息，未超出授权许可范围。

②研发使用数据处理平台、采用技术措施减少数据接触范围、保障数据安全

发行人自行研发一体化数据处理平台、采集软件、工具等，嵌入业务过程的数据安全管理需求，实现除特定项目需要以外的大部分终端采集数据直接上传至发行人服务器，标注工作主要在发行人自有平台上加工处理；采用加密等安全措施存储和传输数据信息；定期对业务系统进行漏洞扫描，及时对业务系统漏洞进行修复；网络环境中部署IPS、IDS 等安全设备等。

③制定相关制度建立个人信息保护、数据安全管理制度和流程

发行人已制定《数据安全管理制度》、《IT 安全管理办法》、《存储服务器使用规范和方法》、《网络安全规范》、《数据备份规范》、《网络应急响应机制》等制度，建立了个人信息保护、数据安全管理制度和流程。发行人设立数据保护官负责全面统筹个人信息数据保护制度的落实，已建立各部门所持有的包含个人信息的数据类型和授权访问策略，定期开展个人信息安全影响评估和安全审计，建立个人数据泄露等应急处理机制。

④制定、公开个人信息保护政策，并建立个人信息保护投诉机制，保障被采集人的个人信息权益

发行人已制定个人信息保护政策并于发行人官方网站公开，被采集人享有查询、更正、删除等权利，并公布投诉联系方式；发行人收到投诉后将分析原因、积极与客户沟通、提出解决方案，并及时将解决方案告知投诉人。

发行人已取得北京赛西认证有限责任公司核发的《管理体系认证证书》（证书编号 01119IS20090R0M），证明发行人信息安全管理体系符合标准：GB/T 22080-2016/ISO/IEC 27001: 2013《信息技术 安全技术 信息安全管理体系要求》，证书覆盖范围为与计算机应用软件的设计与开发、智能语音、计算机视觉、自然语言等人工智能数据资源产品及服务相关信息安全管理活动，有效期至 2022 年 10 月 7 日。

2、原料数据采集供应商

原料数据采集供应商为发行人提供业务数据采集、标注服务，主要通过发行人研发的采集、标注软件、工具及工作平台进行，采集、标注形成的业务数据直接上传至发行人服务器，原料数据采集供应商无法保存、私自使用该类业务数据。因特定项目具体需要，原料数据采集供应商存在少量使用第三方工具/通用工具（即非发行人研发的工具）进行业务数据采集、标注的情形，在该种情形下，发行人要求原料数据采集供应商将业务数据交付给发行人后即予以删除，不再留存备份数据。

发行人主要原料数据采集供应商建立了内部保密制度，与相关人员签署了保密协议，数据存储设置密码，限定接触数据信息的岗位并设置具体权限，不向第三方提供或公开披露数据信息，保障接触的数据信息的安全。

为规范接触数据信息的原料数据采集供应商保护数据安全，发行人对供应商进行管理的主要措施具体见本问询函回复之“问题 7”之“发行人说明事项（1）“发行人及其原料数据采集供应商相关数据的获取方式及其合规性…”之“二、相关授权许可是否存在使用范围、主体或期限等方面的限制，发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形”之“（二）发行人及其原料数据采集供应商是否存在超出上述限制使用数据的情形”之“2、原料数据采集供应商使用数据不存在超出上述限制的情形”之“（3）发行人对原料数据采集供应商使用数据的合规控制措施”相关表述。

综上，发行人及其原料数据采集供应商在个人信息保护、数据安全等方面的情况符合《中华人民共和国民法通则》、《中华人民共和国民法总则》、《信息安全技术 个人信息安全规范》（GB/T 35273—2017）的相关规定。

(二) 《个人信息安全规范》《数据安全法》(草案)等法规的出台对发行人业务的具体影响, 发行人及其原料数据采集供应商是否存在违反上述规定的情形

1、《中华人民共和国民法典》

《中华人民共和国民法典》已于2020年5月28日公布, 自2021年1月1日起施行。《中华人民共和国民法典》将“隐私权和个人信息保护”在第四编人格权设置第六章专门进行规定。其中与发行人业务相关的法律规定主要如下:

法条序号	主要内容	发行人及其原料数据采集供应商是否符合规定
第一千零一十九条	任何组织或者个人不得以丑化、污损,或者利用信息技术手段伪造等方式侵害他人的肖像权。未经肖像权人同意,不得制作、使用、公开肖像权人的肖像,但是法律另有规定的除外。 未经肖像权人同意,肖像作品权利人不得以发表、复制、发行、出租、展览等方式使用或者公开肖像权人的肖像。	符合
第一千零二十三条	对自然人声音的保护,参照适用肖像权保护的有关规定。	符合
第一千零三十二条	自然人享有隐私权。任何组织或者个人不得以刺探、侵扰、泄露、公开等方式侵害他人的隐私权。 隐私是自然人的私人生活安宁和不愿为他人知晓的私密空间、私密活动、私密信息。	符合
第一千零三十三条	除法律另有规定或者权利人明确同意外,任何组织或者个人不得实施下列行为:(一)以电话、短信、即时通讯工具、电子邮件、传单等方式侵扰他人的私人生活安宁;(二)进入、拍摄、窥视他人的住宅、宾馆房间等私密空间;(三)拍摄、窥视、窃听、公开他人的私密活动;(四)拍摄、窥视他人身体的私密部位;(五)处理他人的私密信息;(六)以其他方式侵害他人的隐私权。	
第一千零三十四条	自然人的个人信息受法律保护。 个人信息是以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人的各种信息,包括自然人的姓名、出生日期、身份证件号码、生物识别信息、住址、电话号码、电子邮箱、健康信息、行踪信息等。 个人信息中的私密信息,适用有关隐私权的规定;没有规定的,适用有关个人信息保护的规定。	符合
第一千零三十五条	处理个人信息的,应当遵循合法、正当、必要原则,不得过度处理,并符合下列条件: (一)征得该自然人或者其监护人同意,但是法律、行政法规另有规定的除外; (二)公开处理信息的规则; (三)明示处理信息的目的、方式和范围; (四)不违反法律、行政法规的规定和双方的约定。 个人信息的处理包括个人信息的收集、存储、使用、加工、传输、提供、公开等。	
第一千零三	处理个人信息,有下列情形之一的,行为人不承担民事责任:	符合

十六条	(一) 在该自然人或者其监护人同意的范围内合理实施的行为; (二) 合理处理该自然人自行公开的或者其他已经合法公开的信息, 但是该自然人明确拒绝或者处理该信息侵害其重大利益的除外; (三) 为维护公共利益或者该自然人合法权益, 合理实施的其他行为。	
第一千零三十七条	自然人可以依法向信息处理者查阅或者复制其个人信息; 发现信息有错误的, 有权提出异议并请求及时采取更正等必要措施。 自然人发现信息处理者违反法律、行政法规的规定或者双方的约定处理其个人信息的, 有权请求信息处理者及时删除。	符合
第一千零三十八条	信息处理者不得泄露或者篡改其收集、存储的个人信息; 未经自然人同意, 不得向他人非法提供其个人信息, 但是经过加工无法识别特定个人且不能复原的除外。 信息处理者应当采取技术措施和其他必要措施, 确保其收集、存储的个人信息安全, 防止信息泄露、篡改、丢失; 发生或者可能发生个人信息泄露、篡改、丢失的, 应当及时采取补救措施, 按照规定告知自然人并向有关主管部门报告。	符合

2、《个人信息安全规范》

《信息安全技术 个人信息安全规范》(GB/T35273—2020, 代替 GB/T 35273—2017) 已于 2020 年 3 月 6 日发布, 自 2020 年 10 月 1 日起实施。

目前已实施的 2017 年发布的《信息安全技术 个人信息安全规范》(GB/T 35273—2017) 对个人信息的收集、存储、委托处理、转让等规定了较为细致的原则和安全要求, 作为个人信息安全方面的推荐性国家标准, 并非强制性标准。根据《中华人民共和国标准化法》(中华人民共和国主席令第 78 号) 第二条, “强制性标准必须执行。国家鼓励采用推荐性标准。” 如前所述, 发行人已按照《信息安全技术 个人信息安全规范》(GB/T 35273—2017) 建立个人信息、数据安全保护机制, 在数据收集、存储、使用、删除、转让等数据流转全流程确立个人信息、数据安全的规范要求, 并取得北京赛西认证有限责任公司核发的《管理体系认证证书》(证书编号 01119IS20090R0M), 符合《信息安全技术 个人信息安全规范》(GB/T 35273—2017) 的相关规定。

与目前已实施的 2017 年发布的《信息安全技术 个人信息安全规范》(GB/T 35273—2017) 相比, 《信息安全技术 个人信息安全规范》(GB/T35273—2020) 在个别细节上进行修订, 并新增如下内容:

①新增“个人信息安全工程”, 要求开发具有处理个人信息功能的产品或服务时, 个人信息控制者宜根据国家有关标准在需求、设计、开发、测试、发布等系统工程阶段考虑个人信息保护要求, 保证在系统建设时对个人信息保护措施同步规划、同步建设和同步使用;

②新增“个人信息处理活动记录”，个人信息控制者宜建立、维护和更新所收集、使用的个人信息处理活动记录，记录的内容可包括：a) 所涉及个人信息的类型、数量、来源（如从个人信息主体直接收集或通过间接获取方式获得）；b) 根据业务功能和授权情况区分个人信息的处理目的、使用场景，以及委托处理、共享、转让、公开披露、是否涉及出境等情况；c) 与个人信息处理活动各环节相关的信息系统、组织或人员。

发行人研发使用的一体化数据处理平台，已在数据采集、加工、质检等环节融入数据安全模块，符合上述“个人信息安全工程”的要求；发行人研发使用的全时日志系统对个人信息处理活动进行记录，符合上述“个人信息处理活动记录”的要求。

3、《数据安全法》（草案）

《数据安全法》（草案）已于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，正在面向公众公开征求意见。其中与发行人业务相关的法律规定主要如下：

法条序号	主要内容	发行人及其原料数据采集供应商是否符合规定
第八条	开展数据活动，必须遵守法律、行政法规，尊重社会公德和伦理，遵守商业道德，诚实守信，履行数据安全保护义务，承担社会责任，不得危害国家安全、公共利益，不得损害公民、组织的合法权益。	符合
第二十二 条	国家建立数据安全审查制度，对影响或者可能影响国家安全的数据活动进行国家安全审查。	目前相关法规尚未界定国家安全审查事项范围
第二十三 条	国家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。	目前相关法规尚未界定出口管制范围
第二十五 条	开展数据活动应当依照法律、行政法规的规定和国家标准的强制性要求，建立健全全流程数据安全管理制度，组织开展数据安全教育培训，采取相应的技术措施和其他必要措施，保障数据安全。 重要数据的处理者应当设立数据安全负责人和管理机构，落实数据安全保护责任。	符合
第二十六 条	开展数据活动以及研究开发数据新技术，应当有利于促进经济社会发展，增进人民福祉，符合社会公德和伦理。	符合
第二十七 条	开展数据活动应当加强风险监测，发现数据安全缺陷、漏洞等风险时，应当立即采取补救措施；发生数据安全事件时，应当按照规定及时告知用户并向有关主管部门报告。	符合
第二十八 条	重要数据的处理者应当按照规定对其数据活动定期开展风险评估，并向有关主管部门报送风险评估报告。 风险评估报告应当包括本组织掌握的重要数据的种类、数量，收集、存储、加工、使用数据的情况，面临的数据安全风险及其应对措施等。	虽然目前相关法规尚未界定重要数据的处理者范围，但发行人比照要求定期开展风险评估。
第二十	任何组织、个人收集数据，必须采取合法、正当的方式，不得窃取或	符合

九条	者以其他非法方式获取数据。 法律、行政法规对收集、使用数据的目的、范围有规定的，应当在法律、行政法规规定的目的和范围内收集、使用数据，不得超过必要的限度。	
----	--	--

综上，上述新法规出台对发行人业务不存在不利影响，发行人和原料数据采集供应商不存在违反上述主要规定的情形，待新法规实施及配套制度正式确立后将按照规定要求予以执行。

二、请发行人就相关监管政策出台对行业可能产生的影响进行重大事项提示和风险揭示

就相关监管政策出台对行业可能产生的影响，发行人已在招股说明书“重大事项提示/三、特别风险提示/（五）数据安全相关风险”和“第四节 风险因素/八、数据安全相关风险”中补充重大事项提示和风险揭示如下，

“（五）数据安全相关风险

发行人主要从事训练数据的研发设计、生产及销售业务，所提供的产品和服务主要以数据的形式体现。一方面，随着发行人业务的快速发展和规模的持续扩张，原料数据采集与数据加工的数量持续增长；另外一方面，与数据安全、个人信息保护相关的法律规章体系逐步完善。如果将来公司未能根据法律规章的更新要求及时调整现行业务开展方式、公司的数据安全管理体系研发升级未能跟上业务发展的需要、或客户未能遵守训练数据产品保护相关商业约定，则公司可能面临生产经营不符合法律规章的要求、训练数据产品被泄露、盗版等数据安全相关风险或可能产生诉讼纠纷，进而给公司的经营带来不利影响。”

“八、数据安全相关风险

发行人主要从事训练数据的研发设计、生产及销售业务，所提供的产品和服务主要以数据的形式体现。一方面，随着发行人业务的快速发展和规模的持续扩张，原料数据采集与数据加工的数量持续增长；另外一方面，与数据安全、个人信息保护相关的法律规章体系逐步完善。如果将来公司未能根据法律规章的更新要求及时调整现行业务开展方式、公司的数据安全管理体系研发升级未能跟上业务发展的需要、或客户未能遵守训练数据产品保护相关商业约定，则公司可能面临生产经营不符合法律规章的要求、

训练数据产品被泄露、盗版等数据安全相关风险或可能产生诉讼纠纷，进而给公司的经营带来不利影响。”

中介机构核查事项

(1) 核查程序

保荐机构、发行人律师主要履行了如下核查程序：

1、与发行人数据保护官了解业务经营涉及的数据获取、存储、使用、删除等业务流程中相关数据保护措施，了解发行人内部制度及对供应商的管理措施；

2、抽查发行人业务项目涉及的被采集人签署的授权声明、被采集人或其监护人填写的个人身份信息表；

3、现场查看发行人一体化数据处理平台、终端人管理系统的内容，查看发行人去标识化存储信息的情况；

4、收集、审阅发行人与主要原料数据采集供应商、主要客户之间的协议、发行人《数据安全管理制度》、《个人信息保护政策》等相关制度政策、发行人与员工的保密协议等；

5、与主要原料数据采集供应商访谈了解其数据采集流程、对相关数据的保护措施，并取得相关确认、承诺文件；

6、收集、审阅主要原料数据采集供应商保密制度、与员工的保密协议等；

7、与主要小外包人员、终端人员访谈了解相关采集流程等；

8、网络公开检索发行人主要客户的隐私政策等个人信息保护制度，审阅发行人客户对发行人进行数据安全调查的问卷等文件；

9、网络公开检索发行人、主要原料数据采集供应商与数据收集、使用相关的争议、纠纷或负面报道情况；

10、研究检索《个人信息安全规范》、《数据安全法》（草案）等新法规的相关内容和实施进展情况。

(2) 核查结论

经核查，保荐机构、发行人律师认为：

1、发行人及其原料数据采集供应商相关数据的获取方式合规，发行人或客户享有数据的所有权，已获得相关数据主体的授权许可；相关授权许可存在使用范围、主体、期限等方面的限制，发行人及其原料数据采集供应商不存在超出上述限制使用数据的情形，不存在数据内容侵犯个人隐私或其他合法权益的风险；

2、《个人信息安全规范》《数据安全法》（草案）等法规的出台对发行人业务不存在不利影响，发行人及原料数据采集供应商不存在违反上述主要规定的情形；就相关监管政策出台对行业可能产生的影响，发行人已在《招股说明书》补充重大事项提示和风险揭示。

二、关于发行人业务

8. 关于业务资质

请保荐机构、发行人律师结合相关法律法规规定，就发行人是否取得从事 AI 训练数据的相关业务资质，是否存在无资质、许可或超出规定范围开展业务经营的情形进行核查，并发表明确意见。

答复：

中介机构核查事项

发行人不存在无资质、许可或超出规定范围开展业务经营的情形：

（一）相关法律法规并未对从事 AI 训练数据规定相关业务资质

《国务院关于印发新一代人工智能发展规划的通知》（国发[2017]35号）、《信息安全技术 个人信息安全规范》（GB/T 35273—2017）、《信息安全技术 个人信息安全规范》（GB/T35273—2020）等具体法律法规并未对从事训练数据采集、加工等业务规定相关业务资质。

根据保荐机构、发行人律师与中国人工智能产业发展联盟（该联盟由中国信息通信研究院牵头会同百度、阿里、腾讯、清华、浙大、中科院自动化所等积极推动人工智能产业发展的代表性企业、高校、科研院所参加，共同发起成立，旨在搭建产学研用合作平台，为业界知名的行业组织）工作人员访谈了解，目前法律法规并未对从事训练数据采集、加工等业务规定相关业务资质，发行人从事训练数据业务无需相关业务资质。

发行人承诺，如未来法律法规对从事训练数据业务作出具体资质要求，发行人将依法申请相关资质许可以保障合规经营。

（二）发行人业务经营未超出《营业执照》上核准的经营范围

根据发行人现行有效的《营业执照》，发行人的经营范围为“技术开发、技术服务、技术咨询、技术转让、技术推广；销售计算机、软件；货物进出口、技术进出口；出租办公用房。”发行人目前持有北京市商务局核发的《对外贸易经营者备案登记表》，备案登记表编号02138492。

发行人主要从事训练数据的研发设计、生产及销售业务，未超出《营业执照》上核准的经营范围和经营方式，就其主要业务和经营也已经取得必要的政府审批和资质。

综上，发行人从事训练数据无需相关业务资质，发行人不存在无资质、许可或超出规定范围开展业务经营的情形。

（1）核查程序

保荐机构、发行人律师主要履行了如下核查程序：

- 1、检索查阅相关法律法规；
- 2、与中国人工智能产业发展联盟工作人员进行电话咨询。
- 3、查阅公司的营业执照、对外贸易经营者备案登记表等证照文件；
- 4、通过查阅公司的业务合同、实地查看等了解发行人具体业务经营情况。

（2）核查结论

经核查，保荐机构、发行人律师认为：相关法律法规并未对从事训练数据业务规定相关业务资质，发行人不存在无资质、许可或超出规定范围开展业务经营的情形。

9. 关于销售模式和主要客户

9.1 关于主要服务和产品

招股说明书披露，发行人产品和服务分为训练数据定制服务、训练数据产品和训练数据相关的应用服务，训练数据定制服务和训练数据产品按照应用领域可以分为智

能语音、计算机视觉和自然语言，智能语音可以进一步分为语音识别采集、语音识别转写和语音合成。

请发行人补充披露：（1）训练数据定制服务中各应用领域中采集加工和纯加工服务的营业收入、营业成本和毛利率情况；（2）不同目标客户类型的客户数量、产品销售或服务的具体情况。

请发行人说明：（1）相关产品或服务的分类是否符合行业惯例，是否属于行业通用用语；（2）不同服务形式和应用领域的定制服务从合同签署至最终验收交付的实际产销过程以及采用目前经营模式的原因；（3）三项智能语音细分类别与训练数据定制服务形式之间的对应关系；（4）训练数据相关的应用服务的具体内容、相关合同的主要权利义务、服务成果以及定价依据等，是否独立于训练数据定制服务和训练数据产品；（5）训练数据定制服务和训练数据产品的目标客户是否存在重合及重合情况，两项业务是否构成替代关系。

答复：

发行人补充披露事项

（1）训练数据定制服务中各应用领域中采集加工和纯加工服务的营业收入、营业成本和毛利率情况

发行人已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（三）毛利及毛利率分析/4、训练数据定制服务各服务形式毛利率情况”中补充披露了下述内容：

报告期内，发行人在训练数据定制服务各应用领域均同时提供采集加工服务和纯加工服务。前述两种业务类型分别主要服务于客户不同阶段的需求：

①客户算法模型初始上线前的训练阶段——采集加工服务：在此阶段，客户的 AI 技术应用还未对外提供使用，尚处于算法模型的内部训练和模拟阶段，还未取得算法模型对外上线后能收集到的实际数据。为了使得算法模型取得更好的训练效果、更适用于实际场景，客户可向发行人采购[采集加工]类别的训练数据定制服务，发行人将根据客户算法模型的个性化需求，有针对性地采集原料数据，并进行加工处理，从而更好地模拟实际应用场景数据，使得客户的算法模型获得更佳训练效果。

②客户算法模型上线后的持续优化迭代阶段——纯加工服务：在此阶段，客户的AI算法模型/技术应用已正式上线并对外运营，客户在运营过程中可以收集到大量实际场景下的原料数据。为进一步优化算法模型，客户可向发行人采购[纯加工服务]，将运营过程中收集的实际场景原料数据提供给发行人，由发行人对其执行标注等加工工作，形成源自真实应用场景的训练数据，对算法模型进行进一步的训练、优化、迭代。

③客户算法模型的领域拓展阶段——采集加工服务：在此阶段，客户拟对其AI算法模型进行领域拓展（如应用场景拓展、语种/方言拓展等）。与上述阶段①类似，客户需要先对其算法模型进行相应领域的内部训练，暂未取得相应领域下的实际数据，因此可向发行人采购[采集加工服务]，以对其算法模型加以训练。

报告期内，训练数据定制服务各领域中采集加工和纯加工服务的营业收入、营业成本和毛利率情况具体如下所示：

金额单位：万元

应用领域	业务类别	营业收入				营业成本				毛利率			
		2020年 1-3月	2019年	2018年	2017年	2020年 1-3月	2019年	2018年	2017年	2020年 1-3月	2019年	2018年	2017年
智能语音	语音识别 (即语音识别采集)	57.16	2,059.97	2,809.18	1,451.45	23.72	1,157.17	1,606.83	741.48	58.50%	43.83%	42.80%	48.91%
	纯加工服务 (即语音识别转写)	209.61	2,562.42	2,960.54	1,409.18	146.15	1,740.04	2,131.45	979.43	30.28%	32.09%	28.00%	30.50%
	语音合成 采集加工服务	-	5,653.06	3,613.25	1,541.13	-	2,173.05	1,394.73	674.30	-	61.56%	61.40%	56.25%
	语音合成 纯加工服务	25.53	675.01	506.71	492.22	3.97	235.53	189.32	146.93	84.45%	65.11%	62.64%	70.15%
计算机视觉	采集加工服务	532.28	1,051.77	381.37	623.27	210.77	595.38	157.49	279.64	60.40%	43.39%	58.70%	55.13%
	纯加工服务	38.69	1,026.70	1,287.35	199.12	20.37	573.97	749.25	131.95	47.36%	44.09%	41.80%	33.73%
自然语言	采集加工服务	136.15	288.82	205.24	126.83	54.64	63.45	108.65	68.95	59.87%	78.03%	47.06%	45.64%
	纯加工服务	138.33	915.16	605.91	454.04	77.26	419.08	386.31	260.18	44.15%	54.21%	36.24%	42.70%
合计		1,137.75	14,232.91	12,369.55	6,297.23	536.87	6,957.69	6,724.04	3,282.86	52.81%	51.12%	45.64%	47.87%

2017年、2018年、2019年及2020年1-6月，发行人训练数据定制服务中，各领域采集加工服务收入合计值分别为3,742.68万元、7,009.05万元、9,053.63万元及725.58万元；纯加工服务收入合计值分别为2,554.55万元、5,360.51万元、5,179.28万元及412.16万元。采集加工服务和纯加工服务均为发行人训练数据定制服务业务的重要形式，报告期内收入贡献比例分别约为60%：40%，采集加工服务收入贡献度整体高于纯加工服务。

其中，在智能语音类别中，语音合成领域纯加工服务的毛利率水平高于语音识别领域纯加工服务，主要原因是：语音合成领域的标注工作涉及音素、韵律、词性等多种标注类型，相对复杂度较高，所需技术处理环节较多，因此毛利率相对高于语音识别领域的纯加工服务。

(2) 不同目标客户类型的客户数量、产品销售或服务的具体情况。

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/(二) 主要客户情况/3、各类型目标客户情况”中披露下述相关内容：

发行人的目标客户可分为大型科技公司、人工智能企业、科研机构三类²。截至 2020 年 3 月 31 日，发行人不同目标客户类型的客户数量情况如下：

目标客户类型	客户数量	客户示例
大型科技公司	24 个品牌、80 个客户主体	三星、阿里巴巴、微软、腾讯、亚马逊等
人工智能企业	261 个品牌、338 个客户主体	海康威视、科大讯飞、Naver、Verint 等
科研机构	57 个品牌、64 个客户主体	中国科学院、清华大学等

注：上表统计客户均为历史上曾为发行人贡献业务收入或已与发行人签署业务合作协议/合同并实际开展了业务往来的客户群体。

不同目标客户类型的产品销售或服务的具体情况如下：

金额单位：万元

销售收入	2020 年 1-3 月		2019 年		2018 年		2017 年	
	营业收入	占比	营业收入	占比	营业收入	占比	营业收入	占比
大型科技公司	1,950.40	62.39%	17,003.58	71.58%	14,188.80	73.65%	7,425.81	62.36%
其中：训练数据定制服务	930.91	29.78%	11,626.58	48.94%	9,788.06	50.81%	4,344.48	36.49%
训练数据产品	1,019.50	32.61%	5,376.49	22.63%	4,400.74	22.84%	3,005.69	25.24%
训练数据相关的应用服务	-	-	0.51	0.00%	-	-	75.63	0.64%
人工智能企业	917.65	29.35%	6,049.60	25.47%	4,641.46	24.09%	3,941.94	33.11%
其中：训练数据定制服务	200.37	6.41%	2,139.52	9.01%	2,187.33	11.35%	1,822.42	15.31%
训练数据产品	717.28	22.94%	3,583.13	15.08%	2,160.50	11.21%	2,074.59	17.42%
训练数据相关的应用服务	-	0.00%	326.95	1.38%	293.64	1.52%	44.92	0.38%
科研机构	258.04	8.25%	702.63	2.96%	435.50	2.26%	539.35	4.53%
其中：训练数据定制服务	6.47	0.21%	466.81	1.97%	394.17	2.05%	130.32	1.09%
训练数据产品	251.57	8.05%	216.85	0.91%	40.43	0.21%	409.03	3.44%
训练数据相关的应用服务	-	-	18.98	0.08%	0.91	0.00%	-	-
营业收入合计	3,126.09	100.00%	23,755.81	100.00%	19,265.77	100.00%	11,907.09	100.00%

² 大型科技公司主要为三星、阿里巴巴、微软、腾讯、亚马逊等业务范围较广的知名综合性科技企业，人工智能企业主要为海康威视、科大讯飞等主要专长于人工智能具体技术领域（如智能语音、计算机视觉等）的企业；科研院所主要指中国科学院、清华大学等开展人工智能相关科学研究的研究院、研究所、大专院校等机构。

报告期内，大型科技公司类客户贡献的收入金额最大、占比最高，在 2017 年-2019 年间，发行人主要向其销售训练数据定制服务和训练数据产品，其中训练数据定制服务收入占比更高，一方面是由于阿里巴巴、腾讯、百度、微软、三星、亚马逊等大型科技公司近些年加大了在人工智能领域的投入，相应提升了对训练数据产品和服务的采购规模；另一方面是由于发行人通过长期经营树立了品牌形象和市场地位，对大型客户保持了较高的吸引力和黏性。人工智能企业类别客户贡献收入仅次于大型科技公司，在 2017 年-2019 年间，发行人向该类客户销售呈现训练数据定制服务和训练数据产品并重的现象，训练数据产品整体销售收入略高于训练数据定制服务。科研机构类客户贡献收入占比较低，发行人向其销售的也主要是训练数据定制服务和训练数据产品。

2020 年第一季度，新冠病毒疫情对发行人现场集中进行数据采集、标注造成了一定影响，发行人通过引导客户采购成品训练数据集（即训练数据产品）等方式平抑相关影响，因而当期向三类客户销售的训练数据产品收入均高于训练数据定制服务。

发行人说明事项

（1）相关产品或服务的分类是否符合行业惯例，是否属于行业通用用语；

发行人将主要产品或服务按业务类型及下游应用领域划分符合行业惯例，属于行业通用用语。具体说明如下：

1、主要产品或服务按业务类型划分

发行人将主要产品及服务划分为训练数据定制服务、训练数据产品及训练数据相关的应用服务。前述分类主要是依据为客户提供产品服务的业务模式及交付物的权属进行，主要划分为根据客户需求提供服务、交付的训练数据权属归属于客户的定制服务，以及可反复销售使用授权、发行人拥有知识产权的通用成品训练数据集。以同行业竞争对手标贝科技为例，根据其官方网站披露，其将产品服务划分为“提供服务”及“热销数据”两类，与发行人分类方式相同。

发行人的前述分类符合行业惯例，是行业通用用语。

2、主要产品或服务按下下游应用领域划分

发行人将训练数据定制服务、训练数据产品均按应用领域划分，分为智能语音、计算机视觉及自然语言三大领域。

前述分类主要是按发行人提供的训练数据对应的下游 AI 应用领域划分，智能语音、计算机视觉及自然语言是当前人工智能发展阶段最为主流的三个应用领域，是众多人工智能行业介绍、技术架构介绍中的通行标准。如中国信通院与中国人工智能产业发展联盟共同编纂的《人工智能发展白皮书技术架构篇（2018 年）》、中国电子技术标准化研究院编纂的《人工智能标准化白皮书 2018 版》等，均使用智能语音、计算机视觉、自然语言处理作为人工智能技术的领域划分用语进行呈现。

不同应用领域的训练数据在数据形态方面存在差异，对应采集的原料数据类型、执行的标注工作类型也有所差异，发行人以此分类具备合理性。以同行业竞争对手标贝科技为例，其也按应用领域对其数据产品和服务加以分类，如语音识别数据库、语音合成定制等。

发行人的前述分类符合行业惯例，属于行业通用用语。

（2）不同服务形式和应用领域的定制服务从合同签署至最终验收交付的实际购产销过程以及采用目前经营模式的原因；

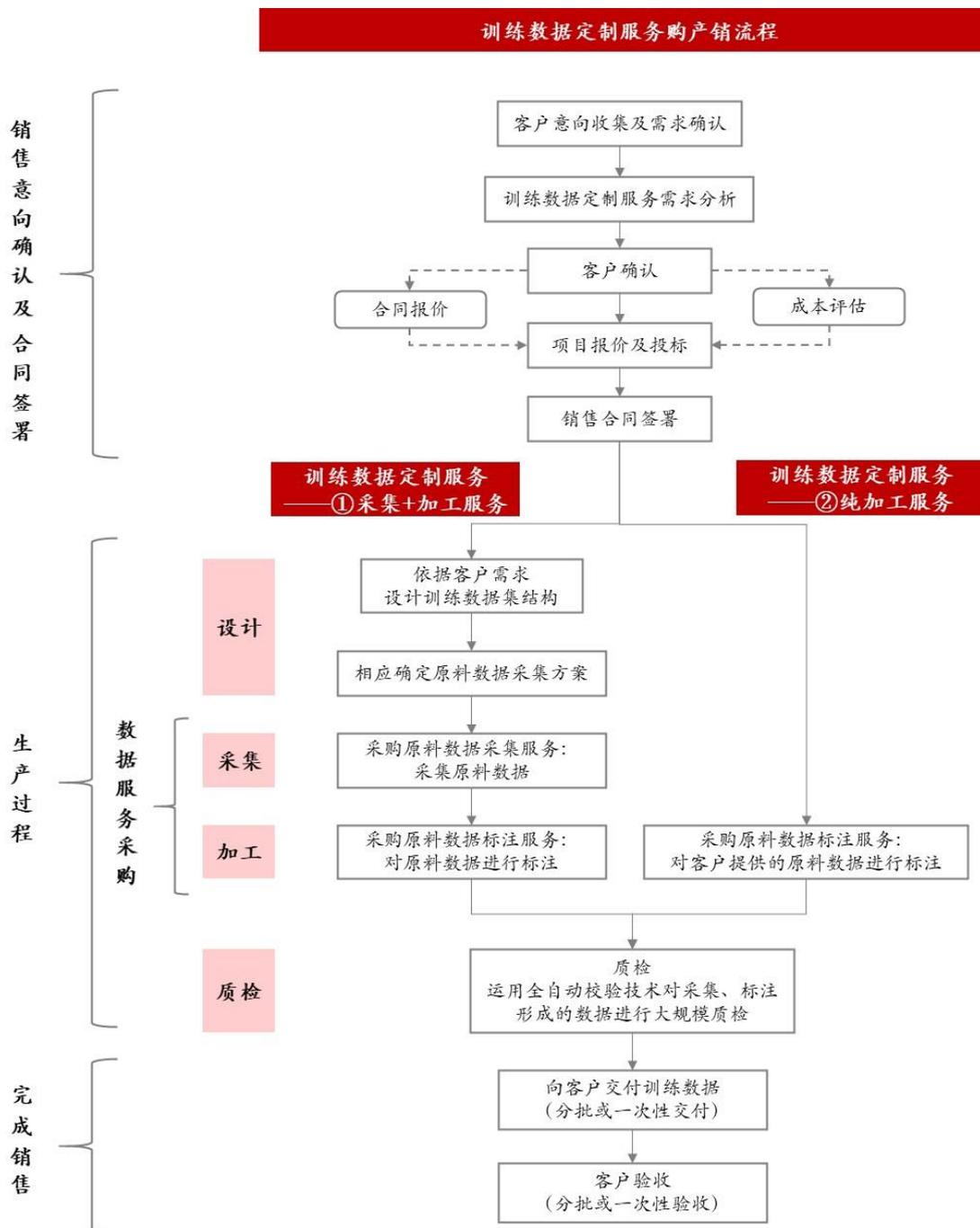
一、不同服务形式和应用领域的定制服务从合同签署至最终验收交付的实际购产销过程

发行人各应用领域的定制服务的购产销过程基本相同。

在采购过程中，因各应用领域涉及的原料数据类型不同，对应采购的采集、标注服务类型也有所不同，例如：智能语音领域需要采集的原料数据主要为语音数据等，对应的标注类型有转写、校对、韵律、词性、音素标注等；计算机视觉领域涉及的原料数据主要为图像、视频数据等，对应的标注类型有属性标注、主观标注等；自然语言领域对应的原料数据主要为对话文本等，对应的标注类型有词法标注、语音、语法、语用相关标注等。

但各应用领域的采购流程整体相同。因此下文区分服务形式说明发行人定制服务的购产销过程，不区分应用领域。

发行人的训练数据定制服务实质为按客户需求生产训练数据集并交付予客户，因此其购产销过程为以销定产的按需生产过程。采集加工、纯加工两种服务形式的购产销过程具体如下所示：



二、采用目前经营模式的原因

发行人的训练数据定制服务致力于以客户需求为导向，在客户自身技术和产品研发的不同阶段满足其不同需求，因此，在该类服务领域，发行人根据客户的不同需要可提供采集加工和纯加工两种服务形式：

1、新产品或应用首次研发并上市的准备阶段，客户可使用发行人提供的训练数据定制服务-采集加工服务，定制更符合其算法模型运用场景的训练数据；

2、客户的算法模型或应用的实际运行阶段，客户的算法模型将可获取实际场景数据，为进一步实现算法模型性能的持续提升，客户可使用训练数据定制服务-纯加工服务，将收集到的实际场景数据交由发行人进行加工，定制训练数据用于进一步训练算法模型；

3、在应用领域拓展及语种拓展阶段，客户可使用发行人提供的训练数据定制服务-采集加工服务等，定制特定领域或语种的训练数据。

（3）三项智能语音细分类别与训练数据定制服务形式之间的对应关系；

三项智能语音细分类别与训练数据定制服务形式之间的对应关系如下：

产品/服务类型	应用领域	产品/服务细分类别		说明
训练数据定制服务	智能语音	语音识别	语音识别采集	采集加工服务
			语音识别转写	纯加工服务
		语音合成		采集加工服务、纯加工服务

发行人将训练数据定制服务及训练数据产品均按下游应用领域进行了划分。在智能语音领域，发行人又根据所提供的训练数据服务的智能语音算法模型类别，划分为语音识别和语音合成两个技术领域。

1、语音识别：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。对于服务语音识别算法模型的训练数据定制服务，发行人根据下游客户需求的实际情况和公司内部精细化经营管理需要，依据是否需要发行人采集原料数据进行了划分，分为语音识别采集（采集加工服务，发行人采集数据并加工）及语音识别转写（纯加工服务，发行人对客户提供的数据进行加工）两类。

2、语音合成：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。发行人将服务于语音合成算法模型的训练数据定制服务划分为语音合成类别，该类别定制服务同时包含采集加工服务及纯加工服务，因纯加工服务收入规模较小，发行人未单独再将该类别进行划分。

综上，发行人依据下游算法模型的技术类型将训练数据服务分为语音合成、语音识别两大类；对于语音识别类别又根据发行人是否需要采集原料数据划分为语音识别采集（采集加工服务）、语音识别转写（纯加工服务）两小类，总共分为语音合成、语音识别采集、语音识别转写 3 个类别。

（4）训练数据相关的应用服务的具体内容、相关合同的主要权利义务、服务成果以及定价依据等，是否独立于训练数据定制服务和训练数据产品；

一、训练数据相关的应用服务的具体内容、相关合同的主要权利义务、服务成果以及定价依据

（一）具体内容

训练数据相关的应用服务是指发行人基于其生产的训练数据为客户提供 AI 算法模型相关的训练服务，运用训练数据研发能力助力下游客户完成其 AI 算法模型的语言拓展、特定算法模块拓展、垂直应用领域拓展等，为客户定制针对特定应用场景的专属 AI 算法模型，提高 AI 技术应用效果。

具体内容主要包括：①依据客户需求，在客户通用算法模型的基础上，为客户定制开发特定语种/方言、应用领域的 AI 算法模型或②使用特定语种/方言、特定应用领域的训练数据，对客户的 AI 算法模型进行训练，使客户的算法模型实现语种/方言覆盖能力的拓展，满足其在特定行业、地区或者特定应用领域的运用需要。

（二）相关合同的主要权利义务：具体约定根据具体合同有所差异，核心约定通常为：客户对其算法模型相关的软件功能提出要求，发行人为客户的 AI 算法模型相关软件系统提供具体应用层面的开发、调试、优化、安装、使用培训等技术支持和实施服务工作，客户提供相应配合并支付项目费用。

（三）服务成果及形式：通常为软件形式或软硬件一体化销售。

（四）定价依据：根据客户的具体需求，评估项目实施难度、周期和成本，针对每一个项目进行单独的价格协商后确定价格。

（五）补充举例说明

发行人为某主要客户提供某特定地区方言的算法模型训练服务，具体情况为：该客户拟开发覆盖多语种/方言的语音识别算法模型，但因涉及的语种/方言类别众多、训练工作量较大，为节省模型开发训练时间、提升工作效率，该客户将某一特定地区方言的少量算法模型的开发、训练工作交由发行人执行。在此服务项目过程中，发行人执行的具体工作内容及义务为：运用自有的特定方言训练数据对客户的算法模型进行训练和调优，使其达到客户要求的语音识别准确率效果，之后将训练、调优完成后的算法模型（软件形式）交付予客户。

发行人为客户开发、训练的算法模型需要依托一定的运行环境、硬件系统，例如服务器等硬件设备需要满足一定的参数、设备要求。为节省时间、提升效率，少数客户在委托发行人为其进行少量算法模型开发、训练工作的同时，也直接向发行人采购符合算法模型运行要求的硬件设备，发行人直接将开发、训练完成的算法模型部署在硬件设备中，以软硬件结合的形式向客户交付算法模型。

二、是否独立于训练数据定制服务、训练数据产品

训练数据相关的应用服务独立于训练数据定制服务、训练数据产品，具体说明如下：

训练数据定制服务、训练数据产品的核心都是为客户提供训练数据；而训练数据相关的应用服务不涉及为客户提供训练数据，而是主要为客户的 AI 算法模型提供应用层面的开发、调试、优化、安装等相关训练服务，与前述两项业务完全独立。

（5）训练数据定制服务和训练数据产品的目标客户是否存在重合及重合情况，两项业务是否构成替代关系。

一、训练数据定制服务和训练数据产品的目标客户范围一致

训练数据定制服务、训练数据产品业务的核心均在于为客户提供可供 AI 算法模型训练使用的训练数据，两类业务目标客户范围一致，均为人工智能产业链上的各类机构，包括阿里巴巴、腾讯、百度、微软、三星等大型科技公司，科大讯飞、商汤科技、云知声、海康威视等人工智能企业，及中国科学院、清华大学等科研机构。对同一客户而言，训练数据定制服务和训练数据产品可运用于不同的模型训练阶段和训练目标。

二、训练数据定制服务及训练数据产品存在一定替代关系，二者共同构成发行人完整的训练数据服务体系

训练数据定制服务与训练数据产品均为公司主营业务的重要组成部分，二者在公司整体销售策略的把控下存在一定的替代关系，共同构成公司完整的训练数据服务体系。阿里巴巴、腾讯、百度、微软、三星等公司主要客户普遍存在同时采购训练数据定制服务与训练数据产品的情况，训练数据定制服务可满足客户设备系统独特性、应用领域特殊性持续提升衍生的差异化训练数据需求；训练数据产品则可满足客户通用型算法模型需要的通用训练数据的快速采购需求。在客户自身技术和产品研发的不同阶段，发行人的训练数据定制服务和训练数据产品都可以针对不同需求提供支持，如新产品或应用首次研发并上市（使用训练数据定制服务或训练数据产品）、持续性能提升（使用训练数据定制服务-纯加工服务）、应用领域拓展（使用训练数据定制服务或针对特定领域的训练数据产品）、语种拓展（使用训练数据定制服务或使用多语种训练数据产品）等。

9.2 关于主要客户和境内外销售

招股说明书披露：（1）报告期各期发行人营业收入分别为 11,907.09 万元、19,265.77 万元和 23,755.81 万元，前五大客户占当期营业收入的比重分别为 51.22%、59.06%和 53.88%；（2）报告期各期境内销售收入占比分别为 70.94%、69.00%和 50.40%，逐年下降；（3）报告期内，发行人分别取得境外收入 3,460.17 万元、5,972.00 万元和 11,782.07 万元，占主营业务收入的比重分别为 29.06%、31.00%和 49.60%。

请发行人补充披露：（1）按照产品类别披露报告期各期前五大客户以及主要客户的获取方式、合作历史，是否存在应履行公开招投标程序而未履行的情形；（2）报告期各期前五大客户的直接签约方或接受服务方名称、销售内容及其对应的营业收入和占比，结合报告期各期对该等客户的项目服务量、产品数量销售情况，分析发行人对上述客户在报告期各期的营业收入变化原因；（3）境内外前五大客户的公司名称、销售内容及其对应的营业收入和占比；（4）报告期内境外销售具体情况，包括但不限于国家地区、产品种类、销售量、销售单价、销售金额及占比，境外销售模式及流程、主要客户，境外经营是否符合当地规定以及进口国同类产品的竞争格局；发行人境外销售收入逐年上升的原因，是否来源于新增客户，发行人海外业务拓展采取了哪些措施，是否具有可持续性，结合当前疫情、国际形势等因素说明对于发行人经营状况的

影响及发行人的应对措施；（5）报告期内训练数据产品的具体销售情况，包括主要内容、资源类型、数量、授权使用使用次数、平均售价等；（6）训练数据产品是否存在数据资源被泄露、盗版等可能，发行人所采取的保护措施及其有效性，相关风险揭示是否充分。

请发行人说明：（1）上述客户的基本情况与合作历史，发行人获取该等客户的渠道及方式；（2）2019 年度境内销售收入 2018 年下降的原因，是否系市场竞争加剧所致。

请保荐机构、申报会计师对上述事项进行核查，并发表明确意见。

答复：

发行人补充披露事项

（1）按照产品类别披露报告期各期前五大客户以及主要客户的获取方式、合作历史，是否存在应履行公开招投标程序而未履行的情形；

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/（二）主要客户情况/4、按产品类别分类的前五大客户情况及 7、涉及公开招投标程序的情况说明”中披露下述回复的相关内容：

一、按照产品类别披露报告期各期前五大客户以及主要客户的获取方式、合作历史

报告期内，训练数据定制服务和训练数据产品为发行人最主要的业务类别，各年收入占比合计均超过 98%；训练数据相关的应用服务收入占比较低，尚处于起步拓展阶段。因此下文主要针对训练数据定制服务、训练数据产品披露前五大客户及其具体情况：

（一）训练数据定制服务

训练数据定制服务各期前五大客户及其收入情况如下：

金额单位：万元

2020 年 1-3 月				
序号	客户名称（品牌）	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	某国际消费电子产品厂商	611.65	53.76%	19.57%

2	阿里巴巴	101.11	8.89%	3.23%
3	微软	64.91	5.71%	2.08%
4	INTER GROUP ³	52.34	4.60%	1.67%
5	地平线 ⁴	51.63	4.54%	1.65%
	合计	881.64	77.49%	28.20%
	训练数据定制服务收入总额	1,137.75	100.00%	36.40%
	营业收入总额	3,126.09	-	100.00%
2019年				
序号	客户名称(品牌)	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	三星	4,189.52	29.44%	17.64%
2	阿里巴巴	2,150.34	15.11%	9.05%
3	微软	1,385.40	9.73%	5.83%
4	腾讯	1,074.74	7.55%	4.52%
5	某国际消费电子产品厂商	920.39	6.47%	3.87%
	合计	9,720.39	68.30%	40.92%
	训练数据定制服务收入总额	14,232.91	100.00%	59.91%
	营业收入总额	23,755.81	-	100.00%
2018年				
序号	客户名称(品牌)	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	阿里巴巴	4,452.41	35.99%	23.11%
2	三星	2,269.89	18.35%	11.78%
3	腾讯	1,193.79	9.65%	6.20%
4	微软	972.26	7.86%	5.05%
5	百度	391.19	3.16%	2.03%
	合计	9,279.52	75.02%	48.17%
	训练数据定制服务收入总额	12,369.55	100.00%	64.20%

³ Inter Group: Inter Group Corp. 成立于1966年, 是日本的一家集团企业, 主要提供活动会议策划及设施管理, 口译、翻译, 语言专家派遣介绍, 自然语言及语音信息处理、IT等相关业务。2020年1-3月该品牌客户与发行人合作的签约主体为 Inter Group Corp.

⁴ 地平线: Horizon Robotics. 边缘人工智能芯片及解决方案设计企业, 主要依托软硬结合产品向行业客户提供“芯片+算法+云”的解决方案。在智能驾驶领域已赋能奥迪、博世、长安、上汽、广汽等国内外厂商。2020年1-3月该品牌客户与发行人合作的签约主体为北京地平线信息技术有限公司、南京地平线集成电路有限公司。

营业收入总额		19,265.77	-	100.00%
2017年				
序号	客户名称(品牌)	营业收入	占训练数据定制服务收入比重	占营业收入比重
1	三星	1,112.18	17.66%	9.34%
2	阿里巴巴	783.47	12.44%	6.58%
3	腾讯	764.41	12.14%	6.42%
4	微软	552.94	8.78%	4.64%
5	某国际消费电子产品厂商	533.03	8.46%	4.48%
合计		3,746.03	59.49%	31.46%
训练数据定制服务收入总额		6,297.23	100.00%	52.89%
营业收入总额		11,907.09	-	100.00%

(二) 训练数据产品

训练数据产品各期前五大客户及其收入情况如下：

金额单位：万元

2020年1-3月				
序号	客户名称(品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	Google	547.75	27.55%	17.52%
2	客户1	272.22	13.69%	8.71%
3	某大学	251.57	12.65%	8.05%
4	客户3	217.42	10.93%	6.95%
5	腾讯	187.69	9.44%	6.00%
合计		1,476.64	74.27%	47.24%
训练数据产品收入总额		1,988.34	100.00%	63.60%
营业收入总额		3,126.09	-	100.00%
2019年				
序号	客户名称(品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	微软	1,490.80	16.25%	6.28%
2	亚马逊	901.45	9.82%	3.79%
3	客户2	666.66	7.26%	2.81%
4	三星	572.03	6.23%	2.41%
5	中国某大型科技公司	525.53	5.73%	2.21%
合计		4,156.47	45.29%	17.50%

训练数据产品收入总额	9,176.47	100.00%	38.63%	
营业收入总额	23,755.81	-	100.00%	
2018年				
序号	客户名称(品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	Google	804.89	12.19%	4.18%
2	阿里巴巴	726.67	11.01%	3.77%
3	中国某大型科技公司	622.51	9.43%	3.23%
4	百度	497.10	7.53%	2.58%
5	亚马逊	422.42	6.40%	2.19%
合计		3,073.60	46.56%	15.95%
训练数据产品收入总额	6,601.67	100.00%	34.27%	
营业收入总额	19,265.77	-	100.00%	
2017年				
序号	客户名称(品牌)	营业收入	占训练数据产品收入比重	占营业收入比重
1	腾讯	1,005.64	18.32%	8.45%
2	微软	505.93	9.22%	4.25%
3	三星	426.75	7.77%	3.58%
4	阿里巴巴	326.84	5.95%	2.74%
5	中国某大型科技公司	239.76	4.37%	2.01%
合计		2,504.92	45.63%	21.04%
训练数据产品收入总额	5,489.31	100.00%	46.10%	
营业收入总额	11,907.09	-	100.00%	

(三) 主要客户的获取方式、合作历史

发行人获取客户的方式主要包括：

情形一：于公司成立初期即开始合作，早期客户开拓主要来自公司创始人贺琳多年积累的人脉基础；

情形二：通过学术会议接触，之后经过客户拜访、意向接触、商务洽谈等步骤建立合作关系；

情形三：因公司在行业内有良好的口碑，合作过的客户向业内其他客户推荐发行人从而建立联系，进而开展合作；

情形四：通过网络挖掘，获得客户的公共联系方式，然后有针对性地向客户发送邮件主动进行宣传推广，进而建立联系、开展合作；

情形五：通过电话销售，获得客户联系人信息，进而通过邮件沟通、客户拜访等方式获取客户需求，进行合作；

情形六：客户通过公司官网主动向发行人留言咨询，后续通过意向接触、商务洽谈等步骤建立合作。

上述主要客户的获取方式、合作历史情况如下：

主要客户名称（品牌）	获取方式	合作起始时间
阿里巴巴	情形二	2010
百度	情形三	2011
三星	情形二	2010
腾讯	情形二、情形三	2012
微软	情形一	2005
中国某大型科技公司	情形二、情形四	2012
地平线	情形三	2015
亚马逊	情形二	2012
某国际消费电子产品厂商	情形四	2010
Google	情形二、情形五	2017
客户 1	情形三、情形六	2016
某大学	情形二、情形六	2019
INTER GROUP	情形三、情形六	2015
客户 2	情形三、情形六	2015
客户 3	情形二	2012

报告期内，发行人的销售模式与行业普遍模式相同，即主要采用直接对接并服务客户的直销模式进行销售。基于多年的行业经验和知识储备，公司以高品质的训练数据产品及服务吸引客户，并在持续服务的过程中提升服务价值和客户黏度。

公司获取客户的具体方式主要是：通过口碑传播（客户介绍）、参与学术会议、官方网站展示等方式建立品牌知名度以及与客户之间的沟通联系，后续按照客户的采购要求以商务谈判、招投标等形式获取业务机会。

初次获取客户后，发行人会通过常规方式维持与客户的稳定交流和持续沟通，不断挖掘新的业务机会、开展合作，具体举措包括：及时响应客户主动下达的订单需求，与客户及时沟通并相应提供产品服务；不定期联系、拜访客户，了解客户业务最新发展动向和潜在的训练数据需求，并向客户推介发行人新增的训练数据产品、服务等。

二、是否存在应履行公开招投标程序而未履行的情形

（一）应履行公开招投标程序的情形说明

根据《中华人民共和国政府采购法》及其实施条例，政府采购是指各级国家机关、事业单位和团体组织，使用财政性资金采购依法制定的集中采购目录以内的或者采购限额标准以上的货物、工程和服务的行为，政府采购方式包括公开招标、邀请招标、竞争性谈判、单一来源采购、询价等；政府采购限额标准，属于中央预算的政府采购项目，由国务院确定并公布；属于地方预算的政府采购项目，由省、自治区、直辖市人民政府或者其授权的机构确定并公布。

报告期内，发行人客户主要为人工智能产业链上的各类知名机构，如阿里巴巴、腾讯、百度、微软、三星、亚马逊等提供服务，仅少量业务涉及根据上述规定应执行政府采购制度的国家机关、事业单位等客户（包含中央预算单位、地方预算单位），该等情形极少且销售额较小，具体参见下述（二）论述。

（二）发行人不存在应履行公开招投标程序而未履行的情形

1、向中央预算单位客户销售的情形

根据《国务院办公厅关于印发中央预算单位 2017-2018 年政府集中采购目录及标准的通知》（国办发〔2016〕96 号）、《财政部关于中央预算单位政府集中采购目录及标准有关问题的通知》（财库〔2019〕2 号）、《国务院办公厅关于印发中央预算单位政府集中采购目录及标准（2020 年版）的通知》（国办发〔2019〕55 号），报告期内，中央预算单位进行政府采购，单项采购金额达到 200 万元以上的，必须采用公开招标方式。

报告期内，发行人涉及向科研院所等中央预算单位销售的销售合同共 63 份，合同金额合计 737.81 万元，单份合同金额最高为 112 万元，未达到依据上述规定需采用公开招标方式的 200 万元单项采购金额标准，不存在应履行公开招投标程序而未履行的情形。

2、向地方预算单位客户销售的情形

报告期内，发行人涉及向地方所属高校等地方预算单位销售的销售合同共 9 份，合同金额合计 42.6 万元，单份合同金额最高为 15 万元。经查询相应各地招投标相关规定，上述合同金额均未达到执行政府采购制度的限额标准，不存在应履行公开招投标程序而未履行的情形。

综上，报告期内发行人不存在应履行公开招投标程序而未履行的情形。

(2) 报告期各期前五大客户的直接签约方或接受服务方名称、销售内容及其对应的营业收入和占比，结合报告期各期对该等客户的项目服务量、产品数量销售情况，分析发行人对上述客户在报告期各期的营业收入变化原因；

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/(二) 主要客户情况/2、前五大客户销售内容及营业收入变动分析”中披露下述回复的相关内容：

报告期内各期，发行人前五大客户的直接签约方或接受服务方名称、销售内容及其对应的营业收入和占比明细具体如下：

单位：万元

2020年1-3月					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1		某国际消费电子产品厂商	611.65	19.57%	计算机视觉、自然语言-训练数据定制服务
2	Google	Google LLC.	547.75	17.52%	智能语音-训练数据产品
3		客户 1	272.22	8.71%	智能语音-训练数据产品
4		某大学	251.57	8.05%	智能语音-训练数据产品
5	腾讯	腾讯科技(成都)有限公司	236.71	7.57%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务
合计			1,919.90	61.42%	
营业收入总额			3,126.09	100.00%	
2019年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	三星	Samsung Electronics Co., Ltd.、北京三星通信技术研究有限公司、Samsung R&D Institute India-Bangalore Pvt. Ltd.	4,762.06	20.05%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务
2	微软	微软(中国)有限公司、Microsoft Corporation、Microsoft India R&D Private	2,876.20	12.11%	智能语音-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务、自然语言-训练数据定制

		Limited			服务及训练数据产品
3	阿里巴巴	淘宝(中国)软件有限公司、浙江天猫技术有限公司、广州神马移动信息科技有限公司、广州阿里巴巴文学信息技术有限公司、阿里云计算有限公司、支付宝(中国)网络技术有限公司、浙江菜鸟供应链管理有限公司、阿里巴巴(中国)网络技术有限公司、上海盒马网络科技有限公司、阿里巴巴丽人购(上海)电子商务有限公司、阿里巴巴科技(北京)有限公司	2,208.26	9.30%	智能语音-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务;自然语言-训练数据定制服务及训练数据产品
4	亚马逊	Amazon Web Services, Inc.	1,492.98	6.28%	智能语音-训练数据产品、智能语音-训练数据定制服务
5	腾讯	腾讯科技(成都)有限公司、腾讯科技(深圳)有限公司、深圳市腾讯计算机系统有限公司、财付通支付科技有限公司	1,459.31	6.14%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务及训练数据产品
合计			12,798.80	53.88%	
营业收入总额			23,755.81	100.00%	
2018年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	阿里巴巴	淘宝(中国)软件有限公司、阿里巴巴(中国)网络技术有限公司、浙江天猫技术有限公司、广州阿里巴巴文学信息技术有限公司、支付宝(中国)网络技术有限公司、阿里云计算有限公司、广州阿里巴巴文学信息技术有限公司北京分公司	5,179.07	26.88%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务
2	三星	Samsung Electronics Co., Ltd.、北京三星通信技术研究有限公司、Samsung R&D Institute India-Bangalore Pvt. Ltd.	2,436.81	12.65%	智能语音-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务;自然语言-训练数据定制服务及训练数据产品
3	腾讯	腾讯科技(深圳)有限公司、深圳市腾讯计算机系统有限公司、腾讯科技(成都)有限公司、腾讯科技(上海)有限公司、腾讯数码(天津)有限公司	1,574.16	8.17%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务
4	微软	微软(中国)有限公司、Microsoft Corporation	1,299.73	6.75%	计算机视觉-训练数据定制服务;智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务
5	百度	北京百度网讯科技有限公司、百度国际科技(深圳)有限公司、百度时代网络技术(北京)有限公司、百度在线网络技术(北京)有限公司	888.29	4.61%	智能语音-训练数据定制服务及训练数据产品;自然语言训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务、及训练数据产品
合计			11,378.06	59.06%	
营业收入总额			19,265.77	100.00%	

2017年					
序号	客户品牌	直接签约方/接受服务方名称	营业收入	占营业收入比重	销售内容
1	腾讯	腾讯科技(深圳)有限公司、深圳市腾讯计算机系统有限公司、腾讯科技(上海)有限公司	1,770.05	14.87%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务及训练数据产品、计算机视觉-训练数据定制服务及训练数据产品
2	三星	北京三星通信技术研究有限公司、Samsung Electronics Co., Ltd.	1,553.08	13.04%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉-训练数据定制服务;训练数据相关的应用服务
3	阿里巴巴	阿里云计算有限公司、淘宝(中国)软件有限公司、浙江天猫技术有限公司、广州神马移动信息科技有限公司、阿里巴巴(北京)软件服务有限公司、支付宝(中国)网络技术有限公司、浙江阿里巴巴机器人有限公司、阿里巴巴(中国)网络技术有限公司	1,132.01	9.51%	智能语音-训练数据定制服务及训练数据产品;自然语言-训练数据定制服务及训练数据产品;计算机视觉训练数据定制服务及训练数据产品;训练数据相关的应用服务
4	微软	微软(中国)有限公司、Microsoft Corporation	1,058.87	8.89%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务及训练数据产品
5	百度	北京百度网讯科技有限公司	585.29	4.92%	智能语音-训练数据定制服务及训练数据产品、自然语言-训练数据定制服务
合计			6,099.29	51.22%	
营业收入总额			11,907.09	100.00%	

对发行人在报告期内向上述客户销售收入变化原因具体分析如下:

1、2020年第一季度时间窗口较短,且受新冠肺炎疫情影响,发行人原料数据的现场集中采集、标注受到一定影响,整体收入规模较小。该季度,前五大客户主要向发行人采购了计算机视觉、自然语言领域及智能语音领域的训练数据定制服务,以及智能语音领域的训练数据产品。

2、2017年-2019年,发行人各年前五大客户较为稳定,包括三星、微软、阿里巴巴、亚马逊、腾讯及百度共计6个人工智能产业链上的知名机构。前述6个主要客户在2017年-2019年间的营业收入变动情况如下:

金额单位:万元

项目	营业收入		
	2019年	2018年	2017年
主要客户合计	13,343.70	11,843.06	6,134.76

前述6个主要客户在2017-2019年间贡献的营业收入合计数持续增长。2017-2018年，前述主要客户营业收入增长较快，主要是其在2018年对智能语音类训练数据采购量迅速增长所致；2019年，主要客户营业收入整体仍呈现增长趋势，但部分客户在2018年购置的训练数据可以在一段时间内支撑其产品或应用拓展和性能迭代，因此2019年的同类训练数据量需求有所减少，导致2019年采购量有所下降，使得主要客户营业收入整体增速有所放缓。

(3) 境内外⁵前五大客户的公司名称、销售内容及其对应的营业收入和占比；

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/(二) 主要客户情况/5、境内外前五大客户及其销售情况”中披露下述回复的相关内容：

一、境内前五大客户及其情况

金额单位：万元

2020年1-3月					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	腾讯	236.71	30.90%	7.57%	智能语音-训练数据产品、智能语音-训练数据定制服务、自然语言-训练数据定制服务
2	阿里巴巴	162.66	21.24%	5.20%	自然语言-训练数据定制服务、智能语音-训练数据产品、智能语音-训练数据定制服务、计算机视觉-训练数据产品、计算机视觉-训练数据定制服务
3	微软	64.91	8.47%	2.08%	自然语言-训练数据定制服务、计算机视觉-训练数据定制服务
4	地平线	51.63	6.74%	1.65%	智能语音-训练数据定制服务
5	中国某大型科技公司	50.96	6.65%	1.63%	计算机视觉-训练数据定制服务
合计		566.87	74.01%	18.13%	
境内收入总额		765.99	100.00%	24.50%	
营业收入总额		3,126.09		100.00%	
2019年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	阿里巴巴	2,208.26	18.44%	9.30%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务
2	腾讯	1,459.31	12.19%	6.14%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品、计算机视觉-训练数据产品、自然语言-训练数据定制服务、计

⁵ 此处以客户主体所在地区划分境内外客户；部分客户品牌在多个国家/地区均有主体分布，因此可能同时为境发行人境内前五大客户及境外前五大客户。

					算机视觉-训练数据定制服务
3	微软	1,415.71	11.82%	5.96%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、自然语言-训练数据定制服务、智能语音-训练数据产品
4	中国某大型科技公司	1,019.94	8.52%	4.29%	计算机视觉-训练数据产品、智能语音-训练数据定制服务、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品
5	百度	544.90	4.55%	2.29%	智能语音-训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据定制服务
合计		6,648.11	55.52%	27.99%	
境内收入总额		11,973.75	100.00%	50.40%	
营业收入总额		23,755.81		100.00%	
2018年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	阿里巴巴	5,179.07	38.96%	26.88%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品、计算机视觉-训练数据定制服务
2	腾讯	1,574.16	11.84%	8.17%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据产品、自然语言-训练数据定制服务、计算机视觉-训练数据定制服务
3	微软	957.75	7.20%	4.97%	计算机视觉-训练数据定制服务、智能语音-训练数据定制服务、自然语言-训练数据定制服务
4	百度	888.29	6.68%	4.61%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品
5	中国某大型科技公司	650.97	4.90%	3.38%	智能语音-训练数据产品、自然语言-训练数据产品、计算机视觉-训练数据产品、自然语言-训练数据定制服务、智能语音-训练数据定制服务
合计		9,250.24	69.58%	48.01%	
境内收入总额		13,293.77	100.00%	69.00%	
营业收入总额		19,265.77		100.00%	
2017年					
序号	客户品牌	营业收入	占境内收入比重	占营业收入比重	销售内容
1	腾讯	1,770.05	20.95%	14.87%	智能语音-训练数据产品、智能语音-训练数据定制服务、自然语言-训练数据产品、计算机视觉-训练数据定制服务、计算机视觉-训练数据产品
2	三星	1,439.12	17.04%	12.09%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品、计算机视觉-训练数据定制服务、其他-训练数据相关的应用服务
3	阿里巴巴	1,132.01	13.40%	9.51%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务、自然语言-训练数据产品、计算机视觉-训练数据产品、计算机视觉-训练数据定制服务
4	百度	585.29	6.93%	4.92%	智能语音-训练数据定制服务、智能语音-训练数据产品、自然语言-训练数据定制服务
5	微软	580.87	6.88%	4.88%	智能语音-训练数据定制服务、自然语言-训练数

					据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品、计算机视觉-训练数据产品
合计	5,507.34	65.20%	46.25%		
境内收入总额	8,446.92	100.00%	70.94%		
营业收入总额	11,907.09		100.00%		

二、境外前五大客户及其情况

金额单位：万元

2020年1-3月					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	某国际消费电子产品厂商	611.65	25.92%	19.57%	计算机视觉-训练数据定制服务、自然语言-训练数据定制服务
2	Google	547.75	23.21%	17.52%	智能语音-训练数据产品
3	客户1	272.22	11.53%	8.71%	智能语音-训练数据产品
4	某大学	256.15	10.85%	8.19%	智能语音-训练数据产品
5	客户3	217.42	9.21%	6.95%	智能语音-训练数据产品
合计		1,905.19	80.72%	60.94%	
境外收入总额		2,360.10	100.00%	75.50%	
营业收入总额		3,126.09		100.00%	
2019年					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	三星	4,269.02	36.23%	17.97%	智能语音-训练数据定制服务、智能语音-训练数据产品
2	亚马逊	1,492.98	12.67%	6.28%	智能语音-训练数据产品、智能语音-训练数据定制服务
3	微软	1,460.49	12.40%	6.15%	智能语音-训练数据产品、自然语言-训练数据产品
4	某国际消费电子产品厂商	920.39	7.81%	3.87%	计算机视觉-训练数据定制服务、自然语言-训练数据定制服务
5	客户2	666.66	5.66%	2.81%	智能语音-训练数据产品、自然语言-训练数据产品
合计		8,809.54	74.77%	37.08%	
境外收入总额		11,782.07	100.00%	49.60%	
营业收入总额		23,755.81		100.00%	
2018年					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	三星	1,833.67	30.70%	9.52%	智能语音-训练数据定制服务、计算机视觉-训练数据定制服务、智能语音-训练数据产品
2	Google	804.89	13.48%	4.18%	智能语音-训练数据产品

3	亚马逊	465.00	7.79%	2.41%	智能语音-训练数据产品、智能语音-训练数据定制服务
4	纽康 ⁶	371.15	6.21%	1.93%	智能语音-训练数据定制服务、训练数据相关的应用服务
5	微软	341.98	5.73%	1.78%	智能语音-训练数据产品、自然语言-训练数据产品
合计		3,816.69	63.91%	19.81%	
境外收入总额		5,972.00	100.00%	31.00%	
营业收入总额		19,265.77		100.00%	
2017年					
序号	客户品牌	营业收入	占境外收入比重	占营业收入比重	销售内容
1	某国际消费电子产品厂商	580.16	16.77%	4.87%	计算机视觉-训练数据定制服务、自然语言-训练数据定制服务、智能语音-训练数据产品
2	微软	477.99	13.81%	4.01%	智能语音-训练数据产品、计算机视觉-训练数据定制服务、自然语言-训练数据定制服务
3	纽康	359.68	10.39%	3.02%	智能语音-训练数据定制服务、自然语言-训练数据定制服务、其他-训练数据相关的应用服务、自然语言-训练数据产品
4	NAVER ⁷	353.33	10.21%	2.97%	智能语音-训练数据定制服务、智能语音-训练数据产品、其他-训练数据相关的应用服务
5	EML ⁸	190.33	5.50%	1.60%	智能语音-训练数据产品
合计		1,961.49	56.69%	16.47%	
境外收入总额		3,460.17	100.00%	29.06%	
营业收入总额		11,907.09		100.00%	

⁶ 纽康：即纽康品牌旗下企业，报告期内与发行人合作的签约主体包括日本纽康株式会社、海南纽康信息系统有限公司。

⁷ Naver：即 Naver 品牌，报告期内该品牌与发行人合作的签约主体包括 NAVER Corporation、LINE Plus Corporation。前述公司与韩国证券交易所上市公司 Naver Corp. (035420) 隶属同一控制。

⁸ EML：即 European Media Laboratory GmbH，成立于 1997 年，由知名软件公司 SAP 创始人、物理学家 Klaus Tschira (1940-2015) 创立。

(4) 报告期内境外销售具体情况，包括但不限于国家地区、产品种类、销售量、销售单价、销售金额及占比，境外销售模式及流程、主要客户，境外经营是否符合当地规定以及进口国同类产品的竞争格局；发行人境外销售收入逐年上升的原因，是否来源于新增客户，发行人海外业务拓展采取了哪些措施，是否具有可持续性，结合当前疫情、国际形势等因素说明对于发行人经营状况的影响及发行人的应对措施；

一、报告期内境外销售具体情况，包括但不限于国家地区、产品种类、销售量、销售单价、销售金额及占比，境外销售模式及流程、主要客户

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/(二) 主要客户情况/6、境外销售具体情况”中披露下述回复的相关内容：

(一) 境外销售主要国家地区、产品服务种类、各国家地区主要客户

金额单位：万元

序号	国家地区	产品服务类别	2020年1-3月		2019年		2018年		2017年		主要客户
			营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	营业收入	占营业收入总额比重	
1	日韩	总收入	655.09	20.96%	5,688.34	23.95%	2,531.45	13.14%	1,491.68	12.53%	三星(韩国)、纽康、Naver、客户2、客户1
		其中：训练数据定制服务	52.34	1.67%	3,973.37	16.73%	2,473.59	12.84%	705.72	5.93%	
		训练数据产品	602.75	19.28%	1,714.97	7.22%	48.96	0.25%	747.34	6.28%	
		训练数据相关的应用服务	-	-	-	-	8.90	0.05%	38.62	0.32%	
2	美国	总收入	1,603.47	51.29%	4,628.51	19.48%	2,496.28	12.96%	1,494.79	12.55%	微软、亚马逊、某国际消费电子产品厂商、Google
		其中：训练数据定制服务	634.59	20.30%	1,521.43	6.40%	342.36	1.78%	679.15	5.70%	
		训练数据产品	968.88	30.99%	3,107.08	13.08%	2,153.92	11.18%	792.85	6.66%	
		训练数据相关的应用服务	-	-	-	-	-	0.00%	22.80	0.19%	
3	其他	总收入	101.54	3.25%	1,465.21	6.17%	944.27	4.90%	473.70	3.98%	Verint ⁹ 、三星(印度)、Nuance ¹⁰
		其中：训练数据定制服务	37.72	1.21%	742.31	3.12%	624.16	3.24%	209.57	1.76%	
		训练数据产品	63.82	2.04%	713.23	3.00%	319.64	1.66%	257.83	2.17%	
		训练数据相关的应用服务	-	-	9.67	0.04%	0.47	0.00%	6.30	0.05%	

⁹ Verint：即 Verint 品牌，报告期内发行人与该品牌合作的签约主体包括 Verint Systems Inc.（已于纳斯达克主板上市，股票代码 VRNT）与 Verint CES Ltd.。

¹⁰ Nuance：即 Nuance 品牌，报告期内发行人与该品牌合作的签约主体包括 Nuance Communications International BVBA、Nuance Communications Ireland Limited、Nuance Communications Services Ireland Limited 及纽昂司软件技术（北京）有限公司，前述主体与纳斯达克上市公司 Nuance Communications Inc. (NUAN) 隶属同一控制。

境外收入合计	2,360.10	75.50%	11,782.07	49.60%	5,972.00	31.00%	3,460.17	29.06%	
营业收入总额	3,126.09	100.00%	23,755.81	100.00%	19,265.77	100.00%	11,907.09	100.00%	

注：训练数据相关的应用服务各年境外收入占比较低，下文销量、单价情况主要列示训练数据定制服务、训练数据产品相关情况。

（二）境外销售主要国家地区的销售量、销售单价情况

境外销售主要国家地区的销量单价情况如下：

1、训练数据定制服务

类别	收入（万元）				销量					单价				
	2020年1-3月	2019年	2018年	2017年	销量单位	2020年1-3月	2019年	2018年	2017年	单价单位	2020年1-3月	2019年	2018年	2017年
智能语音	111.38	5,032.28	3,002.51	915.06	小时	2,072	11,302	17,219	6,884	元/小时	537.61	4,452.74	1,743.67	1,329.26
计算机视觉	477.03	807.02	266.86	531.45	个	5	6	15	8	元/个	954,051.96	1,345,028.75	177,907.56	664,313.80
自然语言	136.24	397.81	170.74	147.93	个	8	55	63	37	元/个	170,306.17	72,329.31	27,102.24	39,979.88

2、训练数据产品

类别	收入（万元）				销量					单价				
	2020年1-3月	2019年	2018年	2017年	销量单位	2020年1-3月	2019年	2018年	2017年	单价单位	2020年1-3月	2019年	2018年	2017年
智能语音	1,614.30	5,292.45	2,477.03	1,705.12	小时	16,483	57,366	46,493	15,991	元/小时	979.37	922.58	532.77	1,066.28
计算机视觉	-	-	2.28	-	个	-	-	1	-	元/个	-	-	22,822.21	-
自然语言	21.15	242.84	43.20	92.89	个	3	16	5	9	元/个	70,497.65	151,775.42	86,398.93	103,210.58

3、说明

在境外销售主要国家/地区，发行人同类业务产品/服务单价基本稳定，但受客户议价情况、具体项目需求、项目复杂程度，以及涉及的语种/方言等多方面因素影响，也存在一定波动。各年计算机视觉、自然语言按项目（销售的训练数据集）个数计算单价，因境外客户上述类别项目规模通常较大，促使销售单价均较高。

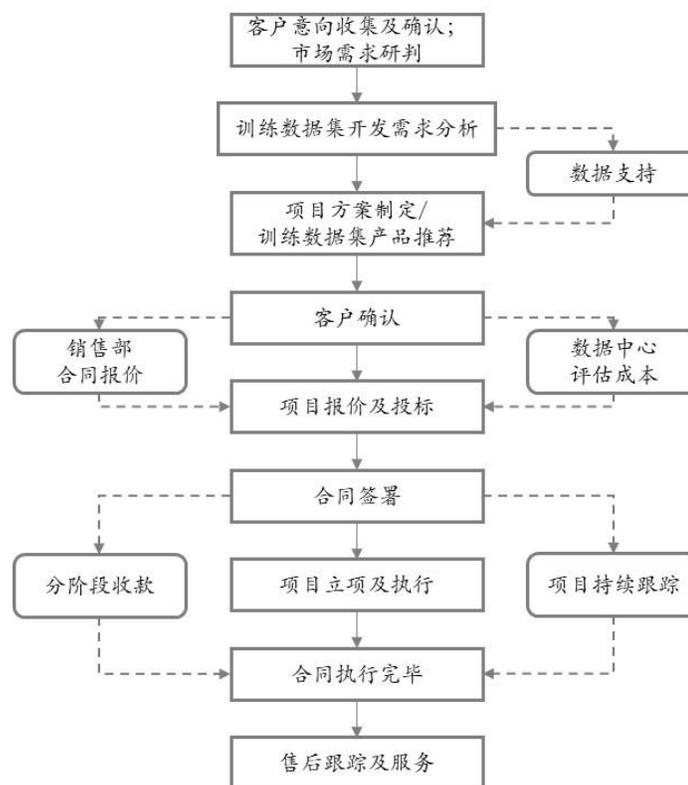
（三）境外销售模式及流程

发行人境外销售的模式、流程与发行人通用的销售模式、销售流程无差异，发行人已在招股说明书“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/（三）主要模式/6、销售模式”中披露了相关内容，具体如下：

1、销售模式

公司采用直接对接并服务客户的直销模式进行营销，符合行业通行惯例。公司以高品质的训练数据产品和服务吸引客户，并在持续服务客户的过程中提升服务价值和客户黏度。公司通过口碑传播、参与学术会议、官方网站展示等方式建立品牌知名度、与客户建立联系，后续再通过商务谈判、招投标等形式获取具体业务机会。

2、销售流程



图：销售流程图

二、境外经营是否符合当地规定及进口国同类产品的竞争格局

发行人已在招股说明书“第六节 业务和技术/七、发行人的境外经营及境外资产情况”、“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（四）发行人产品或服务的市场地位/5、市场地位、竞争格局在报告期内的变化情况及未来变化趋势”中补充披露下述回复的相关内容：

（一）发行人的境外经营符合当地规定

发行人主要在中华人民共和国境内进行生产经营，其全资子公司香港海天瑞声在境外承担发行人的少量境外业务拓展和服务采购职能。除香港子公司外，发行人在境外无经营场所或经营主体；发行人及其子公司为境外客户提供产品/服务，涉及境外销售。

1、香港子公司经营情况

根据君合律师事务所出具的法律意见书，截至2020年5月19日，香港海天瑞声有效登记、存续，其依据与发行人签署的业务协议内容开展业务不需要向香港监管机构或部门申请、取得任何执照或许可。香港子公司经营符合当地规定。

2、境外销售情况

发行人的境外销售主要涉及美国、韩国、日本等地，上述国家对信息技术等货品进口的基本规定主要如下：

国家	法律法规	相关规定
美国	《特别指定国民和被封锁人员清单》	美国人被禁止与特别指定国民清单中的各方进行交易，并且必须封锁其在全球各地占有或控制的特别指定国民清单上的各方的任何资产。任何人寻求与被封锁方进行交易，必须首先向海外资产控制办公室申请特别许可，除非海外资产控制办公室一般豁免许可允许交易。
韩国	《对外贸易法》	1、在属于以下各款之一情况下，产业资源部长可根据总统令规定，采取特别措施，限制或禁止物品的进口或出口：（1）在韩国或韩国的贸易对象国发生战争、事变或自然灾害时；（2）在交易对象国否认一般公认的国际协定中所规定的韩国的权益时；（3）当交易对象国对同韩国的贸易增加不正当或不合理的负担或限制时；（4）为履行依据宪法缔结并公布的有关贸易条约和一般公认的国际法规中规定的维持国际和平与安全的义务必需时；（5）为保护人类健康和生命安全或动植物的健康、生命和环境安全以及保护国内资源所必需时； 2、当国内产业处于以下各项之一的情况时，与该产业有利害关系者或掌管该产业的有关行政机关之长可向贸易委员会提出对该特定产品的进口、贸易及流通服务对国内产业造成的影响进行调查。 （1）特定物品进口增加对生产同类物品或生产有直接竞争关系的物品的国内产业的发展造成严重影响或可能造成影响时；（2）外国人（包括依据大韩民国法

		<p>律设立的法人企业和二分之一股份由外国人持有的法人企业)从事贸易、流通服务的增加对从事同类贸易、流通服务的国内产业或有直接竞争关系的国内产业造成严重影响或可能造成严重影响时；(3)进口商品损害到受国内法律保护的特许权、实用新型权、外观设计权、商标权、著作权、著作邻接权、软件开发权及半导体集成电路的配置设计权,使得生产该产品或与该产品有直接竞争关系的产品的国内产业受到影响或有可能受到影响时；</p> <p>3、贸易委员会通过对国内产业被侵害的调查认定国内产业遭受侵害后,自判定日起45日内可建议有关行政机关之在一定时日内采取以下各项之一的措施。</p> <p>(1)限制进口物品的数量(2)调整进口关税税率(3)根据有关法律对工业农林水产业、矿业的中小企业提供技术支援(4)终止或禁止特定物品或特定贸易交易者的进口(5)为救济其他国内产业实行总统令规定的措施</p>
日本	《外汇与外国贸易法》	<p>1、为对外贸易和国民经济的健康发展而进口货物的业者,依据政令的规定,有义务取得进口许可。</p> <p>2、居住者与非居住者之间拟转让非居住者实施的工业产权及其他有关技术的权利,或订立有关这些权利的使用权,以及签订或更新关于事业经营的技术指导合同,或者变更该合同的条款,如果在这些当中有政令规定的属于需要审查的,依据政令的规定,事先必须把签订该引进技术合同的条款及其他政令规定的事项,向大藏大臣和行业主管大臣申报;在接到申报的情况下,大藏大臣及行业主管大臣,认为有必要审查该申报的引进技术合同,如存在会损害国家安全、妨碍维护公共秩序或妨碍保护人民群众的安全、给予日本经济的顺利发展带来严重的不利影响的引进技术合同等时,可以延长不得签订与该申报有关的引进技术合同的期限。期限为自受理之日起四个月。</p>

经核查,发行人境外销售主要涉及美国、韩国、日本等国家,前述国家主要针对可能严重损害数字经济安全、国家安全、国民健康、产业安全等的情形进行了限制,未直接限制、禁止从发行人进口数据产品或接受发行人提供数据服务。发行人严格遵守境外销售国家/地区的相关规定,为境外客户提供训练数据服务或产品,不存在被列入负面清单或违反当地规定的情形。

综上,发行人香港子公司境外经营符合当地规定;发行人境外销售主要涉及的国家并无禁止或限制数据进口的相关规定,发行人向境外客户销售训练数据产品及服务符合当地规定。

(二) 进口国同类产品的竞争格局

发行人的产品和服务的主要出口地包括美国、日韩等国家或地区。在美国市场,发行人面对的主要竞争对手有 Appen、Globalme、DefinedCrowd、Scale AI 等;在日韩市场,发行人面对的主要竞争对手依然以 Appen、Globalme、DefinedCrowd 为主,同时还有一些本土的数据服务商如: Baobab (日本) 和 AIMMO (韩国) 等。

在美国市场上，由于美国 AI 技术发展比中国早，AI 数据服务商的整体成熟度高于中国。语言文字同源，文化同源等因素的影响，使得澳洲公司、欧美公司在美国市场有着先天的信任优势和沟通优势，发行人在这方面有一定竞争劣势；但发行人也存在以下竞争优势：1、公司现有 600 余个自有知识产权的训练数据产品，相对竞争对手有一定优势，可以帮助客户更快更好地进行算法优化；2、公司已经拥有 150 余个语种/方言的覆盖能力，不仅包括含英法德日意等常见语种，还包括东南亚、一带一路等国家地区的罕见小语种，尤其在亚洲小语种的服务上，具备竞争优势。

在日韩市场上，日韩科技公司或 AI 公司也会使用 Appen 以及部分美国数据供应商。同时，本土的数据服务供应商也会参与竞争。在日韩市场，与欧美竞争对手相比，公司在训练数据产品库和覆盖语言能力，尤其是亚洲语种覆盖方面具备竞争优势；与日韩本土服务商相比，公司在此领域深耕多年，具备多语种服务能力，数据资源来源广泛、储备丰富，具备竞争优势。

三、发行人境外销售收入逐年上升的原因，是否来源于新增客户，发行人海外业务拓展采取了哪些措施，是否具有可持续性，结合当前疫情、国际形势等因素说明对于发行人经营状况的影响及发行人的应对措施

(一) 发行人境外收入逐年上升原因及是否来源于新增客户

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/(一) 营业收入分析/5、按区域分类的主营业务收入分析/(2) 境外收入来源于新增客户的情况”中进行了补充披露：

报告期内各期，发行人境外收入按是否来源于新增客户拆分如下：

境外国家/地区	2020 年 1-3 月		2019 年		2018 年		2017 年	
	新增客户收入占比	原有客户收入占比	新增客户收入占比	原有客户收入占比	新增客户收入占比	原有客户收入占比	新增客户收入占比	原有客户收入占比
日韩	-	100%	5.39%	94.61%	1.11%	98.89%	19.97%	80.03%
美国	-	100%	9.84%	90.16%	21.00%	79.00%	16.86%	83.14%
其他	-	100%	36.37%	63.63%	36.11%	63.89%	13.32%	86.68%

境外收入合计	-	100%	10.99%	89.01%	14.96%	85.04%	17.72%	82.28%
--------	---	------	--------	--------	--------	--------	--------	--------

报告期内，发行人各期境外收入均主要来源于原有客户，而非新增客户。对各年发行人境外收入逐年上升的原因具体分析如下：

2020年第一季度时间窗口较短，且新冠病毒疫情病毒主要在境内流行，对发行人境内业务开展产生了一定影响，当期境外收入占比较高。2017-2019年，发行人境外收入呈现上升趋势，原因主要是：

1、部分客户品牌变更采购签约主体：如大客户品牌三星向发行人采购金额持续增加，且因自身业务经营和管理需要，变更其向发行人执行采购的签约主体，主要签约主体从境内主体中国三星变更为境外主体韩国三星。报告期内，发行人向三星销售收入按签约主体拆分情况如下：

单位：万元

销售收入	2019年	2018年	2017年
三星	4,762.06	2,436.81	1,553.08
境内签约主体	493.04	603.15	1,439.12
境外签约主体	4,269.02	1,833.67	113.96

2、境外客户因自身业务发展需要拓展采购需求：来自境外主要客户亚马逊、微软（境外签约主体）、某国际消费电子产品厂商的销售收入持续增长。报告期内，前述境外主要客户收入增长情况如下所示：

单位：万元

销售收入	2019年	2018年	2017年
亚马逊（境外签约主体）	1,492.98	465.00	35.46
微软（境外签约主体）	1,460.49	341.98	477.99
某国际消费电子产品厂商	920.39	278.97	580.16

亚马逊、微软销售收入增长主要是前述客户拓展其算法模型的语种/方言覆盖，加大对外语种训练数据产品需求和采购力度所致。某国际消费电子产品厂商销售收入增长主要是其对滑行输入训练数据需求持续增长所致。

3、发行人增聘了境外销售人员、加强境外业务开拓力度：发行人高度重视境外业务开拓，持续增聘境外销售人员，增强境外销售团队实力，助力公司境外业务开拓、新客户拓展。发行人注重加强与原有客户的沟通联系、稳固原定境外客户合作关系、拓展业务合作范围；增强新客户开拓和对接，储备业务合作基础。2019年至今，发行人新建立合作关系的境外客户共计30家，覆盖多家知名大型科技公司、人工智能企业及科研院所。

综上，发行人境外收入增加主要是：部分主要客户变更境内外采购签约主体；境外客户因其自身业务发展需要、算法模型的语种覆盖需求拓展、对新兴品类训练数据需求增加等因素增加训练数据采购以及发行人提升海外业务拓展力度所致。

（二）发行人海外业务拓展采取了哪些措施，是否具有可持续性

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/5、按区域分类的主营业务收入分析/（3）发行人海外业务拓展措施”中进行了补充披露：

针对海外业务拓展，发行人采取了下述措施：

1、建设内部团队：增聘境外销售人员

发行人高度重视境外业务开拓，持续增聘境外销售人员，通过境外销售人员进行有规律的信息收集，获得海外市场与客户的一手信息。同时加强对境外销售人员的产品培训力度，增强境外销售团队实力，助力公司境外业务开拓、新客户拓展。

2、跟进境外优质客户的训练数据需求，拓展业务增长点

在业务开展过程中，对于原有客户，发行人注重与客户的长期关系维系和持续服务，持续跟进下游客户业务的发展方向，随之布局、拓展相应的训练数据产品服务类型和应用领域，以持续满足客户需求、拓展营收增长点。

3、加强在海外市场上线线下的宣传投放力度，持续针对海外客户及潜在客户进行品牌宣传

线上推广部分：发行人持续针对海外目标客户所在国家和地区进行搜索引擎广告投放，吸引新老客户的关注及问询。同时发行人还在全球知名的职场社交平台“领英”进行

广告投放，提升品牌的境外知名度、美誉度。同时，发行人还会根据不同的市场活动，在海外学术论坛发布推广信息。

线下推广部分：发行人持续主动参加国际学术会议，增加展位部署，提升在海外客户市场的宣传曝光力度、吸引海外客户。例如：发行人每年都会参加全球人工智能领域的权威学术会议（如：CVPR-计算机视觉/ICASSP-综合/INTERSPEECH-智能语音等），发行人作为参展商向与会的众多人工智能产业机构进行品牌宣传并直接获客；并通过参与学术会议跟进 AI 技术发展中的前沿方向，以此作为后续产品服务开发拓展的重要参考。

通过采取上述措施，发行人对外持续加强针对海外市场的新增客户开拓能力、原有客户新增业务挖掘能力；对内持续提升训练数据产品服务提供能力。同时，下游 AI 行业的持续发展，AI 算法模型的垂直领域、新兴语种覆盖拓展需求预期均将持续带动对相应训练数据需求的持续增长，发行人的海外业务拓展措施、境外收入均具备可持续性。

（三）结合当前疫情、国际形势等因素说明对于发行人经营状况的影响及发行人的应对措施

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/5、按区域分类的主营业务收入分析/（4）当前疫情、国际形势对发行人经营状况的影响及发行人的应对措施”中进行了补充披露：

疫情、国际形势均属发行人业务的短期影响因素：疫情在短期之内对发行人组织原料数据现场集中采集、标注工作造成了一定影响，发行人通过引导客户采购成品训练数据集（即训练数据产品）平抑相关影响；随着我国疫情逐步稳定、各地有序复工，境内疫情影响逐步减弱；但境外疫情仍在蔓延，发行人的境外原料数据采集、标注工作开展会受到一定影响，发行人通过改善和提升工具性能，引导采集、标注人员使用线上远程采集、标注工具平抑相关影响。国际形势在现阶段对发行人所在行业无实质影响，发行人与境外各客户的商业洽谈、合作和拓展均在稳定、持续进行中。

发行人对此采取的应对措施包括：

1、对于疫情：疫情影响发行人原料数据现场采集、标注工作的开展，因此也对训练数据定制服务的开展造成影响，发行人通过引导客户采购成品训练数据集、改善和提升工具性能，引导采集标注人员使用远程采集、标注工具平抑前述短期影响。

2、对于国际形势：国际形势在现阶段对发行人所在行业无实质影响。为了应对国际形势未来可能造成的影响，发行人采取的措施包括：（1）持续跟进服务现有境内外优质客户；（2）拓展新兴优质客户（互联网领域新兴活跃企业，如今日头条系、拼多多系、小米系、美团系客户等），持续跟进服务前述客户对训练数据的需求；（3）跟进AI前沿领域发展趋势、拓展训练数据产品服务布局覆盖，整体稳固现有客户收入、拓展业务增长点。

此外，发行人已在招股说明书中针对疫情、国际形势等进行了风险提示，具体参见招股说明书“重大事项提示/三、特别风险提示/（三）中美贸易摩擦风险”及“重大事项提示/四、新型冠状病毒肺炎疫情对公司经营情况的影响”、“第四节 风险因素/二、疫情风险及四、中美贸易摩擦风险”。

（5）报告期内训练数据产品的具体销售情况，包括主要内容、资源类型、数量、授权使用次数、平均售价等；

发行人已在招股说明书“第六节 业务和技术/三、销售情况和主要客户/（一）报告期内销售情况/3、训练数据产品销售情况/（3）训练数据产品销售具体情况”中披露下述回复的相关内容：

报告期内，发行人训练数据产品的具体销售情况（主要内容、资源类型、数量、授权使用次数、平均售价）如下：

注：发行人的训练数据产品开发完成后可授权予客户使用，一次生产完成后可重复多次销售使用权。下文列示的“授权使用次数合计”指标为各期实现销售的训练数据产品在当期销售使用权的次数合计值。

1、2020年1-3月训练数据产品的具体销售情况

（1）智能语音领域

序号	资源类型		主要内容	2020年1-3月销售情况				发行人积累的 训练数据产品总数 (个,截至2020 年3月31日)
				当期实现销售 的训练数据产品数量 (个)	授权使用次 数合计(次)	销量(小 时)	平均售价 (元)	
1	智能语音	中文	唤醒词、单句、 对话等	16	17	7,613	486	168
		外语种	唤醒词、单句、 对话等	42	47	15,517	1,013	324

(2) 计算机视觉、自然语言领域

序号	资源类型	主要内容	2020年1-3月销售情况			发行人积累的 训练数据产品总数 (个,截至2020 年3月31日)
			当期实现销售 的训练数据产品数量 (个)	授权使用次数 合计(次)	平均售价 (万元)	
1	自然语言	词典	1	3	7.05	83
2	计算机视觉	人脸、人像目标 等	6	6	4.26	66

2、2019年训练数据产品的具体销售情况

(1) 智能语音领域

序号	资源类型		主要内容	2019年销售情况				发行人积累的 训练数据产品总数 (个,截至2019 年12月31日)
				当期实现销售 的训练数据产品数量 (个)	授权使用次 数合计(次)	销量(小 时)	平均售价 (元)	
1	智能语音	中文	唤醒词、单句、 对话等	79	114	55,957	317	166
		外语种	唤醒词、单句、 对话等	171	245	80,728	793	317

(2) 计算机视觉、自然语言领域

序号	资源类型	主要内容	2019年销售情况			发行人积累的 训练数据产品总数 (个,截至2019 年12月31日)
			当期实现销售 的训练数据产品数 量(个)	授权使用次 数合计(次)	平均售价 (万元)	
1	自然语言	词典、语料等	26	30	12.56	82
2	计算机视觉	人脸、手势、活体对抗、 自然场景、人车2D、交 通标志等	17	25	25.17	54

3、2018年训练数据产品的具体销售情况

(1) 智能语音领域

序号	资源类型		主要内容	2018年销售情况				发行人积累的 训练数据产品总数 (个,截至2018 年12月31日)
				当期实现销售的 训练数据产品 数量(个)	授权使用次 数合计(次)	销量(小 时)	平均售 价(元)	
1	智能语音	中文	唤醒词、单句、 对话等	57	88	68,587	364	151
		外语种	唤醒词、单句、 对话等	90	123	47,145	747	256

(2) 计算机视觉、自然语言领域

序号	资源类型	主要内容	2018年销售情况			发行人积累的 训练数据产品总数 (个截至2018年 12月31日)
			当期实现销售的 训练数据产品数 量(个)	授权使用次 数合计(次)	平均售价 (万元)	
1	自然语言	词典、语料等	17	29	13.72	72
2	计算机视觉	人脸、人车2D、行车、 前/后置人像等	19	25	7.45	37

4、2017年训练数据产品的具体销售情况

(1) 智能语音领域

序号	资源类型		主要内容	2017年销售情况				发行人积累的 训练数据产品总数 (个,截至2017 年12月31日)
				当期实现销售的 训练数据产品 数量(个)	授权使用次 数合计(次)	销量(小 时)	平均售 价(元)	
1	智能语音	中文	唤醒词、单句、 对话等	71	98	41,274	517	143
		外语种	唤醒词、单句、 对话等	81	119	27,247	1,022	240

(2) 计算机视觉、自然语言领域

序号	资源类型	主要内容	2017年销售情况			发行人积累的 训练数据产品总数 (个,截至2017 年12月31日)
			当期实现销售的 训练数据产品数 量(个)	授权使用次 数合计(次)	平均售价 (万元)	
1	自然语言	词典、语料等	9	30	10.60	56
2	计算机视觉	人脸、室内场景、车辆特 征等	5	16	15.65	26

(6) 训练数据产品是否存在数据资源被泄露、盗版等可能，发行人所采取的保护措施及其有效性，相关风险揭示是否充分。

一、训练数据产品是否存在数据资源被泄露、盗版等可能，发行人所采取的保护措施及其有效性

发行人已在招股说明书“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/4、训练数据产品的数据资源泄露、盗版风险及发行人采取的保护措施情况”中披露下述回复的相关内容：

发行人的训练数据产品是以二进制形式存储，理论上存在被泄露或者盗版的可能，针对训练数据产品，发行人采取的保护措施如下：

（一）外部：与客户就训练数据产品保护事项在商业层面进行条款约定

在与客户进行合作时，发行人首先会在商业合同条款中约定训练数据产品保护的相关条款，对此进行约束。其次，考虑到训练数据产品的泄露或者盗版会对发行人的客户也可能造成直接的利益损失（如增加其竞争对手等），因此发行人的客户也有动机主动保护训练数据产品不被泄露。

（二）内部：采用技术保密手段、制定数据保密相关制度对训练数据产品加以保护

1、信息安全管理体制层面——已取得了 ISO/IEC 27001 信息安全管理体制认证

发行人构建了完善的信息系统维护管理制度，确保信息系统能够对数据安全起到有力保障：通过信息系统的日常运行维护、系统更新、安全管理等手段保障服务器安全、应用系统安全、网络安全、物理安全，满足开发管理、防病毒管理、应急管理等方面要求；组织开展对信息系统的安全评估，发现危害数据、信息安全的问题及时整改；定期进行数据备份、软硬件更新，保障数据安全、降低数据流失风险；严禁擅自修改系统软件和应用软件，确需修改需履行审批申请程序，由信息管理员负责，并在维护过程中防止数据丢失、失效。

2、技术层面——构建一体化数据处理平台（以下简称“平台”），嵌入业务过程的数据安全管理需求；运用技术手段保障数据安全

对于在海天瑞声的平台上处理的数据，发行人的一体化数据处理平台可实现下述功能，在数据采集、传输、加工、存储过程中确保数据安全：

（1）可实现私有化部署，加工、质检过程中数据源均保存在海天瑞声的服务器中，不允许随意复制、外泄；

(2) 数据加密存储：对所有存储在平台的数据进行加密存储；涉及移动设备（如移动硬盘、U 盘等）外部移动数据存储的也进行加密处理，并制定了严格的移动设备管理制度；

(3) 数据加密传输：对于平台之间，或平台、客户端之间的数据传输，运用行业通行的加密方式对数据进行加密传输；

(4) 嵌入访问权限控制、身份验证、密码管理功能，限制数据的查看、使用范围，保障数据安全。

对于在客户平台上处理的数据：发行人及其团队可依据客户需求运用客户指定的工具、平台对数据加以处理，严格防范数据外泄。

3、内部控制层面——强化职责分离、权限管理

与上述信息系统安全管理相关举措配合，发行人制定、采取了有效的内控制度及举措，防止因内控弱化导致的数据被随意控制、操作的风险，具体包括：加强信息系统的职责分离、权限管理，加强与信息系统使用、授权相关的制度建设，完善授权审批管理流程等。

4、保密制度层面——签署保密协议，制定保密要求、泄密惩治机制

发行人构建了有力的保密制度，对于数据、技术等各类保密信息制定了相应的保密要求，与员工签署保密协议，约定保密义务；辅以保密教育、泄密惩治机制，确保员工遵守保密要求。

综上，发行人的训练数据产品存在数据资源被泄露、盗版等的可能，发行人采取了有效的保护措施对其进行保护，防止相关泄露、盗版情形发生。但若公司的数据安全管理体系研发升级未能跟上业务发展的需要、或客户未能遵守训练数据产品保护相关商业约定，则公司可能面临训练数据产品被泄露、盗版等数据安全相关风险，详见招股说明书“重大事项提示/三、特别风险提示/（五）数据安全相关风险”。

三、相关风险揭示是否充分

发行人已在招股书“重大事项提示/三、特别风险提示/（五）数据安全相关风险”及“第四节 风险因素/八、数据安全相关风险”中对数据安全相关风险进行了充分提示，具体如下：

“数据安全相关风险

发行人主要从事训练数据的研发设计、生产及销售业务，所提供的产品和服务主要以数据的形式体现。一方面，随着发行人业务的快速发展和规模的持续扩张，原料数据采集与数据加工的数量持续增长；另外一方面，与数据安全、个人信息保护相关的法律规章体系逐步完善。如果将来公司未能根据法律规章的更新要求及时调整现行业务开展方式、公司的数据安全管理体系研发升级未能跟上业务发展的需要、或客户未能遵守训练数据产品保护相关商业约定，则公司可能面临生产经营不符合法律规章的要求、训练数据产品被泄露、盗版等数据安全相关风险或可能产生诉讼纠纷，进而给公司的经营带来不利影响。”

发行人说明事项

（1）上述客户的基本情况与合作历史，发行人获取该等客户的渠道及方式；

依据问题 9.2 补充披露事项（1）列示，发行人获取客户的方式主要包括：

情形一：于公司成立初期即开始合作，早期客户开拓主要来自公司创始人贺琳多年积累的人脉基础；

情形二：通过学术会议接触，之后经过客户拜访、意向接触、商务洽谈等步骤建立合作关系；

情形三：因公司在行业内有良好的口碑，合作过的客户向业内其他客户推荐发行人从而建立联系，进而开展合作；

情形四：通过网络挖掘，获得客户的公共联系方式，然后有针对性地向客户发送邮件主动进行宣传推广，进而建立联系、开展合作；

情形五：通过电话销售，获得客户联系人信息，进而通过客户拜访等方式获取客户需求，进行合作；

情形六：客户通过公司官网主动向发行人留言咨询，后续通过意向接触、商务洽谈等步骤建立合作。

补充披露事项涉及客户的基本情况、合作历史、获取方式，列示如下：

客户名称(品牌)	基本情况	获取方式	合作起始时间
阿里巴巴	<p>阿里巴巴集团控股有限公司（已于纳斯达克主板、香港联交所主板及伦敦证券交易所上市，股票/存托凭证代码：BABA、9988、0HCI）成立于1999年，现已成为大型科技公司集团，旗下主体业务覆盖电商、网上支付、物流、云计算、文娱等多个领域。</p> <p>报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。</p>	情形二	2010
百度	<p>百度股份有限公司（已于纳斯达克主板、新加坡证券交易所国际板、伦敦证券交易所国际板上市，股票/存托凭证代码：BIDU、K3SD、OHL1）成立于2000年，是目前全球最大的中文搜索引擎企业，提供以信息和知识为核心的互联网综合服务，业务领域覆盖搜索、人工智能等。</p> <p>报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。</p>	情形三	2011
三星	<p>三星集团成立于1938年，是韩国的大型跨国企业集团，旗下拥有三星电子（已于韩国证券交易所、伦敦证券交易所主板上市，股票/存托凭证代码分别为005930、SMSD等）、三星物产、三星人寿保险等诸多子公司，业务覆盖电子、金融、机械、化学等众多领域。</p> <p>报告期内，该品牌与发行人合作的签约主体均隶属三星集团。</p>	情形二	2010
腾讯	<p>腾讯控股有限公司（已于香港联交所主板上市，股票代码0700）成立于1999年，是中国最大的互联网综合服务提供商之一，也是中国服务用户最多的互联网企业之一。提供包括社交和通信服务、游戏、新闻、网络视频服务等在内的多元化服务；腾讯旗下腾讯音乐已于纽约证券交易所主板上市，股票代码TME。</p> <p>报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。</p>	情形二、情形三	2012
微软	<p>微软公司（Microsoft Corp，已于纳斯达克主板、伦敦证券交易所国际版、香港联交所主板上市，股票代码分别为MSFT、0QYP、4388）成立于1975年，是一家美国的跨国科技公司，是个人计算机软件开发的先驱者，目前是全球最大的电脑软件提供商，业务覆盖电脑软件的研发、制造、授权及相关服务等。</p> <p>报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。</p>	情形一	2005
中国某大型	是我国领先的信息与通信技术解决方案提供商，可为运	情形二、情形四	2012

科技公司	营商、企业及消费者提供有竞争力的 ICT 解决方案、产品和服务。		
地平线	北京地平线信息技术有限公司成立于 2015 年的人工智能企业，拥有 AI 算法和芯片设计能力，主要通过软硬结合设计开发高性能、低成本、低能耗的边缘 AI 芯片及解决方案；在智能驾驶领域客户包括奥迪、博世、长安、上汽广汽等国内外知名厂商。该企业股东为 Horizon Robotics Holdings Limited（香港注册企业），法定代表人为余凯。报告期内，发行人与该品牌合作的签约主体为北京地平线信息技术有限公司、南京地平线集成电路有限公司。	情形三	2015
亚马逊	Amazon.Com Inc.（已于纳斯达克主板、圣地亚哥证券交易所、伦敦证券交易所国际板上市，股票代码分别为：AMZN、AMZN、OR1O）成立于 1994 年，是美国最大的电商公司，也是最早经营电商的公司之一，已发展为业务覆盖电商、互联网、云服务的大型科技公司。报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。	情形二	2012
某国际消费电子产品厂商	是美国的一家大型科技公司。经营范围包括电脑硬件、软件，消费电子产品、数字发布、零售业务等。	情形四	2010
Google	Alphabet Inc.（已于纳斯达克主板、伦敦证券交易所国际板上市，股票代码分别为 GOOG、OHD6），是美国的一家跨国企业，是全球最大的搜索引擎公司，业务覆盖互联网搜索、云计算、广告业务等，也提供基于互联网的产品与服务。报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。	情形二、情形五	2017
客户 1	成立于 2010 年，是日本的一家贸易公司，主要从事汽车及机械零件的进出口贸易、海外商业软件代理、行业相关的新技术、新产业信息的贸易和中介等。客户覆盖京都大学、日本国家航空航天局、厚生劳动省等知名机构。报告期内，该主体是个别最终需求方的指定代理方，发行人与最终需求方就训练数据产品或服务的具体需求进行直接对接沟通，后续最终需求方出于自身经营管理需要，指定客户 1 为其签约/采购主体，发行人完成训练数据相关产品/服务后向客户 1 进行交付。	情形三、情形六	2016
INTER GROUP	即 Inter Group Corp.，成立于 1966 年，是日本的一家集团企业，主要提供活动会议策划及设施管理，口译、翻译，语言专家派遣介绍，自然语言及语音信息处理、IT 等相关业务。	情形三、情形六	2015
客户 2	成立于 1970 年，是日本的一家贸易企业，主要从事重型电气设备及工业的进出口贸易，海外材料、零件、设备的采购以及海外新产品、新兴市场制造产品的销售。报告期内，该主体是个别最终需求方的指定代理方，发行人与最终需求方就训练数据产品或服务的具体需求进行直接对接沟通，后续最终需求方出于自身经营管理需	情形三、情形六	2015

	要，指定客户 2 为其签约/采购主体，发行人完成训练数据相关产品/服务后向客户 2 进行交付。		
客户 3	成立于 1946 年，是日本的一家大型综合性跨国企业集团，业务覆盖视听、电子游戏、通讯产品、信息技术等领域。	情形二	2012
某大学	成立于 1876 年，是美国知名的研究型大学、北美学术联盟美国大学协会（AAU）的 14 所创始院校之一。	情形二、情形六	2019
纽康	日本纽康株式会社是成立于 1991 年的日本企业，主要提供系统解决方案、医疗系统业务、信息服务、客户服务业务等。 报告期内，该品牌与发行人合作的签约主体包括日本纽康株式会社、海南纽康信息系统有限公司。	情形三、情形六	2015
Naver	Naver Corp.（已于韩国证券交易所主板上市，股票代码：035420）是著名社交软件 Line 的母公司，世界第五大搜索引擎网站，韩国最大的搜索引擎和门户网站。韩国股票市场市值最高的互联网公司。 报告期内，该品牌客户与发行人合作的签约主体均隶属前述上市公司体系。	情形四	2013
EML	European Media Laboratory GmbH，成立于 1997 年，由知名软件公司 SAP 创始人、物理学家 Klaus Tschira（1940-2015）创立，主要从事人机交互领域的研发以及自动语音处理，产品服务可覆盖语音分析、媒体转换、语音信息、语音搜索等。	情形六	2013
Verint	Verint Systems Inc.（已于纳斯达克主板上市，股票代码 VRNT）成立于 1994 年，主要提供可操作的智能解决方案和增值服务，帮助客户收集和分析大量的结构化和非结构化信息，如电话交谈、视频交流等。 报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。	情形三	2010
Nuance	Nuance Communications Inc.（已于纳斯达克主板上市，股票代码 NUAN）成立于 1992 年，是专门从事语音识别软件、图像处理软件及输入法软件研发、销售的公司，覆盖医疗、银行、航空、电信以及汽车等诸多领域。 报告期内，该品牌客户与发行人合作的签约主体隶属于前述上市公司体系，或与前述上市公司受同一控制。	情形三	2010

上述客户中，客户 1、客户 2 分别为个别最终需求方指定的签约/采购主体，发行人与最终需求方直接沟通具体产品/服务需求并执行，完成相关产品/服务后向客户 1、客户 2 交付。前述由最终需求方指定其他签约/采购主体的情形极少，发行人在与客户直接对接沟通后通常采取与客户主体直接签约的形式开展业务合作。

（2）2019 年度境内销售收入较 2018 年下降的原因，是否系市场竞争加剧所致。

发行人 2019 年度境内收入为 11,973.75 万元，较 2018 年度的 13,293.77 万元有所下降，主要原因是：

第一，语音识别采集的中文业务市场竞争激烈，市场需求一定程度被分流，销量有所下降，对发行人境内销售收入造成一定影响；

第二，发行人向主要客户阿里巴巴的销售金额下降，由 2018 年的 5,179.07 万元降至 2019 年的 2,208.26 万元，主要是阿里巴巴对训练数据的需求受其研发计划进展节奏影响有所波动所致，如阿里巴巴在 2018 年购置的语音识别转写训练数据可以在一段时间内支撑其产品或应用性能迭代，因此 2019 年的同类中文转写需求量有所减少；

第三，部分跨国企业变更向发行人执行采购的签约主体，其签约主体由境内主体变更为境外主体，使发行人统计的境内销售收入减少，具体情况请参见本问题“发行人补充披露事项（4）/三、发行人境外销售收入逐年上升的原因，是否来源于新增客户，发行人海外业务拓展采取了哪些措施，是否具有可持续性，结合当前疫情、国际形势等因素说明对于发行人经营状况的影响及发行人的应对措施”的回复内容。

中介机构核查事项

（1）核查程序

1、保荐机构、申报会计师访谈了发行人的主要管理人员，取得了报告期内各期发行人的收入明细及主要客户合同，取得了报告期内训练数据产品的销售数据，查阅了报告期内前五大客户（及其签约方）的销售情况、境内外前五大客户及其销售情况、境外销售情况，对相关项目服务量、产品数量及营收变动进行了分析；

2、保荐机构、申报会计师梳理了报告期内发行人涉及公开招投标程序的项目合同，核查发行人是否存在应履行公开招投标程序而未履行的情形；

3、保荐机构、申报会计师对报告期内的主要客户执行了访谈、函证程序，了解各期主要客户与发行人的主要交易内容及交易情况，确认交易额及应收账款余额；

4、保荐机构、申报会计师查阅了发行人主要境外收入来源地区的法律规定及其相关行业研究报告，对发行人境外销售是否合规、境外收入地区的竞争格局情况进行了核查；

5、保荐机构、申报会计师访谈了发行人的主要管理人员、技术人员，了解发行人采取的训练数据产品保密措施，并对发行人的信息系统进行了专项核查，检验包括数据安全在内的信息系统运行情况。

(2) 核查结论

经核查，保荐机构、申报会计师认为：

1、报告期内，发行人不存在应履行公开招标程序而未履行的情形；

2、结合报告期内对各期前五大客户的项目服务量、产品数量销售情况，前述客户在报告期内营业收入变化合理；

3、发行人境外经营符合当地规定。境外销售收入持续上升主要来源于原有客户，具体原因包括：部分客户品牌因自身业务管理需要变更采购签约主体；境外客户因自身业务发展需要增加采购；发行人增聘了境外销售人员，加强了境外业务拓展及储备力度。发行人为海外业务拓展采取了建设内部团队、持续跟进客户需求、加大线上线下宣传推广力度等措施，具备可持续性；

4、疫情、国际形势均属发行人业务的短期影响因素。发行人通过引导客户采集成品训练数据集、引导采集标注人员使用远程采集、标注工具平抑疫情影响；通过持续跟进服务现有优质客户、拓展新兴优质客户、跟进 AI 前沿领域发展需求平抑国际形势影响；

5、发行人的训练数据产品理论上存在被泄露、盗版可能，发行人采取了多层次保护措施对其进行保护，保护措施有效、相关风险揭示充分；

6、发行人 2019 年境内销售收入较 2018 年下降主要是语音识别采集类中文业务竞争加剧、发行人向个别客户销售金额下降、部分跨国企业将向发行人执行采购的签约主体由境内主体变更为境外主体所致。

9.3 关于销售收入核查

根据申报文件，申报会计师将营业收入认定为关键审计事项，招股说明书未披露关键审计事项。

请发行人补充披露：关键审计事项。

请保荐机构和申报会计师说明：（1）与收入确认相关的关键财务报告内部控制的设计和运行有效性的测试情况；（2）主要客户的实地走访和视频访谈情况；（3）抽样程序样本的选取方法，抽样过程、金额和比例；（4）应收账款和营业收入的函证程序和函证比例，未回函和回函差异金额及比例，主要未回函和回函差异客户的情况以及履行的替代程序。

答复：

发行人补充披露事项

关键审计事项

发行人已在招股说明书“第八节财务会计信息与管理层分析/一、报告期经审计的财务报表/（一）审计意见”中披露如下：

毕马威华振会计师事务所（特殊普通合伙）识别出的关键审计事项如下：

收入确认	
请参阅本节“四、报告期内主要采用的会计政策和会计估计/（十七）收入”所述的会计政策及“九、经营成果分析/（一）营业收入分析”。	
关键审计事项	在审计中如何应对该事项
<p>北京海天瑞声科技股份有限公司及子公司（以下简称“海天瑞声集团”）2017年度、2018年度及2019年度及2020年1-3月（以下简称“报告期”）的收入分别为11,907.09万元、19,265.77万元、23,755.81万元及3,126.09万元，主要包括提供训练数据定制服务收入及训练数据产品收入。训练数据定制服务收入主要是指海天瑞声集团根据客户需求为客户提供训练数据定制服务而收取的收入。训练数据产品收入主要是指海天瑞声集团授权客户永久使用本集团已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。</p> <p>于2017年度、2018年度、2019年度，海天瑞声集团以风险报酬转移时点作为收入确认的判断标准。对于训练数据定制服务收入，采用完工百分比法确认收入，并按照提供给客户并经客户验收和确认的服务量占服务总量的比例计算和确定完工进度。对于训练数据产品收入，按照合同约定向客户提交训练数据产品并经客户验收确认后的时点确认收入。</p> <p>自2020年1月1日起，海天瑞声集团执行《企</p>	<p>与评价收入确认相关的审计程序中包括以下： 了解并评价与收入确认相关的关键财务报告内部控制的设计和运行有效性； 选取样本，检查销售合同或订单，识别与风险和报酬转移或控制权转移相关的合同或订单条款，评价海天瑞声集团各类收入的收入确认政策是否符合企业会计准则的相关要求； 选取各期收入的前十大客户和当期新增主要客户，获取相关公开信息（如通过“全国企业信用信息公示系统”），核对客户的股东、董事和监事情况，并和海天瑞声集团提供的信息进行比对，检查是否存在关联关系； 在抽样的基础上，将各期记录的收入核对至相关的销售合同或订单、客户验收文件、发票等相关支持性依据，以评价相关收入是否按照海天瑞声集团的收入确认政策予以确认； 根据客户的交易特点和性质，选取样本，就于资产负债表日的往来款项余额及报告期间的销售交易金额执行函证程序； 选取重要客户，实地走访或视频访谈，询问其</p>

业会计准则第14号——收入（修订）》（以下简称“新收入准则”），以控制权转移作为收入确认时点的判断依据，并综合考虑合同条款和业务安排来评估新收入准则的影响。经评估认为，对于训练数据定制服务收入，单元定制化训练数据被识别为单项履约义务，属于某一时点履行的履约义务，因此在将单元数据向客户交付并经客户验收确认时确认收入。对于训练数据产品收入，因为训练数据产品属于某一时点履行的知识产权许可授权，因此在向客户交付训练数据产品并经客户验收确认后确认收入。

由于收入是海天瑞声集团的关键业绩指标之一，存在管理层为了达到特定目标或期望而操纵收入确认时点的固有风险或收入被确认于不正确的会计期间的错报风险，同时新收入准则的实施也需要管理层根据合同条款和业务安排评估对收入确认会计政策的影响，该评估涉及管理层重大判断，因此毕马威华振事务所将海天瑞声集团的收入确认识别为关键审计事项。

与海天瑞声集团的主要交易情况，及其是否与海天瑞声集团存在关联方关系，检查客户及交易的真实性，识别是否存在异常情况；

从资产负债表日前后的收入交易中选取样本，检查与收入确认相关的销售合同或订单、客户验收文件、发票等相关支持性依据，评价接近资产负债表日前后记录的收入是否记录于恰当期间；

选取符合特定风险标准的收入会计分录（包括期后重大调整分录），核对至相关支持性文件。

保荐机构和申报会计师说明事项

（1）与收入确认相关的关键财务报告内部控制的设计和运行有效性的测试情况

发行人按照所在行业销售与收款业务的流程特点，制定了《销售管理制度》、《销售信用管理规定》、《项目管理办法》和《会计核算制度》等内部控制制度，明确规定了公司的销售主管部门——营销中心的岗位职责分工、权限范围和审批程序，确保办理销售与收款业务的不兼容岗位相互分离、制约和监督；同时对销售业务的定价原则、赊销业务的信用资格审查、合同审批、客户验收、收入确认、应收账款对账、坏账计提等相关内容作出了明确规定。

发行人与收入确认相关的关键财务报告内部控制的内容（关键控制点、控制活动情况）具体如下：

关键控制点	控制活动
信用管理	公司营销中心对客户信用进行统一管理，具体包括新客户审查、授信额度及付款账期评估、设置、变更、延期、中止等。 财务部负责对客户的信用评级进行审核及审批，以及对信用执行进行日常监督管理。 总经理办公会负责对特殊信用额度申请/变更进行审批。
合同审批	销售人员以销售价格体系和销售政策为依据，与客户进行商务沟通和谈判并与客户协商确定销售合同条款，经销售分管负责人、业务部门负责人、财务部、法务人员、总经理逐级审批。重要合同须经总经理办公会审议。

产品服务交付	项目完成或阶段性工作完成后，将产品或服务交付客户。在交付产品及服务时需向客户发送交付通知，并将交付通知留存且在业务管理平台中上传，作为交付产品/服务的证明。
客户验收	交付完成后，由客户按双方约定形式进行验收。在客户对产品及服务进行验收确认后，需在业务管理平台中填写项目验收信息、上传验收证据文件，经业务部门负责人、财务部审核后，将验收数量、验收金额、验收时间记录在业务管理平台中，作为收入确认的基础信息。
收入确认记账	每日，业务管理平台自动提取验收的业务数据，自动传导至财务系统生成记账凭证，确认收入和应收账款。
应收账款对账	公司营销中心定期与客户对账，并将对账记录进行归档保管；财务部每月与营销中心进行核对确认，将核对结果反馈给营销中心，由其对逾期未收款的部分进行催收。
计提坏账	财务人员根据会计政策计提坏账，登记应收账款明细账，经财务部负责人审核后计提坏账记账凭证。

保荐机构和申报会计师针对上述关键内部控制的设计和运行的有效性情况执行了测试/核查程序，具体情况如下：

1、访谈发行人管理层及销售业务主管人员，了解公司的销售业务流程情况；

2、查阅发行人的《销售管理制度》、《销售信用管理规定》、《项目管理办法》和《会计核算制度》等与发行人销售流程管理、内部控制相关的规章制度，结合发行人销售、收款、收入确认流程情况，对照《企业内部控制基本规范》的要求，评估公司相关内部控制制度的设计合理性；

3、执行穿行测试，将收入相关内部控制制度及流程中规定的内部控制措施，与穿行测试抽取的证据进行核对，验证收入相关内部控制是否有效执行；

4、执行控制测试，根据相关内部控制测试点的发生频率抽取样本，核查相关内部控制执行在报告期内是否具备一贯性，并确保获取证据的充分性。

经核查，保荐机构和申报会计师认为，发行人与收入确认相关的关键财务报告内部控制设计有效并得到有效执行。

（2）主要客户的实地走访和视频访谈情况

保荐机构和申报会计师主要通过实地走访形式对发行人主要客户执行了访谈核查程序。因受境内外新冠肺炎疫情影响，有少量客户访谈以视频访谈等形式进行。通过前述访谈，保荐机构和申报会计师向客户询问了解了：客户的规模、业务经营等基本情况，其与发行人之间的合作历史、业务模式、业务往来的发展情况，向发行人采购的具体内容、结

算方式，对发行人的评价、持续合作意愿，以及发行人在业务合作中的合规情况等，并取得签字确认的访谈纪要。

经访谈核查，保荐机构和申报会计师认为，发行人与主要客户已建立了良好的合作关系，相关销售业务真实、具备商业合理性，不存在异常情形。

报告期内，保荐机构、申报会计师访谈的客户覆盖发行人各期营业收入比例情况如下：

金额单位：万元

金额	2020年1-3月	2019年	2018年	2017年
实地走访	1,232.14	14,979.29	14,624.37	8,882.55
视频访谈等非实地走访方式	1,205.36	3,733.03	1,356.11	375.82
访谈客户覆盖收入合计	2,437.49	18,712.32	15,980.48	9,258.38
占比				
实地走访客户占营业收入比重	39.41%	63.06%	75.91%	74.60%
视频走访等非实地走访方式客户占营业收入比重	38.56%	15.71%	7.04%	3.16%
访谈客户占营业收入比重合计	77.97%	78.77%	82.95%	77.76%

注：1、保荐机构、申报会计师对发行人各期主要客户进行了多次走访/访谈，对于同时接受过实地走访、视频访谈等非实地走访的客户，上表计入“实地走访”计算。

2、因受新冠肺炎疫情影响，2019年、2020年第一季度主要客户视频走访等非实地走访方式占比相比2018年、2017年有所提高。

（3）抽样程序样本的选取方法，抽样过程、金额和比例

针对收入这一关键审计事项，保荐机构、申报会计师结合各期营业收入情况，抽样选取样本，与对应的销售合同或订单、客户验收文件、发票等相关支持性依据进行核对，核查验证营业收入的真实性、准确性。

抽样选取方法为个别重大与随机抽样相结合的方法。抽样过程具体如下：

保荐机构、申报会计师根据发行人的实际情况，以经常性业务的税前利润为基准，设定个别重大标准的金额，将当期收入确认金额超过个别重大标准的项目全部抽取作为样本；对于当期确认收入小于个别重大标准的项目，使用非统计抽样，结合可容忍错报金额、错报风险、样本总体等因素随机选取样本。样本金额和比例详见下表：

单位：万元

项目	2020年1-3月		2019年		2018年		2017年	
	数量	金额	数量	金额	数量	金额	数量	金额
个别重大样本	62	2,674.57	43	8,539.94	59	7,729.27	83	6,137.46
非个别重大样本	25	82.93	229	7,817.13	142	3,850.15	118	2,151.31
抽样样本合计	87	2,757.51	272	16,357.07	201	11,579.42	201	8,288.78
收入金额（总体）	165	3,126.09	1,330	23,755.81	1,110	19,265.77	750	11,907.09
抽样样本项目占比	-	88%	-	69%	-	60%	-	70%

（4）应收账款和营业收入的函证程序和函证比例，未回函和回函差异金额及比例，主要未回函和回函差异客户的情况以及履行的替代程序：

一、应收账款和营业收入的函证程序情况

（一）函证程序设计

保荐机构、申报会计师按照《保荐人尽职调查工作准则》、《中国注册会计师审计准则第1312号—函证》的相关规定，综合考虑海天瑞声的经营环境、内部控制的有效性、账户或交易的性质、被询证者处理询证函的习惯做法及回函的可能性等，各自对海天瑞声报告期内的营业收入、应收账款的函证程序进行了恰当的设计，确定了函证的内容、范围、时间和方式。

保荐机构和申报会计师（本题合称“中介机构”）分别对发行人报告期内的主要客户执行了独立函证程序，发函主要通过邮寄纸质函证方式进行。因受境内外新冠病毒疫情影响，对于少量海外客户或受疫情影响无法纸质寄达的国内客户，中介机构同时发送电子邮件函证。

中介机构在询证函中所列示的信息通常包括：截至资产负债表日针对该客户的应收账款（或预收账款）余额、当年向该客户实现的销售额、当年向该客户开具的发票明细、当年收到该客户付款的明细，以核实公司与客户之间的交易金额、期末往来余额信息。

中介机构根据回函差异情况编制回函差异调节表，对回函差异原因予以核实，并对未回函的客户执行相应的替代程序。

（二）发函程序控制情况

对于通过邮寄方式发放的函证，保荐机构、申报会计师会分别独立核对被询证者的名称、地址的真实性，以确保询证函中的名称、地址等内容的准确性。

对于通过电子邮件发放的函证，保荐机构、申报会计师分别独立执行的控制措施包括但不限于：（1）将收件人邮箱的后缀与被询证者官网显示的企业邮箱后缀进行比对；（2）获取发行人与客户在日常沟通时的往来邮件，确认该邮箱地址的使用者确实是被询证者与发行人日常业务往来的联系人且知晓双方的交易情况；（3）将核对后的盖章纸质函证扫描成 PDF 文件作为邮件附件；（4）由保荐机构、申报会计师分别独立直接向被函证的客户发送函证，并要求对方直接回复至发件的中介机构。

（三）回函程序控制情况

对于通过邮寄方式收到的回函，保荐机构、申报会计师分别执行下述独立检查程序：

（1）检查被询证者确认的询证函是否是原件，是否与发出的询证函是同一份；（2）回函是否由被询证者直接寄给发件方（保荐机构、申报会计师）；（3）寄给发函主体（保荐机构、申报会计师）的回邮信封或快递信封中记录的发件方名称、地址是否与询证函中记载的被询证者名称、地址相符；（4）回邮信封上寄出方的邮戳显示发出城市或地区是否与被询证者的地址一致；（5）被询证者加盖在询证函上的印章以及签名中显示的被询证者名称是否与询证函中记载的被询证者名称一致。

对于通过电子邮件收到的回函，保荐机构、申报会计师分别独立核对：（1）发出回函的发件人是否与保荐机构、申报会计师发函时的收件人一致；（2）客户回复的函证附件是否与保荐机构、申报会计师各自独立发送的原始发函文件一致。

二、报告期内应收账款和营业收入发函金额及函证比例，未回函和回函差异金额及比例情况

（一）保荐机构函证情况

1、营业收入

金额单位：万元

项目		2020年1-3月	2019年度	2018年度	2017年度
收入总额	A	3,126.09	23,755.81	19,265.77	11,907.09

1、发函情况					
发函金额	B	2,857.54	21,668.34	18,055.21	10,813.94
发函比例	C=B/A	91.41%	91.21%	93.72%	90.82%
其中：纸质邮寄发函比例		63.80%	89.90%	93.72%	90.82%
电子邮件发函比例		27.61%	1.32%	-	-
2、回函情况					
回函金额	D	2,125.36	17,969.69	13,516.03	9,708.16
回函比例	E=K/A	62.12%	76.72%	71.10%	82.33%
其中：纸质邮寄回函比例		37.73%	69.58%	63.86%	71.25%
电子邮件回函比例		24.39%	7.13%	7.24%	11.08%
回函差异	F=B-D-H	-183.34	254.95	182.30	95.50
回函差异比例	G=F/A	-5.86%	1.07%	0.95%	0.80%
未回函金额	H	915.53	3,443.70	4,356.88	1,010.28
未回函比例	I=H/A	29.29%	14.50%	22.61%	8.48%
3、回函差异情况说明及调节、替代核查程序					
(1) 回函差异原因					
入账时间及暂估税额差异	J	-183.34	254.95	182.30	95.50
(2) 回函差异调节					
经回函差异调节表调整后回函金额	K=D+J	1,942.02	18,224.64	13,698.34	9,803.66
经回函差异调节表调整后回函差异		-	-	-	-
(3) 未回函替代核查					
替代核查金额	L	915.53	3,443.70	4,356.88	1,010.28
(4) 函证核查结果					
调节回函差异、执行替代程序后可确认的金额	M=K+L	2,857.54	21,668.34	18,055.21	10,813.94
调节回函差异、执行替代程序后可确认的金额占比	N=M/A	91.41%	91.21%	93.72%	90.82%

2、应收账款

金额单位：万元

项目		2020年3月 31日	2019年12月 31日	2018年12月 31日	2017年12月 31日
应收账款余额	A	3,883.67	4,660.36	3,096.75	2,707.57
1、发函情况					
发函金额	B	3,474.57	4,200.26	2,991.29	2,585.77
发函比例	C=B/A	89.47%	90.13%	96.59%	95.50%
其中：纸质邮寄发函比例		76.29%	90.13%	96.59%	95.50%
电子邮件发函比例		13.17%	-	-	-
2、回函情况					
回函金额	D	1,728.53	2,318.26	1,621.26	1,243.82
回函比例	E=K/A	69.75%	78.32%	80.45%	79.88%
其中：纸质邮寄回函比例		53.51%	76.58%	74.41%	78.90%
电子邮件回函比例		16.24%	1.74%	6.04%	0.98%
回函差异	F=B-D-H	980.39	1,331.72	870.15	919.05
回函差异比例	G=F/A	25.24%	28.58%	28.10%	33.94%
未回函金额	H	765.65	550.28	499.88	422.91
未回函比例	I=H/A	19.71%	11.81%	16.14%	15.62%
3、回函差异情况说明及调节、替代核查程序					
(1) 回函差异原因					
入账时间及暂估税额差异	J	980.39	1,331.72	870.15	919.05
(2) 回函差异调节					
经回函差异调节表调整后 回函金额	K=D+J	2,708.92	3,649.98	2,491.41	2,162.87
经回函差异调节表调整后 回函差异		-	-	-	-
(3) 未回函替代核查					
替代核查金额	L	765.65	550.28	499.88	422.91
(4) 函证核查结果					
调节回函差异、执行替代	M=K+L	3,474.57	4,200.26	2,991.29	2,585.77

程序后可确认的金额					
调节回函差异、执行替代程序后可确认的金额占比	N=M/A	89.47%	90.13%	96.59%	95.50%

注 1：上表中列示的 2020 年 3 月 31 日应收账款余额中包括合同资产。

注 2：因受新冠肺炎疫情影响，2020 年第一季度电子邮件函证及回函比例有所提高。

（二）申报会计师函证情况

1、 营业收入

金额单位：万元

项目		2020 年 1-3 月	2019 年度	2018 年度	2017 年度
收入总额	A	3,126.09	23,755.81	19,265.77	11,907.09
1、发函情况					
发函金额	B	2,857.54	21,668.34	17,239.82	10,839.94
发函比例	C=B/A	91.41%	91.21%	89.48%	91.04%
其中：纸质邮寄发函比例		63.80%	89.90%	88.63%	89.91%
电子邮件发函比例		27.61%	1.32%	0.85%	1.12%
2、回函情况					
回函金额	D	2,376.92	17,973.03	15,093.41	9,717.38
回函比例	E=K/A	70.17%	76.72%	79.37%	76.91%
其中：纸质邮寄回函比例		37.73%	69.58%	75.21%	70.91%
电子邮件回函比例		32.44%	7.13%	4.16%	6.00%
回函差异	F=B-D-H	-183.34	251.61	198.19	-559.80
回函差异比例	G=F/A	-5.86%	1.06%	1.03%	-4.70%
未回函金额	H	663.96	3,443.70	1,948.22	1,682.36
未回函比例	I=H/A	21.24%	14.50%	10.11%	14.13%
3、回函差异情况说明及调节、替代核查程序					
(1) 回函差异原因					
入账时间及暂估税额差异	J	-183.34	251.61	198.19	-559.80

(2) 回函差异调节					
经回函差异调节表调整后回函金额	$K = D + J$	2,193.58	18,224.64	15,291.60	9,157.58
经回函差异调节表调整后回函差异		-	-	-	-
(3) 未回函替代核查					
替代核查金额	L	663.96	3,443.70	1,948.22	1,682.36
(4) 函证核查结果					
调节回函差异、执行替代程序后可确认的金额	$M = K + L$	2,857.54	21,668.34	17,239.82	10,839.94
调节回函差异、执行替代程序后可确认的金额占比	$N = M / A$	91.41%	91.21%	89.48%	91.04%

2、应收账款

金额单位：万元

项目		2020年3月31日	2019年12月31日	2018年12月31日	2017年12月31日
应收账款余额	A	3,883.67	4,660.36	3,096.75	2,707.57
1、发函情况					
发函金额	B	3,474.57	4,257.16	3,063.41	2,401.82
发函比例	$C = B / A$	89.47%	91.35%	98.92%	88.71%
其中：纸质邮寄发函比例		76.29%	91.35%	98.92%	85.81%
电子邮件发函比例		13.17%	-	-	2.89%
2、回函情况					
回函金额	D	1,728.53	2,375.16	1,590.74	1,156.32
回函比例	$E = K / A$	69.75%	79.54%	79.41%	76.06%
其中：纸质邮寄回函比例		53.51%	77.80%	78.41%	73.17%
电子邮件回函比例		16.24%	1.74%	1.00%	2.89%
回函差异	$F = B - D - H$	980.39	1,331.72	868.31	903.19
回函差异比例	$G = F / A$	25.24%	28.58%	28.04%	33.36%
未回函金额	H	765.65	550.28	604.35	342.31
未回函比例	$I = H / A$	19.71%	11.81%	19.52%	12.64%

3、回函差异情况说明及调节、替代核查程序					
(1) 回函差异原因					
入账时间及暂估税额差异	J	980.39	1,331.72	868.31	903.19
(2) 回函差异调节					
经回函差异调节表调整后回函金额	K =D+J	2,708.92	3,706.88	2,459.05	2,059.51
经回函差异调节表调整后回函差异		-	-	-	-
(3) 未回函替代核查					
替代核查金额	L	765.65	550.28	604.35	342.31
(4) 函证核查结果					
调节回函差异、执行替代程序后可确认的金额	M=K+L	3,474.57	4,257.16	3,063.41	2,401.82
调节回函差异、执行替代程序后可确认的金额	N=M/A	89.47%	91.35%	98.92%	88.71%

注 1：上表中列示的 2020 年 3 月 31 日应收账款余额中包括合同资产。

注 2：因受新冠肺炎疫情影响，2020 年第一季度电子邮件函证及回函比例有所提高。

针对 2020 年 1-3 月营业收入、应收账款，保荐机构、申报会计师发函范围相同；2017 年-2019 年，保荐机构、申报会计师发函范围略微存在差异，但对报告期内合计营业收入、期末应收账款合计余额的覆盖率均超过 90%。2017 年-2019 年发函范围差异情况汇总如下：

单位：万元

营业收入发函情况	2019 年	2018 年	2017 年
保荐机构发函金额	21,668.34	18,055.21	10,813.94
申报会计师发函金额	21,668.34	17,239.82	10,839.94
营业收入发函差异金额 (保荐机构-申报会计师)	-	815.39	-26.00
情形①保荐机构发函，申报会计师未发函	-	1,045.59	110.37
情形②申报会计师发函，保荐机构未发函	-	230.20	136.37
应收账款发函情况	2019 年 12 月 31 日	2018 年 12 月 31 日	2017 年 12 月 31 日
保荐机构发函金额	4,200.26	2,991.29	2,585.77
申报会计师发函金额	4,257.16	3,063.41	2,401.82

应收账款发函差异金额 (保荐机构-申报会计师)	-56.90	-72.11	183.95
情形①保荐机构发函,申报会计师未发函	-	-	210.09
情形②申报会计师发函,保荐机构未发函	56.90	72.11	26.14
发函范围差异涉及的主要客户	2019年 /2019年12月31日	2018年 /2018年12月31日	2017年 /2017年12月31日
情形①保荐机构发函,申报会计师未发函的主要客户	无	Google Ireland Limited、 Amazon.com Services, Inc.、 杭州逾期管家信息技术有限 责任公司	Google Ireland Limited、乐乐 创新智能科技(北京)有限公 司
情形②申报会计师发函,保荐机构未发函的主要客户	上海未来伙伴机器人 有限公司	Inter Group Corp.、Institute for Infocomm Research、广州阿 里巴巴文学信息技术有限公司	ETRI ¹¹ 、国家计算机网络与信息 安全管理中心、 Facebook,Inc.

2017年-2019年发函范围差异主要是因为：在覆盖各期主要客户、同时满足各次发函金额覆盖比率要求的总体原则下，因保荐机构参与了发行人两次申报科创板上市的筹备工作、申报会计师（毕马威华振事务所）仅参与了发行人本次申报科创板上市的筹备工作，导致双方发函次数不同，所以双方针对2017年-2019年函证范围的选取略有差异，总体发函金额也存在略微差异。

三、报告期末回函情况及针对此类情况执行的程序

（一）未回函的原因说明

报告期内，客户未回函的原因主要包括以下4种情况：

情形1：客户根据其内部管理规定不对外部函证进行处理，并以邮件的形式明确告知；

情形2：客户认为业务已结算完毕，没有再配合处理外部函证的义务；

情形3：客户无专门处理函证的人员和流程，或者客户因相关人员和流程发生变更而未处理函证；

情形4：2020年受新冠疫情影响客户延迟复工、人员流动接触受限等原因，因此部分客户未回函。

¹¹ ETRI: Electronics and Telecommunications Research Institute, 韩国人工智能企业。

（二）主要未回函情况

保荐机构、申报会计师分别对报告期内的主要客户独立执行了函证程序，经保荐机构、申报会计师交叉核对，除个别年度、个别客户差异外，未回复保荐机构、申报会计师函证的客户范围、金额基本一致，将主要未回函客户及其未回函原因合并列示如下：

1、营业收入

报告期内，各期营业收入未回函的主要客户涉及的未回函原因主要为情形 1-3，2020 年第一季度个别客户未回函原因为情形 4。

2、应收账款

报告期内，各期应收账款未回函的主要客户涉及的未回函原因主要为情形 1 及情形 3。

（三）履行的主要替代程序

针对未回函客户，保荐机构、申报会计师执行了以下替代程序：

1、了解发行人与相应客户之间的验收、结算和对账流程，并对具体业务环节中的关键控制点进行测试；

2、针对营业收入未回函客户，执行收入核查程序，取得当年/期实现的销售收入的支持性依据，包括销售合同、订单、交付和验收记录、发票、银行进账单等，并核对查验当年/期记录的销售收入是否与上述原始凭据相符；

3、针对应收账款未回函客户，取得相应的银行进账单及账务处理记录，检查应收账款的期后回款情况；

4、对发行人主要客户进行实地走访或视频访谈，了解客户和发行人之间的合作模式、合作业务规模和结算方式。通过访谈确认发行人与客户业务和交易的真实性。

（四）执行替代程序结果汇总情况

针对未回函情况履行替代程序核查结果汇总如下：

1、保荐机构执行替代程序核查情况

金额单位：万元

项目	2020年/2020年3月31日			2019年/2019年12月31日			2018年/2018年12月31日			2017年/2017年12月31日		
	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比
应收账款	765.65	765.65	100%	550.28	550.28	100%	499.88	499.88	100%	422.91	422.91	100%
营业收入	915.53	915.53	100%	3,443.70	3,443.70	100%	4,356.88	4,356.88	100%	1,010.28	1,010.28	100%

2、申报会计师执行替代程序核查情况

金额单位：万元

项目	2020年/2020年3月31日			2019年/2019年12月31日			2018年/2018年12月31日			2017年/2017年12月31日		
	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比	未回函金额	替代核查金额	占比
应收账款	765.65	765.65	100%	550.28	550.28	100%	604.35	604.35	100%	342.31	342.31	100%
营业收入	663.96	663.96	100%	3,443.70	3,443.70	100%	1,948.22	1,948.22	100%	1,682.36	1,682.36	100%

四、报告期回函差异情况及针对此类情况执行的程序

（一）回函差异的原因说明

发行人根据项目验收进度确认对该客户的应收账款和营业收入；而客户存在根据收到发行人发票时间作为确认采购和应付账款的时点的情形。发行人向客户开具发票的时间与项目验收时间并不完全同步，导致发行人根据项目验收确认的营业收入、应收账款与客户根据收到发票确认的采购额和应付账款产生不一致情况。

（二）针对回函差异执行的核查程序

针对营业收入、应收账款的回函差异情况，保荐机构、申报会计师执行了核查程序，具体如下：

1、了解发行人与客户之间的验收、结算和对账流程，并抽样对业务环节中的关键控制点进行测试。

2、访谈发行人管理层，了解回函差异的产生原因，通过函证取得客户可协助确认的信息（包括发票与付款信息），与发行人账面记载的应收账款金额、营业收入金额、开具发票金额、客户验收金额和收款金额进行核对，分析其勾稽关系的合理性，并对差异情况进行调节；获取营业收入、应收账款相关的合同订单、交付和验收记录、发票、银行流水等支持性依据，对差异原因进行核实检查。

3、检查应收账款的期后回款，并检查核对至银行进账单等支持性依据。

4、对发行人主要客户进行实地走访或视频访谈，了解客户和发行人之间的合作模式、合作业务规模和结算方式。通过访谈确认发行人与客户业务和交易的真实性。

保荐机构、申报会计师分别对报告期内的主要客户独立执行了函证程序，经保荐机构、申报会计师交叉核对，除个别年度、个别客户差异外，保荐机构、申报会计师执行函证程序的客户范围、金额基本一致，下文将营业收入、应收账款的主要函证差异及其调节过程/核查程序情况进行了合并列示。

（三）营业收入回函差异调节过程

发行人以项目交付与验收进度为依据确认对客户的营业收入，而客户存在按照收到发行人的发票作为确认当期采购的情形，双方核算的差异导致营业收入的回函存在差异，具体分析如下：

类型	发行人（以验收为依据）			客户（以发票为依据）			是否需要调节
	数据验收时间	开具发票时间	确认收入	数据验收时间	取得发票时间	确认采购	
情形 1	当期	当期	√	当期	当期	√	否
情形 2	当期	前期	√	当期	前期	×	是
情形 3	当期	尚未	√	当期	尚未	×	
情形 4	尚未	当期	×	尚未	当期	√	
情形 5	前期	当期	×	前期	当期	√	

注：情形 1 即发行人确认收入的时间与客户确认采购的时间一致的情形，该情形下无需对营业收入进行调节。除情形 1 外，其余 4 种情形均可能导致发行人确认收入的时间与客户确认采购的时间不一致，均需要对回函差异进行调节分析。

保荐机构、申报会计师对发行人营业收入回函差异的调节步骤一致，均为：以客户回函的金额为基础，结合验收与开具发票各种情形进行调节，最终与发行人的营业收入调节一致，具体过程如下：

	项目	备注
营业收入回函差异调节过程	客户回函的发票金额	-
	减：当期开票、前期验收	情形 5
	减：当期开票、尚未验收	情形 4

	加：当期验收、前期开票	情形 2
	加：当期验收、尚未开票	情形 3
	调节后的营业收入金额	-

（四）主要客户营业收入回函差异调节过程

2017 年、2018 年、2019 年及 2020 年 1-3 月，主要客户营业收入回函差异调节过程汇总如下：

单位：万元

三年一期各期营业收入回函差异主要客户	发函金额	回函金额	营业收入回函差异调节过程					调节后金额	调节结论
			当期开票前期验收	当期开票尚未验收	当期验收前期开票	当期验收尚未开票	小计		
逻辑关系	①	②	③	④	⑤	⑥	⑦=-③ -④+⑤ +⑥	⑧=②+⑦	
2020 年 1-3 月	583.41	713.39	566.02	60.71	115.61	381.14	-129.98	583.41	一致
2019 年	11,381.30	11,086.37	1,400.05	1,600.01	1,961.20	1,333.79	294.93	11,381.30	一致
2018 年	9,741.25	9,562.42	849.6	1,695.50	331.49	2,392.52	178.91	9,741.25	一致
2017 年	5,618.50	6,186.25	578.36	1,815.49	699.01	1,127.11	-567.73	5,618.50	一致

（五）应收账款回函差异核查过程

报告期内各年/期末，发行人根据年初余额、本年含税营业收入、本年收款、本年只开票税金影响等项目变动的逻辑关系计算确认对客户的应收账款余额。其中，开具发票和本年收款可通过客户回函情况确认，本年营业收入根据客户回函情况进行专门调节。只开票税金影响金额较小，主要通过复核未验收已开票项目（包括开票并收预收款和只开票两种类型）金额进行确认。

基于此，保荐机构、申报会计师对于应收账款回函差异的核查方式一致，均主要通过检查应收账款年初余额、本年含税营业收入、本年收款、本年只开票税金影响与年末应收账款余额的会计核算逻辑等式的一致性进行核查。

（六）主要客户应收账款回函差异核查过程

2017 年、2018 年、2019 年及 2020 年 1-3 月，主要客户应收账款回函差异调节过程汇总如下：

单位：万元

时点	期初	本期销售		只开票税金影响	本期收款	期末	逻辑关系检查	核对结论	期后回款 ¹
	应收账款	营业收入	含税销售额			应收账款			
逻辑关系	①	②	③	④	⑤	⑥	⑦=①+③+④-⑤-⑥	—	—
2020年3月31日	1,538.78	439.23	474.27	-3.57	1,034.13	975.35	-	核对一致	782.80
2019年12月31日	1,573.40	4,517.26	4,825.42	-10.38	4,686.99	1,701.45	-	核对一致	1,701.45
2018年12月31日	1,159.62	6,954.09	7,499.07	-3.41	6,667.03	1,988.25	-	核对一致	1,988.25
2017年12月31日	504.95	2,538.56	2,781.52	20.19	2,193.03	1,113.63	-	核对一致	1,113.63

注1：期后回款截止日期为2020年7月31日。

注2：期末预收账款余额在调节过程中以应收账款负数列示。

注3：只开票税金的影响是由于发行人按合同约定节点开具发票，尚未收到款项，且未验收达到收入确认时点，发行人相应确认应收开票税金和应交销项税所导致。

（七）回函差异调节/核查情况汇总

回函差异调节/核查情况汇总如下：

1、保荐机构回函差异调节/核查情况

金额单位：万元

项目	2020年/2020年3月31日			2019年/2019年12月31日			2018年/2018年12月31日			2017年/2017年12月31日		
	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比
应收账款	980.39	980.39	100%	1,331.72	1,331.72	100%	870.15	870.15	100%	919.05	919.05	100%
营业收入	-183.34	-183.34	100%	254.95	254.95	100%	182.30	182.30	100%	95.50	95.50	100%

2、申报会计师回函差异调节/核查情况

金额单位：万元

项目	2020年/2020年3月31日			2019年/2019年12月31日			2018年/2018年12月31日			2017年/2017年12月31日		
	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比	回函差异金额	调节或核查金额	占比
应收账款	980.39	980.39	100%	1,331.72	1,331.72	100%	868.31	868.31	100%	903.19	903.19	100%
营业收入	-183.34	-183.34	100%	251.61	251.61	100%	198.19	198.19	100%	-559.80	-559.80	100%

五、结论

保荐机构、申报会计师在对发行人报告期内的营业收入、应收账款的函证按照《保荐人尽职调查工作准则》、《中国注册会计师审计准则第 1312 号—函证》执行，对被询证者的选择、询证函的设计、询证函的发出和收回均保持了恰当的控制。针对回函不符事项，保荐机构、申报会计师追加了必要的核查程序，并最终与函证金额调节一致；针对未回函事项，保荐机构、申报会计师实施了有效的替代性核查程序，并最终与函证金额印证一致。

10. 关于采购模式和主要供应商

10.1 关于数据服务采购

招股说明书披露：（1）在原料数据采集、数据标注环节，发行人主要向人力资源服务公司（以下简称服务公司）等采购非核心技术环节的数据采集、标注服务；（2）服务公司采用与不同类型终端人员的组织人员（小外包）对接方式来组织终端人员为发行人提供相应服务。根据申报文件，发行人采购数据服务过程中存在少量通过发行人业务人员事先垫付资金的情形。

请发行人补充披露：（1）前五大服务公司名称、采购内容、采购金额及其占比；（2）发行人向自然人和服务公司的采购内容、采购金额及其占比，服务公司采取小外包和直接对接的采购内容、采购金额及其占比；（3）发行人与服务公司、小外包人员和终端人员之间的合作模式、合同签署过程、实际业务执行过程以及资金支付情况；（4）主要原料数据采集、标注服务提供商是否具备必要的业务资质；（5）结合主要原料数据采集、标注服务提供商的规模、员工数量等，说明其是否有足够能力为发行人提供相关服务，相关服务提供商报告期内是否存在以发行人名义招聘实习生或者兼职人员的情形，如存在，请说明具体情况；发行人是否存在以外包名义按劳务派遣用工形式使用劳动者的情形；（6）发行人报告期内是否存在向个人供应商采购原料数据采集、标注服务的情形，如存在，向个人供应商采购的原因及商业合理性；（7）发行人采取何种措施保证供应商提供的数据满足其采集需求，比如如何确定被采集人的年龄、籍贯、性别、口音特点、采集场地以及标注人员的语言背景、专业技能和学历要求等符合采集需求，发行人采取的相关内部控制措施及其有效性；（8）发行人向个人租赁房产的实际用途与其法定用途是否相符。

请发行人说明：（1）发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效；（2）报告期各期服务公司小外包人员的数量及其变动情况，是否为服务公司员工，是否存在相互重合的情形，发行人和服务公司对小外包人员的管理措施，是否均由发行人员直接对小外包人员进行管理和控制；（3）发行人业务人员事先垫付资金的具体原因、后续结算及差异调整情况，是否符合相关内部控制制度的规定，报告期各期事先垫付资金的金额及占比；（4）结合服务公司提供管理服务的内容、过程和必要性，说明发行人向服务公司支付服务管理费的原因及合理性，加成率的定价依据及公允性，不同公司的加成率之间是否存在差异并说明具体原因；（5）采集和标注单价的具体确定依据和过程，同类型项目的采集和标注单价是否存在差异并分析具体原因；（6）采集数据量、标注数据量和标产比相关的确定方法和具体依据，是否存在随意修改或调整的情形，同类型项目的采集数据量、标注数据量和标产比是否存在差异并分析具体原因；（7）劳务数量及数据服务费用变动是否与发行人经营情况相匹配，数据服务费用的采购价格是否公允，劳务用工成本核算是否完整且不存在跨期，发行人建立的相关内部控制措施及执行情况；（8）前述服务公司的基本情况和员工数量，是否与发行人对其的采购规模相匹配，认定同一控制的依据，通过同一控制下不同企业开展业务的原因及合理性，发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务，主要小外包人员的基本情况，服务公司及其关联方和主要小外包人员是否与发行人客户及其关键经办人员、发行人控股股东、实际控制人、发行人关联方、员工及前员工之间存在关联关系、关联交易、资金往来或其他利益安排，是否存在通过服务公司向客户进行商业贿赂的情形；（9）服务公司作为扣缴义务人是否已替小外包人员或终端人员履行了代扣代缴税款义务，是否存在违反相关法律法规的情形。

请发行人律师对披露事项的（4）-（8）进行核查并发表明确意见。

请申报会计师对上述说明事项进行核查并发表明确意见，并说明：（1）未将营业成本认定为关键审计事项的原因；（2）对数据服务采购发生、准确性和截止性认定履行的具体审计程序；（3）对 IT 审计报告中存在的异常项目履行的进一步审计程序；（4）对服务公司、小外包人员和终端人员履行的具体核查程序、核查过程、核查比例和核查结论。

答复：

发行人补充披露事项

(1) 前五大服务公司名称、采购内容、采购金额及其占比

发行人已在招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（二）主要供应商情况/2、主要数据服务供应商情况”中补充披露下述内容：

报告期内发行人向前五名数据服务供应商（受同一控制的供应商合并计算采购额）采购情况如下表所示：

金额单位：万元

2020年1-3月					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	善世(广东)企业服务外包有限公司 ¹²	否	357.00	24.22%	原料数据采集、标注服务
	共青城宜邦人力资源服务有限公司	否	283.80	19.26%	原料数据标注服务
2	北京博睿驰信人力资源有限公司	否	237.46	16.11%	原料数据采集、标注服务；岗位服务
	恩泽尔商贸有限公司	否	93.34	6.33%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	64.03	4.34%	原料数据标注服务
4	Voices.com Inc	否	44.41	3.01%	原料数据采集服务
5	上海佑译信息科技有限公司	否	31.13	2.11%	原料数据采集服务
合计			1,111.18	75.40%	
2019年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	芜湖善慧人力资源服务有限公司	否	1,512.55	19.80%	原料数据采集、标注服务
	善世(广东)企业服务外包有限公司 ¹³	否	1,374.11	17.98%	原料数据采集、标注服务

¹² 曾用名“广州善世人力资源服务有限公司”。

¹³ 曾用名“广州善世人力资源服务有限公司”。

	共青城宜邦人力资源服务有限公司	否	261.18	3.42%	原料数据标注服务
	才星（广州）科技服务有限公司 ¹⁴	否	5.57	0.07%	原料数据标注服务
2	北京博睿驰信人力资源有限公司	否	1,432.77	18.75%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否	140.69	1.84%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	388.90	5.09%	原料数据采集、标注服务
4	Atrium Studio Medienproduktion GmbH	否	138.71	1.82%	原料数据采集服务
5	SAKURA JAPAN LLC	否	114.55	1.50%	原料数据采集服务
合计			5,369.03	70.27%	
2018年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	广州善世人力资源服务有限公司	否	1,755.67	19.39%	原料数据标注服务
	芜湖善慧人力资源服务有限公司	否	1,541.91	17.03%	原料数据采集、标注服务
	广州金社宝网络技术有限公司	否	26.89	0.30%	原料数据标注服务
2	北京博睿驰信人力资源有限公司	否	1,829.37	20.21%	原料数据采集、标注服务，岗位服务
	恩泽尔商贸有限公司	否	249.09	2.75%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	905.61	10.00%	原料数据采集、标注服务
4	TWIN STUDIOS	否	229.97	2.54%	原料数据采集服务
5	THINKWILD STUDIOS, S. L.	否	129.85	1.43%	原料数据采集服务
合计			6,668.36	73.66%	
2017年					
序号	供应商名称	是否为发行人关联方	采购金额	占采购总额比重	采购内容
1	芜湖善慧人力资源服务有限公司	否	1,064.36	21.02%	原料数据采集、标注服务
	广州善世人力资源服务有限公司	否	658.34	13.00%	原料数据采集、标注服务

¹⁴ 曾用名“广州金社宝网络技术有限公司”。

2	北京博睿驰信人力资源有限公司	否	536.68	10.60%	原料数据采集、标注服务, 岗位服务
	恩泽尔商贸有限公司	否	244.78	4.83%	原料数据采集、标注服务
	北京零点起航营销策划有限公司	否	68.11	1.35%	原料数据采集、标注服务
3	北京冠华英才国际经济技术有限公司	否	662.12	13.08%	原料数据采集、标注服务, 岗位服务
4	SAKURA JAPAN LLC	否	150.69	2.98%	原料数据采集服务
5	信阳市联智网络科技有限公司	否	89.08	1.76%	原料数据采集、标注服务
合计			3,474.15	68.62%	

注：（1）善世（广东）企业服务外包有限公司、芜湖善慧人力资源服务有限公司、才星（广州）科技服务有限公司、共青城宜邦人力资源服务有限公司因共同关联股东持股及共同人员管理而被认定为同一控制主体；（2）北京博睿驰信人力资源有限公司、恩泽尔商贸有限公司均受同一股东控制，且该股东拥有北京零点起航营销策划有限公司的委托经营权，上述三个主体被认定为同一控制主体。

（2）发行人向自然人和服务公司的采购内容、采购金额及其占比，服务公司采取小外包和直接对接的采购内容、采购金额及其占比

发行人已在招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（一）报告期内采购情况/3、发行人向自然人、服务公司采购数据服务情况”中补充披露下述内容：

发行人向自然人和服务公司（包括服务公司采取小外包和直接对接两种形式）的采购内容均为原料数据采集、标注服务，其中自然人为直接向发行人提供服务，服务公司为按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员，组织其使用发行人研发的技术、平台、工具等，为发行人提供合乎标准的数据采集、标注服务。

报告期内，发行人向自然人和服务公司（包括服务公司采取小外包和直接对接两种形式）的采购金额情况如下：

单位：万元

采购方式	2020年1-3月		2019年度		2018年度		2017年度	
	采购金额	占比	采购金额	占比	采购金额	占比	采购金额	占比
通过服务公司采购	1,113.03	95.24%	5,912.58	96.93%	7,231.42	98.35%	3,816.55	94.07%
其中：通过小外包采购	857.21	73.35%	4,281.03	70.18%	4,567.55	62.12%	1,949.36	48.05%
直接对接采购	255.82	21.89%	1,631.55	26.75%	2,663.87	36.23%	1,867.19	46.02%

采购方式	2020年1-3月		2019年度		2018年度		2017年度	
	采购金额	占比	采购金额	占比	采购金额	占比	采购金额	占比
直接向自然人采购	55.60	4.76%	187.37	3.07%	121.24	1.65%	240.76	5.93%
数据服务费采购额	1,168.63	100.00%	6,099.95	100.00%	7,352.66	100.00%	4,057.31	100.00%

(3) 发行人与服务公司、小外包人员和终端人员之间的合作模式、合同签署过程、实际业务执行过程以及资金支付情况

发行人已在招股说明书“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 / (三) 主要经营模式/4、数据服务采购情况介绍”中补充披露下述内容：

一、合作模式

发行人与服务公司、小外包人员和终端人员共同合作完成原料数据的采集和标注工作，具体职责划分如下：

1、发行人职责

- (1) 设计训练数据集结构，确定训练数据集开发标准，制定原料数据采集方案；
- (2) 确定数据采集、标注标准和规程并提供数据采集、标注过程中所需的技术、平台和工具；
- (3) 监督采集、标注工作的具体开展；
- (4) 对采集、标注形成的数据质量进行检测、验证。

2、服务公司职责

- (1) 按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员。在部分具体训练数据生产项目中，涉及的终端采集、标注人员数量较多，且覆盖的语种、口音区、年龄特征等要求多种多样，服务公司也采取与不同类型小外包人员对接的方式，来组织终端人员为发行人提供相应服务；
- (2) 组织上述人员使用发行人研发的技术、平台、工具等，并对服务进度与交付进行管理，保证该等人员提供服务的质量；

(3) 对小外包及直接对接的终端人进行人员管理，包括劳务合同签署、进度安排、支付款项、代扣代缴税款等。

3、小外包人员职责

(1) 按照发行人确定的原料数据采集、标注要求（如发音人的性别、年龄、口音地区等需求）寻找匹配的终端采集、标注人员；

(2) 组织终端人员为发行人提供相应服务。

4、终端人员职责

(1) 为发行人提供合乎标准的数据采集、标注服务；

(2) 授权发行人使用其提供的数据用于业务经营。

二、合同签署过程、实际业务执行过程

1、发行人与服务公司之间的合同签署及实际业务执行情况

(1) 发行人与服务公司等长期签订框架性的数据服务采购协议，约定发行人根据业务需要委托服务公司完成原料数据采集、标注工作，双方根据验收结算单确认的服务量和收费标准结算数据服务费；

(2) 发行人根据服务公司的地区覆盖能力、资源匹配能力和管理能力，对于不同的业务类型选择不同服务公司进行合作，例如智能语音业务涉及人员规模大、覆盖地区多，因此优先选择体量规模相对较大的善世系服务公司（包括善世（广东）企业服务外包有限公司、芜湖善慧人力资源服务有限公司、共青城宜邦人力资源服务有限公司等同一控制主体），服务公司根据下属主体划分的覆盖区域确定具体服务主体；

(3) 发行人在实际业务执行时通过会议、邮件等方式向服务公司沟通具体采购需求，确认服务工期要求，原料数据采集、标注服务量，被采集人的年龄、籍贯、性别、口音特点、采集场地等，以及标注人员的语言背景、专业技能和学历要求等；

(4) 服务公司根据寻找的人力资源情况，将价格信息反馈给发行人，发行人参考业务采购价格体系和过往执行的同类项目的价格，并结合当前项目可能存在的特殊要素要求，与数据服务提供商协商确认价格信息，双方达成一致后确定采集和标注单价；

(5) 发行人在数据服务采购过程中，对采集、标注人员或其组织人员（服务公司人员或小外包人员）进行培训，部署所需的平台和工具，并下达采集、标注的具体要求；

(6) 发行人把控项目执行的时间进度，敦促服务公司及其相关人员按时交付数据，并持续监督、关注交付的数据质量。

2、服务公司与小外包人员以及直接对接的终端人员之间的合同签署及实际业务执行情况

(1) 服务公司与小外包人员以及直接对接的终端人员签署劳务服务合同；

(2) 服务公司在实际业务执行时对其进行监督、管理，包括进度时间安排、人员信息登记、直接对接的终端人员授权签署、人员纠纷解决等，确保其提供发行人所需的数据服务，以及提供数据的完整性；

(3) 服务公司在数据服务采购过程中会持续与发行人进行沟通，获取发行人对采标服务及数据质量的反馈情况，及时反馈给小外包人员和直接对接的终端人员，确保其为发行人提供服务的质量。

3、小外包人员与其管理的终端人员之间的实际业务执行情况

(1) 小外包人员根据发行人数据服务采购需求寻找符合条件的终端人员；

(2) 小外包人员组织并管理终端人员提供劳务，包括人员信息登记、终端人授权签署、人员工作时间安排及现场引导、对终端人员的服务进度和数据完整性进行把控等。

三、资金支付情况

发行人按项目进度向服务公司结算采购费用，统一采用银行转账的形式。服务公司收到资金后，向小外包人员和其直接对接的终端人员支付劳务费用，统一采用银行转账的形式。小外包人员收到资金后向终端人员支付劳务费用。

四、数据服务采购执行过程示例

以 2019 年某唤醒词命令词语音识别采集项目为例，该项目采集人数超过 200 人，采集对象要求多地区、可进行标准美国英语发音、年龄结构多样，采集工期为 15 天，发行人与服务公司、小外包人员和终端人员之间的数据服务采购过程如下：

1、发行人与各数据服务提供商签订采购框架协议，包括芜湖善慧人力资源服务有限公司（以下简称“芜湖善慧”），双方约定服务费率为 10%；

2、发行人综合考虑该项目所需要的语音识别采集人员数量较多、覆盖地区较广、工期较紧等实际情况，参考各服务公司的资源匹配能力和管理能力，并结合服务公司对于下属主体覆盖区域的划分，确定选择芜湖善慧作为该项目的供应商；

3、发行人自测标产比，并与芜湖善慧沟通后达成一致，按照项目所需标注的数据量与标产比的乘积计算所需的标注工时，该项目中所需标注的数据量为 200 小时，约定的标产比是 2.625；

4、发行人以电子邮件形式向芜湖善慧下达采购通知，写明项目简介、服务工期要求，并明确项目采购量为采集服务 210 人，标注服务 525 小时，其中标注服务 525 小时=所需标注的数据量 200 小时*标产比 2.625；

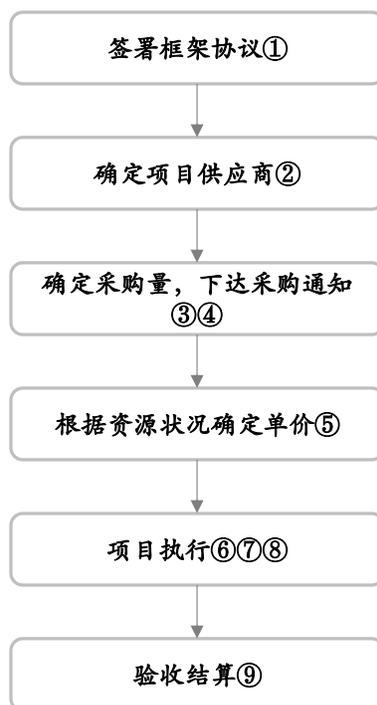
5、芜湖善慧通过小外包人员在上海、深圳、成都等地寻找到足够数量的终端采集人员，同时直接寻找了部分终端标注人员，将报价信息反馈给发行人，经发行人判断符合其业务采购价格体系及同类项目价格区间，双方达成一致确定采集单价和标注单价；

6、服务公司与小外包人员和直接对接的终端人员签订劳务合同，双方正式确定劳务关系；

7、发行人对小外包人员、终端人员进行培训，部署所需的平台和工具，小外包人员组织并管理终端人员开始进行原料数据的采集和标注，服务公司进行全程的监督、管理；

8、服务公司根据原料数据的采集和标注情况，保持与发行人、小外包人员、终端人员等各方的及时沟通，确保采购服务工期进度及数据质量，并向发行人交付采集标注后的数据；

9、发行人对数据进行质量验收后，与服务公司进行结算并出具验收结算单。验收结算单由发行人与服务公司双方盖章确认，记载经发行人验收合格的采集、标注工作量，以及经双方约定一致的采集、标注单价信息。发行人根据验收结算单计算的数据服务费金额（数据服务费=验收工作量*单价）向服务公司进行支付，并按照框架协议中约定的服务费率按比例支付给服务公司。服务公司按照约定将发行人支付的数据服务费部分支付给小外包人员和直接对接的终端人员，小外包人员收到资金后向其寻找的终端人员支付劳务费用。



（4）主要原料数据采集、标注服务提供商是否具备必要的业务资质

发行人已在招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（二）主要供应商情况/2、主要数据服务供应商情况”中补充披露下述内容：

由于相关法律法规并未就原料数据的采集、标注服务规定专门的业务资质许可，主要原料数据采集、标注服务提供商向发行人提供原料数据采集、标注服务不需要取得专门的业务资质。根据中国境内主要原料数据采集、标注服务提供商的营业执照，其为发行人提供数据采集、标注服务未超出其《营业执照》上核准的经营范围。主要原料数据采集、标

注服务提供商与发行人合作，具备提供相关服务的能力，能够按照与发行人的协议约定提供原料数据采集、标注服务。

(5) 结合主要原料数据采集、标注服务提供商的规模、员工数量等，说明其是否有足够能力为发行人提供相关服务，相关服务提供商报告期内是否存在以发行人名义招聘实习生或者兼职人员的情形，如存在，请说明具体情况；发行人是否存在以外包名义按劳务派遣用工形式使用劳动者的情形

一、结合主要原料数据采集、标注服务提供商的规模、员工数量等，说明其是否有足够能力为发行人提供相关服务

发行人已在招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（二）主要供应商情况/3、主要数据服务供应商服务能力、服务提供情况说明”中补充披露下述内容：

（一）主要原料数据采集、标注服务提供商的规模、员工数量等情况如下：

序号	供应商名称	成立日期	注册资本	股权结构	注册地	实际经营地	业务经营情况	合作期间员工数量(人)	合作期间营业收入(万元/年)
1	善世(广东)企业服务外包有限公司	2014年2月	2,100万元	陈玟霏持股40% 白云持股40% 高洁持股20%	广州市白云区齐富路自编5号718房		人力资源外包、劳务承揽、劳务派遣服务等	90-150	13,000-28,000
	共青城宜邦人力资源服务有限公司	2019年3月	200万元	夏建平持股70% 陈礼秋持股30%	江西省九江市共青城市科技一大道66号3楼308室				
	芜湖善慧人力资源服务有限公司	2016年11月	200万元	陈玟霏持股32% 莫天勤持股32% 徐洪超持股30% 欧阳刚持股3% 王宏志持股3%	安徽省芜湖市镜湖区镜湖世纪城绿地新都会办公C座15层077号	安徽省芜湖市镜湖区镜湖世纪城绿地新都会D座16层			
	才星(广州)科技服务有限公司	2016年5月	680万元	王宏志持股40% 莫天勤持股30% 陈玟霏持股30%	广州市海珠区阅江中路832号3605、3606房				
	北京博睿驰信人力资源有限公司	2013年5月	500万元	王驰持股100%	北京市大兴区滨河街27号14层1402				
2	恩泽尔商贸有限公司	2016年9月	5万美元	王驰持股100%	马绍尔群岛	北京市大兴区滨河街27号14层1402	劳务服务、劳务派遣、企业管理咨询等	45-65	2,000-5,200
	北京零点起航营销策划有限公司	2013年12月	200万元	张立文持股50% 刘凡茂持股50%	北京市大兴区滨河街27号14层1402				
	北京冠华英才国际经济技术有限公司	1999年11月	11,000万元	齐宝春持股50% 孙进荣持股50%	北京市平谷区平谷北街甲15号院2号1层101	北京朝阳区北三环安贞桥东胜古家园2号冠华商务会馆2层			
3	北京冠华英才国际经济技术有限公司	1999年11月	11,000万元	齐宝春持股50% 孙进荣持股50%	北京市平谷区平谷北街甲15号院2号1层101	北京朝阳区北三环安贞桥东胜古家园2号冠华商务会馆2层	专业承包、劳务分包、劳务派遣等	约100	19,000-26,000
4	上海佑译信息科技有限公司	2014年4月	19,713.4205万元	杜金林持股26.07%，上海鼎晖麒麟善创业投资中心(有限合伙)持股20.29%，宁波佑译元钛投资合伙企业(有限合伙)持股9.59%，杭州老板实业集团有限公司持	上海市宝山区逸仙路2816号B栋20F		在计算机软硬件专业领域内从事技术开发等	60	5,000

				股 7.83%，上海瞬译信息科技有限公司（有限合伙）持股 7.05%，梁小平持股 6.59%，上海杜晨忠信息科技有限公司（普通合伙）持股 6.34%，北京搜狗信息服务有限公司持股 5.18%，蒋喜文持股 5.07%，深圳市腾讯创业基地发展有限公司持股 2.97%，晋江普发投资有限公司持股 1.10%，杜大永持股 0.96%，杭州海沣投资管理有限公司持股 0.96%				
5	信阳市联智网络科技有限公司	2016年1月	1,000万元	洪琪持股 100%	信阳市羊山新区新八街信阳电子商务产业园大厦二层 205	网络信息技术开发与服务、计算机领域的技术服务等	约 50	180-470
6	SAKURA JAPAN LLC	2016年12月	50 万日元	高见一广持股 60% 劉晓傑持股 40%	栃木県小山市横倉新田 95-208	人力资源服务、技术服务等	约 50	600-1,000
7	THINKWILD STUDIOS, S.L.	2009年5月	3,010 欧元	Rossana Giacomelli 和 Carlos Gomez-Mira 共同持股	C/ Castillo de Fuensaldaña Nº4 Las Rozas, 28232 Madrid Spain	录音制作等	—	—
8	TWIN STUDIOS	1996年1月	48,000 欧元	—	2 AVENUE DU MARECHAL MAUNOURY 75016 Paris 16th Arrondissement	录音制作等	—	—
9	Atrium Studio Medienproduktion GmbH	1998年	—	—	Zielstattstr. 33 81379 München	录音制作等	—	—
10	Voices.com Inc	2003年12月	—	David Ciccarelli 和 Stephanie Ciccarelli 共同持股	100 Dundas St., Suite 700 London, ON N6A 5B6 Canada	录音制作等	约 100	—

注：（1）北京博睿驰信人力资源有限公司、恩泽尔商贸有限公司、北京零点起航营销策划有限公司的合并员工数量及营业收入计算包含与其受同一控制的北京天创共赢科技发展有限公司、北京中惠商客科技有限公司；（2）信阳市联智网络科技有限公司的员工数量及营业收入计算包含与其受同一控制的信阳市晨曦电子科技有限公司；（3）上述部分境外供应商未提供相关资料信息。

（二）发行人对原料数据采集、标注服务提供商的主要需求

发行人从事数据采集、标注所需劳务人员数量较大、临时性较强、对人员响应速度要求较快，且该类工作的执行为非核心技术环节，无需较高技术门槛，而服务提供

商专业从事大量人员招募和管理工作，能够满足发行人相关服务需求，因此发行人主要通过向服务提供商采购的方式完成数据采集、标注的业务环节。

发行人对服务提供商的主要需求如下：

1、接收并传达采集、标注需求：发行人与服务提供商沟通确定数据采集、标注的数量、具体要求、时间进度等，并对服务提供商进行培训，服务提供商接受发行人的服务需求，并将服务要求传达给所联系的劳务人员（包括可以直接提供服务的终端人或具有相应能力的小外包人员）；

2、人员寻找和管理：服务提供商按发行人指定要求寻找上述劳务人员，并对该等劳务人员进行管理，与其签署劳务协议，按服务量向其支付劳务报酬，并为其代扣代缴个税等；

3、工作进度与数据交付管理：服务提供商或小外包人员对终端人进行数据采集、标注的进度等进行监督管理；服务提供商持续与发行人进行沟通，获取发行人对数据服务质量的反馈情况，及时反馈给小外包、直接对接的终端劳务人员，保证其提供服务的质量。

（三）主要原料数据采集、标注服务提供商有足够能力为发行人提供服务

1、主要原料数据采集、标注服务提供商具有相关人员资源

数据采集、标注服务不涉及专门的业务资质。上述表格所列 1-6 项主要原料数据采集、标注服务提供商具体从事对外提供劳务/技术服务的业务经营，积累了相关人员资源；上述表格所列 7-10 项主要原料数据采集、标注服务提供商为专业录音机构，具有专业发音人资源。服务提供商均能够及时响应发行人对数据采集、标注人员的需求。

2、主要原料数据采集、标注服务提供商具有人员管理等业务经验

上述表格所列 1-6 项主要原料数据采集、标注服务提供商具有人员管理、项目管理或技术服务的业务经验，经发行人培训后能够掌握数据采集、标注的规范和流程，能够按发行人要求对数据采集、标注的进度、质量进行管控。上述表格所列 7-10 项主要原料数据采集、标注服务提供商为专业录音机构，具有专业的录音场地和人员及项目管理经验，能够为发行人提供对录音指标、录音环境要求较高的数据采集服务。

上述主要原料数据采集、标注服务提供商在与发行人合作过程中，均按照协议约定提供数据采集、标注服务，未发生因其自身能力问题与发行人停止合作的情形。

综上，主要原料数据采集、标注服务提供商有足够能力为发行人提供相关服务。

二、相关服务提供商报告期内是否存在以发行人名义招聘实习生或者兼职人员的情形，如存在，请说明具体情况

发行人已在招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（二）主要供应商情况/3、主要数据服务供应商服务能力、服务提供情况说明”中补充披露下述内容：

根据相关服务提供商的确认及网络公开检索，相关服务提供商报告期内不存在以发行人名义招聘实习生或者兼职人员的情形。相关服务提供商通过其自身招聘渠道寻找劳务人员，并直接与劳务人员签署劳务协议。根据实际业务需要，发行人存在少量以自身名义临时招聘实习生或兼职人员的情形。

三、发行人是否存在以外包名义按劳务派遣用工形式使用劳动者的情形

发行人已在招股说明书“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 /（三）主要经营模式/5、数据服务采购的合规性说明”中补充披露下述内容：

根据《中华人民共和国劳动合同法》、《中华人民共和国劳动合同法实施条例》、《劳务派遣暂行规定》等的规定，劳务派遣是指用工单位与劳务派遣单位订立劳务派遣协议，劳务派遣单位与被派遣劳动者订立劳动合同，被派遣劳动者在约定的用工单位相应岗位上工作的制度。劳务外包是指企业将部分业务或工作发包给外包服务单位，由外包服务单位安排人员完成相关的业务或工作，企业按照业务或工作的完成情况与外包服务单位进行结算，劳务外包关系适用《中华人民共和国合同法》。

发行人与数据服务提供商之间的业务合作属于劳务外包关系，与劳务派遣的主要区别如下：

区别	劳务派遣	发行人与数据服务提供商之间的业务关系
合同关系	实际用工单位与劳务派遣单位签订劳动派遣协议，劳务派遣单位与劳务派遣人员签订劳动合同。	发行人和数据服务提供商签署《业务服务协议》，对服务提供商采购服务，服务提供商向发行人提供合同约定的服务。服务提供商根据项目需要与劳务人员签署劳务协议

费用 结算	实际用工单位向劳务派遣单位支付劳务派遣费用，具体根据被派遣员工的人数、派遣用工的工作时间和约定的薪酬标准等支付对应费用。	发行人对数据服务提供商提供服务的成果进行验收，验收合格后根据数据量或服务量支付对应的采购费用。
服务人员的 管理	被派遣人员在实际用工单位处工作，接受实际用工单位管理，需要遵守用工单位的相关规章制度，在用工单位的工作时间内专为用工单位提供劳务。	数据服务提供商提供服务对应的终端人员不在发行人处工作，也不接受发行人管理，并非专门为发行人提供劳务。发行人仅按照采购合同约定要求服务提供商提供服务。

综上，发行人与数据服务提供商签署合同，由数据服务提供商提供原料数据的采集、标注等服务，发行人要求数据服务提供商按照约定交付工作成果，并按照服务项目的验收成果进行结算，不对劳务人员具体进行管理，不要求劳务人员遵守发行人规章制度。因此，发行人委托数据服务提供商提供原料数据采集和标注服务构成劳务外包，发行人不存在以外包名义按劳务派遣用工形式使用劳动者的情形。

(6) 发行人报告期内是否存在向个人供应商采购原料数据采集、标注服务的情形，如存在，向个人供应商采购的原因及商业合理性

发行人已在招股说明书“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 / (三) 主要经营模式/4、数据服务采购情况介绍”中补充披露下述相关内容：

报告期内，在个别业务中，发行人存在向个人供应商采购原料数据采集、标注服务的情形。该等向个人供应商采购的金额较低，占比很小，并非发行人主要采用的采购模式，而是作为发行人在个别业务中的少量补充性采购，满足发行人的零散数据服务采购需求，具备商业合理性，并且不对发行人业务经营构成重大影响。发行人向个人供应商采购的具体金额及占比请参见本问题“发行人补充披露事项/(2) 发行人向自然人和服务公司的采购内容、采购金额及其占比，服务公司采取小外包和直接对接的采购内容、采购金额及其占比”。

(7) 发行人采取何种措施保证供应商提供的数据满足其采集需求，比如如何确定被采集人的年龄、籍贯、性别、口音特点、采集场地以及标注人员的语言背景、专业技能和学历要求等符合采集需求，发行人采取的相关内部控制措施及其有效性

发行人已在招股说明书“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 / (三) 主要经营模式/4、数据服务采购情况介绍”中补充披露下述相关内容：

发行人有效执行了一系列质量控制措施，以确保供应商提供的数据满足其采集需求。

1、与供应商充分沟通传达采集、标注需求：发行人在根据项目情况确定采购需求后，即通过会议、邮件等方式与供应商进行沟通，明确被采集人的年龄、籍贯、性别、口音特点、采集场地等，以及标注人员的语言背景、专业技能和学历要求等。对于重要项目，发行人会与供应商就上述要求进行书面确认，确保最终的采集、标注人员符合发行人需求。

2、采集、标注人员属性的事前验证：在原料数据的采集、标注工作开始前，发行人会直接或通过供应商、小外包人员对终端人员进行身份验证，包括身份证明、学历证明、资质证明等，验证被采集人的年龄、籍贯、性别、口音等身份特征和标注人员的专业性。标注人员需要进行试标注，以检查其专业技能，并可能通过多个标注人员进行交叉验证。同时，发行人会对服务公司人员、小外包人员或终端人员进行规则培训，培训合格后正式开始原料数据的采集、标注。

3、数据质量的事中控制：在原料数据的采集、标注过程中，发行人提供技术人员、管理人员，全程监督指导，确保交付的数据质量符合要求。数据采集、标注工作主要在发行人的平台、工具上执行，发行人通过其系统对数据进行自动检验，通过检验的数据才能保留在系统中，进行下一步的生产环节、交付及验收。

4、数据质量的事后验收：在原料数据的采集、标注工作结束后，发行人按照制定的验收标准对采集或标注形成的数据进行验收，包括工具质检、人工复核、抽检等形式，确保其符合发行人需求。对于可在采集、标注后数据中体现的人员属性（如性别、口音等），发行人通过事后对数据进行质检进一步复核是否符合前期提出的采集需求。

上述质量控制措施涉及的发行人内部控制手段为采购前下达采购通知和采购执行过程中的质量检查。发行人采取的相关内部控制措施及其有效性的具体情况请参见本问题“发行人说明事项/（1）发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效/二、与数据服务采购相关内部控制的设计是否合理、执行是否有效”。

综上，发行人采取的质量控制措施能够保证供应商提供的数据满足其采集需求，发行人采取了相关内部控制措施并有效执行。

(8) 发行人向个人租赁房产的实际用途与其法定用途是否相符

发行人已在招股说明书“第六节 业务和技术/五、发行人的主要固定资产和无形资产 / (一) 主要固定资产”中补充披露下述相关内容：

发行人目前向个人租赁的主要房产的用途等情况如下：

序号	出租方	租赁房屋坐落	租赁期限	实际用途	不动产权证记载用途
1	李宏	北京市海淀区成府路28号4座8层801、02、03、06、07、08、09、11	2020年5月17日至2022年5月16日	办公	办公
2	赵峻	北京市海淀区成府路28号3-801、3-803、3-807、3-809	2020年5月16日至2021年5月15日	办公	办公
3	徐小勇、苗英	北京市海淀区成府路28号12层4-1215	2020年4月19日至2021年4月18日	办公	办公

上述出租方提供了租赁房屋的不动产权证，发行人实际使用用途与其产权证上记载房屋用途均为办公，因此，发行人向个人租赁房产的实际用途与其法定用途相符。

发行人说明事项

(1) 发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效

一、发行人采购数据采集和标注服务的具体情况

采购内容	人员管理	场地使用	进度管理
原料数据采集、标注服务	服务公司、小外包负责对参与采集、标注的人员进行管理，具体职责划分如下： ① 服务公司 ：直接监督、管理小外包及其直接对接的终端人员；管理前述人群的劳务关系（包括劳务合同签署、进度时间安排、人员信息登记、终端人授权签署等）、按经核对一致的服务量支付款项，并为其代扣代缴需要缴纳的社保、个税等。 ② 小外包 ：负责寻找符合发行人要求的终端劳务人员；组织并管理终端人提供劳务；包括人员信息登记、终端人授权签署、人员工作时间安排及现场引导等；与其负责的终端人进行结算支付。	依据项目要求确定，有的项目在服务公司、小外包提供的场地进行；有的项目在发行人提供的场所进行	发行人、服务公司、小外包均参与进度监控和管理。具体职责如下： ① 发行人 ：负责在采集、标注的具体执行过程中，对采集、标注人员或其组织人员（服务公司人员或小外包）进行培训，部署采集、标注所用平台、工具，下达采集、标注的具体要求；把控项目整体执行的时间进度，敦促服务公司及其相关人员按时交付数据。 ② 服务公司 ：负责敦促小外包及其直接对接的终端人员按时提供发行人所需的数据服务，

			保证提供数据的完整性。 ③小外包：负责对其负责的终端人的服务进度和数据完整性进行把控等。
--	--	--	---

(续)

采购内容	质量控制措施	验收标准
原料数据采集、标注服务	<p>发行人、服务公司均参与质量控制，具体职责如下：</p> <p>发行人：</p> <p>①订单确认：发行人通过会议或者邮件等方式与服务公司详细沟通需求；如被采集人的年龄、籍贯、性别、口音特点、工期要求、采集场地等；标注人员的语言背景、专业技能和学历要求等。重要项目会对该类要求有书面确认，保证被采集人和标注员的构成符合发行人制定的要求。</p> <p>②原料数据采集和标注过程管理：发行人对服务公司人员、小外包或终端服务人员进行采集、标注规程相关培训，培训合格后开始原料数据的采集、标注工作。采集、标注具体过程中，发行人的技术人员和管理人员会进行监督和指导，确保交付的原料数据或标注数据符合发行人的质量标准。</p> <p>③质检及验收：发行人根据项目工期对服务公司负责采集或标注的数据进行验收。验收前发行人会进行质量验证，数据采集和标注的工作主要在海天瑞声提供的平台、工具上进行，系统提供自动验证方式，搭配人工复核、抽检等方式，通过质量验证的数据即实现交付及验收。根据双方合同约定，根据服务数据量，发行人与服务公司签署验收结算单予以确认，不符合要求的不予验收。</p> <p>服务公司：</p> <p>在采集、标注具体执行过程中，持续与发行人进行沟通，获取发行人对采标服务及数据质量的反馈情况，及时反馈给小外包、终端劳务人员，确保其为发行人提供服务的质量。</p>	<p>根据双方约定来确定验收标准，验收标准主要包括数据数量及根据不同项目设定的质量参数（如：声音参数和声音质量、声音文本一致性、图像像素等级、文本规模等）</p> <p>根据双方约定来确定验收标准，验收标准因项目不同会有所差异（一般验收标准参数包括：标注字准确率、句准确率、图像框位置准确率等）</p>

(续)

采购内容	结算方式	结算对象	结算周期
原料数据采集、标注服务	发行人与服务公司根据验收结算单对账，依据双方认可的数据质量、数量进行结算，并以银行对公转账方式付款。	发行人向服务公司结算	按月/半月/周结算，以周结算为主

二、与数据服务采购相关内部控制的设计是否合理、执行是否有效

（一）发行人数据服务采购相关的内部控制情况

报告期内，发行人建立了《海天瑞声采购管理制度》、《海天瑞声项目管理办法》、《海天瑞声供应商管理制度》、《海天瑞声合同管理制度》和《海天瑞声业务采购实施细则》等规章制度作为对外采购数据服务的制度依据。

针对数据服务采购业务环节涉及的控制点(年度总体预算、供应商选择和持续管理、采购协议签署、采购价格管理、立项及预算管理、采购申请与审批、下达采购需求、供应商提供服务、验收结算及费用报销、供应商付款、项目决算、业财核对)，发行人均制定了、采取了有效的控制措施/活动，并确保数据服务采购管控制度有效执行。

具体控制手段及制度设置情况如下：

数据服务采购主要业务环节/控制点	控制手段	控制频率	内控制度
年度总体预算	年度采购预算	年度采购预算：每年一次	《海天瑞声全面预算管理办法》 《海天瑞声项目管理办法》
供应商选择和持续管理	《合格供应商名录》管理	每年定期评估一次，不定期更新维护	《海天瑞声供应商管理制度》
采购价格管理	采购价格体系	持续维护，每半年更新一次	《海天瑞声采购管理制度》 《海天瑞声业务采购实施细则》
采购协议签署	合同签订审批	每个合同审批一次	《海天瑞声采购管理制度》 《海天瑞声业务采购实施细则》 《海天瑞声合同管理制度》
立项及项目预算管理★	项目预算	每个项目一次	《海天瑞声项目管理办法》
采购申请与审批	审批控制	每类申请审批一次	《海天瑞声采购管理制度》 《海天瑞声业务采购实施细则》
下达采购通知★	采购会议、发送采购需求邮件	每个项目一次	《海天瑞声业务采购实施细则》
供应商提供服务	标准制定、服务过程监督和指导、质检、验收	提供服务全过程跟踪控制	《海天瑞声项目管理办法》
验收结算及费用报销★	报销申请审批、验收结算确认单盖章确认	每次申请审批一次、每次报销与供应商签订确认单	《海天瑞声项目管理办法》
供应商付款	付款审批	每次申请审批一次	《海天瑞声财务报销及借款管理办法》
项目决算★	项目决算	每个项目一次	《海天瑞声项目管理办法》
业财核对★	业财核对	每月一次	《海天瑞声会计核算制度》

注：上述过程中的关键控制节点以★标示。

各控制点及其对应的控制手段具体说明如下：

1、年度总体预算

发行人实施全面预算管理。各年年初，发行人会针对各业务条线制定总体收入计划，结合各业务指导毛利率、一般性支出结构情况计算得到关于数据服务采购金额总量的大致预算情况。

上述年度预算由发行人财务部门会同各部门共同起草制定，经董事会、股东大会审批生效。

2、供应商选择和持续管理

发行人建立“供应商评估标准”，对《合格供应商名录》内的数据服务供应商进行年度评估，重点评估要点包括人员需求完成程度、响应速度、提供数据的数量和质量等，保证其在各方面均可满足公司要求，可作为合格供应商与公司继续合作。若已有《合格供应商名录》内的供应商无法满足需求时，发行人将从外部寻找新供应商，首先针对新供应商进行资质评估，评估范围主要包括基本工商资质、主要产品与服务、售后服务情况和价格水平等，经总经理办公会审批后进入《合格供应商名录》，后续才可启动后续采购。

3、采购价格管理

发行人建立了业务采购价格体系，用于汇总和追踪采集与标注服务的单价实施情况和变动情况，同时作为业务采购的价格指导。该体系每半年更新一次，以确保该价格体系与公司实际采购情况、市场价格相符。采购部对该体系中的价格进行定期公允性检查，以确保采购价格在合理范围内。检查方式包括但不限于同类供应商询价、供应商及其内部人员访谈、发布模拟招聘信息询价、波动原因解析等。对超出价格体系范围 10% 以上的采购需求，需对原因进行说明，并在采购部备案。

4、采购协议签署

发行人与供应商签署框架协议，约定双方的权利与义务等，并约定数据服务管理费率及结算方式。发行人在正式合同签署前，要求由采购部专员提交《合同审批表》，并按公司规定，经过采购部门负责人、业务部门负责人、财务部、法务和总经理审批后，方能正式签署合同。

5、立项及项目预算管理★

在项目执行之前，发行人需要对每个项目编制详细的项目预算。立项及项目预算管理，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。

对于项目预算中数据服务费部分，采集数据量需要参照“以销定采”的业务逻辑，根据销售合同或订单约定的交付数据量并考虑合理的采集损耗率后确定，标注量（即有效工时）根据采集数据量并参照合理的标产比后确定；标产比根据项目启动前的自测标产比或与客户约定的标产比确定；采集单价和标注单价参照公司采购价格体系确定；项目负责人编制预算时需要将数据服务费的采购量、采购单价均按工作环节逐项、单独填报于业务管理平台，经过业务部门、财务部等多层级审批通过后方可执行。

对于上述采购量，发行人执行的主要复核环节如下：①将预计采集数据量、标注数据量与销售合同或订单约定的交付数据量之间进行核对，保证预算采购数据量的真实、完整、准确；②分析采集损耗率的合理性，对比同类项目间的标产比，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，对比采购价格体系，对采集、标注单价进行复核，保证预算中采集、标注单价的合理性；④根据销售合同或订单金额、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。

6、采购申请与审批

业务部门需提出采购申请，以对采集和标注服务的采购行为的发起进行合理性控制，经部门直属负责人、部门分管负责人、采购部三级审批，确保采购需求及供应商选择的合理性。

7、下达采购通知★

发行人通过以下两种方式向供应商下达采购通知：1、与数据服务供应商召开采购会议，形成采购会议纪要；2、发送采购需求邮件。采购通知中对采集、标注需求进行明确规范。采购通知的内容包括项目简介、服务工期、符合项目预算的原料数据采集、标注服务量、采集或标注的人员要求和特别要素（例如被采集人的年龄、籍贯、性别、口音特点、采集场地，标注人员的语言背景、专业技能、学历要求）等。其中原料数据标注服务量根据需要标注的数据量与双方协商确定的标产比相乘计算得出。

8、供应商提供服务（采购执行）

在供应商提供服务过程中，发行人的技术人员和管理人员负责对服务公司人员、小外包人员或终端服务人员进行规则培训，培训合格后开始采集或标注。在原料数据采集、标注过程中，发行人会参与监督和指导，确保其交付的数据符合发行人的要求。

9、验收结算及费用报销★

供应商采集与标注完毕后，业务部门项目负责人进行验收，并将质量验收合格的采集数据量、有效工时等信息录入表格/信息系统（即采购量录入环节），并要求终端劳务人员签字确认其提供的劳务工作量。在上述过程中的采购量录入环节，复核人员针对采集、标注各环节的数据量执行了详细的比对、复核程序，具体如下：

①对于采集：复核人员对录入的采集数据量、交付数据量与采集环节的结算量等业务量信息进行交叉比对，如存在重大差异，需由项目负责人解释原因；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。具体核对内容包括：交付数据量对比采集数据量，是否总体基本一致，采集损耗率是否在合理范围内；设计数据量对比采集数据量，是否总体基本一致；音频文件个数对比采集数据量，是否一致；采集数据量对比采集结算量，是否存在异常结算情况；采集结算量对比预算采购量，是否存在较大偏离；复核实际采集产出比，是否与商务约定或预算方案一致；采集单价对比预算单价，是否出现重大偏离。

②对于标注：复核人员针对标产比、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。具体核对内容包括：对比采集数据量，复核标注工作覆盖的标注数据量是否完整；标产比对比预算，是否出现重大偏离；标产比在同类项目中进行对比，是否出现重大异常或偏离；有效工时的计算是否准确；标注单价对比预算单价，是否出现重大偏离。

业务部门项目负责人将载有采集数据量、有效工时、采购单价等信息的采购支出申请在业务管理平台中提交报销申请，部门分管负责人、财务部通过检查采集损耗率、标产比、项目预算以及生产系统日志记录等信息来验证报销数据的合理性，确认无误后对报销申请进行审批，审批后，系统生成《验收结算确认单》，采购人员将《验收结算确认单》发送至供应商，由供应商确认后盖章，形成双方结算依据。

10、供应商付款

发行人定期发起向供应商进行付款，经部门负责人审批完成后，由财务部进行审核，并最终由财务经理或财务负责人审批完成，确认支付义务并对外进行支付。

11、项目决算★

发行人由项目负责人在业务管理平台中提出“决算申请”，并查阅“支出预算差额百分比”，若差异率超过 10% 以上需填写原因。项目负责人申请决算前，所有支出必须填写和报销完毕，由财务部确认项目预算执行情况和款项付款情况，审批项目决算。项目决算应在项目全部验收后进行，最晚不得晚于验收后 10 个工作日，且季度内验收的项目必须在季度末完成决算、当年验收的项目必须在当年完成决算。项目决算后不得再对项目费用进行报销。

12、业财核对★

2019 年 4 月前，项目采购支出、收入、成本等关键信息通过业务管理平台自动汇总并生成支出汇总表、财务确认表，财务人员根据报销申请单手工在财务用友系统中人工录入制造费用和研发费用，并根据财务确认表在财务用友系统中人工输入营业收入确认和营业成本结转数据。每月末，财务部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。

2019 年 4 月至 10 月，发行人实现了业务管理平台自动对接用友财务系统生成制造费用和研发费用的财务数据，财务人员仍根据财务确认表在财务用友系统中人工输入营业收入确认和营业成本结转数据。

2019 年 11 月起，发行人实现了业务管理平台自动对接用友财务系统生成营业收入和营业成本的财务数据。

13、小结

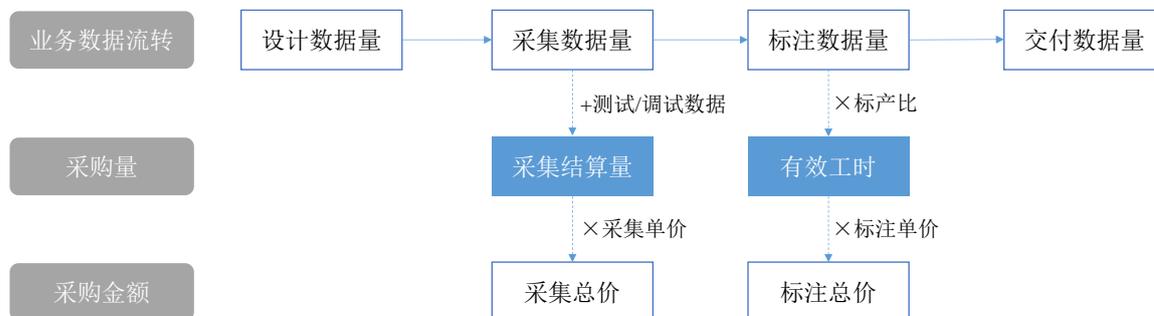
综上所述，发行人根据《企业内部控制基本规范》的规定，针对对外采购原料数据采集和标注服务了完善的内控制度，包括：《海天瑞声采购管理制度》、《海天瑞声项目管理办法》、《海天瑞声供应商管理制度》、《海天瑞声合同管理制度》和《海天瑞声业务采购实施细则》等规章制度，对数据服务采购过程执行严格的内控管理制度。

（二）与数据服务采购相关的内部控制的设计合理，符合发行人业务特征

发行人的上述内部控制制度符合《企业内部控制基本规范》相关规定，并结合了发行人数据服务采购业务流程的实际特征，可针对数据服务采购过程中的各个环节有效进行复核、控制，相关内部控制设计合理。具体说明如下：

1、数据服务采购过程呈现“以销定采”、“同类业务标产比可比”的业务逻辑，可通过前述业务逻辑对数据服务采购量进行复核、确认

发行人训练数据定制服务和训练数据产品业务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：



注：采集结算量、有效工时指发行人与供应商结算的采集环节、标注环节采购量；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。

发行人数据服务（采集服务、标注服务）采购量与交付给客户的数据量紧密挂钩。对于采集服务：发行人采集损耗率保持较低水平，保证了采集服务的结算量可通过对比设计数据量、采集数据量和交付数据量得到有效控制。对于标注服务：发行人以标注数据量为基数，用标产比（通过实际自测或与客户约定）折算有效工时的方式来核算标注工作量并结算，因此可通过对比标注数据量和交付数据量、复核标产比的合理性对标注服务的结算量（有效工时）进行有效控制。

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（交付数据量）实际采集人数为152人（采集数据量、采集结算量）。成品训练数据集需要交付多少数据量，则相应采集对应的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。经试标注测速后，标产比确定为2.7，假设某标注人员完成了10小时的音频数据标注工作，则其有效工时核定为 $10 \times 2.7 = 27$ 小时，同类型项目（即数据的特征与领域基本相当）的标产比应基本一致。

在复核、确认采购量的同时，发行人也通过内部控制设计定期维护更新业务采购价格体系，从而充分适应行业下游训练数据需求的变化，在数据类别、数据形式等方面不断拓展细化，以确保该价格体系与发行人实际采购情况、市场价格相符。

2、发行人的数据服务采购内部控制设计与上述业务逻辑相符，设计有效

发行人设计了完善的内部控制制度，对数据服务采购的各业务环节均施加有效控

制。特别地，与上述业务逻辑相适应，发行人在立项及项目预算管理、验收结算及费用报销、项目决算、业财核对各个关键业务点设置了针对数据服务采购量、采购价格、采购金额、业务数据与财务数据一致等的复核控制，确保数据服务采购核算真实、准确、完整。发行人的数据服务采购内部控制设计符合业务逻辑，设计合理、有效。

（三）与数据服务采购相关的内部控制执行有效，可确保采购信息完整、准确记载，数据服务采购量可复核，财务信息与实际采购业务一致

1、发行人搭建了有效的内部信息系统，可记载完整、准确的采购信息等业务数据

发行人的内部信息系统包含三个组成部分，分别为一体化数据处理平台（前台生产系统）、业务管理平台（中台业务信息流转与管理系统）与用友财务系统（后台记账与报表系统）。各组成部分情况具体说明如下：

（1）一体化数据处理平台。该平台是发行人训练数据生产所需的各类工具、软件的集合，于2014年9月初始上线并持续优化，2018年初开始保留与生产过程相关的基本日志信息，如每个录音人的录音时长信息、录音文件大小信息等，在2019年9月上线搭载了终端人管理系统及日志管理系统，能够更为全面记录并保留生产过程中的详细日志信息，包括终端劳务人员注册信息、生产过程中终端人员行为日志信息等。

（2）业务管理平台。该平台于2017年1月开始上线使用，主要功能包含合同和项目创建、项目执行过程和结算、验收管理等，可覆盖项目管理全生命周期，并可根据系统记录的业务数据自动生成收入、成本、存货、产品研发费用等相关的财务数据。发行人报告期内销售、采购业务涉及的量、价和金额信息均完整记载于业务管理平台。

（3）用友财务系统。该系统于2015年开始上线使用，自上线至2019年4月为用友U8系统，2019年4月至今为用友NC系统，系统自上线以来未发生重大变更。

根据上述说明可知，公司内部信息系统均在报告期初上线运行并覆盖完整报告期间。发行人业务管理平台具有项目核算模块，能够按项目维度完整记录项目预算及支出相关信息，包括不限于每个项目所对应采购的劳务服务数量、结算价格及金额等，组成发行人业务采购的完整信息。

2、发行人业务管理平台记录了完整、准确的采购信息，使得可以通过“以销定采”、“同类业务标产比可比”的业务逻辑对数据服务采购量进行复核、确认

发行人通过业务管理平台对项目进行管理控制，包括项目预算申请和复核，支出的填报和复核，对采集数据量、采集结算量、交付数据量进行交叉比对，对标注产出比、有效工时进行合理性复核，生成供应商验收结算确认单，进行验收和决算信息复核等。前述信息的准确记录，使得发行人可通过“以销定采”、“同类业务标产比可比”的业务逻辑对数据服务采购量进行复核、确认。

同时，发行人在项目预算申请环节参考业务采购价格体系对预算采购价格进行复核、确认，在项目支出报销环节参考预算采购价格对实际采购价格进行复核、确认，从而对采购价格进行有效的内部控制。

3、发行人财务系统中数据服务采购信息均基于业务管理平台产生，可保证实际采购业务与财务信息一致

项目采购支出信息通过业务管理平台自动汇总并生成支出汇总表，人工录入财务系统并经专人复核¹⁵，进而生成存货、研发费用等信息，可保证采购相关财务信息与业务信息的一致性。

（四）总结

综上，报告期内，发行人依据《企业内部控制基本规范》要求，结合数据服务采购的具体流程、特征设置了内部控制体系，相关内部控制体系符合数据服务采购的业务逻辑，设计有效；发行人搭建了有效的内部信息系统和业务数据流转、复核机制，可确保采购信息完整、准确记载，数据服务采购量、采购单价可复核，财务信息和实际采购业务一致，数据服务采购相关内部控制制度有效执行。

（2）报告期各期服务公司小外包人员的数量及其变动情况，是否为服务公司员工，是否存在相互重合的情形，发行人和服务公司对小外包人员的管理措施，是否均由发行人员工直接对小外包人员进行管理和控制

报告期各期，服务公司小外包人员数量如下所示：

	2020年1-3月	2019年	2018年	2017年
--	-----------	-------	-------	-------

¹⁵ 报告期内，发行人持续改进、升级业务管理平台功能，于2019年4月后可实现自动将相关信息传输至财务系统，进一步确保采购相关财务信息与业务信息的一致性。

	小外包个数及覆盖终端劳务人员数量占比	领取劳务费金额(万元)及占比	小外包个数及覆盖终端劳务人员数量占比	领取劳务费金额(万元)及占比	小外包个数及覆盖终端劳务人员数量占比	领取劳务费金额(万元)及占比	小外包个数及覆盖终端劳务人员数量占比	领取劳务费金额(万元)及占比
当年组织终端人达 1,001 人以上	1	29.44	12	806.62	18	1,388.32	17	624.54
覆盖终端劳务人员占比	15.77%	3.56%	50.56%	19.72%	66.72%	31.86%	72.44%	33.51%
当年组织终端人 501-1,000 人	6	67.99	15	674.06	13	522.68	6	119.14
覆盖终端劳务人员占比	26.26%	8.21%	16.46%	16.48%	11.76%	12.00%	8.10%	6.39%
当年组织终端人 101-500 人	31	313.22	61	1,035.21	49	1,451.45	33	483.50
覆盖终端劳务人员占比	42.42%	37.83%	21.10%	25.31%	15.03%	33.31%	12.60%	25.94%
当年组织终端人 100 以下	172	417.39	618	1,574.92	213	994.56	213	636.82
覆盖终端劳务人员占比	15.55%	50.41%	11.88%	38.50%	6.49%	22.83%	6.86%	34.16%
小外包人员数量合计	210		706		293		269	

2017 年、2018 年、2019 年及 2020 年 1-3 月，小外包人数分别为 269、293、706 及 210 名。2019 年小外包人员数量较多，主要原因是：当年执行的项目涉及较多新拓展的语种/方言，需要的被采集人和标注人员更为多样、所在地区更广，服务公司通过对接更多小外包为发行人寻找合适的采集、标注人员。同时，小外包人员规模的扩大也可以降低资源单一、集中带来的风险；2020 年 1-3 月小外包人员数量较少，主要是因为 2020 年第一季度时间窗口较短，同时受新冠病毒疫情对发行人的现场集中采集、标注工作造成了一定影响，服务公司为发行人组织采集、标注工作对接的小外包人数较少。

通过对服务公司及小外包人员访谈确认，上述小外包人员均不属于服务公司员工。由于部分小外包人员专注从事终端人员的寻找和组织工作，可能对接多个服务公司，因此不同服务公司的小外包人员可能存在重合的情形。

服务公司根据发行人的需求，向具备相应能力的小外包人员下达需求，寻找合适的终端人员，并在原料数据的采集、标注过程中，督促小外包人员按项目进度和要求完成既定工作安排，同时与发行人进行沟通，获取发行人对采标服务及数据质量的反馈情况，及时反馈给小外包人员，确保其提供的数据服务质量。服务公司对小外包人员的管理措施包括签订劳务协议明确双方权利义务，建立相应管理制度对小外包人员提供劳务服务的具体过程进行规范等。而发行人负责对小外包人员进行业务方面的培训，不会对小外包人员进行直接的人员方面管理和控制，主要通过与服务公司的沟通合作，间接确保小外包人员按照发行人需求寻找并组织终端劳务人员进行原料数据采集和标注服务。

(3) 发行人业务人员事先垫付资金的具体原因、后续结算及差异调整情况，是否符合相关内部控制制度的规定，报告期各期事先垫付资金的金额及占比

由于发行人的劳务采购存在临时性、一次性的特征，部分情况下需要在劳务提供完成时立即向终端劳务人员支付报酬，对于通过小外包人员对接劳务终端人员的情形下，一般会由小外包人员支付该笔款项，待相关劳务服务经发行人与服务公司验收结算后，由服务公司向小外包进行支付。对于少量由服务公司直接对接终端劳务人员的情形下，一般会由服务公司人员支付，对于服务公司人员未到达劳务服务现场或在现场未支付劳务报酬的极个别情况，存在发行人业务人员事先垫付资金的情形。后续发行人与供应商验收结算后，终端劳务人员向发行人业务人员归还垫付款项并由供应商按劳务协议约定向终端劳务人员支付劳务费用。由于发行人业务人员垫付资金属于个别情形，未出现后续验收不通过的情况，因此不存在结算后需要调整差异的情况。

上述情况存在于 2017 年和 2018 年，且金额较小，不会对发行人内部控制的有效性造成重大不利影响，自 2019 年起，发行人已对上述备用金事项进行整改规范，内部完善了《备用金管理制度》，对于备用金使用范围、内部审批程序，款项归还时间等进行了明确约定，并要求员工对于备用金的支取和使用通过专门银行账户进行核算，同时严格要求服务公司履行对终端人的管理措施，确保其在必要时及时支付劳务报酬，因此不再存在上述由发行人业务人员垫付资金的情形。

报告期内，发行人业务人员通过备用金事先垫付资金的金额及占比情况如下：

单位：万元

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
业务人员垫付资金总额	-	-	39.42	29.10
占服务公司直接对接采购支出的比例	-	-	1.48%	1.56%
占数据服务费采购支出的比例	-	-	0.54%	0.72%

(4) 结合服务公司提供管理服务的内容、过程和必要性，说明发行人向服务公司支付服务管理费的原因及合理性，加成率的定价依据及公允性，不同公司的加成率之间是否存在差异并说明具体原因

发行人向服务公司采购非核心技术环节的数据采集、标注服务，是由于发行人业务的核心技术环节主要体现在训练数据集结构设计，训练数据集开发标准制定，数据采集加工过程中的技术研究、工具开发，训练数据集质量标准制定、质量检测和等方面，

原料数据的具体采集和标注等过程主要为基础性、重复性的人工操作，发行人直接从事该等工作并不能突出发行人的竞争优势，而服务公司等具备在人力市场的资源获取能力，适合原料数据的采集与标注工作所呈现的执行人员属性要求众多、数量较大、临时性较强、流动性较高的特点，能够为发行人整体协调众多人力提供原料数据服务，解决发行人对被采集人、标注人员的寻找和管理需求，并保证服务的按期交付。因此发行人向服务公司采购非核心技术环节的数据采集、标注服务并支付服务管理费，具备商业上的必要性和合理性。

发行人支付给服务公司的加成率是综合考虑了服务内容、市场同类服务价格、服务公司自身成本及合理利润空间等因素，并经双方商务谈判与协商而确定的，符合服务公司所提供服务的价值以及市场同类人力资源服务公司和劳务公司管理服务费率的整体水平，具备公允性。加成率有所差异的原因是税务成本和管理难度的差异所导致的。

对于境内数据服务提供商，由于不同的供应商在各类型项目中组织和管理被采集人员、标注人员的数量规模、管理难度、税务成本等有所差异，因此发行人与供应商协商分别约定采用 10-11.5% 的服务费率，存在小幅差异。

对于境外数据服务提供商，主要考虑到供应商所在地的税务成本和人力资源匹配成本，经发行人与供应商协商约定采用 6-15% 的服务费率，具备商业合理性。

根据广州善世、芜湖善慧、北京博睿、北京零点、信阳联智等主要境内数据服务提供商出具的说明，其向其他客户收取的服务费率区间为 9-13%，根据恩泽尔等主要境外数据服务提供商出具的说明，其向其他客户收取的服务费率区间为 5-8%，与该等数据服务提供商向发行人收取的服务费率基本一致，不存在明显差异。

同时，经核查，根据发行人对市场同类人力资源服务公司和劳务公司的市场询价结果，同类境内服务商报价区间为 9-15%，发行人向供应商支付的 10-11.5% 的服务费率符合境内服务商的普遍报价区间，服务费率的采购定价具备公允性。

(5) 采集和标注单价的具体确定依据和过程，同类型项目的采集和标注单价是否存在差异并分析具体原因

一、采集和标注单价的具体确定依据和过程

采集和标注单价的确定依据均为参照市场价格及特定任务的难度综合确定。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。标注单价根据市场上类似服务价格，并参考标注任务的难度来确定。例如标注专业领域的数据等高难度项目，则标注单价高于一般项目。发行人已经制定并定期更新业务采购价格体系，对采集和标注单价进行规范化管理。

采集和标注单价的确定过程主要为：结合上述确定依据，发行人在下达采购通知后，数据服务提供商根据寻找的人力资源情况，将价格信息反馈给发行人，双方协商一致后确定采集和标注单价，依此进行结算。

发行人数据服务采购所需劳务人员较多，且呈现人员要求多样化的特征，因此在下达采购通知时无法直接确定具体单价，需要服务公司及小外包人员根据采购通知的要求寻找对应的劳务人员时进行询价。对于服务公司反馈的价格信息，发行人参考业务采购价格体系和过往执行的同类项目的价格，并结合当前项目可能存在的特殊要素要求，综合判断价格的合理性并与数据服务提供商协商，双方达成一致后确定采集和标注单价。考虑到业务实际情况和经营管理的便利性，在发行人下达采购通知确定采购量后，采集和标注单价通常采用电子邮件、补充确认单、通讯软件沟通等方式确定，并留存于验收结算单中。

二、同类型项目采集、标注单价情况，存在的差异及具体原因

发行人各年采购的数据服务对应的采集、标注单价按项目类型拆分如下：

单价单位：元/单位

应用领域	项目类型	单价类别	单位	2020年1-3月	2019年	2018年	2017年
智能语音	语音识别采集	采集单价	人	220.55	263.90	202.28	141.79
		标注单价	工作小时	41.42	37.30	25.89	26.92
	语音识别转写	标注单价	工作小时	18.48	17.78	17.60	16.49
	语音合成	采集单价	字/词	0.27	0.46	0.52	0.45
标注单价		工作小时	31.45	28.19	46.04	41.83	
计算机视觉	手写体	采集单价	工作小时	234.13	225.88	42.81	68.39
		标注单价	工作小时	25.54	-	-	24.49
	OCR	采集单价	工作小时	-	-	52.55	11.96
		标注单价	工作小时	24.91	18.02	21.08	15.52
	行车视频	采集单价	工作小时	129.70	75.37	43.09	12.42
		标注单价	工作小时	15.83	15.76	16.82	16.30

自然语言	语义理解	采集单价	工作小时	-	-	23.12	15.50
		标注单价	工作小时	26.19	23.34	19.35	16.03

注：计算机视觉、自然语言类项目特异性较强、非标准化程度较高，涉及采集、标注单价类型及单位众多，此处选取采集、标注类型单位较为统一的主要类型的主要项目进行列示。

同类型项目采集标注单价情况分析：

1、语音识别采集

采集单价：2018年-2020年3月单价整体稳定，有所波动。自2018年开始采集单价较2017年呈现增长趋势，主要原因是发行人业务逐步向采集单价较高的外语种、自由对话类型训练数据拓展，且该趋势在近几年持续，推高了采集的平均单价。2020年1-3月采集单价相比2019年有所下降，一方面是因为该季度执行的中文语音项目主要涉及短时长采集工作，采集单价较低；另一方面是因为当期执行的英语采集项目涉及的细分地区英语种类较多，各细分地区英语采集单价有所差异，结构变动使得当期英语采集总体单价相比2019年有所下降。前述两方面因素共同带动整体采集单价下降。

标注单价：2017-2020年3月标注单价有所差异，整体呈现波动上升趋势，主要原因是：各年语音识别采集项目涉及的语种类型有所差异：与中文标注相比，外语种单价标注价格更高；在外语种标注内部，英国英语、美国英语标注难度相对较低，单价也较低，其他国家语种视具体类型标注单价有所差异。以上语种的结构变化导致了标注单价的波动，对于相同类别语种而言，各年标注单价基本保持稳定。

2、语音识别转写

标注单价：报告期内单价均较为稳定，呈现小幅上升趋势是由于外语种语音识别转写业务的持续拓展使外语种标注占比逐年小幅提高。

3、语音合成

采集单价：2017-2019年单价基本稳定；2020年1-3月单价较低，主要原因是一方面该季度完成的语音合成类原料数据采集以单价相对较低的中文类语音采集工作为主，另一方面，该季度完成的语音合成原料数据采集的项目类型决定了其所涉及的中文类语音采集工作主要为一般人（非专业发音人）参与的多人语音采集工作，与常见的专业发音人采集相比单价更低。

标注单价：2017-2018 年单价较为稳定，自 2019 年起有所下降，2020 年 1-3 月水平与 2019 年基本相当，主要原因是：（1）发行人预标注技术和工具自 2018 年开始逐步运用并普及，提升了原料数据标注能力和效率，对标注人员的个人要求降低，使得标注单价降低；（2）发行人通过服务公司组织更多具备多语种语言能力的中国标注人员对部分外语种语音进行标注，提升性价比、降低标注单价。

4、非智能语音类项目（计算机视觉、自然语言类）

报告期内，发行人的非智能语音类项目（计算机视觉、自然语言类项目）特异性较强、单个项目需求差异化特征较强，同类项目之间具体要求也存在较大差异，造成各年采集、标注单价波动较大。对于各类非智能语音类项目各年采集、标注单价变动情况分析如下：

（1）手写体

采集单价：2017 年-2018 年单价较低，2019 年、2020 年 1-3 月单价较高，主要原因是：发行人各年执行的手写体项目细分类型不同，难度逐步提升，滑行输入数据采集逐步从中文拓展至外语种、手写体项目细分类型从滑行输入数据采集拓展至复杂文本手写采集（含字体、左右手等细分要求），采集要求、难度逐步提升，采集单价随之提升。

标注单价：报告期内标注单价较为稳定。2018 年、2019 年无标注单价是因为对应项目为按客户需求采集手写体原料数据，无需执行标注工作。

（2）OCR

报告期内，发行人执行的 OCR 项目个性化特征较强，采集、标注类型、单位众多，差异较大，此处选取以工作小时为单位衡量的 OCR 项目进行列示及分析。此类 OCR 项目受采集资源特定要求的影响，采集单价存在一定波动，2018 年采集试题数据，对作业本、采集对象年级等均有要求，采集单价较高；报告期内标注单价相对稳定。

（3）行车视频

采集单价：报告期内采集单价逐步提升，主要原因是发行人执行的行车视频采集项目采集要求和难度逐步提升，例如逐步从静止的泊车边位线图片采集拓展至动态的人车 2D 矩形框视频采集、城市交通路况路标的视频采集，采集环境要求（光线要求、角度要求等）也逐步拓展，采集单价逐步提升。

标注单价：报告期内标注单价较为稳定。

（4）语义理解

采集单价：2019年、2020年1-3月，发行人未开展语义理解采集类项目；2017年、2018年语义理解项目采集规模较小，采购金额较低，采集单价易受个别项目影响呈现波动。

标注单价：报告期内标注单价整体呈现逐步上升趋势，主要原因是：随着自然语言类项目逐步复杂化，文本结构、具体内容、标注规范均呈现难度逐渐加大的趋势，对标注人员的要求逐步提升，使得标注单价逐年提高。

综上，对于智能语音类项目，同类项目采集、标注单价相对稳定，但受具体语种/方言结构变动、发行人技术提升等因素影响存在差异；对于非智能语音类项目，因发行人项目非标准化、采标难度随发行人业务发展逐步提升等，同类型项目的采集和标注单价也存在一定差异。

（6）采集数据量、标注数据量和标产比相关的确定方法和具体依据，是否存在随意修改或调整的情形，同类型项目的采集数据量、标注数据量和标产比是否存在差异并分析具体原因

一、采集数据量、标注数据量和标产比的确定方法和具体依据，是否存在随意修改或调整的情形

发行人采集数据量、标注数据量与交付给客户的数据量之间有密切联系。交付给客户的数据量由发行人的销售合同或销售订单确定，且最终需要发行人客户进行验收确认。在实际业务执行过程中，发行人会根据经验判断，在满足客户交付数据量要求的基础上留出少量冗余，例如某“200人新加坡英语手机3通道语音识别项目”，客户订单要求录音人数为200人，实际采集、结算人数为206人。报告期内，发行人采集损耗率，即采集数据量相对于交付给客户的数据量的损耗比例，总体控制在5%以内，而标注数据量介于采集结算量和交付给客户的数据量之间，因此采集数据量和标注数据量不存在随意修改或调整的情形，具体请参见下述“二、同类型项目的采集数据量、标注数据量和标产比的差异情况及原因分析”。

发行人采购原料数据标注服务的有效工时由两方面因素构成：标注数据量和标产比（单位产出量所需的标注工作时长）。发行人通过自行测量的方式形成标产比后，与数

据服务提供商沟通标产比的可行性，双方达成一致后发行人按照自测标产比折算有效工时，向数据服务提供商下达采购通知，确定标注服务采购量，并依此与数据服务提供商进行结算，主要留痕证据为采购通知和验收结算单。发行人自测标产比的具体方法为：

（1）发行人将待标注数据拆分为若干任务后，选取少量任务由标注人员进行试标注，测量在连续工作状态下的标产比；（2）自测标产比与供应商达成一致后确定为结算所用标产比。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。综上所述，标产比一经确定，即按照这一比例与供应商结算标注工时，不存在随意修改或调整的情形。报告期内，发行人标产比的变动情况具体请参见下述“二、同类型项目的采集数据量、标注数据量和标产比的差异情况及原因分析”。

二、同类型项目的采集数据量、标注数据量和标产比的差异情况及原因分析

同类型项目由于项目规模不同，因此所需的采集数据量、标注数据量必然有所差异。以下通过分析报告期各期发行人训练数据定制服务完成项目和新增训练数据产品项目¹⁶的采产比和标产比来说明同类型项目的采集数据量、标注有效工时相对水平的差异情况，其中训练数据定制服务完成项目包括语音识别采集、语音识别转写、语音合成、计算机视觉主要细分类别（手写体、光学字符识别、行车相关视频）和自然语言主要细分类别（语义理解），新增训练数据产品包括语音识别采集、语音合成两类发行人最主要的产品类型。

（一）训练数据定制服务

1、语音识别采集

项目	2020年1-3月	2019年度	2018年度	2017年度
采产比 (每小时产出所需人数)	1.0	0.6	0.8	1.0
标产比 (每小时产出所需标注时长)	0.7	3.5	3.0	2.5

报告期内，发行人语音识别采集训练数据定制服务的采产比分别为 1.0、0.8、0.6 和 1.0，2018 年和 2019 年较低的原因是部分项目存在客户特殊要求，单个发音人的录

¹⁶ 新增训练数据产品指当期研发费用的数据服务费所对应的训练数据产品产出量，其中部分训练数据产品可能并未开发完成，仅计入完整数据产品的一部分时长

音时长由 1 小时左右增加到 2-3 小时左右，以及发行人为应对各种发音人资源稀缺的情况，采用了延长单个发音人录音时长的设计方式，该等项目的具体情况如下：

分类	时间	销量（小时）	采集数量（人）	采产比
客户特殊要求	2019 年	15,965.40	5,753	0.4
	2018 年	18,330.00	8,238	0.5
发音人资源稀缺	2019 年	2,200.00	959	0.4
	2018 年	10,800.00	7,767	0.7
其他一般情况	2019 年	15,251.20	14,968	1.0
	2018 年	29,211.00	27,891	1.0

客户特殊要求：部分客户对采集环节的实施有特殊要求，包括同一发音人覆盖多种细分场景、表达多种情感、涉及多个话题等，因此部分发音人的采集时间需要达到 2-3 个小时甚至更多，从而使得平均一个发音人产出的小时量增长，单位产出的采集量（即采集人数）下降。

发音人资源稀缺：为了解决部分项目中符合质量要求的发音人资源稀缺的问题，例如重口音类、儿童类、非母语外语种录音项目，捕捉同一发音人更多语言现象，发行人采用了延长单个发音人录音时长的设计方式，适度提高了单个发音人的录音时长，从而更加有效的利用发音人资源，达到利用同样数量发音人得到更多有效语音时长的效果。

扣除上述特殊项目，其他语音识别采集项目采产比在 2018 年和 2019 年均为 1.0，与 2017 年和 2020 年 1-3 月相比保持稳定，不存在明显差异。

报告期内，发行人语音识别采集训练数据定制服务的标产比分别为 2.5、3.0、3.5 和 0.7，2018 年和 2019 年较高的主要原因是项目需求涉及较难语种或较难口音，同时客户对数据标注要求也较高，使得标产比增加，2020 年 1-3 月完成的项目主要为标注难度很低的唤醒词项目以及部分不需要进行标注的项目，因此标产比较低，该等项目的具体情况如下：

分类	时间	销量（小时）	标注数量（工作小时）	标产比
特殊项目	2020 年 1-3 月	1,522.00	880.94	0.6
	2019 年	4,511.00	40,495.18	9.0
	2018 年	13,137.00	65,727.55	5.0
其他一般项目	2020 年 1-3 月	90.00	192.09	2.1

	2019 年	28,905.60	76,599.23	2.6
	2018 年	45,204.00	111,928.89	2.5

扣除上述特殊项目，其他语音识别采集标产比在报告期内分别为 2.5、2.5、2.6 和 2.1，2020 年 1-3 月由于其他一般项目很少，收入不足 5 万元，因此标产比受个别项目影响有所波动。

2、语音识别转写

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
标产比 (每小时产出所需标注时长)	12.5	11.4	12.5	11.1

报告期内，发行人语音识别转写训练数据定制服务的标产比分别为 11.1、12.5、11.4 和 12.5，总体较为平稳。2018 年标产比略高是由于长句音频项目当年占销量比重较高，长句通常语速较快且变化更为复杂，人物角色和领域较为多样，多为客服等领域，因而难度较高，增加了语音标注的难度。2020 年 1-3 月标产比略高是由于个别特殊项目涉及网络用语、地名等内容，转写难度较大，因此标注工时较高，该等项目的具体情况如下：

分类	时间	销量（小时）	标注数量 (工作小时)	标产比
特殊项目	2020 年 1-3 月	2,791.34	39,164.92	14.0
其他一般项目	2020 年 1-3 月	2,337.63	24,738.92	10.6

扣除上述特殊项目，2020 年 1-3 月其他语音识别转写项目标产比为 10.6，与 2017 年和 2019 年相比不存在明显差异。

3、语音合成

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
采产比 (每小时产出所需万字数)	-	0.6	1.0	1.1
标产比 (每小时产出所需标注时长)	39.9	45.1	44.8	67.3

2017-2019 年，发行人语音合成训练数据定制服务的采产比分别为 1.1、1.0 和 0.6，2019 年较低的原因是由于语音合成业务承接向多元化趋势发展，当年承接了无采集环节的语音合成数据标注业务较多，其占比从 2018 年的 31% 提升至 2019 年的 47%，因此造成采产比下降。

报告期内，发行人语音合成训练数据定制服务的标产比分别为 67.3、44.8、45.1 和 39.9，整体呈现下降趋势的原因是随着发行人音素边界自动预测、基于语音特征的韵律预测等高效数据处理技术的成熟和普遍应用，提高了发行人数据处理工作效率，有效减少了标注过程中的人工工作量，使得单位产出的标注量有所下降。另外，2020 年 1-3 月由于周期较短，项目数量较少，受个别项目影响较为明显，由于部分项目的标注只进行校对不进行其他标注环节，因此投入标注工时较少。

4、计算机视觉

选取计算机视觉类业务中销量单位、采集数量单位、标注数量单位可比的部分主要业务细分类别进行分析，包括手写体、光学字符识别、行车相关视频。

(1) 手写体

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
采产比 (千词组产出所需采集工时)	2.4	2.6	2.5	2.6
标产比 (千词组产出所需标注时长)	0.8	-	-	1.3

报告期内，发行人手写体训练数据定制服务的采产比分别为 2.6、2.5、2.6 和 2.4，折算为每词组采集所需时间约为 9 秒左右，基本保持稳定。

2017 年和 2020 年 1-3 月，手写体训练数据定制服务的标产比分别为 1.3 和 0.8，折算为每词组标注所需时间约为 4.7 秒和 3.0 秒，主要是语种不同导致的标注标准复杂度与难度差异所致，2017 年手写体项目是语种为日语、德语等外语种的手写输入项目，而 2020 年 1-3 月手写体项目是语种为简体中文和繁体中文的手写输入项目。2018 年和 2019 年，手写体项目不涉及标注环节。

(2) 光学字符识别

①有采集的光学字符识别

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
采产比 (百张产出所需百张数)	1.1	1.0	1.0	1.0
标产比 (百张产出所需标注时长)	1.7	-	0.5	0.5

报告期内，发行人有采集的光学字符识别训练数据定制服务的采产比分别为 1.0、1.0、1.0 和 1.1，基本保持稳定。

2017年、2018年和2020年1-3月，发行人有采集的光学字符识别训练数据定制服务的标产比分别为0.5、0.5和1.7，折算为每百张标注时间分别为32分钟、31分钟和99分钟，2019年该业务无标注。2020年1-3月标产比较高的原因是当期主要项目采集对象为实况图片，检查内容较多，质检难度较大，花费的标注工时显著高于其他项目。

②无采集的光学字符识别

项目	2020年1-3月	2019年度	2018年度	2017年度
标产比 (百张产出所需标注时长)	19.2	1.2	0.9	1.0

报告期内，发行人无采集的光学字符识别训练数据定制服务的标产比分别为1.0、0.9、1.2和19.2，折算为每百张标注时间分别为61分钟、52分钟、70分钟和1,153分钟，2020年1-3月标产比高的原因是当期该业务均为例如试题、货架、报表等单图标注内容很多，字符多而密集，标注难度很高的项目，投入工时较长。

无采集的光学字符识别的标注环节涉及对字符进行定位和框选，而有采集的光学字符识别的标注环节主要为检查工作，难度相对较低，因而无采集的光学字符识别单位产出的标注数量高于有采集的光学字符识别。

(3) 行车相关视频

项目	2020年1-3月	2019年度	2018年度	2017年度
标产比 (十个目标产出所需标注时长)	0.01	0.07	0.09	-

行车相关视频业务为2018年新拓展业务，由客户提供原料数据，发行人提供标注服务，因此不涉及采集环节。2018年、2019年和2020年1-3月，发行人行车相关视频训练数据定制服务的标产比分别为0.09、0.07和0.01，折算为每分钟标注目标数量分别为1.8、2.2和21.2，2020年1-3月标产比较低的原因是当期该业务的标注图片内容简单，平均每张图对应目标个数较少，识别速度快，标注难度较小。

5、自然语言——语义理解

选取自然语言类业务中销量单位、采集数量单位、标注数量单位可比的主要业务细分类别语义理解进行分析。

项目	2020年1-3月	2019年度	2018年度	2017年度
标产比 (每小时产出所需标注时长)	0.9	1.0	1.2	1.2

语义理解业务主要是基于客户提供的文本进行分词、词性、情感等标注，通常不存在采集环节。报告期内，发行人语义理解训练数据定制服务的标产比分别为 1.2、1.2、1.0 和 0.9，整体略有下降趋势是由于 2019 年开始发行人采取了机器辅助标注手段，进行预标注后再由人工校对，降低了人工投入，使标产比小幅下降。

（二）训练数据产品

1、语音识别采集

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
采产比 (每小时产出所需人数)	0.7	0.6	0.7	1.1
标产比 (每小时产出所需标注时长)	3.3	5.3	3.9	8.8

报告期内，发行人语音识别采集训练数据产品的采产比分别为 1.1、0.7、0.6 和 0.7，2018 年和 2019 年较低的原因与前述训练数据定制服务类似，发行人在设计训练数据产品时考虑特殊场景的配置，以及发音人资源稀缺等原因，延长了单个发音人的录音时长，2020 年 1-3 月较低的原因是当期存在部分训练数据产品的完善，仅进行标注但无采集环节，因此拉低了采产比。该等项目的具体情况如下：

分类	时间	新增时长（小时）	采集数量（人）	采产比
无采集环节	2020 年 1-3 月	1,532.00	-	-
采集特殊要求	2019 年	4,805.48	1,533	0.3
发音人资源稀缺	2019 年	11,284.97	5,003	0.4
	2018 年	15,077.02	8,965	0.6
其他一般情况	2020 年 1-3 月	2,637.98	3,003	1.1
	2019 年	10,461.16	9,079	0.9
	2018 年	8,794.21	8,836	1.0

扣除上述特殊项目，其他语音识别采集项目采产比在 2018 年、2019 年和 2020 年 1-3 月分别为 1.0、0.9 和 1.1，与 2017 年相比保持稳定，不存在明显差异。

报告期内，发行人语音识别采集训练数据产品的标产比分别为 8.8、3.9、5.3 和 3.3，2017 年和 2019 年较高的原因是由于当年发行人基于对行业和市场的判断，开发了部分自由对话类训练数据产品，由于自由对话类项目是由发行人预设场景，采集环节无预设文本，而由发音人自由发挥，发行人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，从而提高标产比，该等项目的具体情况如下：

分类	时间	新增时长 (小时)	标注数量 (工作小时)	标产比
自由对话类	2019年	10,096.25	82,941.48	8.2
	2017年	7,658.11	136,491.58	17.8
预设文本类	2019年	16,455.36	57,713.72	3.5
	2017年	13,288.65	48,260.41	3.6

自由对话类项目标产比在 2017 年明显高于 2019 年的原因是，2017 年为保证产品质量，对转写环节进行了 100% 质检，并进行了二次抽检，因此投入标注工时较多，后续公司引入语音识别算法技术进行部分辅助转写，且能够对产品库的错误率进行自动检测，因此 2019 年进行的自由对话类项目直接进行抽检即可，节省了标注工时，降低了标产比。

扣除上述特殊项目，其他语音识别采集标产比在报告期内分别为 3.6、3.9、3.5 和 3.3，不存在明显差异。

2、语音合成

项目	2020年1-3月	2019年度	2018年度	2017年度
采产比 (每小时产出所需万字数)	0.1	0.5	1.0	-
标产比 (每小时产出所需标注时长)	33.8	61.1	58.9	-

2018 年、2019 年和 2020 年 1-3 月，发行人语音合成训练数据产品的采产比分别为 1.0、0.5 和 0.1，2019 年和 2020 年 1-3 月较低的原因是当年进行了部分语音合成训练数据产品的完善或扩建工作，主要针对声音清晰度、标注准确率等进行完善，产生新增时长但无对应的新增采集量或新增采集量很低。该等项目的具体情况如下：

分类	时间	新增时长 (小时)	采集数量 (万字数)	采产比
产品完善与扩建项目	2020年1-3月	48.18	0.32	0.01
	2019年	28.79	0.71	0.02
其他一般项目	2020年1-3月	3.41	3.28	1.0
	2019年	28.10	30.04	1.1

扣除上述特殊项目，其他语音合成标产比在 2019 年和 2020 年 1-3 月分别为 1.1 和 1.0，与 2018 年相比保持稳定，不存在明显差异。

2018 年、2019 年和 2020 年 1-3 月，发行人语音合成训练数据产品的标产比分别为 58.9、61.1 和 33.8。2018、2019 年标产比较为平稳，高于训练数据定制服务的原因是由于训练数据产品通常标注内容全面，且对标注的质量、质检的比例有统一和较高的要求，因此标注投入较多。2020 年 1-3 月，发行人进行了部分语音合成产品的完善工作，只进行单一标注任务，所需标注工时较少，使得标产比偏低。

(7) 劳务数量及数据服务费用变动是否与发行人经营情况相匹配，数据服务费用的采购价格是否公允，劳务用工成本核算是否完整且不存在跨期，发行人建立的相关内部控制措施及执行情况

一、劳务数量及数据服务费用变动是否与发行人经营情况相匹配

报告期内，数据服务费用的变动由劳务数量的变动和采购单价的变动构成，发行人采购单价在报告期内与发行人的经营情况相匹配（采购单价的具体分析参见本问题“发行人说明事项/（5）采集和标注单价的具体确定依据和过程，同类型项目的采集和标注单价是否存在差异并分析具体原因”的回复内容），因此发行人采购的数据服务费与发行人经营情况的匹配程度，也反映了发行人采购的劳务数量与发行人的经营情况的匹配程度。数据服务费同发行人主营业务收入的匹配情况如下：

单位：万元

项目	2020 年 1-3 月	2019 年	2018 年	2017 年
采购的数据服务费	1,168.63	6,099.95	7,352.66	4,057.31
主营业务收入	3,126.09	23,755.81	19,265.77	11,907.09
采购的数据服务费占营业收入的比例	37.38%	25.68%	38.16%	34.07%

报告期内，发行人数据服务费采购占发行人营业收入的比例基本稳定，2019 年数据服务费占营业收入的比例较低主要是由于部分验收周期较长的语音合成项目在 2019 年验收，贡献收入。而这些项目的原料数据采集工作主要在 2018 年完成，在 2018 年 12 月 31 日的存货中进行归集，因此 2018 年末存货余额中的数据服务费有明显增加，从 2017 年末的 830.95 万元增加至 2018 年末的 1,827.06 万元，而 2019 年原料数据采集相应减少，数据服务费采购占收入比例也相应降低。

二、数据服务费的采购价格是否公允

（一）数据服务费的采购价格构成

报告期内，发行人采购定价由两部分组成，一部分是供应商支付给终端服务人员的费用，一部分是供应商就其提供的组织和管理服务收取的管理服务费。

第一部分的定价是随着整体市场人力成本的变动而变动，发行人按人或按工作小时等支付的单价整体上符合市场上对类似服务价格的定位。发行人在业务实施过程中，未发生在现行价格体系下，因价格过低造成服务资源不足的情况。发行人内部管理制度《海天瑞声业务采购实施细则》对数据服务费采购价格的公允性做出了规定：采购部对业务价格体系中的价格进行定期公允性检查，以确保供应商价格在合理范围内，检查方式包括但不限于同类供应商询价、供应商及其内部人员访谈、发布模拟招聘信息询价、波动原因解析等。

第二部分的定价是根据供应商提供服务的能力不同而有所差异，管理服务费按照原料数据采集、标注服务采购金额乘以管理费率计算，目前的管理费率是提供类似服务的人力资源公司或技术服务公司通行的收费水平，主要涵盖其项目人员薪酬、人员招募与管理、公司税费等成本，符合该类公司提供服务的技术含量和定位。对于境内数据服务提供商，由于不同的供应商在各类型项目中组织和管理被采集人员、标注人员的数量规模和管理难度有所差异，因此发行人与供应商分别约定采用 10-11.5% 的服务费率，存在小幅差异。对于境外数据服务提供商，主要考虑到供应商所在地的税务成本和人力资源匹配成本，发行人与供应商协商约定采用 6-15% 的服务费率。

（二）支付给终端服务人员的费用与市场上类似服务费用的比较

数据服务提供商支付给终端服务人员的费用与市场上类似服务费用的比较情况如下：

1、采集服务

市场上类似采集服务信息相对较少，从公开网站找到类似采集服务价格情况如下：

类别	发布网站	招聘职位	报价	服务地点	折算单价（元/人次）
采集	58 同城	数据采集员	150 元/天	北京	150
	大街网	数据采集员	120-150 元/天	北京	120-150

	BOSS 直聘	数据采集员	80-200 元/天	北京	80-200
	发行人		智能语音		0.27 元/字或词
					220.55 元/人

注1：以上市场信息均来自相应发布网站；

注2：由于市场上类似服务可比信息较少，特别是细分领域后更加难以准确找到合适信息进行对比，因此未进一步进行细分领域对比。

由于发行人的训练数据定制服务以及训练数据产品的研发，均有较强的单个项目需求差异化特征，采集业务受语种、发音人身份、地域、专业度、原料数据采集周期、原料数据特征需求等多种因素影响，而造成采购采集服务的价格有所差异。发行人会根据不同客户不同订单的特定需求，描绘被采集人的概况和特征，并承担不同水平的采购单价。2020年1-3月，发行人中文语音识别采集业务采集服务的平均单价（扣除个别项目因采集时长较短而导致采集服务单价偏低的影响后）为102元/人，与市场类似采集服务报价不存在明显差异。由于外语种资源的稀缺性，发行人外语种智能语音类业务采集服务平均采购单价较高，2020年1-3月外语种语音识别采集业务采集服务的平均单价为429元/人，提升了智能语音类业务采集服务的整体平均采购单价。

2、标注服务

根据标注服务的类别，选取了市场类似服务报价情况进行比较，标注服务采购单价对比情况如下：

类别	发布网站	招聘职位	报价	服务地点	折算单价（元/小时）
标注	58 同城	人工智能文本	100-220 元/天	北京	12.5-27.5
		数据标注员			
	实习僧	AI 数据标注员	120-200 元/天	北京	15-25
	大街网	中文标注员	100-200 元/天	北京	12.5-25
	BOSS 直聘	数据标注员	100-200 元/天	北京	12.5-25
	发行人平均价格		智能语音		21.64
			自然语言		23.83
	58 同城	图片数据标注	110-180 元/天	北京	13.75-22.5
	实习僧	图片标注员	100-250 元/天	北京	12.5-31.25
	大街网	图像处理员	100-200 元/天	北京	12.5-25
	BOSS 直聘	图片标注员	100-150 元/天	北京	12.5-18.75
发行人平均价格		计算机视觉		17.30	

注1：以上市场信息均来自相应发布网站；

注2：发行人的折算单价为2020年1-3月相关类别数据服务提供商支付给终端服务人员的平均含税单价；

注3：发行人智能语音、自然语言领域的标注都是在文本载体上进行的，因此统一归类为语音标注，与市场类似服务的价格进行对比，计算机视觉归类为图片标注与市场类似服务的价格进行对比。

如上表所示，发行人采购标注服务的价格与市场上类似服务的价格相比处于同一水平，发行人采购标注服务的价格与市场上类似服务的价格不存在明显差异。

（三）支付的管理服务费与市场上类似服务费用的比较

根据广州善世、芜湖善慧、北京博睿、北京零点、信阳联智等主要境内数据服务提供商出具的说明，其向其他客户收取的服务费率区间为 9-13%，根据恩泽尔等主要境外数据服务提供商出具的说明，其向其他客户收取的服务费率区间为 5-8%，与该等数据服务提供商向发行人收取的服务费率基本一致，不存在明显差异。

同时，根据发行人对市场同类人力资源服务公司和劳务公司的市场询价结果，同类境内服务商报价区间为 9-15%。发行人向供应商支付的 10-11.5%的服务费率符合境内服务商的普遍报价区间，服务费率的采购定价具备公允性。

三、劳务用工成本核算是否完整且不存在跨期

发行人在报告期内采购的劳务费支出主要反映在采购的数据服务费支出中。发行人在报告期内对劳务用工成本是采用了项目制的方式来进行归集和核算，即按照项目需求进行劳务用工采购，并按照项目归属来核算相关劳务用工成本。对于项目采购的劳务用工成本，发行人主要通过制定和执行严格的项目预算和项目决算制度来保证劳务用工成本核算的完整性。

对于每个项目，发行人在项目开始前制定详细的项目预算。对于项目预算中数据服务费部分，采集量需要参照“以销定采”的逻辑，并考虑合理的采集损耗率后确定，标注量需要参照合理的标产比后确定；采集单价和标注单价参照公司采购价格体系确定；项目负责人编制预算时需要将数据服务费的采购量、采购单价均按工作环节逐项、单独填报于业务管理平台，经过业务部门直属负责人、财务部审核审批通过后方可执行。

在项目进行过程中，数据服务费的采购严格按照项目预算执行，如果超出预算，则需要经过业务部门分管负责人、财务负责人批准后方可执行。

项目完工后必须及时进行项目决算。由项目经理提出决算申请，并说明预算完成情况及分析差异原因，并由财务部确认项目预算执行情况和款项付款情况，并对项目决算进行审批。项目决算应在项目全部验收后进行，最晚不得晚于验收后 10 个工作日，且

季度内验收的项目必须在季度末完成决算、当年验收的项目应在当年完成决算。项目决算后不得再对项目费用进行报销。

因此通过上述内部控制管理措施，发行人可以有效保证了劳务用工成本核算的完整性且不存在跨期。

另外，发行人与数据服务提供商之间的结算周期较短，报告期内应付账款周转天数平均为 6.45 天，因此相关劳务用工成本存在重大跨期结算和跨期入账的风险较低。从报告期内发行人账面记录的实际劳务成本金额来看，其与发行人的经营情况也是基本相符的，因此报告期内发行人的劳务成本核算是完整的，且不存在跨期现象。

四、发行人建立的相关内部控制措施及执行情况

报告期内，发行人建立的与数据服务采购和劳务用工成本核算相关的内部控制措施及执行情况请参见本问题“发行人说明事项/（1）发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效”的回复内容。

（8）前述服务公司的基本情况和员工数量，是否与发行人对其的采购规模相匹配，认定同一控制的依据，通过同一控制下不同企业开展业务的原因及合理性，发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务，主要小外包人员的基本情况，服务公司及其关联方和主要小外包人员是否与发行人客户及其关键经办人员、发行人控股股东、实际控制人、发行人关联方、员工及前员工之间存在关联关系、关联交易、资金往来或其他利益安排，是否存在通过服务公司向客户进行商业贿赂的情形

一、前述服务公司的基本情况和员工数量，是否与发行人对其的采购规模相匹配，认定同一控制的依据，通过同一控制下不同企业开展业务的原因及合理性，发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务

（一）前述服务公司的基本情况和员工数量，是否与发行人对其的采购规模相匹配

前述服务公司的基本情况和员工数量等情况见前述第本问询函回复第 10 题“发行人补充披露事项”之“(5)结合主要原料数据采集、标注服务提供商的规模、员工数量等，说明其是否有足够能力为发行人提供相关服务……”部分所述。

主要原料数据采集、标注服务提供商主要组织劳务人员来完成数据采集、标注服务，数据采集、标注服务并不涉及专门的业务资质，服务提供商具有劳务人员资源及劳务人员管理等业务经验，为包括发行人在内的客户群体提供服务。通过对比前述服务公司的经营财务数据与发行人对其的采购金额，以及就发行人对服务公司的业务需求、服务公司的人员资源及业务等经验进行分析，前述服务公司有足够能力为发行人提供相关服务，与发行人对其的采购规模相匹配。

(二) 认定同一控制的依据，通过同一控制下不同企业开展业务的原因及合理性，发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务

1、善世（广东）企业服务外包有限公司（曾用名“广州善世人力资源服务有限公司”，以下简称“广州善世”）、芜湖善慧人力资源服务有限公司（以下简称“芜湖善慧”）、才星（广州）科技服务有限公司（曾用名“广州金社保网络技术有限公司”，以下简称“广州才星”）、共青城宜邦人力资源服务有限公司（以下简称“共青城宜邦”）

(1) 认定同一控制的依据

该等公司系根据其股权结构及主要人员情况、人员之间关系等认定为同一控制，具体情况如下：

序号	公司名称	股权结构	主要人员
1	广州善世	陈玦霏持有 40% 股权，白云持有 40% 股权，高洁持有 20% 股权	陈玦霏任执行董事、经理，白云任监事
2	芜湖善慧	陈玦霏持有 32% 股权，莫天勤持有 32% 股权，徐洪超持有 30% 股权，王宏志持有 3% 股权，欧阳刚持有 3% 股权	陈玦霏任执行董事，莫天勤任监事
3	广州才星	王宏志持有 40% 股权，莫天勤持有 30% 股权，陈玦霏持有 30% 股权	王宏志任执行董事，陈玦霏任经理，莫天勤任监事
4	共青城宜邦	2019 年 12 月王宏志、莫天勤将合计持有的 70% 股权转让给夏建平；上述股权转让后，夏建平持有 70% 股权，陈礼秋持有 30% 股权，夏建平为广州善世股东高洁的母亲，陈礼秋为广州善世股东陈玦霏的妹妹	2019 年 12 月股权转让前，王宏志任董事，莫天勤任经理；2019 年 12 月股权转让后，夏建平任执行董事、经理，陈礼秋任监事

(2) 通过同一控制下不同企业开展业务的原因及合理性

上述该企业通过同一控制下不同企业开展业务主要基于分地域进行业务开拓和区域管理的需要，具有合理性。其中，广州善世主要负责华南地区业务，芜湖善慧主要负责华东地区业务，广州才星分担广州善世部分业务，业务转型为“互联网+”相关劳务业务后与发行人终止合作，芜湖善慧的业务于 2019 年下半年逐渐由共青城宜邦承接。

发行人实际向其进行采购时亦主要根据项目涉及地域、业务需要等向不同主体发送采购需求。

(3) 发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务

报告期内，发行人一直与广州善世开展业务合作，自报告期初至 2019 年 12 月与芜湖善慧开展业务合作，自 2018 年 4 月至 2019 年 5 月与广州才星开展业务合作，自 2019 年 9 月至今与共青城宜邦开展业务合作。报告期内各期发行人向广州善世、芜湖善慧、广州才星、共青城宜邦采购金额及占其营业收入比例具体如下：

年度	发行人采购金额（万元）	供应商营业收入（万元）	占比（%）
2017	1,722.70	12,965.02	13.29
2018	3,324.46	15,551.30	21.38
2019	3,153.41	28,063.01	11.24
2020 年 1-3 月	640.80	12,586.68	5.09

报告期内，发行人向广州善世及其受同一控制同类业务主体采购额占其营业收入的比例较低，其从事业务经营与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

2、北京博睿驰信人力资源有限公司（以下简称“北京博睿”）、恩泽尔商贸有限公司（以下简称“恩泽尔”）、北京零点起航营销策划有限公司（以下简称“北京零点”）

(1) 认定同一控制的依据

基于股权结构及委托经营管理的情况，北京博睿、恩泽尔、北京零点以及未和发行人开展业务合作的北京天创共盈科技发展有限公司、北京中惠商客科技有限公司为王驰同一控制，具体情况如下：

①北京博睿、恩泽尔为王驰的一人公司，王驰持有其 100% 股权；

②北京零点、北京零点的股东与王驰签署了《公司委托经营协议》，约定北京零点委托王驰统一经营管理，委托经营期限自 2015 年 6 月 1 日至 2020 年 5 月 31 日；

③北京天创共盈科技发展有限公司与王驰签署《公司委托经营协议》，约定北京天创共盈科技发展有限公司委托王驰统一经营管理，委托经营期限自 2017 年 1 月 1 日至 2021 年 12 月 31 日；

④北京中惠商客科技有限公司与王驰签署了《公司委托经营协议》，约定北京中惠商客科技有限公司委托王驰统一经营管理，委托经营期限自 2016 年 10 月 1 日至 2021 年 9 月 30 日。

(2) 通过同一控制下不同企业开展业务的原因及合理性

王驰持有 100% 股权的北京博睿、恩泽尔开展业务主要系按照地域划分，北京博睿负责境内业务，恩泽尔负责境外业务。因王驰具有业务拓展能力和管理劳务公司的经验，北京零点、北京天创共盈科技发展有限公司及北京中惠商客科技有限公司委托王驰进行经营管理。因此，上述企业同受王驰控制具有合理性。

报告期初，发行人与北京零点仅发生少量采购。发行人一直根据境内、境外劳务服务的需求分别向北京博睿、恩泽尔进行采购。

(3) 发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务

报告期内，发行人一直与北京博睿、恩泽尔开展业务合作，2017 年初与北京零点发生少量合作。报告期内各年度发行人向北京博睿、恩泽尔、北京零点等同受王驰控制的同类业务主体采购金额及占其营业收入比例具体如下：

年度	发行人采购金额（万元）	供应商营业收入（万元）	占比（%）
2017	849.57	1,999.85	42.48
2018	2,078.46	4,417.41	47.05
2019	1,573.46	5,271.08	29.85
2020 年 1-3 月	330.80	1,274.67	25.95

备注：为计算占比，恩泽尔各期美元营业收入按照当期 1 月 1 日、12 月 31 日或 3 月 31 日的汇率的平均值折算为人民币。

报告期内，发行人向北京博睿及其受同一控制的同类业务主体采购额占其营业收入的比例低于 50% 且总体呈下降趋势，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

3、北京冠华英才国际经济技术有限公司（以下简称“北京冠华”）

报告期内，发行人一直与其开展业务合作，各年度发行人向北京冠华采购金额及占其营业收入比例具体如下：

年度	发行人采购金额（万元）	供应商营业收入（万元）	占比（%）
2017	662.12	19,260.30	3.44
2018	905.61	21,757.42	4.16
2019	388.90	25,976.91	1.50
2020年1-3月	64.03	7,475.92	0.86

报告期内，发行人向北京冠华的各年采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

4、上海佑译信息科技有限公司（以下简称“上海佑译”）

发行人报告期内与其开展业务合作，2018、2019年度未发生采购，2017年度、2020年1-3月发行人向上海佑译采购金额分别为8.13万元、31.13万元，分别占其营业收入比例为0.16%、6.92%。

报告期内，发行人向上海佑译采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联；因此不存在专门或主要为发行人服务的情形。

5、信阳市联智网络科技有限公司（以下简称“信阳联智”）

（1）认定同一控制的依据，通过同一控制下不同企业开展业务的原因及合理性

信阳联智与未和发行人开展业务合作的信阳市晨曦电子科技有限公司（以下简称“信阳晨曦”）同受洪琪控制。洪琪持有信阳联智100%股权。基于洪琪的业务开拓能力和管理劳务公司的经验，信阳晨曦股东黄晨与洪琪签署《委托经营协议》，委托洪琪对信阳晨曦统一经营管理，委托经营期限自2017年9月15日至2020年12月31日。因此，信阳联智和信阳晨曦同受洪琪控制具有合理性。

（2）发行人采购额占其同类收入的比例，是否专门或者主要为发行人服务

报告期内，发行人一直与信阳联智开展业务合作向其采购数据服务，各年度发行人向信阳联智等同受洪琪控制的同类业务主体采购金额及占其营业收入比例具体如下：

年度	发行人采购金额（万元）	供应商营业收入（万元）	占比（%）
2017	89.08	184.53	48.27

2018	98.66	472.62	20.88
2019	104.29	407.77	25.58
2020年1-3月	1.89	62.71	3.01

报告期内发行人向信阳联智及其同一控制同类业务主体采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

6、SAKURA JAPAN LLC（以下简称“SAKURA”）

2017、2018、2019年度以及2020年1-3月，发行人从SAKURA采购金额分别为150.69万元、84.81万元、114.55万元、8.70万元，分别约占其营业收入的13.90%、11.82%、18.70%、3.12%。

报告期内发行人向SAKURA采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

7、THINKWILD STUDIOS, S.L.（以下简称“THINKWILD”）

发行人自2018年开始向THINKWILD采购数据服务。2018年度、2019年度，发行人采购金额分别为129.85万元、12.62万元，约占其营业收入的37%、3%，2020年1-3月未发生采购。

报告期内发行人向THINKWILD采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

8、TWIN STUDIOS

发行人自2018年开始向TWIN STUDIOS采购数据服务。2018年度、2019年度，发行人采购金额分别为229.97万元、125.08万元，2020年1-3月未发生采购。由于TWIN STUDIOS并非上市公司，无公开披露的财务数据，且未提供其营业收入数据，因此无法计算发行人采购金额占其营业收入的比例。根据TWIN STUDIOS确认，除向发行人提供劳务服务外，TWIN STUDIOS也为其他客户提供服务，且从事业务经营活动与发行人不存在任何关联，因此不存在专门或主要为发行人服务的情形。

9、Atrium Studio Medienproduktion GmbH（以下简称“Atrium”）

2017、2018、2019年度，发行人采购金额为8.23万元、125.54万、138.71万元，由于Atrium并非上市公司，无公开披露的财务数据，且未提供其营业收入数据，因此无法计算发行人采购金额占其营业收入的比例。根据Atrium确认，除向发行人提供劳务服务外，Atrium也为其他客户提供服务，且从事业务经营活动与发行人不存在任何关联，因此不存在专门或主要为发行人服务的情形。

10、Voices.com Inc

2017、2018、2019年度、2020年1-3月，发行人采购金额分别为25.01万元、60.12万元、91.29万元、44.41万元，根据Voices.com Inc的确认，报告期内发行人各期采购金额约占其当期营业收入的1%。

报告期内发行人向Voices.com Inc采购额占其营业收入的比例较低，且从事业务经营活动与发行人不存在其他关联，因此不存在专门或主要为发行人服务的情形。

二、主要小外包人员的基本情况

发行人报告期内数据采集、标注服务公司涉及的主要小外包人员约60余人。小外包人员主要负责根据服务公司的安排寻找符合发行人要求的终端人，组织并管理终端人提供劳务，具体包括人员信息登记、终端人授权书签署、人员时间安排及现场引导、对终端人的服务进度和数据完整性进行把控等，并与其负责的终端人进行结算支付。由于采集、标注服务采购涉及的采集、标注人员众多，且具备一次性、临时性、需求多样性等特征，以及对人员响应速度要求较快，仅单纯依靠服务公司直接寻找终端人员难以满足发行人业务扩张所需要的人员寻找和管理的需求，服务公司通过小外包人员寻找符合需求的人员能够提高寻找和管理效率，提高响应速度，满足发行人的需求。

上述主要小外包人员并非专职为服务公司提供劳务，其主要为自由职业者、企业人力资源岗位人员、销售岗位人员、在校大学生等具有人脉资源的岗位身份，主要利用其自由时间根据服务公司需求寻找联系组织终端人员为发行人提供劳务，并获取劳务报酬。上述工作的时间较灵活，主要通过人脉网络进行宣传，不需要较强的专业技术能力，不需要花费成本进行专门投资或投入专业设备，因此该项工作对上述人员具有吸引力。

三、服务公司及其关联方和主要小外包人员是否与发行人客户及其关键经办人、发行人控股股东、实际控制人、发行人关联方、员工及前员工之间存在关联关系、

关联交易、资金往来或其他利益安排，是否存在通过服务公司向客户进行商业贿赂的情形；

根据发行人及其控股股东、实际控制人、发行人主要关联方、关键岗位人员（主要采购、销售及财务人员）、服务公司以及主要小外包人员书面确认并经核查服务公司、发行人及其控股股东、实际控制人、发行人主要关联方、发行人关键岗位人员报告期内的银行流水，基于上述核查，服务公司及其关联方和主要小外包人员与发行人客户及其关键经办人员、发行人控股股东、实际控制人、发行人关联方、员工及前员工之间不存在关联关系、关联交易、资金往来或其他利益安排。

根据发行人及其实际控制人、总经理、主要销售人员、主要采购人员以及主要服务公司书面确认，不存在通过服务公司向客户进行商业贿赂的情形。

（9）服务公司作为扣缴义务人是否已替小外包人员或终端人员履行了代扣代缴税款义务，是否存在违反相关法律法规的情形

一、小外包人员和终端人员应税收入的范围

发行人通过服务公司采集和标注业务数据。资金支付方面，发行人按项目进度向服务公司结算采购费用。服务公司收到资金后，向小外包人员和其直接对接的终端劳务人员支付劳务费用。小外包人员收到资金后，向其对接的终端劳务人员支付费用。

在发行人采集和标注数据的业务开展过程中，小外包人员及终端劳务人员各自对其获得的收入承担个人所得税的纳税义务，但由于业务开展中的职责不同，二者的应税收入所有区别：终端劳务人员因其向发行人提供采集、标注劳务服务而获取收入，小外包人员因其向服务公司提供寻找、管理相关终端人员的劳务服务而获得收入。因此小外包人员作为终端劳务人员的组织者，其收到的服务公司付款中，大部分是代为收取的终端人员的劳务收入，该部分收入纳税义务主体为终端劳务人员，小外包仅对其自身取得的劳务报酬部分承担纳税义务。

二、发行人不同业务类型中小外包人员及终端人员的应税收入及缴纳情况

根据发行人的业务类型，对应的小外包人员及终端人员的应税收入及缴纳情况如下：

（一）采集服务

关于数据采集业务，存在以下两种情形：

(1) 语音识别采集项目、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，上述项目的人均结算单价（包含小外包人员及项目对应全部终端人员）分别是 171.07 元、254.74 元、218.38 元和 236.06 元，因此，此类个体（包含小外包人员和终端人员）平均收入较低，平均而言未达到劳务报酬个人所得税起征点。

(2) 语音合成项目

该等项目特点为采集人数较少，人均结算单价（包含小外包人员及项目对应全部终端人员）较高，终端人员（即发音人）单次收入平均而言达到了个人所得税缴纳标准。2017 年、2018 年、2019 年和 2020 年 1-3 月，发行人为采购此类采集服务向服务公司支付的数据服务费金额分别为 293.49 万元、380.42 万元、224.40 万元和 78.16 万元。假设终端人员（即发音人）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，服务公司需代扣代缴的个人所得税总额分别不超过 58.70 万元、76.08 万元、44.88 万元和 15.63 万元，服务公司所负有的代扣代缴义务对应的税款金额较低。

(二) 标注服务

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服务公司的标注服务。

标注人员主要为临时性兼职人员，单次劳务服务收入较低，与供应商签署劳务协议，按照劳务量领取劳务报酬。报告期内，上述标注项目的平均单个标注人员的年均劳务收入（包含小外包人员及项目对应全部终端人员）分别是 2,154.48 元、2,458.80 元、2,363.35 元和 1,847.02 元，金额较低，推断标注人员涉及个人所得税纳税义务的可能性较小，相关个人所得税扣缴义务人为服务公司。

发行人律师核查事项

(1) 核查程序

发行人律师主要履行了如下核查程序：

- 1、检索相关法律法规，对原料数据采集和标注所需要的专业资质进行研究。

2、查阅主要原料数据采集、标注服务提供商的官方网站及其提供的人员花名册/员工数量说明、财务报表等；在国家企业信用信息公示系统查询其基本信息；与主要供应商进行访谈，取得其确认函等；网络检索以发行人名义开展的招聘信息；查阅发行人与主要服务提供商签署的《业务服务协议》、验收结算确认单，查阅主要服务提供商与部分劳务人员签署的劳务协议；

3、查阅发行人报告期内向自然人采购、通过服务公司采购的采购金额及占比，了解发行人直接向自然人采购的原因；

4、和发行人相关工作人员沟通，了解发行人对供应商提供数据质量采取的相关内部控制措施，查阅发行人相关内部控制制度，现场查看采集、标注过程；

5、查阅租赁房产的租赁合同、产权证书，实地查看发行人租赁房产用途情况。

(2) 核查结论

经核查，发行人律师认为：上述服务公司为发行人提供有关原料数据的采集和标记的劳务服务，不需要具备专业资质；主要原料数据采集、标注服务公司有足够能力为发行人提供相关服务，相关服务公司报告期内不存在以发行人名义招聘实习生或者兼职人员的情形，发行人不存在以外包名义按劳务派遣用工形式使用劳动者的情形；报告期内，发行人向个人供应商采购的金额较低，占比很小，并非发行人主要采用的采购模式，而是作为发行人在个别业务中的少量补充性采购，具备商业合理性；发行人采取的质量控制措施能够保证供应商提供的数据满足其采集需求，发行人采取了相关内部控制措施并有效执行；发行人向个人租赁房产的实际用途与其法定用途相符。

申报会计师核查事项

(1) 核查程序

对于上述说明事项，会计师已经执行了与数据服务采购发生、准确性和截止性认定相关的审计程序及核查程序，具体请参见本问题的申报会计师核查事项。另外，针对上述发行人说明事项，申报会计师进一步执行了如下核查程序：

1、获取发行人业务人员事先垫付资金的具体明细，了解具体原因，检查资金流转的过程是否符合发行人的相关内控制度；

2、查阅中国境内主要服务公司及其同一控制主体营业执照、公司章程、报告期各期末员工花名册及财务报表、委托经营管理协议等，通过国家企业信用信息公示系统查询中国境内主要服务公司的基本信息；

3、获得发行人报告期内发生采购交易的主要服务公司的股东名单和主要小外包人员名单，与发行人报告期内的员工花名册进行核对，与发行人股东、实际控制人、发行人关联方进行核对，与发行人报告期内前十大客户的关键经办人员的名单进行核对，检查是否存在发行人服务公司及其股东和主要小外包人员，与发行人员工和前员工之间、与发行人股东、实际控制人、发行人关联方之间，与发行人客户关键经办人员之间的关联关系；

4、检查报告期内发生采购交易的主要服务公司的工商登记等公示信息，检查股东信息、办公地址与关联方信息，并与发行人报告期内前十大客户的工商登记信息等公开信息进行核对，检查主要服务公司与发行人主要客户之间是否存在关联关系；

5、取得发行人实控人个人银行账户、其他关联方（包括实控人控制的法人主体（中瑞安和创世联合）银行账户、发行人董监高个人银行账户）以及发行人关键岗位人员，包括主要采购人员、主要销售人员及主要财务人员银行账户在2017年、2018年、2019年的银行账户流水，检查是否存在与发行人服务公司及其股东和主要小外包人员的资金往来；

6、取得主要服务公司确认函，确认服务公司与发行人股东、实际控制人、发行人关联方、员工及前员工不存在关联关系、关联交易、资金往来或其他利益安排；

7、取得发行人及其实际控制人、总经理、主要销售人员、主要采购人员的书面确认函，确认不存在通过服务公司向发行人客户进行商业贿赂的情形；

8、获得发行人报告期内主要服务公司对外公示的工商登记信息及其他公开信息，检查主要服务公司及其股东是否存在因商业贿赂及其他违法行为造成受到处罚或刑罚的情况；

9、取得报告期内境内前五大服务公司的税务合规证明、涉税征信情况、涉税信息查询结果等当地税务局提供的信息，确认是否存在重大的税务违规情况。并取得境内前五大服务公司关于履行代扣代缴义务的书面确认文件。

(2) 核查结论

经核查，申报会计师认为：

1、报告期内，发行人披露的采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况与发行人实际情况相符，发行人与数据服务采购相关内部控制的设计合理、执行有效；

2、报告期内，服务公司小外包人员的数量变动情况合理，小外包人员不属于服务公司员工，不存在相互重合的情形；发行人不会对小外包人员进行直接的人员方面管理和控制，主要通过与服务公司的沟通合作实现劳务外包，并间接通过服务公司把控小外包对终端人员的寻找和组织管理工作；

3、发行人披露的事先垫付资金的具体原因、后续结算及差异调整情况与发行人实际情况相符，2017年、2018年垫付资金的金额较小，不会对发行人内部控制的有效性造成重大不利影响，2019年发行人对事先垫付资金的事项进行整改规范，不再存在上述由发行业务人员垫付资金的情形，符合发行人相关内部控制制度的规定；

4、发行人向服务公司支付服务管理费原因合理，加成率定价依据充分、公允，不同公司加成率之间的差异具备合理原因；

5、发行人采集和标注单价的确定依据和过程合理，同类型项目的采集和标注单价的差异原因合理；

6、发行人采集数据量、标注数据量和标产比相关的确定方法和具体依据合理，不存在随意修改或调整的情形，同类型项目的采集数据量、标注数据量和标产比的差异原因合理；

7、劳务数量及数据服务费用变动与发行人经营情况相匹配，数据服务费用的采购价格公允，劳务用工成本核算完整且不存在跨期，发行人建立的相关内部控制措施完善，且得到有效执行；

8、服务公司的基本情况和员工数量与发行人的采购规模相匹配，服务公司认定为同一控制的依据合理，通过同一控制下不同企业开展业务的原因合理，不存在服务公司专门或主要为发行人服务的情形；服务公司及其关联方和主要小外包人员，与发行人客户及其关键经办人员、发行人控股股东、实际控制人、发行人关联方、员工及前员工之

间不存在关联关系、关联交易、资金往来或其他利益安排；发行人及其实际控制人、总经理、主要销售人员以及主要采购人员不存在通过服务公司向客户进行商业贿赂的情形；

9、报告期内，服务公司已经作为扣缴义务人替小外包人员或终端人员履行了代扣代缴税款义务，不存在违反相关法律法规的情形。

申报会计师说明事项

（1）未将营业成本认定为关键审计事项的原因

根据《中国注册会计师审计准则第 1504 号——在审计报告中沟通关键审计事项》第九条规定：“注册会计师应当从与治理层沟通过的事项中确定在执行审计工作时重点关注过的事项。在确定时，注册会计师应当考虑下列方面：（一）按照《中国注册会计师审计准则第 1211 号——通过了解被审计单位及其环境识别和评估重大错报风险》的规定，评估的重大错报风险较高的领域或识别出的特别风险；（二）与财务报表中涉及重大管理层判断（包括被认为具有高度不确定性的会计估计）的领域相关的重大审计判断；（三）本期重大交易或事项对审计的影响。”第十条规定：“注册会计师应当从根据本准则第九条的规定确定的事项中，确定哪些事项对本期财务报表审计最为重要，从而构成关键审计事项。”

根据《中国注册会计师审计准则第 1504 号——在审计报告中沟通关键审计事项》应用指南 29 条规定：在确定某一与治理层沟通过的事项的相对重要程度以及该事项是否构成关键审计事项时，下列考虑也可能是相关的：（1）该事项对预期使用者理解财务报表整体的重要程度，尤其是对财务报表的重要性。（2）与该事项相关的会计政策的性质或者与同行业其他实体相比，管理层在选择适当的会计政策时涉及的复杂程度或主观程度。（3）从定性和定量方面考虑，与该事项相关的由于舞弊或错误导致的已更正错报和累积未更正错报（如有）的性质和重要程度。（4）为应对该事项所需要付出的审计努力的性质和程度，包括：①为应对该事项而实施审计程序或评价这些审计程序的结果（如有）在多大程度上需要特殊的知识或技能。②就该事项在项目组之外进行咨询的性质。（5）在实施审计程序、评价实施审计程序的结果、获取相关和可靠的审计证据以作为发表审计意见的基础时，注册会计师遇到的困难的性质和严重程度，尤其是当注册会计师的判断变得更加主观时。（6）识别出的与该事项相关的控制缺陷的严重

程度。(7) 该事项是否涉及多项可区分但又相互关联的审计考虑。例如, 长期合同可能在收入确认、诉讼或其他或有事项等方面需要重点关注, 并且可能影响其他会计估计。

根据审计准则的规定, 申报会计师在审计工作中评价具有重大错报风险的领域包括: (1) 收入确认; (2) 营业成本的核算。申报会计师将上述两个领域评估作为重大错报风险较高的审计领域予以重点关注。

其中, 对于营业成本的核算这一重点领域, 申报会计师考虑与其相关的重要会计科目和认定主要包括营业成本和存货的完整、真实和准确。对该领域相关的重要会计科目和认定, 评估其固有风险主要包括存货和成本的金额未被完整记录以及结转成本的不及时和不准确, 固有风险较高。申报会计师对发行人相关的内部控制进行了了解、测试和评价, 由于发行人与营业成本核算相关的内部控制存在且运行有效, 因此申报会计师将与营业成本核算相关的控制风险认定为低风险, 并依赖控制测试的结果设计和执行适当的实质性审计程序和其他核查程序, 主要包括抽凭检查、分析性复核、函证、走访和访谈、银行流水核查等。通过执行适当的审计程序和核查程序, 将营业成本核算领域的重大错报风险降低至可接受的水平。

根据审计准则的相关规定, 申报会计师需要从重点关注事项中选取对财务报表审计最为重要的领域和事项, 作为关键审计事项。申报会计师认为收入确认事项是对财务报表审计最为重要的事项, 因此将收入确认事项作为关键审计事项予以披露。

申报会计师考虑重点关注事项是否构成关键审计事项时, 考虑的因素主要包括重要程度和重要性、相关会计政策的复杂程度或主观程度、以及审计程序的特殊性和复杂性等方面。申报会计师认为, 收入是发行人的关键业绩指标之一, 存在管理层为了达到特定目标或期望而操纵收入确认时点的固有风险或收入被确认于不正确的会计期间的错报风险, 同时新实施的收入准则也需要管理层根据合同条款和业务安排评估对收入确认会计政策的影响, 该评估涉及管理层重大判断, 因此将收入确认认定为关键审计事项。对于营业成本核算, 申报会计师考虑到与其相关的会计处理相比较而言并不复杂, 不涉及管理层重大判断, 因此未将营业成本核算认定为关键审计事项。

(2) 对数据服务采购发生、准确性和截止性认定履行的具体审计程序

针对数据服务采购发生、准确性和截止性认定, 申报会计师执行的具体审计程序及核查程序总结如下:

（一）与数据服务采购支出相关的内部控制测试

通过询问了解报告期内与数据服务采购相关的具体业务流程及关键财务报告相关内部控制，包括与签署框架协议、立项及预算管理、采购申请的审批、下达采购通知、验收结算及费用报销、费用支付、项目成本归集、成本结转、项目决算等关键业务环节相关的内部控制情况。报告期内分别选取样本执行穿行测试，获取与上述关键业务环节的内部控制相关的各项单据，包括采购框架协议、项目预算文件、采购通知、验收结算单、费用报销申请、付款凭据、财务记账记录等信息，检查关键内部控制是否设计有效并得到实施。按照相关内部控制发生的频率，在报告期内选取样本执行控制有效性测试，检查相关内部控制是否在报告期内运行有效。另外，选择目前发行人在执行的项目，观察项目的实际采集和标注的流程与操作情况。

根据上述控制测试的结果，发行人在报告期内保持了与采购支出相关的有效的内部控制。

（二）数据服务采购的平均单价分析

获取发行人报告期内的不同业务领域的数据服务采购总量，包括采购的采集数据量与标注有效工时，与各个业务领域的采集与标注的采购支出金额进行对比，分析报告期内的不同业务领域的采集与标注的平均单价是否存在波动，并询问管理层价格波动原因，分析其合理性。将最近一期的发行人的采集与标注的采购平均单价与市场上可获得的类似服务的公开价格进行比较，检查是否存在差异，并询问管理层差异原因，分析其合理性。

具体计算的不同业务领域的采集与标注的采购平均单价及波动分析，请参见本问题“发行人说明事项/（5）采集和标注单价的具体确定依据和过程，同类型项目的采集和标注单价是否存在差异并分析具体原因”。

采集与标注的平均单价与市场上可获得的类似服务的公开价格进行比较分析的具体情况，请参见本问题“发行人说明事项/（7）劳务数量及数据服务费用变动是否与发行人经营情况相匹配，数据服务费用的采购价格是否公允，劳务用工成本核算是否完整且不存在跨期，发行人建立的相关内部控制措施及执行情况”。

根据上述对报告期内的采集和标注的平均采购单价的计算及分析，以及与市场类似服务的公开价格比较，发行人报告期内的不同业务领域的采集与标注的平均采购单价合理，不存在重大异常情况。

（三）数据服务采购的采集损耗率和标注产出比分析

抽取项目将项目的数据服务采购量核对至该项目的销售量，通过二者的比较核对来分析采购量是否合理，是否与销售量配比。其中对于采集环节，统计分析采购端的供应商结算量与销售端的客户验收量的差额占后者的比例（“采集损耗率”），分析项目的采集损耗率是否合理；对于标注环节，统计采购端的标注有效工作时长与销售端的客户验收量的比例（“标注产出比”，简称“标产比”），分析项目的标产比是否合理。

1、采集环节的损耗率分析

抽取项目的收入覆盖比例如下：

年份	核查项目收入(万元)	收入总额(万元) (注)	同类收入覆盖比例
2020年1-3月	745.71	907.91	82.13%
2019年	8,630.42	11,751.23	73.44%
2018年	6,371.44	8,221.99	77.49%
2017年	4,200.87	5,652.71	74.32%
合计	19,948.44	26,533.84	75.18%

注：同类收入的统计口径为当年或当期实现收入的训练数据定制服务和训练数据产品（不含重复销售）的收入，扣除了未包含采集环节的项目收入。

具体执行程序：

对于采集而言，发行人以客户合同数量基础，按照总体约5%以内的冗余度控制实际采集的数量。因此通过计算和分析采集环节的损耗率，来确认发行人采集环节采购数量的完整性、准确性。

核查中，抽取报告期内不同业务类型下的不同项目，获取其采购的采集量和销售量，计算报告期内的抽取项目的平均的采集损耗率。对于语音识别采集项目，检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、供应商结算单等外部证据资料，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；对于语音合成项目，检查报告期内项目的客户合同或订单、客户验收文件、与客户之间的语料设计沟通邮件、供应商结算单等外部证据资料，统计核对向供应商采集的字词数量、发行人设计文本字词数量、交付客户并验收字词数量；对于计算机视觉类项目，检查报告期内项目的客户合同或订单、客户验收文件、供应商结算单等外部证据资料，统计核对客户验收数量、供应商结算数量等数据。

具体按照各个业务类型计算与分析所抽取项目的采集环节的损耗率情况如下：

(1) 语音识别类项目

对于语音识别类型的抽取项目,通过将采购端的供应商结算人数与销售端的客户验收发音人人数进行对比分析,计算报告期内的抽取项目的平均的采集损耗率。采集损耗率= (供应商结算总人数-客户验收发音人总人数)/客户验收发音人总人数。各年抽查项目中,存在少数项目由于采集难度大、设备稳定性较差、部分发音人员多次录制等特殊因素,使得采集损耗率异常的情况,该等项目报告期各期分别为12个、17个、13个、2个。扣除该等项目影响后,其他抽样项目在报告期内的平均采集损耗率的计算结果如下,较为稳定,保持在5%的合理范围之内。

年份	客户验收发音人数(人)	供应商结算人数(人)	采集损耗率
2020年1-3月	1,606	1,670	3.99%
2019年	31,424	32,712	4.10%
2018年	41,378	43,074	4.10%
2017年	33,811	35,184	4.06%

(2) 语音合成类项目

对于语音合成类型的抽取项目,通过将采购端的供应商结算字数与销售端的客户验收字数进行对比分析,计算报告期内的抽取项目的平均的采集损耗率。采集损耗率= (供应商结算字数-客户验收字数)/客户验收字数。抽取项目在报告期内的平均的采集损耗率的计算结果如下,较为稳定,保持在5%的合理范围之内。

年份	客户验收字数(字/词)	供应商结算字数(字/词)	采集损耗率
2020年1-3月	220,087	230,676	4.81%
2019年	11,065,038	11,273,625	1.89%
2018年	6,001,700	6,184,489	3.05%
2017年	4,663,749	4,765,123	2.17%

(3) 计算机视觉类项目

对于计算机视觉类型的抽取项目,由于项目的数量单位较多,不同数量单位之间难以合并计算,因此抽查各年使用相对常见的数量单位(例如“张”或“人”)的项目进行分析。

抽取项目在报告期内的平均的采集损耗率的计算结果如下。采集损耗率= (供应商结算数量-客户验收数量)/客户验收数量。个别期间的采集损耗率略高于5%的一般正

常范围，原因分析如下：（1）2019年以“人”为数量单位的采集损耗率略高，主要是由于个别项目涉及唇语、人物动作准确率要求高等难度较大的采集，因此采集环节的作废率相对较高所致。（2）2020年1-3月以“人”为数量单位和以“张”为数量单位的采集损耗率略高，主要是由于部分项目客户需求中途变更或细化导致作废率较高所致。除此以外，报告期其他各个期间的计算机视觉类项目的平均采集损耗率较为稳定，保持在5%的合理范围之内。

年份	数量单位	客户验收数量	供应商结算数量	采集损耗率
2020年1-3月	张	242,122	255,640	5.58%
	人	600	635	5.83%
2019年	张	103,697	105,029	1.28%
	人	6,850	7,287	6.38%
2018年	张	1,023,069	1,023,069	0.00%
2017年	人	1,200	1,205	0.42%

对于上述采集损耗率超出5%的少数项目，申报会计师执行了进一步的审计程序，包括与项目负责人进行访谈，了解损耗率超出一般水平的原因，获取项目文件、人员邮件往来、文件记录等支持性依据，分析判断原因的真实性与合理性。

程序执行结果：

报告期内，抽取的各类项目的采集损耗率整体合理，发行人原料数据采集量与销售相匹配，采集环节的数据服务采购交易相关的采购数量真实、准确。

2、标注环节的标产比分析

抽取项目的收入覆盖比例如下：

年份	核查项目收入（万元）	收入总额（万元）（注）	同类收入覆盖比例
2020年1-3月	908.40	1,320.07	68.81%
2019年	14,032.38	16,930.51	82.88%
2018年	10,982.16	13,582.50	80.86%
2017年	6,553.50	8,228.96	79.64%
合计	32,476.44	40,062.04	81.07%

注：同类收入的统计口径为当年或当期实现收入的训练数据定制服务和训练数据产品（不含重复销售）的收入。

具体执行程序：

对于标注而言，发行人以客户合同销售数量基础，按照与供应商约定的标产比（即单位产出所需的标注工作量）进行采购结算。因此，通过计算各个业务类型中抽取项目的平均标产比并分析报告期内的变化情况，并通过选取同类型项目进行重新测速，与报告期内的抽取项目的平均标产比进行对比，来分析报告期内的标产比是否合理，分析和确认发行人标注工作量与销售量是否匹配。

核查中，抽取报告期内不同业务类型下的不同项目，获取其采购端的标注有效工作小时数量和销售端的客户验收量，按照不同业务类型分别计算抽取项目的平均的标产比，并进行报告期内的对比分析。项目采购端的标注有效工作小时数量的数据来源是供应商的验收结算单等外部依据，销售端的客户验收数量的数据来源是与客户的验收文件或销售合同或订单等外部依据。

具体按照各个业务类型计算与分析所抽取项目的标注环节的标产比情况如下：

（1）语音识别采集类项目的标产比分析

对于语音识别采集类型的抽取项目，通过将采购端与供应商结算的标注有效工作小时数与销售端的客户验收数量，即音频小时数，进行对比分析，计算报告期内的抽取项目的平均的标产比。标产比=标注量/销售量。计算结果显示如下。

2020年1-3月			
类型	标注量（小时）	销售量（小时）	标产比
读文本	6,134.24	2,811.50	2.2
自由对话	-	-	-
其他	-	-	-
合计	6,134.24	2,811.50	2.2
2019年度			
类型	标注量（小时）	销售量（小时）	标产比
读文本	92,916.52	40,978.00	2.3
自由对话	148,303.43	16,262.00	9.1
其他	12,859.45	2,155.10	6.0
合计	254,079.40	59,395.10	4.3
2018年度			
类型	标注量（小时）	销售量（小时）	标产比
读文本	215,925.88	72,914.40	3.0

自由对话	25,070.99	2,231.60	11.2
其他	3,291.57	2,997.00	1.1
合计	244,288.45	78,143.00	3.1
2017 年度			
类型	标注量（小时）	销售量（小时）	标产比
读文本	99,617.35	34,272.12	2.9
自由对话	135,656.99	6,635.00	20.4
其他	857.72	1,260.00	0.7
合计	236,132.06	42,167.12	5.6

报告期内，发行人语音识别采集项目以读文本和自由对话类项目为主，其他类型项目销售量较低。自由对话标注的复杂程度及难度明显高于读文本，因此标产比较高。发行人 2017 年语音识别采集自由对话项目中，存在部分语种稀缺、辨识难度较高、客户对数据要求高的项目，提高了 2017 年语音识别采集自由对话项目的标产比。2018 年开始，发行人逐步引入工具进行预标注和自动检测，标注速度有所提升。

为了进一步分析报告期内标产比的合理性，抽取 2020 年语音识别采集读文本项目和语音识别采集自由对话项目，分别进行测速得出标产比，并与上述抽取项目计算的平均标产比进行比较。测得语音识别采集读文本项目标产比为 2.4，语音识别采集自由对话项目标产比为 9.7。由于 2018 年开始，发行人逐步引入工具进行预标注和自动检测，标注速度有所提升，因此测速得出的标产比与 2019 年和 2020 年 1-3 月的数据可比。经比较，测速得到的标产比与 2019 年、2020 年 1-3 月抽取的语音识别采集项目的平均标产比基本一致。

（2）语音合成类项目的标产比分析

对于语音合成类型的抽取项目，通过将采购端与供应商结算的标注有效工作小时数与销售端的客户验收数量，即音频小时数，进行对比分析，计算报告期内的抽取项目的平均的标产比。标产比=标注量/销售量。计算结果显示如下。

年度	标注量（小时）	销售量（小时）	标产比
2020 年 1-3 月	3,102.69	49.79	62.3
2019 年	125,504.13	2,748.54	45.7
2018 年	67,795.49	1,513.86	44.8
2017 年	71,470.29	1,006.69	71.0

语音合成项目的细分标注环节较多，包括词性、韵律、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所差异，中文的标产比低于外语种。因此项目的复杂性造成了上述报告期内计算的抽取项目的平均标产比存在差异的情况。

为了进一步分析报告期内标产比的合理性，申报会计师抽取 2020 年语音合成项目，分别按词性标注、韵律标注、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
词性标注	13.2	14.7
韵律标注	6.7	10.0
音素边界标注	40.6	54.6
校对	7.7	14.4

对于上述报告期内抽取的项目，按照包含的明细标注环节，分为不同类型的标注项目。对于其中的两种主要类型计算平均标产比，并将其与测速项目的标产比进行对比如下：

序号	项目包含的明细标注环节	样本均值 2020年 1-3月	样本均值 2019年	样本均值 2018年	样本均值 2017年	中文项目 测速标产 比	外语种项 目测速标 产比
1	韵律+词性+音素边界+校对	80.8	106.5	67.6	97.9	68.2	93.7
2	韵律+音素边界+校对	40.0	61.9	53.1	82.8	55.0	79.0

由上表可以看出，发行人语音合成项目标产比基本介于测速得到的中文和外语种标产比数值之间，不存在显著异常。

另外，上述两种类型的语音合成项目的标产比就总体而言，自 2018 年开始随着发行人预标注技术和工具的普遍应用而呈现降低趋势。但是上述主要类型 1 在 2019 年标产比较高的原因是部分重点客户的语音合成项目标注质量要求高、难度大所致。上述主要类型 2 的标产比还受到项目语种结构的影响；2017 年标产比较高是由于当年主要为外语种业务，而外语种项目标产比相对较高所致；2018 年均为中文业务；2019 年主要为中文业务，因此标产比相对较低；2020 年 1-3 月标产比偏低是由于当期项目较少，其中有特殊项目存在部分数据仅需进行校对的情形，因此拉低了整体标产比。

(3) 语音识别转写类项目的标产比分析

对于语音识别转写类的抽取项目，通过将采购端与供应商结算的标注有效工作小时数与销售端的客户验收数量，即音频小时数，进行对比分析，计算报告期内的抽取项目的平均的标产比。标产比=标注量/销售量。计算结果显示如下。

2020年1-3月			
类型	标注量(小时)	销售量(小时)	标产比
短音频	55,719.70	4,490.76	12.4
长音频	8,184.14	638.21	12.8
合计	63,903.85	5,128.97	12.5
2019年度			
类型	标注量(小时)	销售量(小时)	标产比
短音频	649,548.80	61,797.70	10.5
长音频	249,456.59	16,911.46	14.8
合计	899,005.39	78,709.16	11.4
2018年度			
类型	标注量(小时)	销售量(小时)	标产比
短音频	847,732.21	70,774.67	12.0
长音频	288,698.46	20,148.72	14.3
合计	1,136,430.67	90,923.39	12.5
2017年度			
类型	标注量(小时)	销售量(小时)	标产比
短音频	337,673.87	36,166.29	9.3
长音频	199,813.38	12,318.81	16.2
合计	537,487.25	48,485.10	11.1

报告期内，发行人语音识别转写项目分为短音频类项目和长音频类项目。长音频通常语速较快且变化更为复杂，人物角色和领域较为多样，标注难度高于短音频，因此标产比较高。报告期内语音识别转写项目标产比相对较为稳定。语音识别转写短音频项目2017年标产比较低的原因是当年存在一定规模的短音频项目要求简单，仅需整理数据，因此标产比很低，拉低了短音频项目整体标产比。语音识别转写长音频项目2020年1-3月标产比较低的原因是当期部分项目的音频转写规则、转写内容较为简单，所需标注工时较少。

进一步抽取 2020 年语音识别转写短音频项目和语音识别转写长音频项目，分别进行测速得出标产比，测得上述短音频项目的标产比区间为 10.7-11.8，长音频项目的标产比区间为 14.6-19.5，与上述语音识别转写项目的标产比基本一致。

(4) 计算机视觉项目和自然语言项目的标产比分析

抽取项目中，如果是与客户确认标产比的项目，则根据客户邮件确认的标产比×交付数据量，计算得出客户认可的工作小时，作为销售量。同时以统计的供应商完成标注等业务环节的有效工时，作为标注量。

抽取项目中，如果不是与客户确认标产比的项目，则直接根据客户验收确认的销售量和与供应商结算的标注有效工作小时来作为销售量和标注量。

将上述与客户确认标产比的项目的标注量与销售量进行对比，计算抽取项目的平均的标产比。标产比=标注量/销售量。计算结果显示如下。

年份	标注量（小时）	销售量（小时）	标产比
2020 年 1-3 月	26,662.18	22,515.58	1.2
2019 年	256,082.18	237,135.98	1.1
2018 年	71,137.00	53,879.00	1.3
2017 年	34,740.00	27,402.00	1.3

上述与客户确认标产比的抽样项目实际采购标注工作小时大于客户认可的销售工作小时，即标产比大于 1，主要原因是与客户约定的工作量中不包含质检环节，而发行人需要进行质检保证标注质量，因此实际采购标注工作小时略大于客户认可的销售工作小时。

此外，进一步抽取 2020 年的计算机视觉简体中文手写体项目和繁体中文手写体项目分别进行测速。对于 2020 年 1-3 月计算机视觉的简体中文和繁体中文手写体项目，分别将其标产比与抽取项目的测速结果进行了对比，简体中文手写体项目标产比 0.50，与测速结果 0.48 基本一致，繁体中文手写体项目标产比 0.99，与测速结果 0.96 基本一致。

程序执行结果：

报告期内，发行人抽取的各类项目的标产比整体合理，发行人原料数据标注量与销售量相匹配，标注环节的数据服务采购交易相关的采购数量真实、准确。

（四）核查主要数据服务供应商的采购支出的主要支持性依据

发行人与数据服务供应商的采购交易的主要支持性依据包括以下几种：（1）发行人与供应商签署的框架合作协议，约定双方的权利与义务等，并约定数据服务管理费费率和结算方式；（2）发行人向供应商下达的采购通知，采购通知的内容包括项目简介、服务工期、符合项目预算的原料数据采集、标注服务量（标注服务量根据需要标注的数据量与双方协商确定的标产比相乘计算得出）、采集或标注的人员要求和特别要素等；（3）报销明细，发行人对供应商提供的劳务服务验收合格后进入采购结算环节，项目人员会在系统中进行报销，相关信息会录入报销明细，报销明细会载明项目财务编号、项目支出的类型，对应的项目任务环节和采购的任务需求，并记录对应采购的数量、单价和报销金额；（4）验收结算单，系统根据报销明细表生成验收结算单，并发送给供应商，由供应商确认后作为双方结算依据。验收结算单是发行人与供应商的结算依据，验收合格后双方签订验收结算单。验收结算单会载明项目财务编号、采购劳务内容、采购劳务的数量及对应金额，并由双方加盖公章；（5）供应商支付依据，供应商在对外付款时，一般会按发行人要求将银行回执进行截图并邮件发送给发行人。截图的银行回执会载明收款人信息、付款时间、付款金额等。

因此对于报告期内的主要服务公司（各期采购金额占比5%以上）的数据服务费采购交易，检查至其支持性依据，包括采购框架合同、下达的采购通知、经审核的报销明细、经双方盖章确认的验收结算单、供应商付款依据等，并核对供应商名称与结算金额等信息是否与账面记录一致。

上述核查采购交易金额的覆盖比例如下：

年份	核查数据服务费采购支出金额（万元）	数据服务费采购支出总额（万元）（注）	覆盖比例
2020年1-3月	1,039.52	1,168.63	88.95%
2019年	5,110.19	6,099.95	83.77%
2018年	6,262.67	7,352.66	85.18%
2017年	3,159.40	4,057.31	77.87%
合计	15,571.78	18,678.55	83.37%

上述核查程序执行结果显示，抽取检查的数据服务采购交易的金额与各项支持性证据的金额核对一致。

（五）采购支出在资产负债表日前后的截止性测试

从资产负债表日前后的数据服务费采购支出明细中选取样本执行截止性测试，通过检查双方盖章确认的验收结算单，来评价接近资产负债表日前后入账的数据服务费采购交易是否记录于恰当期间。

根据上述测试的执行结果，未发现数据服务费采购支出的重大截止性差异。

（六）信息系统相关测试

引用信息系统专家对报告期内的业务与财务相关的信息系统进行测试。相关信息系统包括一体化数据处理平台、业务管理平台 and 用友财务系统，测试内容包括信息系统一般性控制测试、信息系统应用控制测试、业务数据与财务数据相结合进行的数据分析。其中，信息系统的一般性控制测试和应用控制测试的结果均显示未发现重大异常，发行人信息系统的一般性控制和应用控制总体有效。

对于业务数据与财务数据的结合进行的数据分析程序，包括两类程序，一类是系统间的数据核对，将发行人报告期内业务管理平台的项目收入和成本信息与用友财务系统中确认收入与成本的信息进行比较，未发现重大差异。另一种类型是将业务数据与财务数据结合进行项目指标分析，通过分层方法随机抽样选取 2018 年、2019 年和 2020 年 1-3 月的项目执行该程序。项目指标分析中采用的具体抽样方法及结果如下：

分层方法		总体项目收入金额（注）			抽样核查项目收入金额			收入覆盖率
项目收入金额分层	抽样比例	2020年1-3月（万元）	2019年（万元）	2018年（万元）	2020年1-3月（万元）	2019年（万元）	2018年（万元）	合计
200万元以上	100%	-	4,562.15	1,137.57	-	4,562.15	1,137.57	100%
200万元以下	50%以上	1,320.07	12,368.36	12,444.93	879.20	8,254.94	6,219.42	59%
合计		1,320.07	16,930.51	13,582.50	879.20	12,817.09	7,356.99	66%

注：总体项目收入金额为发行人当年训练数据定制服务收入和训练数据产品销售收入（不含重复销售）的金额。

对于上述抽样选取的 2018 年、2019 年和 2020 年 1-3 月的项目，结合业务数据与财务数据进行数据分析，计算各项指标数据，包括项目的设计/导入/交付等环节文件数量一致性、废弃率、单位长度语音标注字节数量、项目标注效率、采集单价、单位语音时长采购成本、项目毛利率等，并进一步分析各项目的指标数据的波动或整体分布情况，其中识别出部分项目的指标数据与该类项目指标数据的一般情况相比存在偏离的情况，并执行进一步分析和核查程序，了解和确认这些被识别的项目指标存在偏离的主要原因

因，是否与实际情况相符，是否存在异常。经核查，所抽查项目的数据分析结果未发现异常，业务与财务指标在核查期间的变动趋势或表现在重大方面呈现一致性。

对于指标分析过程中所识别的指标偏离项目的具体的情况及执行的进一步审计程序，请参见本问题“会计师核查事项/（3）对 IT 审计报告中存在的异常项目履行的进一步审计程序”。

（七）主要数据服务供应商的函证程序

对报告期内的主要数据服务供应商进行函证，就于资产负债表日的往来款项余额及报告期间的采购金额进行确认。具体执行的供应商函证程序的情况请参见本问题会计师核查事项的第四项内容“对服务公司、小外包人员和终端人员履行的具体核查程序、核查过程、核查比例和核查结论”。

（八）主要数据服务供应商的访谈走访程序

选取主要数据服务供应商进行实地走访或其他形式的访谈，询问其与发行人的主要交易情况，是否与发行人存在关联关系，检查重要供应商及其交易的真实性，识别是否存在异常情况。具体执行的供应商走访程序的情况请参见本问题会计师核查事项的第四项内容“对服务公司、小外包人员和终端人员履行的具体核查程序、核查过程、核查比例和核查结论”。

（九）对服务公司、小外包人员和终端人员执行的进一步的穿透核查程序

对服务公司、小外包人员和终端人员执行进一步的穿透核查，具体核查程序包括：
（1）对 2017 年、2018 年和 2019 年的主要服务公司的主要银行账户的银行流水进行核查，检查是否存在发行人实际控制人、其他关联方及关键岗位人员向服务公司支付款项的情形；（2）抽取样本将 2017 年、2018 年和 2019 年发行人内部留存的支付小外包人员和终端人员的项目付款依据，核对至主要服务公司的主要银行账户的银行流水支出，对小外包和终端人员的款项支付进行穿透核查；（3）抽取样本对 2017 年、2018 年和 2019 年为发行人提供服务的小外包人员进行访谈；（4）抽取样本对 2019 年为发行人提供服务的终端人员进行访谈；（5）抽取样本对 2019 年终端人员就其所提供劳务工作量等信息进行签字确认的单据进行核查。具体执行的上述穿透核查程序的情况请参见本问题会计师核查事项的第四项内容“对服务公司、小外包人员和终端人员履行的具体核查程序、核查过程、核查比例和核查结论”。

上述执行的审计程序和核查程序的结果汇总如下：

序号	审计程序和核查程序		程序目的	结果及覆盖比例	
1	与数据服务采购支出相关的内部控制测试		确认与数据服务采购支出相关的关键内部控制存在且执行有效	报告期内与采购支出相关的关键内部控制存在且得到有效执行。	
2	数据服务采购的平均单价分析		通过检查数据服务采购单价的合理性，确认数据服务采购支出的真实性与完整性	报告期内的不同业务领域内全部原料数据采集与标注的平均采购单价整体合理。	
3	数据服务采购的采集损耗率和标注产出比分析		通过检查数据服务采购量与销售量是否配比，确认数据服务采购支出的真实性与完整性	报告期内进行采集损耗率分析的抽取项目的收入合计占报告期训练数据业务总体收入（不含训练数据产品重复销售）比例为 75.18%，进行标产比分析的抽取项目的收入覆盖比例为 81.07%。抽取项目的采集损耗率和标注产出比整体合理，采集量、标注量与销售量相匹配，数据服务采购交易相关的采购数量真实、准确。	
4	核查主要数据服务供应商的采购支出的主要支持性依据		检查与确认数据服务采购支出的真实性与准确性	报告期内核查采购支出的主要支持性依据的金额合计占发行人报告期数据服务采购支出金额合计比例为 83.37%，核查结果显示抽查的交易金额与各项支持性依据的金额核对一致。	
5	采购支出在资产负债表日前后的截止性测试		检查与确认数据服务采购支出的完整性	未发现在资产负债表日前后数据服务费采购支出存在重大截止性差异。	
6	信息系统相关测试		确认与数据服务采购支出相关的信息系统运行有效，并通过数据分析程序确认系统记录数据的真实性和准确性	信息系统的一般性控制测试和应用控制测试的结果显示发行人信息系统总体有效。数据指标分析的抽查项目收入覆盖比例为 66%，系统记录数据真实、准确。	
7	主要数据服务供应商的函证程序		确认服务公司采购支出的真实性、完整性和期末余额的准确性	报告期各期间回函确认相符的数据服务费采购金额占比均在 85% 以上，核查结果显示，与服务公司的采购交易真实、完整、准确。	
8	主要数据服务供应商的访谈走访程序		核实与服务公司采购交易的真实性，识别是否存在未披露的关联交易	报告期各期间走访的供应商的数据服务费采购金额占比均在 85% 以上，核查结果显示，与服务公司的采购交易真实、完整。	
9	对服务公司、小外包人员和终端人员执行的进一步的穿透核查程序	核查主要服务公司 2017 年、2018 年和 2019 年的银行流水	核实与服务公司、小外包人员和终端人员的采购交易的真实性，识别是否存在未披露的关联交易，识别是否存在发行人关联主体体外承担成本的情形	识别是否存在未披露的关联交易，是否存在关联主体体外承担成本的情形	核查银行流水的服务公司的数据服务费采购金额占发行人 2017 年至 2019 年数据服务采购合计金额的比例为 71.82%。核查结果显示，发行人不存在未披露的关联交易，不存在发行人关联主体体外承担成本费用情形。
		核查 2017 年、2018 年和 2019 年主要服务公司对小外包和终端人员的款项支付情况		核实与小外包和终端人员的采购交易的真实性	核查至服务公司银行流水的支付小外包和终端人员的金额占发行人 2017 年至 2019 年数据服务采购合计金额（不含数据服务管理费）的比例为 63.62%，核查结果显示，与小外包和终端人员的采购交易真实、准确。
		抽样对 2017 年、2018 年和 2019 年的小外包人员进行电话访谈和深度访谈		核实与小外包采购交易的真实性	访谈核查的小外包人员在 2017 至 2019 年的人数覆盖比例为 88%；覆盖通过小外包途径采购的数据服务金额比例为 89%，核查结果显示，与小外包的采购交易真实、准确。
		抽样对 2019 年完成项目的终端劳务人员进行电话		核实与终端人员的采购交易的真实性	访谈核查的终端人员在 2019 年的人数覆盖比例 62%，核查结果显示，与终端人员的采购交易真实发生。

		访谈			
		抽样对 2019 年终端人员提供劳务的签字文件进行核查		核实与终端人员的采购交易的真实性	核查签字文件的相关项目在 2019 年覆盖训练数据业务收入（不含训练数据产品重复销售）金额比例为 74%，核查结果显示，与终端人员的采购交易真实发生。

(3) 对 IT 审计报告中存在的异常项目履行的进一步审计程序

对于业务数据与财务数据相结合进行的数据分析程序，信息系统专家选取主要项目计算各项指标数据，包括项目的设计/导入/交付等环节文件数量一致性、废弃率、单位长度语音标注字节数量、项目标注效率、采集单价、单位语音时长采购成本、项目毛利率等，并进一步分析各项目的指标数据的波动或整体分布情况，其中识别出部分项目的指标数据与该类项目指标数据的一般情况相比存在偏离的情况，并执行进一步分析和核查程序，了解和确认这些被识别的项目指标存在偏离的主要原因是由于项目的特殊要求或出现的特殊情况导致的（具体分析如下），其指标偏离情况反映了项目的实际情况，因此数据分析程序的整体结论是所抽查项目的数据分析未发现异常。

对于上述数据分析中识别的项目指标出现偏离的情况，其原因主要是由于以下因素导致：1) 客户对项目的质量要求较高，如语音采集的高准确率要求、图片标注的高贴合度要求等，造成项目的废弃率、成本占比等指标偏高，标注效率指标偏低；2) 项目有一定的特殊专业性要求，例如歌曲、小语种、重口音、地域、医学特殊专业词等，造成项目的采集单价、单位语音时长采购成本等指标偏高，标注效率指标偏低；3) 项目有特殊环境要求，如录音棚环境、车载/噪音环境等，造成项目的采集单价、单位语音时长采购成本等指标偏高；4) 项目时间紧急，如交付时间比较紧张等，造成项目的采集单价、单位语音时长采购成本等指标偏高；5) 项目有特殊场景要求，如唤醒词/指令、自由对话等，造成项目的采购单价、单位语音时长的平均字数、标注效率等指标偏低或偏高；6) 项目在执行过程中出现了特殊情况，如设备问题、文件存储格式问题、客户提供的文件与项目设计不一致等情况，造成项目的设计、导入、交付等环节文件数量不一致。

针对上述识别出的指标偏离项目，申报会计师执行了进一步的审计和核查程序：

1、根据项目编号，检查项目的收入和成本支出的相关依据，包括项目合同、客户验收文件、采购支付明细、与供应商的采购验收结算单等，确认交易的真实性。

2、根据系统记录的客户名称，检查客户背景信息，了解是否存在异常客户。项目的客户大部分是发行人十大客户品牌相关客户的项目，这些客户在人工智能领域属于开拓性创新企业，对定制服务项目的专业性、环境、场景等存在更多的特殊要求，易导致相关项目的指标出现不同的偏离情况。

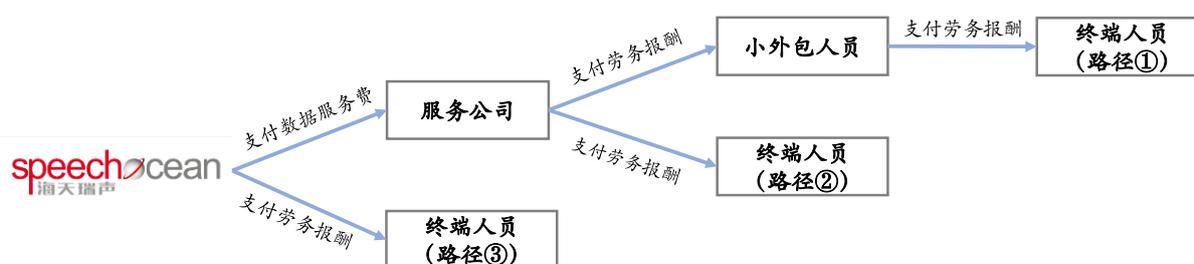
3、访谈项目负责人，了解项目负责人对相关项目的理解、项目的管理情况、执行过程的特殊情况及处理等。查阅其与客户联系人、项目执行人员/机构的沟通邮件或记录，以核实项目的要求及执行情况，了解项目指标出现偏离的原因，判断原因的合理性。

4、对于训练数据产品，查阅数据文件，如语音、图片、文本等，并与项目的具体要求进行比对。例如，抽取其中的录音，听取录音是否符合项目的小语种或重口音方言的要求、听取录音的内容是否与对应的文本信息一致、是否与项目的特殊场景一致；抽取其中的图片，观察其图片及标注是否与项目要求一致等。

通过执行上述的进一步审计和核查程序，对于上述数据分析程序中识别的指标偏离项目，申报会计师未发现舞弊情形，且不存在会计确认和处理方面的重大错报。

(4) 对服务公司、小外包人员和终端人员履行的具体核查程序、核查过程、核查比例和核查结论

发行人与服务公司、小外包人员和终端人员之间的采购路径如下：



报告期内，发行人向自然人和服务公司（包括服务公司采取小外包和直接对接两种形式）的采购金额情况如下：

单位：万元

采购方式	2020年1-3月		2019年度		2018年度		2017年度	
	采购金额	占比	采购金额	占比	采购金额	占比	采购金额	占比
通过服务公司采购	1,113.03	95.24%	5,912.58	96.93%	7,231.42	98.35%	3,816.55	94.07%
其中：通过小外包采购	857.21	73.35%	4,281.03	70.18%	4,567.55	62.12%	1,949.36	48.05%
直接对接采购	255.82	21.89%	1,631.55	26.75%	2,663.87	36.23%	1,867.19	46.02%

采购方式	2020年1-3月		2019年度		2018年度		2017年度	
	采购金额	占比	采购金额	占比	采购金额	占比	采购金额	占比
直接向自然人采购	55.60	4.76%	187.37	3.07%	121.24	1.65%	240.76	5.93%
数据服务费采购额	1,168.63	100.00%	6,099.95	100.00%	7,352.66	100.00%	4,057.31	100.00%

注：以下核查手段覆盖比例均以本数据表中采购金额作为计算基础

（一）核查程序

对服务公司、小外包人员和终端人员履行的核查程序的目的、结果及覆盖比例，请参见前述“（2）对数据服务采购发生、准确性和截止性认定履行的具体审计程序”中（七）（八）（九）的内容。

（二）核查过程和核查比例

1、对主要服务公司的函证程序

申报会计师对发行人报告期内的主要服务公司执行了函证程序。函证主要通过邮寄方式发放，对于少部分的海外服务公司或受疫情影响无法纸质寄达的国内服务公司采用电子邮件函证，采用电子函证的服务公司采购支出占采购支出总额的比例为0.67%，占比较低。询证函中所列示的信息通常包括：截至资产负债表日针对该服务公司的应付账款（或预付款项）当年/当期向该服务公司实现的采购额，以核实公司与服务公司之间的交易金额、期末往来余额信息。对未回函服务公司执行了相应的替代程序。

对于通过邮寄方式发放的函证，核对被询证者的名称、地址的真实性，以确保询证函中的名称、地址等内容的准确性。

对于通过电子邮件发放的函证，控制措施包括但不限于（1）将收件人邮箱的后缀与被询证者官网显示的企业邮箱后缀进行比对；（2）获取海天瑞声与服务公司在日常沟通时的邮件，确认该邮箱地址的使用者确实是被询证者与发行人日常业务的联系人，其对双方的交易情况知情及了解；（3）将盖章纸质函证扫描成PDF文件作为电子邮件附件。

对于通过邮寄方式收到的回函（1）核对被询证者确认的询证函是否是原件，是否与注册会计师发出的询证函是同一份；（2）核对回函是否由被询证者直接寄给注册会计师；（3）核对寄给注册会计师的回邮信封或快递信封中记录的发件方名称、地址是否与询证函中记载的被询证者名称、地址一致；（4）核对回邮信封上寄出方的邮戳显

示发出城市或地区是否与被询证者的地址一致；（5）核对被询证者加盖在询证函上的印章以及签名中显示的被询证者名称是否与询证函中记载的被询证者名称一致。

对于通过电子邮件收到的回函，（1）核对收到函证的发件人是否与发函时的收件人一致；（2）核对服务公司回复函证附件是否与发函文件一致。

对于回函金额核对至发函金额，如果存在差异，通过询问管理层了解差异原因，并抽查支持性依据确认财务记录的准确性。

执行函证程序的数据服务费采购金额及覆盖比例如下：

单位：万元

年度	服务公司发函情况			服务公司回函情况	
	采购金额	发函金额	发函比例	回函且回函相符金额	回函且回函相符比例
	A	B	B/A	C	C/A
2020年1-3月	1,168.63	1,137.71	97%	1,137.71	97%
2019年	6,099.95	5,874.03	96%	5,840.31	96%
2018年	7,352.66	7,143.03	97%	6,990.60	95%
2017年	4,057.31	3,571.23	88%	3,489.15	86%

2、对主要服务公司的访谈走访程序

申报会计师对发行人主要服务公司进行了实地走访、视频访谈或邮件访谈，了解服务公司与发行人之间的合作时间、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、与外部劳务人员是否签订协议、与发行人是否存在关联关系等情况，并取得对方签字确认的访谈纪要。通过上述访谈，未发现发行人与服务公司之间采购业务存在异常情形。

报告期内，实地走访或视频访谈的服务公司对应的数据服务费采购金额及覆盖比例如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
实地走访服务公司数量	4	6	6	6
视频访谈服务公司数量	3	4	4	4
邮件访谈服务公司数量	0	1	1	1
实地/视频/邮件访谈服务公司数据服务费采购额	1,066.77	5,529.27	6,820.37	3,474.15

数据服务费采购总额	1,168.63	6,099.95	7,352.66	4,057.31
覆盖比例	91%	91%	93%	86%

3、核查主要服务公司 2017 年、2018 年和 2019 年的银行流水

获取 2017 年、2018 年、2019 年为发行人提供数据服务的主要服务公司（各年数据服务采购金额占比 5% 以上）提供的主要银行账户流水，同时取得发行人实控人个人银行账户、其他关联方（包括实控人控制的法人主体（中瑞安和创世联合）银行账户、发行人董监高个人银行账户）以及发行人关键岗位人员，包括主要采购人员、主要销售人员及主要财务人员银行账户在 2017 年、2018 年、2019 年的银行账户流水，并执行以下核查程序：

（1）在服务公司的银行流水交易明细中检查是否出现发行人实控人、其他关联方以及关键岗位人员的银行账号，检查交易对手方是否出现发行人三年内的前十大销售客户名称。

（2）在发行人实控人、其他关联方以及关键岗位人员上述三年银行账户流水中，对于单笔支出超过 10 万元的交易对手方，检查该交易对手方是否出现在服务公司的银行流水中，即判断是否存在上述主体将款项支付给中间方后由其转入给发行人供应商的情形。

（3）在发行人实控人、其他关联方以及关键岗位人员银行流水支出项中检查是否出现供应商支出项交易对手方的银行账号信息，以判断上述主体与终端劳务提供者领款人之间是否存在资金往来的情形。

已执行上述核查程序的服务公司的银行账户情况及核查比例如下：

单位：万元

年份	已核查服务公司数量（个）	已核查服务公司银行账户数量（个）	已核查银行账户的服务公司当年采购额	数据服务费采购支出总额	已核查银行账户的服务公司的采购额占比
2019 年	5	20	4,721.30	6,099.95	77.40%
2018 年	5	10	5,357.06	7,352.66	72.86%
2017 年	5	8	2,497.28	4,057.31	61.55%
总计			12,575.64	17,509.92	71.82%

4、核查 2017 年、2018 年和 2019 年主要服务公司对小外包和终端人员的款项支付情况

获取 2017 年、2018 年、2019 年发行人的数据服务费支出明细，以及发行人按照项目逐笔归集的服务费支出的支持性依据，包括供应商验收结算单、付款明细表、报销明细表。同时获取 2017 年、2018 年、2019 年主要服务公司（各年数据服务采购金额占比 5% 以上）提供的主要银行账户流水及相关支付信息，并执行以下核查程序：

（1）逐笔检查发行人数据服务费支出的报销依据，包括报销明细表和验收结算单，确认其支出性质均是与具体项目相关，其发生具有业务真实性。

（2）由于服务公司在将款项支付给小外包和终端人后，会向发行人提交付款明细表或银行付款回执截图，确认已将服务费款项支付给小外包和终端人员。因此进一步逐笔检查发行人数据服务费支出已经发放给小外包和终端人员的支付依据，包括付款明细表或银行付款回执截图，确认其支出已经发放至小外包或终端人员。

（3）根据上述数据服务费的支付信息，进一步检查是否与服务公司的银行账户流水信息，核对支付给小外包或终端人员的金额是否相符。

已执行上述核查程序的数据服务费支出的情况及核查比例如下：

单位：万元

年度	已核查至银行流水的数据服务费支出金额	数据服务费采购支出总额	数据服务费采购支出总额（不含支付给服务公司的管理费）	已核查至银行流水的金额占数据服务采购支出金额的比例（不含管理费）
2019 年	4,092.24	6,099.95	5,891.65	69.46%
2018 年	4,619.22	7,352.66	7,075.89	65.28%
2017 年	2,036.31	4,057.31	3,926.82	51.86%
总计	10,747.77	17,509.92	16,894.36	63.62%

5、抽样对 2017 年、2018 年和 2019 年的小外包人员进行电话访谈和深度访谈

根据发行人业务部门留存的 2017 年、2018 年和 2019 年小外包人员的名称及联系方式，对全体小外包人员随机抽样进行电话访谈核查，就其组织终端人员提供服务的情况，领款金额等信息进行确认；同时对于重点小外包人员（各年覆盖数据服务费支出金额在 100 万元以上、或当年组织终端劳务人员数量达 500 人以上的重点小外包人群），采取深度访谈方式，深入了解其组织终端人员为发行人提供服务的具体业务模式及其必要性，了解小外包通过供应商为发行人提供服务的定价情况、结算方式，并向其核实各年其为海天瑞声寻找的终端采集/标注人员的数量、结算金额。

2017年、2018年和2019年发行人三年小外包数量分别为269人、293人及706人，通过①电话等直接联系方式及②深度访谈方式开展访谈的覆盖情况如下：

(1) 人数覆盖情况

项目	小外包人员总数			核查小外包人员数量			核查人数覆盖率
	2019年	2018年	2017年	2019年	2018年	2017年	三年合计
100万元以上	6	8	1	6	8	1	100%
50-100万元之间	12	18	3	12	16	3	94%
10-50万元之间	75	66	44	69	62	42	94%
10万元以下	613	201	221	578	148	167	86%
合计	706	293	269	665	234	213	88%

(2) 金额覆盖情况

金额单位：万元

年度	数据服务费采购总额	其中：通过小外包途径采购的数据服务费金额	核查金额覆盖	核查金额覆盖率
2019年	6,099.95	4,281.03	3,966.96	92.66%
2018年	7,352.66	4,567.55	4,004.37	87.67%
2017年	4,057.31	1,949.36	1,667.92	85.56%
合计	17,509.92	10,797.94	9,639.26	89.27%

6、抽样对2019年完成项目的终端劳务人员进行电话访谈

由于发行人采购的劳务服务大多数属于一次性交易，除少量直接对接终端人员外，发行人主要通过服务公司→小外包→终端人的形式对终端人进行分层管理，2017、2018年留存的终端人有效联系方式数量较少。因此，针对2019年确认了收入（不含训练数据产品重复销售）的项目，对其数据服务采购涉及的终端人员按照领取劳务费用金额区间进行了分层随机抽样，对于抽样样本执行了电话访谈程序。

前述核查范围的终端劳务人员合计约为5.14万人，包括发行人直接对接的终端人员以及通过服务公司→小外包→终端人和服务公司→终端人三种类型。通过对全量总体样本分层抽样合计对3.18万名终端人员进行了电话访谈。

具体为：对于领取劳务费用金额在9万元以上的终端人，按80%的覆盖率执行随机抽样核查；对于领取劳务费用金额在1-9万元区间的终端人，按70%的覆盖率执行随机

抽样核查；对于领取劳务费用金额低于 1 万元的终端人，按 60% 的覆盖率执行随机抽样核查。

领取劳务费用金额区间	抽样比例	终端劳务人员总体数量（万人）	抽样核查的终端劳务人员数量（万人）	抽样核查比例
		①	②	②/①
9 万元以上	80%	0.004	0.003	80%
1-9 万元	70%	0.06	0.04	70%
1 万元以下	60%	5.07	3.13	60%
合计	62%	5.14	3.18	62%

中介机构通过聘请外部第三方机构等方式对抽样选择的 3.18 万终端人员进行电话访谈，并进行录音，电话访谈的内容主要包括：是否确认为发行人提供过劳务及工作内容，是否领取报酬及具体金额。申报会计师对通话过程进行了观察和抽样监听。

已执行上述核查程序的 2019 年终端人员的核查情况如下：电话接通的终端劳务人员数量约为 1.28 万人，根据接听者的性别、方言、环境等因素可以判断为不同终端自然人；前述接通人群合计覆盖核查人群的人数比率约为 40%。未接通原因主要为：电话访谈过程中往往存在呼出号码被认定为陌生电话或被误认为骚扰电话而遭到拒接、号码占线或信号较差等原因导致无法接听的情形；也存在少量手机号码显示停机等，主要因为发行人数据服务采购终端人员流动性较大，存在其提供劳务服务后更换联系方式的情形，导致此前预留的号码出现异常状态。具体核查比例如下：

领取劳务费用金额区间	抽样核查的终端劳务人员数量（万人）	接通范围的终端劳务人员数量（万人）	接通比率人数
	①	②	②/①
9 万元以上	0.003	0.002	66%
1-9 万元	0.04	0.02	55%
1 万元以下	3.13	1.25	40%
合计	3.18	1.28	40%

7、抽样对 2019 年终端人员提供劳务的签字文件进行核查

对于提供数据服务的终端劳务人员，发行人 2019 年要求其对于提供的劳务工作量等信息进行签字确认，以完善数据服务采购管理并充分留痕。对于发行人一体化数据处理平台上线终端人管理系统前，终端劳务人员签字确认主要为线下纸质形式；上线终端

人管理系统后，采用线上电子签名的形式进行确认。因此会计师核查了 2019 年终端人员的线下签字单据和线上签字单据，检查上述签字文件是否包含了项目编号、劳务内容、劳务提供时间，是否有终端劳务人员的签字或电子签名，以确认是否与发行人的业务相关，是否经终端劳务人员签字确认。

已执行上述核查程序的 2019 年终端人员的核查比例如下：

核查了 2019 年终端人员确认所提供劳务工作量的签字文件合计 80,761 份，并根据项目编号查找与核查的终端人员签字文件相关的项目收入金额。已核查的终端人员签字文件相关的项目收入占发行人总体收入（不含训练数据产品重复销售）的比例为 74%。

（三）核查结论

经核查，申报会计师认为：发行人与服务公司、小外包和终端人员之间的采购交易真实存在，发行人对数据服务费采购的核算真实、准确、完整。发行人不存在未披露的关联交易，不存在发行人关联主体为发行人体外承担成本费用情形。

10.2 关于采购情况

2017 年、2018 年、2019 年，发行人采购总额分别为 5,651.62 万元、9,662.35 万元及 8,890.41 万元。

请发行人结合说明事项修改招股说明书的相关披露内容。

请发行人说明：（1）发行人各类采购额计入的会计科目及核算情况、核算依据，与财务报表的勾稽关系；（2）项目租赁费用支出的具体内容、原因以及与具体项目之间的匹配关系；（3）其他的具体采购内容及金额，相关采购支出的原因；（4）办公房租与租赁房产之间的对应关系，相关租赁房产的具体用途，出租方与服务公司及其关联方和主要分包人员之间是否存在关联关系、关联交易、资金往来或其他利益安排。

请申报会计师对上述事项进行核查并发表明确意见。

发行人说明事项

（1）发行人各类采购额计入的会计科目及核算情况、核算依据，与财务报表的勾稽关系

一、发行人各类采购额计入的会计科目及核算情况、核算依据

单位：万元

类别	2020年 1-3月	2019年	2018年	2017年	计入的会计 科目	核算内容	核算依据
数据服务费	1,168.63	6,099.95	7,352.66	4,057.31	存货 营业成本 研发费用	主要核算采购的数据服务费，其中（1）存货中核算的是当期未完工的训练数据定制服务项目累计发生的数据服务费；（2）营业成本中核算的是当期完工的训练数据定制服务项目当期结转的数据服务费；（3）研发费用中核算的是当期发生的训练数据产品开发项目的数据服务费。	项目报销申请单、验收 结算单
办公房租	210.05	694.66	545.17	415.48	存货 其他流动资 产 营业成本 管理费用 研发费用	主要核算支付的租用办公室房租的费用，其中（1）其他流动资产中核算的是预付的房租费用，再按照使用情况计入存货、营业成本、管理费用和研发费用；（2）存货中核算的是当期未完工的训练数据定制服务项目累计发生的房租费用；（3）营业成本中核算的是当期完工的训练数据定制服务项目当期结转的房租费用；（4）管理费用中核算的是管理部门和人员使用的租赁房屋的当期房租费用；（5）研发费用中核算的是研发部门和人员使用的租赁房屋当期的房租费用以及当期发生的训练数据产品开发项目的房租费用。	租房合同、房租支付申 请单、项目报销申请单
资产采购	17.76	287.12	552.90	203.72	预付款项 固定资产 无形资产	主要核算采购的电脑、录音设备、软件等固定资产和无形资产支出，其中预付款项核算的是已支付的与资产采购相关的预付款项；固定资产和无形资产核算的是采购新增的资产，在获得正式发票后直接计入固定资产或无形资产，或从预付款项转入该类科目。	资产采购合同及订单、 发票、支付申请单
项目租赁费	18.59	274.06	345.08	159.36	存货 营业成本 研发费用	主要核算为训练数据定制服务项目和训练数据产品开发项目专门租赁的采集或标注场地和设备等发生的费用支出。其中（1）存货中核算的是当期未完工的训练数据定制服务项目累计发生的项目租赁费；（2）营业成本中核算的是当期完工的训练数据定制服务项目当期结转的项目租赁费；（3）研发费用中核算的是当期发生的训练数据产品开发项目的项目租赁费。	项目报销申请单、租赁 合同、发票
岗位服务费	58.74	284.89	256.62	226.95	销售费用 管理费用 研发费用	主要核算根据语种扩展的需求聘请外部语言学家所发生的劳务费，及根据日常经营需要采购的临时岗位外包服务费，如保洁、临时招聘服务、少量实习生招聘等。其中（1）外部语言研究学家的费用均在研发费用中核算，（2）岗位外包服务的费用根据服务部门在销售费用、管理费用和研发费用中核算。	合同、发票、支付申请 单

采购总额	1,473.77	7,640.69	9,052.42	5,062.82			
------	----------	----------	----------	----------	--	--	--

二、发行人各类采购额与财务报表的勾稽关系

1、数据服务费与财务报表之间的勾稽关系

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
存货-数据服务费余额的增加/（减少）（扣除存货跌价准备的转销）	459.49	-990.51	996.11	242.36
营业成本-数据服务费	432.73	5,245.49	5,399.51	2,618.57
研发费用-数据服务费	276.41	1,844.97	957.04	1,196.38
合计	1,168.63	6,099.95	7,352.66	4,057.31
采购的数据服务费	1,168.63	6,099.95	7,352.66	4,057.31
差异	-	-	-	-

2、办公房租与财务报表之间的勾稽关系

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
存货-其他-房租费余额的增加/（减少）（扣除存货跌价准备的转销）	5.67	-30.76	-7.00	17.90
其他流动资产-待摊房屋租赁费的增加/（减少）	29.33	31.25	7.76	-44.01
营业成本-其他-房租费	1.29	71.04	42.58	18.54
管理费用-租赁费	154.90	548.80	426.50	417.70
研发费用-房租费	18.86	74.34	75.33	5.35
合计	210.05	694.66	545.17	415.48
采购的办公房租	210.05	694.66	545.17	415.48
差异	-	-	-	-

3、资产采购与财务报表之间的勾稽关系

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
预付款项-其他-资产采购的增加	-	50.47	-	-
固定资产-本期/年购置（不含购置的房屋建筑物）	17.46	166.92	408.00	200.80
无形资产-本期/年增加金额	0.30	69.73	144.90	2.92

合计	17.76	287.12	552.90	203.72
资产采购	17.76	287.12	552.90	203.72
差异	-	-	-	-

4、项目租赁费与财务报表之间的勾稽关系

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
存货-其他-项目租赁费余额的增加/（减少）（扣除存货跌价准备的转销）	13.18	-105.75	83.73	105.72
营业成本-其他-项目租赁费	5.41	373.49	251.17	35.96
研发费用-其他-项目租赁费	-	6.32	10.18	17.68
合计	18.59	274.06	345.08	159.36
采购的项目租赁费	18.59	274.06	345.08	159.36
差异	-	-	-	-

5、岗位服务费与财务报表之间的勾稽关系

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
销售费用-职工薪酬-劳务费	12.96	45.25	32.20	3.82
管理费用-职工薪酬-劳务费	18.57	98.67	81.00	60.18
研发费用-语言研究-劳务费	27.20	140.97	143.42	162.95
合计	58.74	284.89	256.62	226.95
采购的岗位服务费	58.74	284.89	256.62	226.95
差异	-	-	-	-

(2) 项目租赁费用支出的具体内容、原因以及与具体项目之间的匹配关系

项目租赁费支出核算训练数据定制服务和训练数据产品开发过程中临时需要的由发行人直接租赁场地和设备所发生的费用。报告期内项目租赁费用支出的具体内容、原因及与具体项目间的匹配关系见下表：

2020年1-3月：

单位：万元

序号	项目名称	项目租赁费	项目租赁支出的具体内容	租赁原因
1	中文女生语音合成	15.57	租赁录音棚	外地采集项目，租赁当地录音棚
2	音阶录音校对语音合成	1.22	租赁录音棚	配合发音人特殊要求、提升录制效率，租赁录音棚
3	新加坡中文（简体）手写体采集	0.66	租赁设备	同时进行的项目较多，租赁设备
4	台湾中文（繁体）手写体采集	0.65	租赁设备	同时进行的项目较多，租赁设备
5	中文女声语音合成及肢体采集	0.39	租赁设备	需要专业剪辑处理，租赁剪辑设备
6	中文女生语音合成	0.09	租赁录音棚	项目要求特殊，租赁专业录音棚
	项目租赁费合计	18.59		

2019年：

单位：万元

序号	项目名称	项目租赁费	项目租赁支出的具体内容	租赁原因
1	法语女声语音合成	57.66	租赁录音棚	国外采集项目，租赁当地录音棚
2	法语男声语音合成	57.66	租赁录音棚	国外采集项目，租赁当地录音棚
3	德语女声语音合成	48.21	租赁录音棚	国外采集项目，租赁当地录音棚
4	德语男声语音合成	47.78	租赁录音棚	国外采集项目，租赁当地录音棚
5	中文女生语音合成	11.30	租赁录音棚	外地采集项目，租赁当地录音棚
6	中文女生语音合成	8.52	租赁录音棚	项目要求特殊，租赁专业录音棚
7	中文女童语音合成	6.48	租赁录音棚	外地采集项目，租赁当地录音棚
8	中文全消音室语音识别采集	5.78	租赁录音棚	项目要求特殊，租赁专业录音棚
9	人脸车载视频数据采集标注	5.75	租赁汽车和设备	车载视频采集项目，租赁汽车、专业设备
10	多设备多场景语音识别采集	5.30	租赁汽车和录音场地	多场景采集，租赁汽车及场地
11	其他项目	19.60	租赁汽车、录音场地、设备	车载采集项目，租赁汽车；项目要求特殊，租赁专业录音棚、专业设备
	项目租赁费合计	274.06		

2018年：

单位：万元

序号	项目名称	项目租赁费	项目租赁支出的具体内容	租赁原因
1	德语女声语音合成	49.11	租赁录音棚	国外采集项目，租赁当地录音棚

2	德语男声语音合成	49.11	租赁录音棚	国外采集项目，租赁当地录音棚
3	中文女生语音合成	41.58	租赁录音棚	外地采集项目，租赁当地录音棚
4	英国英语男声语音合成	34.44	租赁录音棚	国外采集项目，租赁当地录音棚
5	法语女声语音合成	32.36	租赁录音棚	国外采集项目，租赁当地录音棚
6	法语男声语音合成	29.96	租赁录音棚	国外采集项目，租赁当地录音棚
7	中文手写体采集	29.71	租赁场地	项目要求采集城市，租赁当地专业场地
8	英国英语女声语音合成	23.53	租赁录音棚	国外采集项目，租赁当地录音棚
9	中文男声语音合成	6.94	租赁录音棚	工期紧张，自有录音棚调配不开，租赁录音棚
10	中英文男女声语音合成	6.69	租赁录音棚	外地采集项目，租赁当地录音棚
11	其他项目	41.63	租赁汽车、录音场地、设备	车载采集项目，租赁汽车；项目要求特殊，租赁专业录音棚、专业设备
	项目租赁费合计	345.08		

2017年：

单位：万元

序号	项目名称	项目租赁费	项目租赁支出的具体内容	租赁原因
1	中文唤醒词及家居环境语音识别采集	35.91	租赁场地	项目要求特殊场景录音，租赁场地
2	上海方言车载语音识别采集	17.29	租赁汽车	车载语音采集项目，租赁汽车
3	中文女生语音合成	14.99	租赁录音棚	外地采集项目，租赁当地录音棚
4	英国英语女声语音合成	12.96	租赁录音棚	国外采集项目，租赁当地录音棚
5	儿童唤醒词录音棚语音识别采集	8.37	租赁录音棚	工期紧张，自有录音棚调配不开，租赁录音棚
6	中文普通话语音识别采集	7.17	租赁录音棚	外地采集项目，租赁当地录音棚
7	英国英语男声语音合成	6.89	租赁录音棚	国外采集项目，租赁当地录音棚
8	中文车载语音识别采集	6.08	租赁汽车	车载语音采集项目，租赁汽车
9	中英文女声语音合成	4.62	租赁录音棚	工期紧张，自录音棚调配不开，租赁录音棚
10	中文男生语音合成	2.69	租赁录音棚	工期紧张，自录音棚调配不开，租赁录音棚
11	其他项目	42.40	租赁汽车、录音场地、设备	车载采集项目，租赁汽车；项目要求特殊，租赁专业录音棚、专业设备
	项目租赁费合计	159.36		

(3) 其他的具体采购内容及金额，相关采购支出的原因

单位：万元

类别	2020年1-3月	2019年	2018年	2017年	采购内容	采购原因
中介服务费及专利申请费等	62.98	286.52	170.75	31.88	管理咨询服务、招聘服务及安保服务、信息系统咨询服务、专利申请服务等	管理咨询费主要用于企业采购绩效管理体系优化设计等咨询服务，招聘费用于企业聘请专业机构进行人才招聘，安保服务费用于保安服务，信息系统咨询费主要用于企业信息系统平台的升级优化等，专利申请服务费用于协助企业申请专利等
市场宣传费	12.62	44.76	52.82	29.71	采购的市场宣传服务、展位展览服务及宣传资料印刷等	用于企业品牌形象的宣传推广，扩大企业知名度
押金及保证金	2.14	25.26	59.36	135.08	房租押金、投标保证金等	房租押金用于房屋租赁，投标保证金用于项目投标
装修费	-	64.83	44.24	15.91	办公室装修改造费用等	办公室装修
办公费	3.07	38.17	52.38	32.82	日常办公用品采购	用于维持企业日常办公经营的必要支出
水电费	8.54	36.78	31.08	20.55	日常水电费	用于维持企业日常办公经营的必要支出
邮寄费	2.40	8.76	14.24	9.47	日常快递物流服务	用于维持企业日常办公经营的必要支出
网络服务费	8.57	25.48	35.20	36.98	采购服务器、宽带支持服务	用于维持企业日常办公经营的必要支出
其他	33.93	142.14	104.48	113.37	通讯费、劳保用品等	用于维持企业日常办公经营的必要支出
采购支出其他小计	134.25	672.70	564.56	425.77		

由于上述采购不属于与发行人经营直接相关的采购支出，因此将其从报告期披露的采购明细中删除，并同步更新招股说明书“第六节 业务和技术/四、采购情况和主要供应商/（一）报告期内采购情况/1、报告期内的分类采购明细数据”的相关披露。

(4) 办公房租与租赁房产之间的对应关系，相关租赁房产的具体用途，出租方与服务公司及其关联方和主要分包人员之间是否存在关联关系、关联交易、资金往来或其他利益安排

发行人办公房租与主要租赁房产间的对应关系，及租赁房产的具体用途列示如下：

2020年1-3月：

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积(m ²)	2020年1-3月房租支出(万元)	用途
1	发行人	李宏	北京市海淀区成府路28号4座8层801、02、03、06、08	2017/06/17-2020/05/16	847.50	102.15	办公
2	发行人	李宏	北京市海淀区成府路28号4座8层807、09、11	2017/06/28-2020/05/16	408.13	49.19	办公
3	发行人	赵峻	北京市海淀区成府路28号3-801、3-803、3-807、3-809	2018/10/23-2020/05/15	500.31	41.60	办公
4	发行人	徐小勇、苗英	北京市海淀区成府路28号12层4-1215	2019/04/19-2020/04/18	196.98	- (注1)	办公
5	发行人	荣成市科学技术局	荣成经济开发区海湾南路86号科技企业孵化器内	—	约400	- (注2)	办公
6	发行人	三河市城美房地产开发有限公司	维多利亚D座1801、1803	2019/12/25-2020/12/24	152.71	4.20	办公
7	发行人	三河市城美房地产开发有限公司	维多利亚D座1805	2019/06/25-2021/06/24	78.54	2.10	办公
其他(注3)						10.81	
合计						210.05	

注1：2020年1-3月发行人未向徐小勇、苗英支付办公房租，是由于发行人按照双方租赁合同的约定，于2019年4月和2019年9月将该合同下的全部办公房租支付完毕。

注2：2017年荣成市科学技术局为支持发行人进驻荣成科技企业孵化器内创办软件开发企业，向发行人提供办公场地租赁服务，同时给与发行人租赁费优惠，租赁期自2017年7月1日起计算，前三年荣成市科学技术局给与发行人免缴租金的优惠，满三年后租赁价格另行商定。

注3：其他租赁房产为发行人由于临时性办公需求的而进行短期租赁的房产。

2019年：

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积(m ²)	2019年房租支出(万元)	用途
----	-----	-----	--------	------	-----------------------	---------------	----

1	发行人	李宏	北京市海淀区成府路28号4座8层801、02、03、06、08	2017/06/17-2020/05/16	847.50	289.72	办公
2	发行人	李宏	北京市海淀区成府路28号4座8层807、09、11	2017/06/28-2020/05/16	408.13	139.18	办公
3	发行人	赵峻	北京市海淀区成府路28号3-801、3-803、3-807、3-809	2018/10/23-2020/05/15	500.31	168.70	办公
4	发行人	徐小勇、苗英	北京市海淀区成府路28号12层4-1215	2019/04/19-2020/04/18	196.98	59.67	办公
5	发行人	荣成市科学技术局	荣成经济开发区海湾南路86号科技企业孵化器内	—	约400	-	办公
6	发行人	三河市城美房地产开发有限公司	维多利亚D座1801、1803	2018/12/25-2019/12/24&2019/12/25-2020/12/24	152.71	8.40	办公
7	发行人	三河市城美房地产开发有限公司	维多利亚D座1805	2018/06/25-2019/06/24&2019/06/25-2021/06/24	78.54	4.20	办公
其他(注)						24.8	
合计						694.66	

注：其他租赁房产为发行人由于临时性办公需求的而进行短期租赁的房产。

2018年：

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积(m ²)	2018年房租支出(万元)	用途
1	发行人	李宏	北京市海淀区成府路28号4座8层801、02、03、06、08	2017/06/17-2020/05/16	847.50	276.93	办公
2	发行人	李宏	北京市海淀区成府路28号4座8层807、09、11	2017/06/28-2020/05/16	408.13	133.36	办公
3	发行人	赵峻	北京市海淀区成府路28号3-801、3-803、3-807、3-809	2018/10/23-2020/05/15	500.31	53.61	办公
4	发行人	荣成市科学技术局	荣成经济开发区海湾南路86号科技企业孵化器内	—	约400	-	办公
5	发行人	三河市城美房地产开发有限公司	维多利亚D座1801、1803	2017/12/25-2018/12/24&2018/12/25-2019/12/24	152.71	3.63	办公
6	发行人	三河市城美房	维多利亚D座1805	2017/06/25-2018/6/24&20	78.54	2.10	办公

		地产开 发有限 公司		18/06/25-201 9/06/24			
7	中瑞智	陈治宇	北京市海淀区成府路 28 号 4 座 1002 号房间	2017/01/03-2 019/01/02	163.52	40.49	办公
其他 (注)						35.04	
合计						545.17	

注：其他租赁房产为发行人由于临时性办公需求的而进行短期租赁的房产。

2017 年：

序号	承租方	出租方	租赁房屋坐落	租赁期限	租赁面积 (m ²)	2017 年房租支出 (万元)	用途
1	发行人	北京臣仕物业管理 有限公司	北京市海淀区成府路 28 号 4 座 8 层 801、02、 03、06、08	2015/09/01-2 017/06/16	847.50	70.61	办公
2	发行人	李宏	北京市海淀区成府路 28 号 4 座 8 层 801、02、 03、06、08	2017/06/17-2 020/05/16	847.50	150.23	办公
3	发行人	李宏	北京市海淀区成府路 28 号 4 座 8 层 807、09、 11	2017/06/28-2 020/05/16	408.13	68.32	办公
4	发行人	荣成市科 学技术局	荣成经济开发区海湾 南路 86 号科技企业孵 化器内	-	约 400	- (注 1)	办公
5	发行人	三河市城 美房地产 开发有限 公司	维多利亚 D 座 1801、 1803	2016/12/25-2 017/12/24&2 017/12/25-2 018/12/24	152.71	3.63	办公
6	发行人	三河市城 美房地产 开发有限 公司	维多利亚 D 座 1805	2016/06/25-2 017/06/24&2 017/06/25-20 18/6/24	78.54	2.02	办公
7	发行人	北京邦泰 摩尔资产 管理有限 公司	北京市海淀区成府路 28 号 4 座 13 层 1315	2015/08/10-2 017/06/16&2 017/06/17-20 17/08/03	196.98	23.71	办公
8	中瑞智	陈治宇	北京市海淀区成府路 28 号 4 座 1002 号房间	2017/01/03-2 019/01/02	163.52	49.19	办公
其他 (注)						47.77	
合计						415.48	

注：其他租赁房产为发行人由于临时性办公需求的而进行短期租赁的房产。

经核查主要服务公司 2017-2019 年的银行流水，同时出租方与服务公司均已出具确认函，主要小外包人员已接受访谈并确认，出租方与服务公司及其关联方和主要小分包人员之间不存在关联关系、关联交易、资金往来或其他利益安排。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

- 1、了解发行人与采购相关的关键内部控制制度，评价内部控制的设计是否得到执行，并测试相关内部控制的运行有效性；
- 2、获取发行人的采购支出明细，核对至账务记录和财务报表，检查是否一致；
- 3、获取项目租赁费的支出明细，了解项目租赁费的支出原因，检查相关销售合同和租赁合同，查看是否存在异常，
- 4、获取房屋租赁的支出明细，了解租赁房产的用途，检查是否存在异常；
- 5、获取服务公司、出租方出具的确认函，核查主要服务公司 2017-2019 年的银行流水，并对主要小外包人员进行访谈，就服务公司和出租方及其关联方和主要小分包人员是否存在关联关系、关联交易、资金往来和其他利益安排进行确认。

(2) 核查结论

经核查，申报会计师认为：

- 1、发行人的采购额、核算情况和核算依据符合发行人的实际情况，采购额计入的会计科目核算准确，采购额与财务报表相关科目的余额及变动勾稽相符；
- 2、发行人的项目租赁费用支出原因合理，与具体项目之间存在匹配关系，符合发行人的实际情况；
- 3、发行人的其他类别采购支出原因合理；
- 4、发行人的办公房租采购额与租赁房产之间存在对应关系，相关租赁房产的具体用途符合发行人的实际情况，租赁房产出租方与服务公司及其关联方和主要小分包人员不存在关联关系、关联交易、资金往来和其他利益安排。

三、关于发行人核心技术

11. 关于技术先进性

招股说明书披露，发行人已将其核心技术体系运用于设计、采集、加工、质检等训练数据生产相关的各个环节，构成了发行人业务持续经营的技术基础；发行人构筑了较为明显的核心技术先进性。

请发行人披露：（1）核心技术与训练数据定制服务和训练数据产品之间的关系，是行业通用技术还是发行人特有技术，如属特有技术，与行业通用技术的异同；（2）发行人在训练数据生产过程中是如何运用大数据和算法的，如何检查训练数据集对AI 算法模型的训练效果，定量描述数据处理能力和数据处理效率；（3）结合与主要竞争对手在经营情况、技术指标、产品侧重点、研发投入、市场占有率、客户结构和客户数量等方面的比较，说明发行人技术竞争优势和劣势；列表说明发行人设计的数据集结构较同行业竞争对手或下游客户自身的训练数据集结构之间的差异，如何体现发行人数据集结构设计的优势；（4）发行人相关技术是否成熟或存在快速迭代的风险，下游市场空间是否接近饱和情形；相关风险揭示是否充分；（5）发行人已构筑较为明显的核心技术先进性的依据。

答复：

发行人披露事项

（1）核心技术与训练数据定制服务和训练数据产品之间的关系，是行业通用技术还是发行人特有技术，如属特有技术，与行业通用技术的异同；

一、核心技术与训练数据定制服务、训练数据产品之间的关系

发行人已在招股说明书“第六节 业务和技术/六、发行人的研发和技术/（三）核心技术与训练数据定制服务、训练数据产品之间的关系”针对下述内容进行了补充披露：

发行人从事的训练数据定制服务、训练数据产品相关业务，实质均在于为客户提供可供AI 算法模型训练使用的专业训练数据集。发行人训练数据定制服务的业务过程、训练数据产品的生产过程，就是发行人生产训练数据并交付给客户的过程。

截至目前，发行人共积累了12项核心技术，覆盖基础研究、平台工具及训练数据生产三个层次，均应用于训练数据生产全流程（设计、采集、加工、质检）之中；上

述核心技术构成了发行人生产经营的基础，发行人通过应用上述核心技术、实现成果转化，形成了基于核心技术体系的训练数据相关产品及服务。

发行人的各层次核心技术在训练数据生产环节中的应用情况具体如下：

核心技术层次	核心技术	核心技术项下的细分技术示例	核心技术在人工智能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		√		
	★数据同步技术	音频对齐技术			√	
	实时采集质量控制技术	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				√

二、发行人的核心技术是发行人特有技术；发行人特有技术与通用技术的差异点

发行人已在招股说明书“第六节 业务和技术/六、发行人的研发和技术/（六）发行人核心技术与通用技术的差异情况说明”针对下述内容进行了补充披露：

（一）发行人的核心技术以通用技术为理论基础，经持续的原始创新、自主研发和实践积累形成，是发行人的特有技术

通用技术无法直接满足训练数据生产研发的生产经营需求，发行人以其为理论基础，进行了持续的原始创新和自主研发，形成自己的特有技术。

发行人生产经营所涉及的技术主要包括人工智能算法类技术、语音语言学基础研究、训练数据集设计等基础技术，训练数据生产过程中使用的直接技术群（设计、采集、加工、质检等环节的多项具体核心技术），以及训练数据研发生产过程所需的工具、平台等。

前述领域的通用技术为一般性的数据采集、处理环节提供了基础，但在训练数据领域，不同语种/方言、领域、场景的训练数据在设计、采集、加工、质检各个环节均具备特异性要求，需要运用在相应领域积累的技术诀窍（know-how），例如在设计阶段，发行人就需要运用训练数据集设计技术，在设计中充分考虑基础语料的结构、内容比例，才能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，考虑不同的发音习惯、语言特点，最终实现更好的语料覆盖效果，从而生产出的训练数据集能够帮助 AI 算法模型获得更好的训练结果。

（二）发行人的特有技术与通用技术的差异点

发行人将多领域的技术创新性、有机性地整合在一起，并与研发生产中积累的实践经验相结合，构建了适应自身生产经营所需的多项特有技术，同时还在特定领域、场景对前述特有技术进行了综合性地创新运用，提升了训练数据开发过程各环节的生产效率。截至本问询函回复出具日，发行人已经取得 20 项发明专利授权及 1 项实用新型专利授权，充分说明了发行人的技术实力和技术独特性。

发行人特有技术与通用技术的差异点具体如下：

发行人核心技术层次	发行人的核心技术	行业通用技术简述	发行人核心技术与行业通用技术的具体差异点	在各核心技术领域取得的发明专利情况
基础研究	语音识别算法	语音识别算法指的是 Automatic Speech Recognition, 是让机器能够“听懂”人类语音的技术, 通过机器自动将语音信号转换为文本, 是实现智能人机交互的关键技术之一。	与行业通用技术相比, 发行人核心技术的差异点主要体现在: 基于对训练数据各生产环节细分场景的理解, 发行人对通用的语音识别算法进行了自主深度调优, 充分适配并嵌入运用至训练数据生产环节的各个工具中, 提升数据质量和生产效率。在语音识别算法领域, 发行人通过原始创新取得了包括语音数据库质量预估技术在内的专利技术。	已取得 2 项发明专利授权, 正在申请 1 项发明专利
	语音合成算法	语音合成算法指的是 Text to Speech 或者 Speech Synthesis, 是让机器能够“说出”人类语音的技术, 能将任意文字信息转化为流畅的语音朗读出来, 相当于给机器装上了人工嘴巴。	发行人对通用的参数合成、基于神经网络的端到端等语音合成技术进行了自主深度调优和落地化开发, 形成了特有的核心技术, 运用至训练数据生产过程中。在语音合成算法领域, 发行人通过原始创新取得积累的专利技术可有效提升端到端算法对语音合成训练数据集的运用效果, 实现更好的训练成果、提升发行人训练数据集的质量。	已取得 1 项发明专利授权, 正在申请 2 项发明专利
	计算机视觉算法	计算机视觉指的是 Computer Vision, 是一门研究如何使机器“看”的科学, 是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等	发行人基于通用的计算机视觉算法进行了原始创新和落地研发, 积累了适用于训练数据生产过程的技术诀窍 (know-how)。发行人在该领域进行了大量原始创新, 掌握了可应用于训练数据生产过程中的核心技术, 具体如下: 1、发行人运用人脸识别相关计算机视觉算法技术研发了短视频中相似人脸检查技术, 在该细分领域取得了 3 项发明专利授权, 可用于检查、保证计算机视觉训练数据集中人脸数据的独立性、提升训练数据质量。 2、发行人创新性结合人脸识别相关的计算机视觉算法和语音处理算法, 研发了短视频音频聚类技术, 并在该技术领域已取得 1 项发明专利授权, 可对多模态训练数据集执行数据关联性、重复性检测, 为提升训练数据质量提供了特有技术手段。 3、在物体识别和场景分割等技术领域研发了多项可用于图像分割、图像标注等计算机视觉训练数据生产的核心技术。	已取得 4 项发明专利授权, 正在申请 2 项发明专利
	★语音语言学基础研究	语音语言学基础研究主要指各个语种的高质量发音词典的构建能力, 包括音素集设计、发音规则制定、常见词标音等细分技术。	发行人通过积累语音语言学基础研究成果, 积累了 150 余个语种/方言的发音词典构建能力, 特有核心技术包括: 成熟的发音词典构建技术和流程积累, 可针对不同语言的特点, 从语音识别训练数据集、语音合成训练数据的模型训练角度出发, 优化细分技术和算法模型。	无
	训练数据集设计技术	训练数据集设计技术主要涉及的通用技术包括语言模型相关技术和文本正则化技术。语言模型是对语句概率分布的建模, 用来衡量一句话是否合理存在的方式, 对语言模型的研究和应用能帮助人工智能算法找到最可能出现的语句。文本正则化技术在语音合成领域的应用是将文本中的数字、符号、缩写等转换成语言文字的技术, 为了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。	在训练数据的设计环节, 与通用技术相比, 发行人核心技术的差异点主要在于: 1、发行人的设计流程和自有算法融入了发行人在多年训练数据生产过程中积累的实践经验和技术诀窍 (know-how, 如关于音素均衡、语料主题、覆盖领域、发音人属性、设备特征、自然环境等因素分布、设计的经验化技术积累); 2、发行人对语言模型、文本正则化等通用技术自主进行了系统性实现与落地优化, 将其嵌入自主开发的一体化数据处理平台中, 使公司具备高质量训练数据集的设计能力。例如, 在文本正则化技术领域, 发行人可处理 7 种主流语言、26 种不同种类非标准词; 具备超过 2,000 万句文本的累计处理经验, 积累了大量技术诀窍 (know-how) 和经验。	已取得 1 项发明专利授权
平台工具	一体化数据处理平台	为发行人自主开发的训练数据生产过程使用的数据处理平台, 无直接对应的通用技术。	该平台由发行人自主开发, 整合了发行人多年的训练数据开发生产实践经验, 完整覆盖训练数据生产环节, 拥有统一的人员管理、统一的数据管理、统一的风格管理、统一的流程管理等, 整合贯通了发行人各类核心算法、技术, 高度适用发行人训练数据开发的实际业务场景, 是技术与生产实践的高度结合。可支持大规模并发数据处理的能力、支持严格的数据安全和隐私保护 (如统一的加密数据存储、数据传输, 对所有数据操作人员试行统一、严格的账号和密码管理等), 可运用生产过程中的实时数据和行为日志等大数据对算法进行即时反馈和优化, 最大化算法对生产效率	已取得 2 项发明专利授权, 正在申请 2 项发明专利

			和数据质量的提升作用。	
训练数据生产	★多语种多模态训练数据设计技术	此技术点包括的行业通用技术一般包括：拼写检查、敏感词过滤、语种检测等基础的文本处理技术，和发音预测、音素均衡等基础的语音处理技术。	与通用技术相比，发行人核心技术的差异点主要体现在： 1、将拼写检查等十多种文本相关技术整合在自主研发的语料清洗工具中，并加入自主研发的易读性检测技术等重要模块，提高清洗后语料数据的质量；同时，发行人将前述语料清洗技术与发行人的语种/方言覆盖能力相结合，可支持约 70 种不同语种/方言的数据清洗。同时发行人将语音、视觉、文本等领域积累的丰富的训练数据集设计能力整合起来，融会贯通后形成了高质量多模态训练数据的设计能力。 2、训练数据集的设计是一个研究与工程相结合的综合问题，存在诸多实践所需的技术诀窍（know-how），发行人已在该领域积累的专利技术分别创新性地解决了：中英混合训练数据集的设计问题、最优句长分布与音素覆盖的问题、中文稀缺短句的设计问题、个性化语料设计以优化朗读效果的问题。	已取得 4 项发明专利授权
	多模态多通道数据采集技术	此技术点包括的行业通用技术一般包括：手机录音 APP 的开发、计算机连接麦克风进行录音等基本的语音采集技术，手机拍照及照片管理技术、专业摄像头拍摄采集技术、手写体与光学字符采集技术等图片视频类采集技术。	与行业通用技术相比，发行人核心技术的差异点主要体现在：有机整合了基础的单通道数据采集技术，解决了多设备数据冲突、多来源数据协调一致等关键技术难点并积累了 2 项专利技术，实现多通道（多录音设备、多视频设备等）同时采集的能力。	已取得 2 项发明专利授权及 1 项实用新型专利授权
	★数据同步技术	此技术点仅适用于特定应用场景，行业多以人工标注的方式处理，一般不认为有行业通用技术。	数据同步性是衡量训练数据集质量的重要指标。在多通道数据采集过程中，发行人自主研发了 1 项专利技术，解决语音数据采集过程中多通道语音数据的一致性问题。发行人的专利音频对齐技术，可将数据对齐误差控制在 1 毫秒以内，并支持每小时处理 100 小时以上的双通道音频文件对齐需求。	已取得 1 项发明专利授权
	实时采集质量控制技术	此技术点包括的行业通用技术一般都是语音信号处理相关的技术，包括：声音能量检测、音频截幅检测、静音检测、信噪比检测等技术。	与行业通用技术相比，发行人自主研发的实时采集质量控制技术综合运用了声音能量检测、音频截幅检测、静音检测、信噪比检测等技术，解决了数据采集的实时反馈难点，可支持多个采集工具几乎零延时地对数据瑕疵给出即时反馈，保证原料数据的质量。	已取得 1 项发明专利授权及 1 项实用新型专利授权
	★大数据驱动的高效数据处理技术	此技术点涵盖了训练数据加工环节的众多算法、技术和工具，和这些相关的行业通用技术一般包括：统计学习、语音信号处理、SVM 等基础技术，以及各语种拼写检查、词性预测、分词处理等常见应用技术。	发行人积累的大数据主要是指：大量自有知识产权的训练数据产品积累、大量生产过程中的行为日志数据积累等。发行人基于前述自有大数据积累，对生产经营中的生产模式、流程进行针对性地优化，技术的差异点主要在于运用前述大数据训练自有算法模型、积累自有技术，实现数据生产过程的优化和产品质量、生产效率的提升，具体如下： 1、以自有训练数据产品对初始模型进行训练，将前述算法模型运用于预标注过程中，结合后续人工标注过程的修改痕迹对前述模型持续迭代优化，逐步得到一系列实用的特定算法模型和特有工具。发行人自主开发的中文词性预测工具准确率经逐步优化已从 86%提升至 96%；音素边界预测工具简化了标注人员的人力成本；基于语音特征的韵律预测技术将人工标注工作量降低了 25%。 2、发行人对生产过程积累的日志数据和日常生产管理的反馈信息等大数据进行分析，从而辅助优化数据加工任务类型的划分、数据任务包的分配和标注人员的管理，提升生产效率。 3、发行人基于生产过程中的自有日志数据积累，分析一体化数据处理平台上的操作行为模式，辅助优化工具界面和交互，打造更优的工具集，进一步提升生产效率。	已取得 4 项发明专利授权
	★分布式高性能自动校验技术	此技术点相关的行业通用技术包括两类：一类是高性能分布式集群计算技术；一类是语音信号处理等基础技术。	与行业通用技术相比，发行人的核心技术差异点主要在于自主研发了多种关键质检技术，并将其运行于分布式高性能质检平台上，具体包括： 1、针对训练数据具体质检需求研发多种实用的算法技术，例如相似说话人筛查技术将人工处理工作量降低了 98%以上；音素标注正确率校验技术缩减了音素错误标注的检测时间，可满足音素标注 99%的准确率要求等。发行人在前述技术领域均已获得发明专利授权。 2、发行人在通用的分布式集群系统上针对数据质检的特定场景和需求进行了特有的整体架构优化和升级；并将前述核心算法的计算功能以及大量质检工具部署到该集群上，已达到日质检 660G 训练数据的能力。	已取得 4 项发明专利授权，正在申请 1 项发明专利

(2) 发行人在训练数据生产过程中是如何运用大数据和算法的，如何检查训练数据集对 AI 算法模型的训练效果，定量描述数据处理能力和数据处理效率；

发行人已在招股说明书“第六章 业务和技术/六、发行人的研发和技术/(五) 大数据和算法在训练数据生产过程中的运用情况说明”中对大数据、算法在训练数据生产中的运用情况进行了披露，具体如下：

一、大数据和算法在训练数据生产过程中的运用情况

(一) 大数据在训练生产中的主要运用情况

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，在训练数据的设计、采集和处理环节涉及大量数据，具体包括采集的原料数据、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。

发行人对前述大数据的运用主要体现在：积累了大数据驱动的高效数据处理技术，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，设计开发出高效率的数据处理平台与工具集，并充分运用上述技术、平台、工具集提高数据加工效率和训练数据质量。具体包括：

1、对原料数据及加工后数据的异同点进行分析，运用统计模型归纳出一般性规律，将之运用于开发和优化采集、标注过程所运用的辅助工具与算法，也运用于优化数据加工任务类型的划分和流程管控，从而提升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；

2、对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等大数据进行分析，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。

通过大数据积累和分析运用，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，并利用上述工具提升了数据加工过程的效率及准确率。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%¹⁷。

（二）人工智能算法在训练生产中的主要运用情况

发行人通过运用语音识别算法、语音合成算法、计算机视觉算法，对训练数据集的算法模型训练效果加以检查，从而反哺指导训练数据集的设计；同时将前述算法运用于训练数据生产环节，构建了训练数据设计、加工相关的核心技术。

对于各人工智能算法的运用情况具体说明如下：

1、语音识别算法：发行人将语音识别算法运用于（1）检验语音识别训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别算法模型进行训练，测试训练效果；（2）执行自动化数据加工、质检环节，程序化操作结合人工检查提升训练数据生产效率——通过语音识别算法对智能语音数据进行预标注，结合人工检查，降低人工工作量、提高数据处理能力；（3）直观展示语音识别训练数据的训练效果，为客户提供评判质量优劣的直观标准——运用不同训练数据集对同一语音识别算法模型进行训练，直观展现训练数据质量水平，方便客户选择更符合其预期的语音识别训练数据。

2、语音合成算法：发行人将语音合成算法运用于（1）检验语音合成训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别算法模型进行训练，测试训练效果；（2）运用语音合成算法模型展示训练数据的合成效果、帮助客户选择合意的训练数据——运用不同训练数据集对自有语音合成进行训练和效果测试，使得客户可以直观听取合成样音，进而选择风格、音色更符合要求的语音合成训练数据。

3、计算机视觉算法：发行人将计算机视觉算法运用于训练数据生产过程的降本增效之中，通过将计算机视觉算法技术嵌入至自主研发的分布式标注平台之中，全面提升了计算机视觉领域的原料数据加工效率。例如运用计算机视觉算法对采集的原料数据短视频进行批量识别检查，确认是否存在采集内容重复的情况；以此为基础再进行人工排查。

¹⁷ 上述数据来自发行人实验室模拟运算分析。

(三) 大数据和算法在训练数据生产过程情况的具体说明

类别	大数据、算法在训练数据生产各环节的具体应用情况				运用举例及对数据处理能力、数据处理效率提升的量化指标说明 ¹⁸
	设计	采集	加工	质检	
大数据	<p>大数据的运用主要表现在，发行人基于生产经营积累的大数据，总结出多项技术诀窍（know-how），助力发行人设计训练数据集结构、为生产优质的训练数据集奠定基础：</p> <p>1、发音预测技术：发行人基于大数据积累了核心技术“语音语言学基础研究”项下的词典与模型的发音预测技术，为后续实现智能语音训练数据集的音素均衡奠定基础。</p> <p>2、语料清洗相关技术：发行人基于大数据积累了拼写检查、敏感词过滤、语种预测等多项技术并持续优化，运用于设计环节中的语料清洗步骤。</p> <p>3、文本易读性评判相关技术：基于长期积累的录音采集过程中的行为日志，积累下文本易读性打分技术，可对设计的语料的易读性进行评判，提升语料易读性、降低发音人朗读错误概率，提升训练数据集质量。结合发行人的“个性语料获取方法及个性语料获取装置”这一专利技术，可围绕特定发音人的基础属性、行为数据，进一步提升语料适配发音人的能力、提升训练数据集录制质量。</p>	无	<p>大数据的运用主要体现在：发行人大数据积累，开发训练数据生产工具、优化生产过程，实现产品质量的提升和生产成本降低：</p> <p>1、数据加工工具开发：以自有训练数据产品积累对初始算法模型进行训练，将前述算法模型运用至预标注过程汇总，结合后续人工标注过程中的修改痕迹对算法模型进行迭代优化，逐步得到一整套完整的数据加工工具群。</p> <p>2、生产过程中的任务分配优化：基于生产过程中的日志数据，结合日常生产管理的反馈信息，辅助优化数据加工任务类型的划分、数据任务包的分配和标注人员的管理，提升生产效率。</p> <p>3、行为模式分析驱动工具、平台优化：基于生产过程中的日志数据，分析一体化平台上的操作行为模式，辅助优化工具界面和交互，打造更优的工具集，进一步提升生产效率。</p>	<p>大数据在质检环节的运用主要体现在“分布式高性能自动校验技术”的运用之中，包括以下方面：</p> <p>1、多质检点高效质检：发行人的质检平台可实现多个质检点的自动高效质检，前述功能的实现即来源于：发行人通过分析大数据中的内在规律，基于统计学习和分布预测技术，找到离群点，从而对数据瑕疵进行定位。</p> <p>2、标注人员反常行为数据检测：质检平台上的部分质检点会对数据加工过程中记录的标注人员行为日志大数据进行分析，从而发现“反常”的数据处理行为，以判断其关联的数据是否存在质量问题。</p>	<p>发行人通过大数据分析积累，开发了25个可运用于标注过程之中的预测算法或工具，提升了数据加工过程的效率及准确率。例如：将中文词性预测工具的准确率逐步从86%提升至96%；运用音素边界预测工具简化标注人员的人力成本；基于语音特征的韵律预测技术相比行业通用的基于文本的韵律预测技术降低了25%的人工标注工作量。</p>
语音识别算法	<p>训练效果模拟、指导设计： 检验语音识别训练数据的训练效果、反哺指导训练数据生产——用所生产的训练数据对自有的语音识别算法模型进行训练，测试训练效果。</p>	<p>实时识别、检测采集情况： 将语音识别算法运用于实时采集质检工具中，实时比对录制的语音数据与原始设计的文本语料是否相符。</p>	<p>执行预标注、提升标注效率： 执行自动化数据加工、质检环节，程序化操作结合人工检查提升训练数据生产效率——通过语音识别算法对智能语音数据进行预标注，结合人工检查，降低人工工作量、提高数据处理能力，例如：</p> <p>1、智能语音训练数据集的音字一致性检查：音字一致性即训练数据集中语音文件和文本文件的一致性，是衡量训练数据集质量的重要指标。因此需要在加工环节对音字一致性进行多轮校对。发行人将语音识别算法模</p>	<p>自动化质检运用： 将语音识别相关算法中的部分技术嵌入运用到质检技术中，以适配多个质检点的质检任务。例如：质检环节中运用了相似说话人技术，从而筛选数据集中同一个说话人可能录制的多份文本，该技术将语音识别算法的方法论运用其中。</p>	<p>在运用了实时语音识别算法的采集项目中，原始语音数据的准确性（语音文件与原始设计文本的相同率）相比不包含实时语音识别的项目的准确率提高了15.9%。</p>

¹⁸ 该部分量化指标均来自发行人测试结果。

			型引入上述校对环节，运用语音识别算法将语音文件转换为对应的文字，并与文本文件核对，提高了校对工作的准确率和效率。 2、嵌入运用至加工环节的几个核心技术中，进行场景适配和修改创新，如发行人的专利技术：基于语音特征的韵律预测技术等。		
语音合成算法	无	无	无	自动检测数据质量： 发行人发现语音合成数据库的音色一致性和基频稳定性对于最终算法训练效果的影响很大，故定向研发了相关质检工具，进一步提高了语音合成训练数据的质量	无
计算机视觉算法	无	无	计算机视觉算法在发行人加工环节的应用主要为： 1、部分标注工作的自动处理：运用于原料数据的脱敏处理之中：当所采集的图片或者视频数据的需求和特定人群无关时，发行人会利用计算机视觉算法自动对原始数据中采集到的人脸进行快速定位，并执行脱敏（图片模糊化）处理。 2、数据预测、辅助人工标注：发行人深度结合通用计算机视觉算法和具体数据加工场景，研发了数据预测工具，运用至加工环节中，对人工标注提供辅助，提升数据质量。	自动检测数据质量： 发行人运用人脸识别相关计算机视觉算法技术研发了短视频数据中相似人脸检查技术，可用于检查、保证计算机视觉训练数据集中人脸数据的独立性、提升训练数据质量。	运用于人脸数据的重复性检查，相比人工抽检的方式，运用了计算机视觉算法的处理效率提升约 96%。

二、大数据、算法对训练数据集训练效果检查方式说明

发行人运用 AI 算法检查训练数据集的训练效果的具体方式包括：

1、语音识别算法：基于开源或自有的语音识别算法，以发行人生产的语音识别训练数据集中 90%的数据作为训练数据，对该算法模型进行训练，再以同一训练数据集中剩余的 10%数据作为测试数据集，检验前述经过训练的算法模型对该 10%数据进行语音识别的正确率情况，从而对语音识别训练数据集的质量进行检查和验证。

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，通过算法模型训练结果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/词错误率均低于 5%，反映了发行人训练数据产品较好的训练效果。

产品编号	产品名称	训练数据产品中的音频数据总时长(小时)	测试集比例	字符错误率/词错误率
King-ASR-137	韩语语音识别库(手机)	342.00	10%	0.95%
King-ASR-124	俄语语音识别库(手机)	341.30	10%	0.94%
King-ASR-151	德语语音识别库(手机)	444.90	10%	1.04%
King-ASR-054	英式英语语音识别库(手机)	428.40	10%	2.83%
King-ASR-142	西班牙语西班牙语语音识别库(手机)	496.20	10%	1.83%
King-ASR-241	广州粤语语音识别库(手机)	3,220.80	10%	4.89%
King-ASR-329	加拿大法语语音识别库(手机)	405.00	10%	1.70%
King-ASR-072	法国法语语音识别库(手机)	496.50	10%	2.38%

2、语音合成算法：与语音识别算法类似，发行人运用语音合成算法模型，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。以发行人生产的语音合成训练数据集对语音合成算法模型进行训练，通过样音合成效果检验数据集对语音合成算法模型的训练效果。

三、发行人的数据处理能力、数据处理效率的定量描述情况

经过多年经营，发行人已形成了覆盖多应用领域、多应用场景、多语种的产品和服务体系，具备可覆盖智能语音、计算机视觉和自然语言 3 个应用领域，150 余个语种/方言（包括多个小语种）的数据处理能力。发行人数据处理能力、数据处理效率情况如下：

整体数据处理能力指标	定量描述	说明
应用领域覆盖能力	3种	智能语音领域、计算机视觉领域、自然语言领域
语言覆盖能力	150余个语种/方言	-
小语种覆盖能力	多种	南非荷兰语、拉脱维亚语、斯洛文尼亚语等

通过大数据、算法并用和驱动，发行人开发了多项与训练数据生产相关的核心技术，可运用于数据采集、标注过程中，提升了数据处理能力和效率。报告期内发行人数据处理能力、效率量化指标如下：

应用领域分类	数据处理能力/数据处理效率定量描述情况
语音识别	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
语音合成	报告期内平均每月采集137万字/词语音数据 每月最高可处理15万句语音数据
计算机视觉	每月最高可在图片数据中标注超过500万个目标对象
自然语言	该类业务占比较小、非标准化、类型多样 发行人每月标注自然语言数据可达1.8-2万工作小时

(3) 结合与主要竞争对手在经营情况、技术指标、产品侧重点、研发投入、市场占有率、客户结构和客户数量等方面的比较，说明发行人技术竞争优势和劣势；列表说明发行人设计的数据集结构较同行业竞争对手或下游客户自身的训练数据集结构之间的差异，如何体现发行人数据集结构设计的优势；

一、与主要竞争对手的比较；发行人技术竞争优势劣势情况

发行人已在招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（五）发行人与同行业可比公司在经营情况、市场地位、技术实力、衡量核心竞争力关键业务数据、指标等方面的比较情况/1、整体情况及 2、技术竞争优势”针对下述回复相关内容进行了补充披露：

发行人的同行业竞争对手主要包括 Appen、慧听科技以及标贝科技。结合前述竞争对手的公开披露信息，截至本问询函回复出具日，发行人与主要竞争对手在经营情况、技术指标、产品侧重点、研发投入、市场占有率、客户结构和客户数量等方面的比较情况如下：

项目	海天瑞声	Appen	慧听科技	标贝科技
经营情况				

成立年份	2005 年	1996 年	2011 年	2016 年
市场地位概述	我国领先的人工智能训练数据产品服务专业提供商，是我国最早从事训练数据产品服务研发销售的企业之一	较早从事数据资源开发的数据资源产品服务提供商，经营历史较长，规模、体量较大	-	-
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”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
员工数量	142 人 (截至 2020 年 3 月 31 日)	781 人	未公开披露	未公开披露
市场占有率	中国 AI 基础数据服务行业前五大企业之一，前五大企业市场份额合计为 26.2%，其中海天瑞声约占 8%。	无公开信息披露	无公开信息披露	无公开信息披露
客户结构及客户数量				
主要客户/合作伙伴情况	大型科技公司，如阿里巴巴、腾讯、百度、微软、三星等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学等	微软、谷歌等大型科技公司、汽车厂商及政府	未公开披露	大型科技公司，如阿里巴巴、腾讯、百度、网易、京东等；人工智能企业，如出门问问、旷视等；科研机构，如北京大学等。
客户数量	400 余家	未公开披露	未公开披露	100 余家
技术指标				
技术实力概述	海天瑞声自主开发了一体化人工智能数据处理技术支持平台，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、医疗健康和政府等各个领域。	采用全程质量监控流程，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 评测系统，为客户提供高质量的数据服务。
语种/方言覆盖能力	150 余个	180 余个	19 个	19 个

应用领域覆盖	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
已取得专利授权	22项(20项发明专利+1项实用新型专利+1项外观设计专利)	2项	无	1项
正在申请专利数量	6项	未公开披露	2项	10项
计算机软件著作权数量	117项	未公开披露	13项	23项
产品/服务及其侧重点				
出售的成品训练数据集数量	641个 (截至2020年3月31日)	300个	23个	101个
其他经营指标				
主要财务指标(2019年度/2019年末)				
营业收入	23,756万元	53,550万澳元	未公开披露	未公开披露
营业收入增长率	41.00%	79.00%	未公开披露	未公开披露
净利润	8,159万元	4,161万澳元	未公开披露	未公开披露

数据来源及说明:

1、Appen、慧听科技、标贝科技数据:截至2020年8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,均为截至本问询函回复出具日数据。考虑到可比性、可参考性,主要财务指标选取2019年度/2019年末指标进行比较。

结合上述比较情况,分析如下:

(一) 整体情况

与Appen相比,发行人规模较小,营收及净利润规模、员工数量等均低于Appen,在语种/方言覆盖能力方面也具有一定劣势;训练数据的应用领域覆盖能力相当,在成品训练数据集储备情况以及我国客户的覆盖程度方面具有一定优势——发行人在训练数据服务领域深耕多年,目前与人工智能产业链上的各类机构都建立了长期的战略合作伙伴关系,国内主要合作伙伴包括阿里巴巴、腾讯、百度、科大讯飞、商汤科技、云知声、海康威视、中国科学院、清华大学等知名机构。

与慧听科技、标贝科技相比，根据艾瑞咨询数据，发行人为中国 AI 基础数据服务行业前五大企业之一，市场份额领先于慧听科技、标贝科技。发行人在训练数据覆盖的语种/方言覆盖能力、成品训练数据集储备数量、计算机软件著作权数量方面均高于慧听科技、标贝科技（限于公开信息可查询范围），优势明显；在应用领域覆盖方面，除智能语音、计算机视觉、自然语言外，标贝可覆盖音乐类训练数据；发行人未单独列示音乐类训练数据，但其语音合成类业务可提供歌曲合成类服务，也已覆盖音乐类训练数据，发行人在应用领域覆盖方面不存在明显劣势。

在市场地位方面，Appen 是较早从事训练数据开发的训练数据提供商，经营历史较长，规模、体量等相比发行人均具备优势；而发行人是我国最早从事训练数据研发、生产、销售的企业之一，在我国市场具备领先地位。多年来，发行人深耕训练数据服务领域，伴随了众多国内客户在人工智能领域特别是智能语音领域的开拓、成长，为其持续提供了全球语种语音训练数据的高质量的本土服务，降低了对国外同类训练数据的依赖。

在主要财务指标方面，慧听科技、标贝科技未公开披露其财务数据信息。根据艾瑞咨询数据，发行人为中国 AI 基础数据服务行业前五大企业之一，市场份额领先于慧听科技、标贝科技，发行人收入规模高于慧听科技、标贝科技；发行人在营业收入、净利润规模及营业收入增长率方面相比 Appen 存在一定劣势，综合毛利率则高于 Appen。

在专利储备方面，通过公开信息渠道可获悉的 Appen、慧听科技、标贝科技的专利储备数量较少，发行人在专利技术储备方面具备明显优势。在计算机软件著作权方面，慧听科技及其子公司共拥有计算机软件著作权 13 项，标贝科技及其子公司共拥有计算机软件著作权 23 项；发行人及其子公司共拥有计算机软件著作权 117 项，远高于公开信息可查询的慧听科技、标贝科技的计算机软件著作权数量，具备一定优势。

在语种/方言覆盖能力方面，根据 Appen、慧听科技、标贝科技官方网站的信息，发行人的产品和服务可以覆盖超过 150 余个语种/方言，覆盖的语种/方言数量少于 Appen，但发行人在自有知识产权训练数据产品数量上具备一定优势。与慧听科技、标贝科技相比，发行人在产品和服务覆盖的语种/方言个数、自有知识产权训练数据产品数量方面均高于公开信息可查询的慧听科技、标贝科技的相关数量，具有明显的优势。

（二）技术竞争优劣势

1、竞争优势

结合上述比较可以看出，相比主要竞争对手，发行人的竞争优势主要体现在：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竞争对手。在专利储备方面具备显著优势，截至本问询函回复出具日，发行人已取得 22 项专利授权（包含 20 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。

2、竞争劣势

相比同行业竞争对手，发行人的技术竞争劣势主要体现在：与国内竞争对手相比，发行人无明显劣势；在语种/方言覆盖能力方面，发行人的覆盖能力略低于 Appen。

二、列表对比发行人设计的数据集结构较同行业竞争对手或下游客户自身的训练数据集结构之间的差异¹⁹，如何体现发行人数据集结构设计的优势

发行人已在招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/（五）发行人与同行业可比公司在经营情况、市场地位、技术实力、衡量核心竞争力的关键业务数据、指标等方面的比较情况/3、发行人设计的训练数据集结构与同行业竞争对手或下游客户自身数据集结构之间的差异情况，及发行人数据集结构设计的优势说明”针对下述回复相关内容进行了补充披露：

（一）与同行业竞争对手的对比情况及优势体现

项目	海天瑞声	Appen	慧听科技	标贝科技
1、智能语音类训练数据产品				
数量	492	122	19	86
覆盖语种/方言数量	85	85	19	8

¹⁹ 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，数据截至本问询函回复出具日；发行人的训练数据产品信息为截至 2020 年 3 月 31 日之情况。

录音场景类别	9	11	4	3
采集设备种类	8	2	5	3
采集通道覆盖数	6	5	7	1
2、自然语言类训练数据产品				
数量	83	170	2	12
覆盖语种/方言数量	36	100	2	4
3、计算机视觉训练数据产品				
数量	66	8	2	3

针对各应用领域的训练数据产品，对其结构具体差异进一步对比如下：

1、智能语音训练数据产品结构对比

对于智能语音类训练数据而言，其在内容、分布方面的结构属性体现在文本内容、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在训练数据结构方面比较情况如下：

(1) 通用语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
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	1~5	1~3	未公开披露
噪声环境类型	车内、办公室、家庭、医院、地铁、游戏场、餐	办公室、家庭、街边、车内、其他	办公室/宿舍/家、餐厅/咖啡厅、街道	未公开披露

	厅/咖啡厅、商场、街道			
单一训练数据产品覆盖的录音通道数量	1~6	1~5	1~7	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。

(2) 车载语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
文本内容类型总数	55类	4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	数字串、街道名称、控制词、长句	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	7	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。

(3) 特色语音识别训练数据产品

训练数据集结构属性	海天瑞声	Appen	慧听科技	标贝科技
稀有语种/方言覆盖数量	67	62	4	8

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在结构方面的差异主要体现在：覆盖的噪声类型、录音通道数量、录音文本内容类型等方面。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结构可覆盖更丰富的录音文本内容类型、噪声环境类型、录音通道数量、稀有语种/方言数量，具备广泛性、强适用性。

2、计算机视觉训练数据产品结构对比

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设计结构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，将其在结构方面的差异梳理对比如下：

训练数据产品结构属性	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）	包含300多名儿童不同角度的面部轮廓以及16种场景下的国外人体轮廓数据
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在结构方面的差异主要体现在：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。

3、自然语言训练数据产品结构对比

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，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。不同自然语音训练数据产品的结构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相关属性信息，因此无法直接对具体训练数据产品的细节结构信息进行比较。

4、发行人数据集结构设计的优势说明

结合上述对比情况，发行人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，发行人设计的训练数据产品的差异之处主要体现在：

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。

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略低于 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。

上述训练数据产品结构差异表明，与竞争对手相比发行人有能力设计覆盖多语种/方言、多场景、多领域的，采集方案更为复杂的训练数据产品，体现了发行人在训练数据产品设计领域的全面性、专业性优势。

（二）与下游客户自身的训练数据产品之间的差异及体现的数据集结构设计优势

发行人的下游客户主要为人工智能应用层、技术层的各类机构，下游客户自身的训练数据产品通常服务于与其算法模型或产品应用相关的特定语种/方言及应用场景，而发行人在多场景、多领域、多语种/方言数据资源的全面性、综合性设计能力方面更有优势。此外，下游客户也常提供其获取的各类原料数据给予发行人进行专业化处理，形成可供算法模型训练使用的数据资源，这也反映了发行人在数据处理方面的全面性、专业性优势。

(4) 发行人相关技术是否成熟或存在快速迭代的风险，下游市场空间是否接近饱和和情形；相关风险揭示是否充分；

一、公司相关技术是否成熟或存在快速迭代的风险

发行人已在招股说明书“第六节 业务和技术/六、发行人的研发和技术/(八) 公司技术是否成熟或存在快速迭代风险的情况说明”针对下述回复相关内容进行了补充披露：

发行人已将其核心技术体系运用于设计、采集、加工、质检等训练数据生产相关的各个环节，构成了发行人业务持续经营的技术基础，并在报告期内保障了发行人业务规模和经营业绩的成长性。总体而言发行人的核心技术体系整体已较为成熟。

为了保持并加强公司的技术实力、提升核心技术竞争力，发行人采取了一系列战略举措、构建起技术创新机制，促进公司技术的持续升级和创新。发行人积极通过文化建设营造鼓励创新的工作氛围，结合一系列机制保障开拓创新的工作环境，充分利用薪酬、奖励、表彰等多种方式激发员工的主观能动性，结合公司的主营业务，形成勇于开拓、积极创新的创新环境。报告期内，发行人每年平均完成 10 余次平台升级，涉及百余个功能新增拓展。

但由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，发行人进行训练数据生产所需的各项技术也面临着快速迭代更新的风险。

二、下游市场空间是否接近饱和

发行人已在招股说明书“第六节 业务和技术/二、发行人所处行业的基本情况和竞争状况/(三) 所属行业发展情况和未来发展趋势/4、发行人下游行业尚未接近饱和”针对下述回复相关内容进行了补充披露：

报告期内，发行人主要从事训练数据的研发设计、生产及销售业务。为下游人工智能产业链上的各类机构提供训练数据定制服务、训练数据产品和训练数据相关的应用服务，所在细分行业为人工智能基础数据服务行业，行业发展情况与下游人工智能技术与应用行业发展情况息息相关。结合下游人工智能整体行业发展情况、未来发展趋势及对训练数据的需求情况，发行人下游市场空间尚未接近饱和。具体说明如下：

（一）全球和中国的 AI 产业规模持续增长，下游 AI 行业发展方兴未艾

全球、我国 AI 产业均持续快速发展：

截至 2019 年年末，全球共有 5,386 家活跃 AI 企业，27,400 名高级研究人员，20 座重点发展城市。2019 年，全球人工智能行业共产生约 374 亿美元融资，其中，自动驾驶、药物医疗、人脸识别、视频内容和金融反欺诈是获得融资最多的领域。根据国际数据公司（IDC）的数据，2019 年，全球人工智能产业规模达到 375 亿美元，预计 2023 年全球人工智能市场规模将达到 979 亿美元，年复合增长率为 27.11%。

2019 年，中国人工智能行业的融资总额约为 166 亿美元，占全球人工智能行业融资总额的 44.39%，位居全球首位。根据国际数据公司（IDC）的数据，2019 年，中国人工智能产业规模将达到 44.80 亿美元，预计 2023 年中国人工智能市场规模将达 174.68 亿美元，年复合增长率达到 40.52%。

（二）训练数据是当前 AI 算法发展和演进的“燃料”，对训练数据的需求持续增长

1、有监督的深度学习算法是实现 AI 技术的主要路径

在 AI 产业链中，算法、算力和数据共同构成 AI 技术发展的三大核心要素，而机器学习，特别是有监督的深度学习是现阶段实现 AI 技术的主要路径：一方面，深度学习可以通过对低层特征的组合，形成更加抽象的高层属性类别，自动从信息中学习有效的特征并进行分类，而无需人为选取特征。凭借自动提取特征、神经网络结构、端到端学习等优势，深度学习在图像和语音领域学习效果最佳；另一方面，监督学习需要大量训练数据对学习结果进行反馈，能够有效降低算法的错误率。因此，兼具学习效果优异及错误率较低两大优势的有监督深度学习已成为主流的算法架构。

2、大量训练数据是有监督的深度学习算法实现的重要基础

AI 算法模型从技术理论到应用实践的落地过程依赖于大量的训练数据，2012-2016 年期间，人工智能行业不断优化算法增加深度神经网络层级，利用大量的数据集训练提高算法精准性，例如，ImageNet 数据集超过 1,400 万张训练图片和 1,000 余种分类便在其中起到重要作用。然而，从自然数据源简单收集取得的原料数据并不能直接用于有监督的深度学习算法训练，必须经过专业化的采集、加工，形成相应的工程化训练数据集后才能供深度学习算法等训练使用。目前，应用有监督学习的 AI 算法对于训

练数据的需求远大于现有的标注效率和投入预算，人工智能基础数据服务将持续释放其对于 AI 算法模型的基础支撑价值。

3、AI 产业对训练数据服务的需求持续产生

AI 产业对训练数据的需求主要来源于成熟 AI 算法模型的拓展性需求和新生 AI 算法模型的前瞻性需求。在成熟的拓展性需求方面，Mckinsey Global Institute 的研究报告表明：

深度学习模型对训练数据的数据量、多样性和更新速度方面提出较高要求。为充分发挥人工智能技术的潜能，深度学习模型需要海量且涵盖图像、视频及语音在内等多种类型的训练数据进行模型训练。此外，人工智能技术要求算法模型根据潜在的应用场景变化而持续更新，因此，人工智能算法模型所使用的训练数据亦需要定期更新。具体而言，约 1/3 的算法模型每月至少更新一次，约 1/4 的算法模型每日至少更新一次，人工智能算法模型持续更新的特点将进一步拓展各领域训练数据的需求空间。

而在新生的前瞻性需求方面，随着人工智能商业化进程的演进，新兴 AI 应用场景如物联网 AIoT、AI PaaS、产业互联网等将展现出巨大的发展潜力，并逐步促进 AI 技术和算法模型的优化和创新。因此，在创新应用场景和新型算法的带动下，具有前瞻性的训练数据产品和高定制化的训练数据服务需求将逐步成为主流。

(三) 小结

发行人服务的下游智能行业发展前景广阔，训练数据是人工智能算法发展演进的燃料，下游市场对训练数据的需求正在持续增长。发行人的下游市场空间未接近饱和。

三、相关风险揭示是否充分

发行人已在招股说明书“重大事项提示/三、特别风险提示及”第四节“风险因素”中充分揭示了核心技术存在快速迭代的风险，具体如下：

“发行人已将其核心技术体系运用于设计、采集、加工、质检等训练数据生产相关的各个环节，构成了发行人业务持续经营的技术基础，并在报告期内保障了发行人业务规模和经营业绩的成长性。

由于人工智能行业整体发展较为迅速，应用领域和应用场景不断拓展，下游需求随着行业发展不断发生变化，发行人进行数据开发所需的各项技术也面临着快速迭代更新

的风险。如果发行人无法保持对行业发展趋势的洞察能力以及对研发人才的有效组织和研发经费的经济投入，则可能导致发行人在行业技术快速迭代过程中无法持续保持技术先进性和技术优势，进而对发行人的客户拓展、产品创新和经营业绩造成不利影响。”

（5）发行人已构筑较为明显的核心技术先进性的依据

发行人已在招股说明书“第六节 业务和技术/六、发行人的研发和技术/（一）核心技术整体概述”针对下述回复相关内容进行了补充披露：

海天瑞声拥有对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 辅助标注及精细标注能力。前述核心科技创新实力，使发行人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。发行人的核心技术先进性主要体现在：

1、算法与数据处理技术并用：公司已掌握 12 项核心技术，通过算法与数据处理技术并用，可以满足 AI 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。截至本招股说明书签署日，公司已积累超过 600 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，融入项目流程管理、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 AI 训练生产，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。截至本招股说明书签署日，公司的产品/服务已覆盖 150 余个语种/方言，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，可构建高质量的智能语音训练数据。

截至目前，公司已经取得 **20 项发明专利**、**1 项实用新型专利授权**及 **1 项外观设计专利授权**，**117 项计算机软件著作权**，另有 **6 项专利**已提交申请。专利储备情况与同行业竞争对手相比具备突出优势。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 AI 算法模型的改善、优化效果。

发行人的核心技术先进性得到了业内广泛认可：自 2009 年至今，海天瑞声连续荣获“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。公司是“中国人工智能产业发展联盟”、“中国语音产业联盟”的理事单位，参与了行业的标准的起草工作，在训练数据集的架构设计、开发标准、质检评测等领域，以专业性和创新性获得了行业的普遍认可。2017年，发行人被中国信息协会、信息化观察网评选为“年度人工智能数据服务领域领军企业”；2018年，中国语音产业联盟向发行人颁发了“智能语音创新大赛行业先锋奖”，中国信息协会、信息化和软件服务网向发行人颁发了“2018年人工智能行业最佳产品”奖项；2019年，发行人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”，技术实力得到行业的高度认可。

四、关于公司治理与独立性

12. 关于关联方

招股说明书披露，报告期内，发行人注销或转让的关联方较多。

请发行人说明报告期内注销或转让的关联企业的主要财务指标，与发行人客户和供应商之间是否存在交易、资金往来或其他利益安排，是否存在代垫成本或费用的情形。

请保荐机构、发行人律师及申报会计师对上述事项进行核查，并发表明确意见。

答复：

发行人说明事项

(1) 报告期内注销或转让的关联企业的主要财务指标

一、报告期内发行人注销或转让的关联企业

发行人报告期内注销或转让的关联企业的有关情况如下：

序号	关联方名称	关联关系	备注	注销/转让原因
1	Speechocean Ltd.	发行人实际控制人贺琳持股 100% 并任董事	2017 年 10 月已注销	终止业务经营
2	北京中科海讯软件科技有限公司	发行人实际控制人贺琳及其配偶蔡惠智合计持有 50% 股权的北京中科海讯科技有限公司持股 100%	2018 年 11 月已注销	无实际业务经营
3	北京中科海讯数字设备有限公司		2018 年 11 月已注销	无实际业务经营
4	北京中科海讯声学科技有限	发行人实际控制人贺琳的配偶蔡惠智持股 24%，并任执行董事、经理	2019 年 1 月已注销	无实际业务经营

序号	关联方名称	关联关系	备注	注销/转让原因
	公司			
5	南京智亚源电力科技有限公司	发行人董事志鹏的配偶钟山曾持股0.98%并任董事	2018年12月已转让全部股权并离任	投资退出
6	重庆中康文化传播有限公司	由发行人董事志鹏的配偶钟山间接控制	2017年7月已注销	无实际业务经营
7	西藏微霜夜影视文化有限公司		2019年9月已注销	终止业务经营
8	天津卓腾文化传媒有限公司		2019年12月已注销	无实际业务经营
9	霍尔果斯耐飞影视有限公司		2020年1月已注销	无实际业务经营
10	西藏新七组影视文化有限公司		已于2020年3月决议解散,正在办理注销	终止业务经营
11	西藏雄关峻卡影视有限公司		已于2020年3月决议解散,正在办理注销	终止业务经营
12	飞米光年(北京)科技有限公司	发行人监事方彦彬持股60%,并任执行董事、经理	2019年11月已注销	无实际业务经营
13	杭州行止数据科技有限公司	飞米光年(北京)科技有限公司持股100%,发行人监事方彦彬任执行董事、总经理	2020年7月已注销	终止业务经营
14	上海兴丹投资管理中心(有限合伙)	发行人原持股5%以上的股东上海兴富为有限合伙人,持有50.07%的出资	2017年1月已注销	投资项目退出

二、报告期内注销或转让的关联企业的主要财务指标

1、Speechocean Ltd.

Speechocean Ltd.于2016年12月启动注销,2017年1月注销银行账户,并于2017年10月完成注销。截至2017年10月6日/2017年1月1日—2017年10月6日,其经审计总资产为5.16万美元,净资产为4.44万美元,营业收入为0,净利润为-0.08万美元。

2、北京中科海讯软件科技有限公司

北京中科海讯软件科技有限公司于2018年11月完成注销,其主要财务数据如下:

单位:万元

项目	2018年9月30日/2018年1-9月	2017年12月31日/2017年度
----	----------------------	--------------------

总资产	1.15	1.16
净资产	1.15	1.16
营业收入	0	0
净利润	-0.01	-0.01

3、北京中科海讯数字设备有限公司

北京中科海讯数字设备有限公司于 2018 年 11 月完成注销，其主要财务数据如下：

单位：万元

项目	2018 年 9 月 30 日/2018 年 1-9 月	2017 年 12 月 31 日/2017 年度
总资产	1.27	1.28
净资产	1.27	1.28
营业收入	0	0
净利润	-0.01	-0.01

4、北京中科海讯声学科技有限公司

北京中科海讯声学科技有限公司于 2019 年 1 月完成注销。根据其财务报表，该公司 2017 年度、2018 年度总资产、净资产、营业收入、净利润均为 0。

5、重庆中康文化传播有限公司

重庆中康文化传播有限公司于 2017 年 7 月完成注销。根据其股东北京联合网视文化传播有限公司确认，该公司财务文件保留不完整，根据报税单记载 2017 年 1 月至 2017 年 6 月未实际开展业务，无营业收入。

6、南京智亚源电力科技有限公司

2018 年 12 月，发行人董事志鹏配偶钟山将其持有的南京智亚源电力科技有限公司 0.98% 的股权对外转出并不再担任董事职务，截至 2018 年 12 月 31 日/2018 年度，其主要财务数据为：总资产 60.71 万元、净资产 64.62 万元、营业收入 119.28 万元、净利润 15.66 万元。

7、西藏微霜夜影视文化有限公司

西藏微霜夜影视文化有限公司于 2019 年 9 月完成注销，其主要财务数据如下：

单位：万元

项目	2019 年 9 月 30 日/2019 年 1-9 月	2018 年 12 月 31 日/2018 年度	2017 年 12 月 31 日/2017 年度

总资产	5.22	5.22	2.63
净资产	0	5.22	-5.12
营业收入	0	0	0
净利润	-5.87	-4.57	-5.12

8、天津卓腾文化传媒有限公司

天津卓腾文化传媒有限公司于 2019 年 10 月设立，于 2019 年 12 月完成注销。根据天津卓腾文化传媒有限公司注销前法定代表人确认，该公司自设立至注销无实际经营，无财务报表。

9、霍尔果斯耐飞影视有限公司

霍尔果斯耐飞影视有限公司于 2017 年 12 月设立，于 2020 年 1 月完成注销。根据霍尔果斯耐飞影视有限公司注销前法定代表人确认，该公司自设立至注销无实际经营，无财务报表。

10、西藏新七组影视文化有限公司

西藏新七组影视文化有限公司于 2020 年 3 月决议解散，正在办理注销，其主要财务数据如下：

单位：万元

项目	2019 年 12 月 31 日/ 2019 年度	2018 年 12 月 31 日/ 2018 年度	2017 年 12 月 31 日/ 2017 年度
总资产	3.45	5.20	2.58
净资产	3.45	5.20	-4.92
营业收入	0	0	0
净利润	-0.35	-4.54	-4.92

11、西藏雄关峻卡影视有限公司

西藏雄关峻卡影视有限公司于 2020 年 3 月决议解散，正在办理注销，其主要财务数据如下：

单位：万元

项目	2019 年 12 月 31 日/ 2019 年度	2018 年 12 月 31 日/ 2018 年度	2017 年 12 月 31 日/ 2017 年度
总资产	3.27	5.08	2.94
净资产	3.31	5.08	-5.44

营业收入	0	0	0
净利润	-0.35	-0.35	-5.44

12、飞米光年（北京）科技有限公司

飞米光年（北京）科技有限公司设立于 2019 年 4 月，于 2019 年 11 月完成注销。根据飞米光年（北京）科技有限公司注销前法定代表人确认，该公司自设立至注销未实际开展业务，无营业收入。

13、杭州行止数据科技有限公司

杭州行止数据科技有限公司设立于 2019 年 4 月，于 2020 年 7 月完成注销，其主要财务数据如下：

单位：万元

项目	2020 年 6 月 30 日/2020 年 1-6 月	2019 年 12 月 31 日/2019 年度
总资产	2.20	2.20
净资产	2.20	-47.85
营业收入	0	0
净利润	-26.77	-47.85

14、上海兴丹投资管理中心（有限合伙）

上海兴丹投资管理中心（有限合伙）于 2017 年 1 月完成注销，其自 2017 年 1 月 1 日至注销前没有财务数据。

（2）与发行人客户和供应商之间是否存在交易、资金往来或其他利益安排，是否存在代垫成本或费用的情形

根据上述 13 个注销关联方报告期内的主要财务数据，该等企业报告期内均未发生营业收入，财务报表上记载总资产、净资产较小。根据相关方确认，2017 年 1 月 1 日至注销前，该等企业与发行人主要客户和供应商之间不存在交易、资金往来或其他利益安排，不存在代垫成本或费用的情形。

上述 1 个已转让关联方南京智亚源电力科技有限公司主要从事电力设备研发，报告期内营业收入、总资产、净资产较小。根据该公司确认，2017 年 1 月 1 日至 2018 年 12 月发行人董事志鹏的配偶钟山转让股权时，该公司与发行人主要客户和供应商之间不存在交易、资金往来或其他利益安排，不存在代垫成本或费用的情形。

根据发行人主要客户和供应商的访谈，报告期内与上述转让、注销的关联方之间不存在交易、资金往来。

中介机构核查事项

(1) 核查程序

保荐机构、申报会计师、发行人律师主要履行了如下核查程序：

1、查阅发行人控股股东、实际控制人、董事、监事、高级管理人员出具的调查表，梳理其关联方情况；

2、查阅转让、注销的关联方的营业执照、公司章程、股权转让协议、注销文件等资料，在国家企业信用信息公示系统查询转让、注销的关联方的基本情况；

3、查阅转让、注销的关联方的财务报表或关于财务数据的说明等确认文件，取得相关主体的确认函；

4、与主要客户和供应商进行访谈。

(2) 核查结论

经核查，保荐机构、申报会计师、发行人律师认为：

报告期内注销或转让的关联企业 with 发行人主要客户和供应商之间不存在交易、资金往来或其他利益安排，不存在代垫成本或费用的情形。

五、关于财务会计信息与管理层分析

13. 关于营业收入

13.1 关于收入确认政策

根据申报文件，申报会计师将营业收入认定为关键审计事项，招股说明书未披露关键审计事项。

招股说明书披露，发行人采用完工百分比法确认训练数据定制服务的收入，完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。

请发行人补充披露：不同交付和验收方式下训练数据定制服务取得的验收确认凭据和收入确认时点，完工进度和收入金额的确定方法以及对应定制服务成本的结转方法。

请发行人说明：（1）结合与主要客户的合同约定、提供数据定制服务的实际过程，说明发行人是否满足采用完工百分比法确认收入的条件，是否符合企业会计准则的规定；（2）完工进度的确认方法是否合理、准确，对应定制服务成本的核算流程和结转方法，是否符合企业会计准则的相关规定；（3）结合相关销售合同的具体约定，说明新收入准则下将每单元定制化训练数据识别为单项履约义务且属于某一时点履行的履约义务的原因及依据，认定新收入准则对首次执行日前各年度财务报表无影响的依据是否充分；（4）同行业可比公司类似业务采用的收入确认政策，与发行人之间是否存在显著差异；（5）报告期各期末未完工项目已确认的完工进度、营业收入、营业成本和毛利率，期后是否存在重大调整的情形。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人补充披露事项

不同交付和验收方式下训练数据定制服务取得的验收确认凭据和收入确认时点，完工进度和收入金额的确定方法以及对应定制服务成本的结转方法

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/四、报告期内主要采用的会计政策和会计估计/（十七）收入”中进行了补充披露：

在训练数据定制服务业务类型下，发行人为客户提供训练数据集定制服务，最终形成的训练数据集成品的知识产权由客户享有。该类型业务包含三类不同的交付和验收方式：

方式一：项目完成后向客户一次交付，客户一次性验收；

方式二：项目执行过程中分批交付数据给客户，客户收到最后一批数据后就所有数据整体验收；

方式三：项目执行过程中分批交付数据给客户，客户收到分批交付的数据后分次进行验收。

发行人之所以存在上述三种交付和验收的方式，主要是由项目特点以及根据客户的具体需求决定的。对于中小型的合同或订单，客户通常会约定一次性交付完成的定制化数据，客户也会一次性进行验收。对于较大额的合同或订单，客户通常会要求分

批次提交定制化数据，一般客户在收到分批次提交的定制化数据后，会进行分批验收。但是也存在发行人分批提交定制化数据后，客户由于数据体量较大，质检验收耗费时间较长，会在收到全部数据后进行一次性验收的情况。

上述三种交付和验收方式下训练数据定制服务的验收确认凭据、收入确认时点、完工进度、收入金额的确定方法、对应成本的结转方法具体如下表所示：

类型	交付和验收方式	取得的验收确认凭据	收入确认时点	完工进度	收入金额的确定方法	对应定制服务成本的结转方法
方式一	一次交付，一次验收	项目全部完成后，客户对全部交付训练数据进行验收后出具的验收确认文件	客户对提交的训练数据验收确认的时点	100%	收入金额为按照合同或订单约定的服务交易总额	该项目累计发生的支出全部结转至当期营业成本
方式二	多次交付，一次验收					
方式三	多次交付，多次验收（分次验收）	项目执行过程中，客户每次对分批交付的训练数据进行验收后出具的验收文件	客户对分批交付的训练数据每一次验收确认的时点	按照已经提交并验收的服务量占服务总量的比例确定	按照合同或订单约定的服务交易总额乘以完工进度扣除以前会计期间累计已确认提供的服务收入金额，确认当期的服务收入。	按照提供服务预计总成本乘以完工进度扣除以前期间累计已确认服务成本后的金额，结转当期营业成本

发行人以上三种交付和验收方式对应的具体合同条款举例如下：

交付和验收方式	具体合同条款摘录	
	交付	验收
一次交付，一次验收	卖方应在本合同签订之日起【】个工作日内向买方一次性交付全部合同货物。	系统软件全部达到指定现场后【】个工作日内，买方将根据双方确认的验收测试标准进行验收，卖方应在指定日期自费派遣人员到买方现场参加验收。
	乙方应在合同签订后的【】个工作日内完成第一条所约定的语音数据的录制工作并以硬盘或者 FTP 下载形式将数据交付给甲方。	甲方自项目录音交付之日起【】个工作日内根据数据验收规则对本次服务的质量进行验收。
多次交付，一次验收	乙方应按照附件中工期及里程碑中约定的工期，对甲方提供数据服务；并按照附件中的时间点之前将可交付物交付给甲方。	甲方自收到乙方提交的语音库所有数据之日起【】个自然日内对数据的质量进行验收确认。
	乙方应按约定的开发成果交付日期（“交付日”）如期交付各阶段的开发成果。	甲方应自乙方所有数据交付之日起【】日内，在甲方所在地或甲方指定的其他地点，依照附件中约定的验收标准和验收方法对开发成果进行验收。

多次交付，多次验收 (分次验收)	交付和验收 A 语种文本转语音数据库（女声），预计完成时间【】； 交付和验收 B 语种文本转语音数据库（女声），预计完成时间【】；……； 交付和验收 C 语种文本转语音数据库（女声）的全部资料，预计完成时间【】	交付和验收 A 语种文本转语音数据库（女声），预计完成时间【】； 交付和验收 B 语种文本转语音数据库（女声），预计完成时间【】；……； 交付和验收 C 语种文本转语音数据库（女声）的全部资料，预计完成时间【】
	交付时间以附件《项目任务书》为准，实际交付周期按实际收到的批次数据量经双方邮件确认为准。	甲方应自收到数据之日起【】个工作日内对数据的质量进行确认和验收。

注：【】为合同条款中具体的日期信息。

发行人说明事项

(1) 结合与主要客户的合同约定、提供数据定制服务的实际过程，说明发行人是否满足采用完工百分比法确认收入的条件，是否符合企业会计准则的规定；

1、发行人训练数据定制服务的主要合同约定中对服务总量、服务金额、验收标准、付款条款等与收入确认相关的核心要素均进行了明确约定

根据发行人与主要客户签订的合同或者订单的约定，发行人向客户提供定制化的用于人工智能训练的训练数据。在合同或者订单中双方的主要约定内容包括：（1）提供的训练数据的数量及要求，数量一般会以小时及句或图片数等作为数据单位；（2）合同总价格或者单位数据单价与总价格；（3）项目工期与验收标准；（4）付款条款，一般的付款节点包括合同签订时点、交付 50%或其他比例数据量的数据、项目完成时点等多个节点；（5）交付物，例如智能语音业务交付物一般包括音频文件、标注文本文件和说明及技术文档。

2、发行人提供训练数据定制服务的实际过程中，可以以小时数、图片张数等单位衡量的每个单元训练数据作为交付和客户验收的最小单位，并以其作为计算完工百分比的依据

单元训练数据是指按不同数据形态呈现的每计量单位的训练数据，如 1 小时语音、1 张图片等。例如，对于 10 小时中文方言语音识别采集项目，发行人向客户交付的训练数据集包括 10 小时中文方言音频和其对应的标注文本，其中，每 1 小时中文方言音频及其标注文本构成一个单元训练数据。

发行人提供数据定制服务的实际生产过程主要包括四个环节：（1）设计：一般根据客户要求确定采集数据环节的采集内容和采集要求；（2）采集：获取语音等原料数

据；（3）加工：对语音等原料数据进行标注；（4）质检：数据质量的检测。由于单元训练数据之间不存在高度关联关系（例如发音人 A 的语音录制、标注、质检，不需要依赖发音人 B 的相关工作而进行，发音人 A 录制的前 50 句话也不需要依赖发音人 A 录制的后 50 句话的相关工作而进行），因此在设计阶段之后，采集、加工和质检的环节均可以以单元训练数据为基础进行，并以单元训练数据为基础进行数据交付。

发行人完成上述四个环节的数据生产流程后，将形成的训练数据进行交付和验收。例如智能语音业务的交付物一般包括音频文件、标注文本文件和说明及技术文档。其中音频文件和标注文本文件是核心交付物，也是客户用来进行人工智能模型训练的数据文件。音频文件和标注文本文件一般以小时及句或字词作为数量标准，每小时或每句、每字词的音频文件和其对应的标注文件构成单元训练数据。发行人可以以单元训练数据为基础对定制服务进行分拆交付、分拆验收，客户也可以分拆用于人工智能训练。说明及技术文档一般作为附加文件同时提供给客户，作为音频文件、标注文本文件等数据文件相关标准、规范、使用的说明，不直接应用于客户的人工智能模型训练。发行人完成合同约定的全部数据量的交付和验收后即该定制服务全部提供完毕。

3、发行人满足采用完工百分比法确认收入的条件，符合企业会计准则的规定

发行人对训练数据定制服务收入，从性质上属于提供劳务，采用完工百分比法确认收入，符合企业会计准则的要求，具体分析如下：

根据《企业会计准则第 14 号——收入》第十条：企业在资产负债表日提供劳务交易的结果能够可靠估计的，应当采用完工百分比法确认提供劳务收入。

根据《企业会计准则第 14 号——收入》第十一条：提供劳务交易的结果能够可靠估计，是指同时满足下列条件：收入的金额能够可靠地计量；相关的经济利益很可能流入企业；交易的完工进度能够可靠地确定；交易中已发生和将发生的成本能够可靠地计量。

结合上述发行人与主要客户的合同约定、提供数据定制服务的实际过程来分析训练数据定制服务是否满足上述准则规定的完工百分比法确认收入的条件：

需满足条件	具体分析	结论
收入的金额能够可靠地计量	根据准则要求，该条件是指提供劳务收入的总额能够合理地估计。发行人通过合同或者订单同客户约定提供的服务内容、服务总量以及交易总额，因此收入的金额能够根据合同约定可靠计量。	满足

相关的经济利益很可能流入企业	<p>根据准则要求，该条件是指提供劳务收入总额收回的可能性大于不能收回的可能性。同时准则进一步解释，“通常情况下，企业提供的劳务符合合同或协议要求，接受劳务方承诺付款，就表明提供劳务收入总额收回的可能性大于不能收回的可能性。”</p> <p>发行人根据合同或者订单约定来提供服务，并经客户验收后确认收入，因此可以合理估计交易相关的收入总额收回的可能性是大于不能收回的可能性。同时结合客户的信誉、以前的交易经验等因素分析，收回收入的可能性大于不能收回的可能性，因此，相关的经济利益很可能流入企业。</p>	满足
交易的完工进度能够可靠地确定	<p>根据准则要求，该条件是指交易的完工进度能够合理估计。</p> <p>发行人根据已验收的服务量占服务总量的比例来确定完工进度。由于客户的验收是实质性验收，在验收中如果发现数据存在问题，或者未达到客户的验收标准，发行人需要一直对交付数据进行改进及修正，直至满足客户的验收标准。因此在客户验收之前，发行人无法合理确认其向客户交付的训练数据是否符合合同或订单规定的条件和要求，因此在客户验收之前也无法合理可靠地估计交易的完工进度。</p> <p>已验收的服务量根据发行人获得的客户的验收文件来确定。客户的验收文件会明确说明已验收的数据量，例如小时数、图片张数等，而且发行人与客户签订的合同或订单中也明确规定了总数据量，因此发行人可以根据客户的验收文件合理确认已验收的服务量占服务总量的比例。因此交易的完工进度能够合理估计。</p>	满足
交易中已发生和将发生的成本能够可靠地计量	<p>根据准则要求，该条件是指交易中已经发生和将要发生的成本能够合理地估计。</p> <p>由于发行人实际提供定制数据服务的过程中，工作内容较为明确，因此可以合理预计项目需要投入的对外采购数据服务数量、人工工时和其他需要投入的资源等。另外项目需要采购的数据服务、人工投入、租赁费用等成本的价格均比较稳定，因此可以做到合理预计项目成本。同时，发行人已经建立了良好的内部项目成本核算制度和财务预算及报告制度（如工时系统及项目预算机制），在项目正式启动前，编制相应的项目支出预算。从历史上的项目预算执行情况来看，交付数据后被要求反复修改的情况较少，实际项目成本与预算成本之间一般不存在重大差异。因此交易中已发生和将发生的成本能够可靠的计量。</p>	满足

综上所述，结合发行人与主要客户的合同约定、提供数据定制服务的实际过程，发行人对于训练数据定制服务采用完工百分比法符合企业会计准则的规定。

（2）完工进度的确认方法是否合理、准确，对应定制服务成本的核算流程和结转方法，是否符合企业会计准则的相关规定；

1、完工进度的确认方法采用准则中规定的“已经提供的劳务占应提供劳务总量的比例”，符合发行人业务特点，合理、准确

根据《企业会计准则第 14 号——收入》第十二条，企业确定提供劳务交易的完工进度，可以选用下列方法：已完工作的测量；已经提供的劳务占应提供劳务总量的比例；已经发生的成本占估计总成本的比例。

训练数据定制服务的收入确认中，发行人采用上述准则中规定的“已经提供的劳务占应提供劳务总量的比例”来确定完工进度，具体来讲就是根据已验收的服务量占服务总量的比例来确定完工进度。由于客户的验收是实质性验收，在验收中如果发现数据存在问题，或者未达到客户的验收标准，发行人需要一直对交付数据进行改进及修正，直至满足客户的验收标准，因此在客户验收之前，发行人无法合理确认其向客户交付的训练数据是否符合合同或订单规定的条件和要求，因此在客户验收之前也无法合理可靠地估计交易的完工进度。发行人根据获得的客户的验收文件来确定验收的服务量及其占服务总量的比例。由于客户的验收文件会明确说明已验收的数据量，例如小时数，而且发行人与客户签订的合同或订单中也明确规定了总数据量，因此发行人可以根据客户的验收文件合理确认已验收的服务量占服务总量的比例。

综上所述，发行人完工进度的确认方法是合理和准确的，而且符合企业会计准则的相关规定。

2、对应定制服务成本的核算流程和结转方法，符合企业会计准则的相关规定

报告期内发行人的定制服务成本的核算流程和结转方法如下：

报告期内，发行人训练数据定制服务的各项支出，包括职工薪酬、数据服务费、折旧与摊销等，均按照所属项目在“制造费用”中进行归集与核算。在项目分批验收或一次全部验收并相应确认收入后，将该项目归集的项目成本中与确认的收入对应的部分，由制造费用结转至营业成本。

按项目归集成本：发行人严格按照项目为单位进行成本归集，在项目执行之前，每个项目需要编制项目预算，根据训练数据定制服务的具体内容，对项目所需人员及每人投入工时等直接人工内容，采集人数、采集单价、标注工时、标注单价等数据服务采购内容，折旧与摊销、租赁等项目所需场地、设备内容，差旅、交通、办公等其他项目相关支出内容制定详细预算。项目执行过程中，所有支出填写和报销均比照预算类型执行，对未及预算、超预算等情况设定完整的审批流程，并制定了严格的项目决算要求，在项目申请决算前，所有支出必须填写和报销完毕，经专人审核后项目进行项目决算。决算后，项目不允许发生额外支出。通过以上内部控制手段，保证每笔支出均可以按项目准确、完整归集。

成本结转：具体表现为两种情形，一种是在项目分批验收并确认收入的情况下，在每次验收并确认收入的同时，按照已经验收的服务量占服务总量的比例乘以预计项目总成本并扣除以前期间累计已结转成本后的金额来结转当期营业成本；另一种是在项目一次全部验收并确认收入的情况下，直接进行项目决算，按照实际发生和归集的项目总成本全部结转至营业成本。制造费用中未结转至营业成本的部分，在财务报表中列示和披露为“存货-在产品”。

根据《企业会计准则第 14 号——收入》第十三条，企业应当在资产负债表日按照提供劳务收入总额乘以完工进度扣除以前会计期间累计已确认提供劳务收入后的金额，确认当期提供劳务收入；同时，按照提供劳务估计总成本乘以完工进度扣除以前会计期间累计已确认劳务成本后的金额，结转当期劳务成本。

报告期内，发行人已经将定制服务项目的成本进行合理归集，并按照完工进度和预计项目总成本结转定制服务成本，因此相关会计处理符合企业会计准则的规定。

(3) 结合相关销售合同的具体约定，说明新收入准则下将每单元定制化训练数据识别为单项履约义务且属于某一时点履行的履约义务的原因及依据，认定新收入准则对首次执行日前各年度财务报表无影响的依据是否充分；

一、新收入准则下将每单元定制化训练数据识别为单项履约义务的原因及依据

根据《企业会计准则第 14 号——收入（修订）》（以下简称“新收入准则”）第九条，履约义务，是指合同中企业向客户转让可明确区分商品的承诺。履约义务既包括合同中明确的承诺，也包括由于企业已公开宣布的政策、特定声明或以往的习惯做法等导致合同订立时客户合理预期企业将履行的承诺。企业为履行合同而应开展的初始活动，通常不构成履约义务，除非该活动向客户转让了承诺的商品。

根据新收入准则第十条，企业向客户承诺的商品同时满足下列条件的，应当作为可明确区分商品：（一）客户能够从该商品本身或从该商品与其他易于获得资源一起使用中受益。（二）企业向客户转让该商品的承诺与合同中其他承诺可单独区分。下列情形通常表明企业向客户转让该商品的承诺与合同中其他承诺不可单独区分：1、企业需提供重大的服务以将该商品与合同中承诺的其他商品整合成合同约定的组合产出转让给客户。2、该商品将对合同中承诺的其他商品予以重大修改或定制。3、该商品与合同中承诺的其他商品具有高度关联性。

根据前述发行人与客户签订的合同或订单的分析,发行人向客户提供的定制化训练数据集可以拆分为若干个单元训练数据。发行人根据上述准则规定将每个单元训练数据识别为单项履约义务,具体理由如下:

1、客户可以将每一单元训练数据用于训练人工智能模型,因此能够从该单元训练数据的使用中受益;

2、发行人转让每一单元训练数据的承诺在合同中可以明确区分,主要是因为(1)在转让每一单元训练数据之前,发行人无须提供重大服务以对单元训练数据之间进行整合;(2)发行人转让的每一单元训练数据不会影响其他单元训练数据或其他单元训练数据做出重大修改;(3)不同的单元训练数据之间不存在高度关联性,发行人可以对不同的单元训练数据分批提交,客户可以分批验收,而且分批提交已验收完成的数据可以用于客户的模型训练。

综合上述分析,发行人根据新收入准则的规定,将每单元定制化训练数据识别为单项履约义务。

二、属于某一时点履行的履约义务的原因及依据

根据新收入准则第十一条,满足下列条件之一的,属于在某一时段内履行履约义务;否则,属于在某一时点履行履约义务:(一)客户在企业履约的同时即取得并消耗企业履约所带来的经济利益。(二)客户能够控制企业履约过程中在建的商品。(三)企业履约过程中所产出的商品具有不可替代用途,且该企业在整个合同期间内有权就累计至今已完成的履约部分收取款项。

根据前述发行人与客户签订的合同或订单的分析,发行人根据上述准则规定,认为对识别的单项履约义务,即单元定制化训练数据,不满足一段时间内履行履约义务的条件,属于某一时点履行的履约义务。具体理由分析如下:

1、发行人在履约的同时,客户无法取得并消耗企业履约所带来的经济利益。因为在发行人交付和验收训练数据之前,发行人履行合同义务所进行定制数据的生产活动,包括设计、采集、加工、质检等,客户并未同时取得也无法消耗发行人履约所带来的经济利益;

2、发行人生产数据过程处于发行人控制之下,且验收后交付物的所有权才归客户所有,因此客户无法控制发行人履约过程中产生的数据;

3、由于训练数据是定制的，发行人所生产的数据只能提供给相关的客户，满足具有不可替代用途，但是如果合同由于客户的原因被终止，在合同终止时点，发行人已生产但未交付完成验收的工作量，发行人无权主张补偿截止该时点为止的已发生成本以及合理利润，因此不满足“该企业在整个合同期间内有权就累计至今已完成的履约部分收取款项”的要求。

综合上述分析，发行人根据新收入准则的规定，认为识别的单项履约义务，即每单元定制化训练数据，不满足一段时间内履行履约义务的条件，属于某一时点履行的履约义务。

三、认定新收入准则对首次执行日前各年度财务报表无影响的依据是否充分

财政部于 2017 年颁布了《企业会计准则第 14 号——收入（修订）》（“新收入准则”），发行人自 2020 年 1 月 1 日起执行新收入准则，对会计政策相关内容进行调整。新收入准则取代了财政部于 2006 年颁布的《企业会计准则第 14 号——收入》及《企业会计准则第 15 号——建造合同》（统称“原收入准则”或“旧准则”）。

发行人根据业务条款和合同模式评估了新收入准则对各类业务的收入确认时点和确认金额的影响。具体分析如下：

业务类型	原收入准则	新收入准则	新旧准则下是否存在差异
训练数据定制服务	<p>训练数据定制服务收入主要是指公司根据客户需求为客户提供训练数据定制服务而收取的收入。公司与客户通过合同或订单约定提供的服务内容、服务总量及交易总额。</p> <p>公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。</p>	<p>每单位定制化训练数据被识别为单项履约义务，属于某一时点履行的履约义务，公司在将定制化训练数据向客户交付并经客户验收确认时确认收入。</p>	<p>新旧收入准则下的收入确认时点均为训练数据的验收时点，因此收入确认时点无差异。</p> <p>由于合同或者订单中规定了训练数据的单价和数量，或者规定了收入总额和数量，因此可以按照验收的训练数据数量和单价来确认新准则下的收入金额，该金额也同时是旧准则下按照完工进度确认的收入金额，因此新旧准则下收入确认金额也无差异。</p>
训练数据产品	<p>训练数据产品收入主要是指公司授权客户永久使用公司已经开发完成并拥有所有权的训练数据产品而向客户收取的收入。公司与客户签订合同约定授权客户使用的训练数据产品及其交易金额。</p>		

	公司按照合同约定向客户提交训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。	公司不从事对训练数据产品有重大影响的后续活动，因此训练数据产品属于某一时点履行的知识产权许可授权，公司在向客户交付合同中约定的训练数据产品并经客户验收确认后，按照合同约定的交易金额确认训练数据产品收入。	新旧收入准则下的收入确认时点均为训练数据产品（数据库）的交付和验收时点，金额均为合同约定的价格，因此收入确认时点和确认金额均无差异。
训练数据相关的应用服务	训练数据相关的应用服务收入主要是指公司根据客户需求为其提供应用软件的定制开发服务或者是应用软件的授权使用，以及配套硬件的销售（如有）等，而向客户收取的收入。公司与客户通过签订合同，约定向客户提供的应用软件的定制开发或授权使用，以及配套硬件销售（如有）等服务内容及交易金额。		
	公司按照合同约定向客户交付开发完成或者授权使用的应用程序和配套硬件（如有），并经客户验收确认后，按照合同约定的交易金额确认收入。	应用程序的定制开发或者授权使用、配套硬件的销售属于可明确区分的单项履约义务，并且属于在某一时点履行的履约义务；应用程序的定制开发或者授权使用、以及配套硬件的销售，均是在向客户交付并经客户验收确认时确认收入，并按照合同约定的单项履约义务分摊的交易价格确认收入金额。	新旧收入准则下的收入确认时点均为应用程序和配套硬件（如有）的交付和验收时点，金额均为合同约定的价格，因此收入确认时点和确认金额均无差异。

根据上述分析，发行人认为新收入准则对首次执行日前各年度财务报表均无影响的依据是充分合理的。

(4) 同行业可比公司类似业务采用的收入确认政策，与发行人之间是否存在显著差异；

代码	信息技术服务类上市公司	收入确认政策
300290.SZ	荣科科技	<p>本公司提供劳务主要包括定制化软件开发收入和技术服务收入，其收入的具体确认原则如下：</p> <p>(1) 定制化软件开发收入是指公司接受客户委托，针对客户提出的软件需求进行研究开发所获得的收入。定制化软件一般是公司基于自主研发的软件平台基础上，按照客户的特定需求，进行定制化开发而形成的应用程序。该类软件一般不具有通用性，该类业务实质上属于提供劳务。</p> <p>对于专业程度较高且能够取得客户确认的完工进度依据的定制化软件开发收入，提供定制软件劳务交易的结果在资产负债表日能够可靠估计的，采用完工百分比法确认提供劳务的收入，并按已经发生的成本占估计总成本的比例确定提供劳务交易的完工进度；提供定制软件劳务交易的结果在资产负债表日不能够可靠估计的，若已经发生的劳务成本预计能够得到补偿，按已经发生的劳务成本金额确认提供劳务收入，并按相同金额结转劳务成本；若已经发生的劳务成本预计不能够得到补偿，将已经发生的劳务成本计入当期损</p>

		<p>益，不确认劳务收入。</p> <p>对于其他的定制化软件开发收入，本公司在按照合同约定内容实施了开发工作，为购买方完成了安装、调试工作，取得购买方的验收，并同时满足上述提供劳务收入确认条件时确认定制化软件开发收入。</p>
300245.SZ	天玑科技	<p>数据中心 IT 基础设施服务收入确认的一般原则是：根据合同约定提供了相应服务，在相关的经济利益能够流入公司，收入的金额能够可靠地计量时，相关成本能够可靠计量时确认收入。软硬件销售及自有产品销售收入确认的一般原则是：在销售产品时的主要风险和报酬已转移给买方，不再保留与之相联系的管理权和控制权，在相关的经济利益能够流入公司，收入的金额能够可靠地计量时，相关成本能够可靠地计量时确认收入。根据合同约定条款的不同，分别按以下原则确认收入：</p> <p>(1) 合同明确约定服务期限的，在合同约定的服务期限内按时间进度确认收入。</p> <p>(2) 合同明确约定验收条款的，按约定条款验收并确认收入。</p> <p>(3) 合同明确约定开发项目进度的确认条款的，期末按照提供定制软件开发收入的合同金额乘以经客户确认的开发项目进度的量化百分比扣除以前会计期间累计已确认提供劳务收入后的金额，确认当期提供定制软件开发收入；同时，按照提供劳务估计总成本乘以经客户确认的开发项目进度的量化百分比扣除以前会计期间累计已确认劳务成本后的金额，结转当期劳务成本。</p>
300634.SZ	彩讯股份	<p>提供劳务收入确认的原则，本公司在资产负债表日提供劳务交易的结果能够可靠估计的，按照完工百分比法确认提供劳务收入。</p> <p>B、定制工程化软件</p> <p>合同中明确约定合同总金额中包含质保金的，在确认收入时，先扣除质保金部分，然后按如下方法确认收入：质保期届满后，再将该质保金确认收入。</p> <p>a.取得初验报告开始确认收入。</p> <p>定制工程化软件一般包括软件功能的开发、现场安装调试、旧系统数据迁移、系统培训、用户测试、初验（系统试运行）和终验等环节，其中初验和终验是客户对系统运行情况作出的评价（用户也可直接终验），也是公司与客户结算的主要依据和环节。在初验阶段完成时，围绕项目的大部分工作均已完成，之后只是在运行期间可能进行部分系统功能的修改，所涉工作内容很少、工作量也不大。因此，公司在取得初验报告后开始确认收入。</p> <p>b.初验确认收入的比例。</p> <p>由于公司定制工程化软件主要依靠项目实施人员及自有软件技术的整合来进行开发，因此，公司以投入工时为计算基础，以完工进度作为确认收入的依据。根据历年已完工典型合同统计，初验时项目投入工时约为 80%左右，因此，公司按照以下标准孰低的原则确定初验收入的确认比例：</p> <p>I、初验时合同约定的付款比例等于或高于 80%，则以 80%确认收入；</p> <p>II、初验时合同约定的付款比例低于 80%，则以合同约定的比例确认收入。</p> <p>公司每年年末会根据当年完工合同情况，对初验完工进度进行复核。</p> <p>c.终验时的收入确认：</p> <p>取得客户终验报告确认时，公司扣除终验前已确认的比例后，将</p>

		其余部分确认收入，如合同中明确约定合同总金额中包含质保金的，先扣除质保金部分，再按前述办法确认收入。
300468.SZ	四方精创	<p>(1) 软件开发收入的确认原则及方法</p> <p>软件开发收入是指接受客户委托，根据客户的需要，对应用软件技术进行研究开发，由此开发出来的软件为定制软件、不具有通用性。</p> <p>软件开发收入的确认原则及方法分为两大类：</p> <p>①根据合同按每月实际工作量（人/天数）收费，经客户确认实际工作量后确认软件开发收入。</p> <p>②公司软件开发业务按照合同约定提交软件开发成果并经客户验收后确认收入，该业务若存在分阶段多次验收情况，则按该业务完成各阶段工作并经客户验收时，分别确认相应阶段工作成果的收入。</p>
	发行人	公司采用完工百分比法确认训练数据定制服务的收入。完工进度按照已经提供的劳务占应提供劳务总量的比例来确定。在资产负债表日，公司统计已经按照合同或订单提供给客户并经客户验收和确认的服务量占服务总量的比例，计算和确定完工进度。

上述内容是对原收入准则下发行人与同行业公司的收入确认政策的对比情况。

通过对比发现，从事信息技术服务类业务的上市公司中，通常将定制化软件开发收入定义为提供劳务收入，在资产负债表日提供劳务交易的结果能够可靠估计的，按照完工百分比法确认提供劳务收入。完工百分比的比例通常采取工作量法、已发生成本占预计总成本的比例确定、已经提供的劳务占应提供劳务总量的比例确定。发行人根据自身业务情况确定的验收及收入确认方式符合行业惯例，与同行业公司之间不存在显著差异。

发行人与同行业公司自 2020 年 1 月 1 日开始实施新收入准则。其中彩讯股份和四方精创已经披露了上市公司 2020 年半年度报告，其他两家同行业公司尚未披露半年度报告。彩讯股份在半年度报告中未披露新收入准则下具体业务相关的收入确认的具体方法。四方精创在半年度报告中披露了新收入准则下的软件开发收入的确认原则与方法，其与原收入准则下的软件开发收入的确认原则与方法相同，对收入确认时点与确认金额均未产生影响。由于发行人实施新收入准则后收入确认时点和收入确认金额也未发生变化，因此与同行业公司不存在显著差异。

(5) 报告期各期末未完工项目已确认的完工进度、营业收入、营业成本和毛利率，期后是否存在重大调整的情形。

报告期各期末未完工项目已确认的营业收入、营业成本情况如下：

单位：万元

期间	累计确认收入	累计确认成本
2020年3月31日	1,179.80	412.14
2019年12月31日	1,170.79	418.01
2018年12月31日	883.54	400.59
2017年12月31日	73.44	33.15

报告期各期末未完工项目已确认的完工进度、毛利率以及期后（截至2020年6月30日）完工情况如下：

2020年3月31日：

单位：万元

2020年3月31日	完工百分比	毛利率	是否完工	完工毛利率	毛利率变动比例	差异大于>20%的原因分析
		A		B	C=(B-A)/A	
2020年一季度末未完工项目 1	91.03%	80%	否	—	—	—
2020年一季度末未完工项目 2	46.30%	57%	否	—	—	—
2020年一季度末未完工项目 3	83.98%	60%	否	—	—	—
2020年一季度末未完工项目 4	90.00%	50%	否	—	—	—
2020年一季度末未完工项目 5	58.09%	55%	否	—	—	—
2020年一季度末未完工项目 6	39.35%	78%	否	—	—	—
2020年一季度末未完工项目 7	41.61%	72%	否	—	—	—
2020年一季度末未完工项目 8	5.03%	61%	否	—	—	—
2020年一季度末未完工项目 9	90.00%	51%	是	53%	4%	—
2020年一季度末未完工项目 10	5.58%	34%	是	6%	-82%	标注雷达 3D 点云，标注属性较多（如地面信息、道线信息、周边各类物体信息等属性标注）、规则复杂，持续反复投入的规则沟通、培训等工时较多
2020年一季度末其他 24 个未完工项目	68.05%	44%	—	—	—	—
合计	71.19%	65%	—	—	—	—

注：由于春节假期及疫情影响，2020年一季度末的未完工项目较2019年末相比进度不大，两期累计确认收入及成本的差异主要是由于汇率波动导致。

2019年12月31日：

单位：万元

2019年12月31日	完工百分比	毛利率	是否完工	完工毛利率	毛利率变动比	差异大于>20%的原因分析
-------------	-------	-----	------	-------	--------	---------------

					例	
		A		B	C=(B-A)/A	
2019 年末未完工项目 1	91.03%	79%	否	—	—	—
2019 年末未完工项目 2	46.30%	57%	否	—	—	—
2019 年末未完工项目 3	83.98%	56%	否	—	—	—
2019 年末未完工项目 4	90.00%	50%	否	—	—	—
2019 年末未完工项目 5	58.09%	54%	否	—	—	—
2019 年末未完工项目 6	39.35%	78%	否	—	—	—
2019 年末未完工项目 7	41.61%	73%	否	—	—	—
2019 年末未完工项目 8	5.03%	61%	否	—	—	—
2019 年末未完工项目 9	90.00%	47%	是	53%	13%	—
2019 年末未完工项目 10	95.45%	45%	是	48%	7%	—
2019 年末其他 19 个未完工项目	64.63%	44%	—	—	—	—
合计	73.69%	64%	—	—	—	—

2018 年 12 月 31 日：

单位：万元

2018 年 12 月 31 日	完工百分比	毛利率	是否完工	完工毛利率	毛利率变动比例	差异大于>20%的原因分析
		A		B	C=(B-A)/A	
2018 年末未完工项目 1	10.00%	33%	是	47%	42%	优化渠道和项目流程，缩短了工期，减少了采集费用和项目租赁费等成本
2018 年末未完工项目 2	40.07%	46%	是	53%	15%	—

2018 年末未完工项目 3	35.57%	47%	是	56%	19%	——
2018 年末未完工项目 4	40.70%	66%	是	70%	6%	——
2018 年末未完工项目 5	40.71%	69%	是	71%	3%	——
2018 年末其他 7 个未完工项目	95.48%	52%	是	52%	0%	——
合计	37.19%	55%	——	58%	——	——

2017 年 12 月 31 日：

单位：万元

2017 年 12 月 31 日	完工百分比	毛利率	是否完工	完工毛利率	毛利率变动比例	差异大于>20%的原因分析
		A		B	C=(B-A)/A	
2017 年末未完工项目 1	87.36%	84%	是	82%	-2%	——
2017 年末未完工项目 2	86.94%	20%	是	34%	70%	由于发音人语速的原因，后期采集字数减少，相应的标注支出也有节省
2017 年末未完工项目 3	50.07%	55%	是	54%	-2%	——
2017 年末未完工项目 4	73.89%	23%	是	16%	-30%	期后客户数据量减少，但发生少量人工支出，因此毛利率降低
合计	76.60%	55%	——	61%	——	——

发行人报告期各期末未完工项目以语音合成项目为主，已确认的毛利率与发行人报告期各期语音合成业务毛利率较为接近。同时，报告期各期末未完工项目在该期末按完工百分比确认的毛利率与期后实际完工毛利率基本一致，不存在重大差异，且期后确认的完工收入合理，不存在冲回等异常情形。

此外，上述未完工项目期后实际结转成本与预算成本基本一致，不存在重大差异，报告期各期末前十大未完工项目的实际完工成本与预算成本的差异情况请参见本问询函回复问题 20“发行人说明事项/（2）报告期各期末主要未完工项目的客户名称、项目内容、成本结构、合同金额和预计总成本，分析各项目成本结构的差异原因，说明预计总成本的确认方法，与实际完工成本之间是否存在差异，是否存在应计提而未计提存货跌价准备的情形”。

综上，发行人在各报告期的期后不存在对项目已确认的收入和成本进行重大调整的情形。

申报会计师核查事项

（1）核查程序

申报会计师执行的主要程序如下：

- 1、了解并评价与收入确认相关的关键财务报告内部控制的设计和运行有效性；
- 2、选取样本，检查销售合同或订单，识别与风险和报酬转移或控制权转移相关的合同或订单条款，评价发行人的各类收入的收入确认政策是否符合企业会计准则的相关要求；对比公司收入确认原则和收入确认具体方法在原收入准则下和新收入准则下的差异，分析新收入准则实施后对公司的业务模式、合同条款、收入确认等方面的影响；
- 3、查阅同行业上市公司的收入确认政策，核查发行人收入确认政策是否符合行业惯例；
- 4、选取重要客户，实地走访或视频访谈，了解客户的客户的规模、业务经营等基本情况、其与发行人之间的业务模式、结算方式等；
- 5、在抽样的基础上，将各期记录的收入核对至相关的销售合同或订单、客户验收文件、发票等相关支持性依据，以评价相关收入是否按照发行人的收入确认政策予以确认；

6、从资产负债表日前后的收入交易中选取样本，检查与收入确认相关的销售合同或订单、客户验收文件、发票等相关支持性依据，评价接近资产负债表日前后记录的收入是否记录于恰当期间；检查接近资产负债表前后发生的项目实际成本，核对至相关支持性文件，包括经管理层审批的人员工时记录，费用报销单据、结算单等，以评价相关成本费用是否记录于恰当的会计期间；

7、对于报告期各期末未完工项目的完工进度和营业收入，检查相应的合同或订单，及验收文件，检查已验收的服务量和合同约定的服务总量及交易总额，复核完工百分比和营业收入计算的准确性；

8、对于报告期各期末未完工项目的营业成本，检查报告期经管理层审批的项目预算总成本计算表及预算的实际执行情况，检查期后实际完工结算成本与期末预算总成本是否存在重大差异，期后预算是否存在重大调整。

(2) 核查结论

经核查，申报会计师认为：

1、发行人对训练数据定制服务收入，采用完工百分比法确认收入，符合《企业会计准则 14 号——收入》的规定；

2、发行人完工进度的确认方法是合理和准确的，符合企业会计准则的相关规定；发行人已经将定制服务项目的成本进行合理归集，并按照完工进度和预计项目总成本结转定制服务成本，相关会计处理符合企业会计准则的规定；

3、采用新收入准则之后，将每单元定制化训练数据识别为单项履约义务且属于某一时点履行的履约义务的原因合理，依据充分，新收入准则对首次执行日前各年度财务报表无影响的依据充分；

4、公司根据自身业务情况确定的验收及收入确认方式符合行业惯例，与同行业公司之间不存在显著差异；

5、发行人在各报告期的期后不存在对项目已确认的收入和成本进行重大调整的情形。

13.2 关于营业收入波动

报告期内，训练数据定制服务的营业收入 6,297.23 万元、12,369.55 万元和 14,232.91 万元，训练数据产品的营业收入 5,489.31 万元、6,601.67 万元和 9,176.47 万元，逐年上升。

请发行人补充披露：（1）各类别服务和产品的销量、价格与结构变化的原因，进一步分析营业收入变动的原因；（2）各期合同数量、完成量和未完成量与营业收入的匹配关系；（3）营业收入的季节性分布，与同行业可比公司之间是否存在差异。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人补充披露事项

（1）各类别服务和产品的销量、价格与结构变化的原因，进一步分析营业收入变动的原因

发行人训练数据定制服务和训练数据产品的三大应用领域为智能语音、计算机视觉和自然语言。

智能语音领域的训练数据定制服务和训练数据产品的销量、价格与结构变化情况，发行人已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/4、主要产品和服务的销量和价格分析/（1）智能语音领域”中进行了分析和披露。

计算机视觉和自然语言领域的服务和产品涉及多种类型数量单位，例如点、线、框格、词、句等，因此发行人按照销售的单个数据库作为销量单位，在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/4、主要产品和服务的销量和价格分析/（2）计算机视觉和自然语言领域”中对计算机视觉和自然语言领域的训练数据定制服务和训练数据产品的销量、价格与结构变化情况进行了分析和披露。以下对计算机视觉和自然语言领域内部分主要细分类别的销量、价格与结构变化情况进行分析，进一步说明这两类应用领域的营业收入变动情况。

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/4、主要产品和服务的销量和价格分析”中进行了补充披露：

一、计算机视觉领域

计算机视觉领域的三类主要细分类别为手写体、光学字符识别、行车相关视频。

报告期内上述三类细分类别的销量和价格情况如下：

细分类别	2020年1-3月		2019年度		2018年度		2017年度	
	销量	平均售价(元)	销量	平均售价(元)	销量	平均售价(元)	销量	平均售价(元)
手写体	1,350.00	2,317.28	28,005.45	285.78	2,740.00	494.90	3,875.00	1,238.89
光学字符识别	1,595.27	1,104.02	83,532.02	54.42	47,419.86	65.13	13,238.04	48.40
行车相关视频	11,892.90	2.51	1,630,134.90	2.50	1,188,754.20	4.77	-	-

注：手写体、光学字符识别、行车相关视频的销量单位分别为千词组、百张、十个目标。

1、手写体

报告期内，公司手写体业务2019年销量显著增长，主要来源于某国际消费电子产品厂商的移动设备滑行输入业务。

报告期内，公司手写体业务平均售价在2017-2019年逐年下降，2020年1-3月较高，主要是由于该类业务的细分类别差异及语种差异所导致的。2017年，手写体业务语种涉及日语、英语、德语等，且为手写输入方式。2018年和2019年，该项业务主要涉及滑行输入数据，较传统手写输入方式的数据采集和标注难度更低，且主要语种开始从外语种转向简体中文，采集难度明显降低，因此价格也显著下降。2020年1-3月，该类业务为多地区繁体中文手写体，对被采集人年龄、性别、学历、设备使用经验等均有要求，难度较大，且为华语项目，发行人议价能力较强，因此单价较高。

2、光学字符识别

报告期内，随着光学字符识别的应用场景得到了较大规模的发展，公司在该类业务方面也实现了较快拓展，销量持续增长。

报告期内，公司光学字符识别业务在2017-2019年平均售价呈现小幅波动，2020年1-3月平均售价非常高，是由于季度期间较短，项目数量少，业务收入90%以上来源于某大型光学字符采集项目，该项目相较于一般光学字符识别项目环节更多，且客户指定特定采集设备拍摄指定领域的实况图片，采集难度大，后期需质检图片多项内容信息，质检流程复杂，因此单价非常高。

3、行车相关视频

行车相关视频业务是公司在 2018 年新拓展的业务类别。2019 年，公司继续拓展该类业务，销量有所增长，同时由于竞争较为激烈，为保持市场占有率，平均售价有所降低。2020 年 1-3 月，该项业务平均售价基本保持稳定。

二、自然语言领域

自然语言领域的一类主要细分类别为语义理解。报告期内上述三类细分类别的销量和价格情况如下：

细分类别	2020 年 1-3 月		2019 年度		2018 年度		2017 年度	
	销量 (小时)	平均售价 (元)	销量 (小时)	平均售价 (元)	销量 (小时)	平均售价 (元)	销量 (小时)	平均售 价(元)
语义理解	4,412.88	106.99	86,505.88	55.28	110,407.88	45.90	79,522.18	47.04

报告期内，公司语义理解业务销量在 2018 年较高，主要是由于个别重要客户根据其业务需要明显提升了采购规模所致；2019 年，虽然个别重要客户在 2019 年仍保持了较高的采购规模，但由于其他客户采购规模有所减少，因此 2019 年销量较 2018 年下降。

报告期内，公司语义理解业务平均售价在 2019 年上升的原因是当年新增了医疗专业领域标注、文本编写、粤语分词、词性标注等高难度项目，因此售价较高；2020 年 1-3 月，语义理解项目较少，受个别项目影响较大，由于存在英文标注项目客户要求英语母语人士进行标注，因此售价很高，显著提升了平均售价。

(2) 各期合同数量、完成量和未完成量与营业收入的匹配关系

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 / (一) 营业收入分析/6、合同及订单数量与营业收入的匹配关系”中进行了补充披露：

报告期内，发行人执行的合同数量、合同对应的具体订单数量及对应的营业收入情况如下：

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
合同数量 (个)	202	511	466	406
合同对应具体订单数量 (个)	544	1,552	1,378	926
其中：完成数量 (个)	160	1,341	1,111	709
未完成数量 (个)	384	211	267	217
营业收入 (万元)	3,126.09	23,755.81	19,265.77	11,907.09

报告期各期，发行人收入的变动趋势与完成的具体订单个数变动趋势基本相符，单个完成的具体订单贡献的收入约为 16-20 万元，较为稳定，合同完成数量与营业收入之间具备匹配关系。

(3) 营业收入的季节性分布，与同行业可比公司之间是否存在差异

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 /（一）营业收入分析/7、营业收入的季节性分布”中进行了补充披露：

报告期内，发行人营业收入的季节性分布如下：

单位：万元

项目	2020 年 1-3 月		2019 年度		2018 年度		2017 年度	
	金额	占比	金额	占比	金额	占比	金额	占比
第一季度	3,126.09	100%	3,124.09	13.15%	2,848.45	14.79%	1,224.00	10.28%
第二季度	-	-	6,977.98	29.37%	3,784.80	19.65%	4,146.00	34.82%
第三季度	-	-	3,473.16	14.62%	5,483.29	28.46%	2,759.39	23.17%
第四季度	-	-	10,180.58	42.86%	7,149.23	37.11%	3,777.72	31.73%
合计	3,126.09	100%	23,755.81	100%	19,265.77	100%	11,907.09	100%

同行业可比公司 2019 年各季度营业收入分布情况如下：

	荣科科技	彩讯股份	天玑科技	四方精创	恒生电子	用友网络	万兴科技	泛微网络
第一季度	19.00%	20.73%	18.31%	20.38%	15.29%	14.72%	24.43%	15.97%
第二季度	21.73%	23.06%	22.97%	19.97%	24.07%	24.22%	23.78%	23.29%
第三季度	22.91%	24.03%	19.58%	22.02%	19.97%	19.93%	25.02%	21.77%
第四季度	36.36%	32.18%	39.14%	37.63%	40.67%	41.14%	26.78%	38.96%

由上述表格可以看出，发行人作为软件和信息技术服务行业公司，与同行业可比公司在营业收入的季节性分布方面具有类似的特征，即根据下游客户的年度预算制定及结算需求的节奏，第一季度实现收入普遍较低，第四季度实现收入普遍较高，第二、三季度普遍居中。

申报会计师核查事项

(1) 核查程序

除上述 13.1 题有关收入的程序外，申报会计师执行的与营业收入波动有关的主要程序如下：

- 1、获取发行人报告期内收入明细表，汇总各类别服务和产品的销售收入、销量、单位价格等信息，按主要细分类别对销量、单位售价及其变化进行分析；
- 2、获取发行人报告期内合同信息表，汇总统计各期合同数量、已完成量和未完成量；并选取样本检查合同执行进度和收入确认情况；
- 3、查阅同行业可比公司招股说明书、定期报告，检查同行业可比公司营业收入的季节性分布情况。

(2) 核查结论

经核查，申报会计师认为：

- 1、发行人各类别服务和产品的销量、价格与结构变化具备合理原因；
- 2、发行人各期合同完成数量与营业收入之间具有匹配关系；
- 3、发行人报告期营业收入季节性波动与同行业可比公司之间不存在重大差异，且符合行业惯例。

14. 关于营业成本

报告期各期，公司主营业务成本金额分别为 3,303.47 万元、6,786.78 万元和 7,067.52 万元，主要为训练数据资源定制服务。

请发行人补充披露：（1）训练数据资源定制服务成本构成及变动原因；（2）结合项目复杂程度、采集标注量和单位价格变化等因素，分析各类别定制服务成本构成及变动原因。

请发行人说明：（1）定制服务成本和研发费用的核算流程和结转方法，相关费用和人工的归集和分配方式，是否符合企业会计准则的规定；（2）计入主营业务成本的人员部门、员工数量及其变动和人均薪酬，相关人员的具体职责以及在相关项目中的具体作用，人均薪酬与同行业可比公司的对比情况。

请保荐机构和申报会计师核查并发表明确意见。

答复：

发行人补充披露事项

(1) 训练数据资源定制服务成本构成及变动原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 / (二) 营业成本分析/4、按支出类型分类的主营业务成本分析”中进行了补充披露：

报告期内，公司按支出类型分类的主营业务成本情况如下：

单位：万元

项目	2020年1-3月		2019年度		2018年度		2017年度	
	金额	占比	金额	占比	金额	占比	金额	占比
数据服务费	432.73	80.60%	5,245.49	74.22%	5,399.51	79.56%	2,618.57	79.27%
职工薪酬	82.18	15.31%	1,185.95	16.78%	971.87	14.32%	537.23	16.26%
折旧摊销	4.57	0.85%	141.62	2.00%	58.14	0.86%	45.25	1.37%
差旅费	10.03	1.87%	41.74	0.59%	44.87	0.66%	38.15	1.15%
其他	7.36	1.37%	452.72	6.41%	312.39	4.60%	64.27	1.95%
合计	536.87	100.00%	7,067.52	100.00%	6,786.78	100.00%	3,303.47	100.00%

报告期内，公司按支出类型分类的训练数据定制服务成本情况如下：

单位：万元

项目	2020年1-3月		2019年度		2018年度		2017年度	
	金额	占比	金额	占比	金额	占比	金额	占比
数据服务费	432.73	80.60%	5,245.49	75.39%	5,395.72	80.25%	2,616.47	79.70%
职工薪酬	82.18	15.31%	1,161.02	16.69%	940.99	13.99%	520.55	15.86%
折旧摊销	4.57	0.85%	129.79	1.87%	47.27	0.70%	45.24	1.38%
差旅费	10.03	1.87%	41.74	0.60%	43.96	0.65%	36.40	1.11%
其他	7.36	1.37%	379.65	5.46%	296.09	4.40%	64.18	1.96%
合计	536.87	100.00%	6,957.69	100.00%	6,724.04	100.00%	3,282.86	100.00%

报告期内，公司主营业务成本绝大部分为训练数据定制服务成本，因此主营业务成本与训练数据定制服务成本的构成及变动情况一致。报告期内，公司主营业务成本主要由数据服务费和职工薪酬构成，另外存在少量折旧摊销、差旅费和其他类型的成本项目。其中，数据服务费是公司最主要的成本项目，是公司在提供训练数据服务过程中所需的、非核心技术环节的原料数据采集、标注服务，与公司主营业务收入变动总体保持一致，占报告期各年主营业务成本的比例分别为 79.27%、79.56%、74.22%和 80.60%，占比较为稳定。同时，职工薪酬占报告期各年主营业务成本的比例分别为 16.26%、14.32%、16.78%和 15.31%，总体波动较小。

(2) 结合项目复杂程度、采集标注量和单位价格变化等因素，分析各类别定制服务成本构成及变动原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 / (二) 营业成本分析/4、按支出类型分类的主营业务成本分析”中进行了补充披露：

报告期内，公司按支出类型分类的各应用领域训练数据定制服务成本情况如下：

单位：万元

应用领域	项目	2020年1-3月		2019年度		2018年度		2017年度	
		金额	占比	金额	占比	金额	占比	金额	占比
智能语音	数据服务费	153.77	88.46%	3,957.92	74.60%	4,286.41	80.54%	2,069.47	81.41%
	职工薪酬	17.46	10.04%	869.85	16.39%	674.46	12.67%	356.56	14.03%
	折旧摊销	0.46	0.27%	109.46	2.06%	41.39	0.78%	27.65	1.09%
	差旅费	0.34	0.19%	24.80	0.47%	32.97	0.62%	32.47	1.28%
	其他	1.81	1.04%	343.77	6.48%	287.10	5.39%	55.99	2.20%
	合计	173.84	100.00%	5,305.80	100.00%	5,322.33	100.00%	2,542.14	100.00%
计算机视觉	数据服务费	174.10	75.32%	919.30	78.62%	734.90	81.05%	330.88	80.39%
	职工薪酬	38.16	16.51%	179.13	15.32%	150.60	16.61%	64.04	15.56%
	折旧摊销	3.89	1.68%	19.69	1.68%	4.36	0.48%	13.58	3.30%
	差旅费	9.58	4.14%	15.74	1.35%	10.63	1.17%	1.58	0.38%
	其他	5.41	2.34%	35.50	3.04%	6.24	0.69%	1.50	0.37%
	合计	231.14	100.00%	1,169.35	100.00%	906.74	100.00%	411.59	100.00%
自然语言	数据服务费	104.86	79.50%	368.27	76.32%	374.41	75.64%	216.12	65.67%
	职工薪酬	26.57	20.14%	112.05	23.22%	115.93	23.42%	99.95	30.37%
	折旧摊销	0.21	0.16%	0.63	0.13%	1.52	0.31%	4.01	1.22%
	差旅费	0.12	0.09%	1.21	0.25%	0.36	0.07%	2.35	0.72%
	其他	0.14	0.11%	0.38	0.08%	2.75	0.55%	6.69	2.03%
	合计	131.90	100.00%	482.53	100.00%	494.96	100.00%	329.13	100.00%

由上表可以看出，公司各应用领域的训练数据定制服务成本构成与训练数据定制服务总体成本构成基本一致，主要为数据服务费和职工薪酬，其中各应用领域各期数据服务费占比均在65%以上，是最重要的构成部分。

一、智能语音领域

(一) 数据服务费变动分析

报告期内，发行人智能语音领域训练数据定制服务成本中的数据服务费具体情况如下：

成本项目	单位	2020年1-3月			2019年度		
		数量	单价(元)	金额(元)	数量	单价(元)	金额(元)
采集	人	1,582	83.08	131,434	21,680	208.17	4,513,130
	字/词	-	-	-	16,812,923	0.63	10,519,499
标注	小时	65,877	20.51	1,351,365	1,139,448	20.72	23,608,395
成本项目	单位	2018年度			2017年度		
		数量	单价(元)	金额(元)	数量	单价(元)	金额(元)
采集	人	43,895	174.99	7,681,443	28,458	137.54	3,914,176
	字/词	14,327,056	0.53	7,539,168	10,138,911	0.36	3,645,430
标注	小时	1,381,279	19.21	26,535,366	670,046	17.86	11,966,256

由于发行人训练数据定制服务“以销定采”的业务逻辑，上述采集数据量和标注数据量与销量相匹配，用于衡量采集数据量、标注有效工时相对水平的采产比和标产比合理，具体分析请参见本问询函回复问题10.1“发行人说明事项/（6）采集数据量、标注数据量和标产比相关的确定方法和具体依据，是否存在随意修改或调整的情形，同类型项目的采集数据量、标注数据量和标产比是否存在差异并分析具体原因/二、同类型项目的采集数据量、标注数据量和标产比的差异情况及原因分析”。

上述以“人”为单位的采集单价在2017-2019年逐年上升，主要是由于：（1）采集类型由相对简单的读文本向相对复杂的自由对话转变，采集难度提升；（2）语种向多元化方向拓展，中文逐步由普通话拓展到方言，外语种逐步由常见语种拓展到稀有语种，资源稀缺度提升。采集单价在2020年1-3月较低，主要是由于当期完成项目以唤醒词项目为主，其录音时长短且较为简单，因此单价较低。

上述以“字/词”为单位的采集单价在2017-2019年逐年上升，主要是由于该类采集为语音合成项目，随着大额外语种项目在2018年和2019年完成，且该类项目中客户要求的发音人专业性和特定性越来越高，使得采集单价上升。

上述标注单价在报告期内较为稳定，逐年略有变动主要是由于语种拓展、标注精细度提高、以及对标注人员能力要求的变化等因素所带来的标注难度的变化造成的。

（二）职工薪酬变动分析

发行人智能语音领域训练数据定制服务成本中的职工薪酬变动情况，与发行人营业成本中的职工薪酬变动情况基本一致，具体请参见本问题“发行人说明事项/（2）计入主营业务成本的人员部门、员工数量及其变动和人均薪酬，相关人员的具体职责以及在相关项目中的具体作用，人均薪酬与同行业可比公司的对比情况”。

（三）差旅费变动分析

报告期内，发行人智能语音领域训练数据定制服务成本中的差旅费与项目数量的对应关系如下：

项目	2020年1-3月	2019年度	2018年度	2017年度
差旅费（万元）	0.34	24.80	32.97	32.47
差旅项目数量（个）	1	49	35	33
单个差旅项目平均差旅费（万元）	0.34	0.51	0.94	0.98

由于发行人项目执行范围涉及全国和境外部分国家和地区，对于采集要求较为复杂或采集场景、地区等较新的项目，项目负责人需要出差到现场进行监督、指导和沟通。2017年和2018年单个差旅项目平均差旅费较为接近，2019年单个差旅项目平均差旅费较低的原因是大部分差旅项目规模较小，同时远程沟通方式有效性增强，使得差旅周期缩短，因此费用较低，2020年1-3月受疫情影响，差旅项目很少且差旅周期很短，费用较低。

（四）成本结构变动分析

对于智能语音领域，2017年和2018年成本构成无明显变动；2019年数据服务费占比下降而职工薪酬占比上升，主要原因是：语音识别类业务收入下降，语音合成类业务收入虽然上升但其毛利率较高，综合导致营业成本中的数据服务费有所下降；同时项目数量上升使得发行人人工投入增长，且发行人持续看好智能语音领域业务的发展前景，在行业竞争加剧、业务规模未显著增长的情况下进行了必要的项目人才储备，并提升人均薪酬增强竞争力，以充分把握未来的业务机会，导致人工投入持续增长。2020年1-3月数据服务费占比上升而职工薪酬占比下降，主要原因是当期完成验收的短周期训练数据定制服务项目成本的职工薪酬中不包括年终奖金，因此职工薪酬相对较低，拉升了数据服务费的占比。

二、计算机视觉和自然语言领域

报告期内，发行人计算机视觉和自然语言领域训练数据定制服务成本中的数据服务费与其收入规模基本保持了一致趋势，符合“以销定采”的业务逻辑，其中自然语言领域 2019 年由于稀有语种词典项目增多，毛利率提升，因此较 2018 年训练数据定制服务成本中的数据服务费基本持平而收入规模有所增长。

对于计算机视觉领域，2017-2019 年成本构成无明显变动；2020 年 1-3 月数据服务费占比下降而差旅费占比上升，主要原因是当期主要项目涉及境外采集，因此发行人差旅费用较高所致。

对于自然语言领域，报告期内数据服务费占比逐年上升而职工薪酬占比逐年下降，主要原因是该领域出于自身特点，项目规模普遍偏小，人工投入占比较高，在报告期内随着公司在该领域的深耕与拓展，单个项目规模呈现上升趋势，公司人工投入逐渐形成规模效应，使得职工薪酬占比下降。

发行人说明事项

(1) 定制服务成本和研发费用的核算流程和结转方法，相关费用和人工的归集和分配方式，是否符合企业会计准则的规定；

一、训练数据定制服务业务的成本核算流程和结转方法

报告期内，发行人训练数据定制服务的各项支出，包括职工薪酬、数据服务费、折旧与摊销等，均按照所属项目在“制造费用”中进行归集与核算。在项目分批验收或一次全部验收并相应确认收入后，将该项目归集的项目成本中与确认的收入对应的部分，由制造费用结转至营业成本。

按项目归集成本：发行人严格按照项目为单位进行成本归集，在项目执行之前，每个项目需要编制项目预算，根据训练数据定制服务的具体内容，对项目所需人员及每人投入工时等直接人工内容，采集人数、采集单价、标注工时、标注单价等数据服务采购内容，折旧与摊销、租赁等项目所需场地、设备内容，差旅、交通、办公等其他项目相关支出内容制定详细预算。项目执行过程中，所有支出填写和报销均比照预算类型执行，对预算变更、未及预算、超预算等情况设定完整的审批流程，并制定了严格的项目决算要求，在项目申请决算前，所有支出必须填写和报销完毕，经专人审核后项目进行项目决算。决算后，项目不允许发生额外支出。通过以上内部控制手段，保证每笔支出均可以按项目准确、完整归集。

成本结转：具体表现为两种情形，一种是在项目分批验收并确认收入的情况下，在每次验收并确认收入的同时，按照已经验收的服务量占服务总量的比例乘以预计项目总成本并扣除以前期间累计已结转成本后的金额来结转当期营业成本；另一种是在项目一次全部验收并确认收入的情况下，直接进行项目决算，按照实际发生和归集的项目总成本全部结转至营业成本。制造费用中未结转至营业成本的部分，在财务报表中列示和披露为“存货-在产品”。

报告期内，各相关费用和人工成本在各个项目上的归集和分配的方法如下：

（一）职工薪酬

职工薪酬按照工时积分制归集到各个项目。每月业务人员按其实际工作情况将工时积分填写到相应定制服务项目上，每人每月填写至所有项目（包括定制服务项目和训练数据产品开发项目等）的工时积分共计为8分/日*月工作日。各个项目应分摊的人工成本的计算方式为：各业务人员填写在该项目的积分*该人员所属级别的人工成本单价=该项目的人工支出。

（二）数据服务费

数据服务费指训练数据定制服务项目所需的原料数据采集与标注服务采购支出，业务人员按项目实际发生情况进行支出申请和报销。在进行支出申请和报销时，必须填报项目编号、涉及环节、采购量、采购单价、报销金额、供应商名称等关键信息，保证发生额直接归集到单个项目上，不得与其他项目存在交叉。因此数据服务费支出根据上述项目编号信息直接归集和分配至各个项目上。

（三）折旧与摊销

对于各项目使用的设备和软件的折旧和摊销费用，按各项目领用该设备、使用该软件时间的长短，由系统自动计算在各个项目上的分摊金额。计算方式为：项目折旧与摊销支出=项目领用天数/365天*设备全年折旧额或无形资产全年摊销额。

（四）差旅费等其他费用

差旅费等其他费用按项目实际发生情况报销，在报销时，必须选择项目编号及费用类别，费用发生额直接归集到单个项目上，不得与其他项目存在交叉。因此差旅费等其他费用支出也是根据项目编号信息直接归集和分配至各个项目上。

报告期内，发行人的定制服务发生的相关支出已经在相关项目中进行了适当的归集和分摊，采用了适当的方法进行成本结转，符合企业会计准则的规定。在 2017 年、2018 年和 2019 年的报告期内，发行人执行财政部于 2006 年颁布的《企业会计准则第 14 号——收入》，根据该准则的规定，发行人采用完工百分比法确认收入，成本结转也应该“按照提供劳务估计总成本乘以完工进度扣除以前会计期间累计已确认劳务成本后的金额，结转当期劳务成本”，发行人的上述成本结转方法按照已验收服务量占服务总量的比例来结转预计项目总成本，因此也反映了已验收的单位定制化训练数据的成本，与确认的收入金额相配比，发行人的上述成本结转方法符合该准则的规定。在 2020 年 1-3 月的报告期内，发行人执行财政部于 2017 年颁布的《企业会计准则第 14 号——收入（修订）》的规定，根据该准则的规定，发行人将训练数据定制服务的合同或订单拆分为若干个履约义务，其中每一个单位定制化训练数据构成一个履约义务，并在提交单位定制化训练数据并经客户验收的时点确认收入，并应同时结转确认与该完成的履约义务相关的成本支出，符合企业会计准则的规定。

二、研发费用的核算流程和结转方法

报告期内，研发费用由训练数据产品开发支出及基础研发支出两部分构成。训练数据产品开发支出为发行人开发训练数据产品发生的各项支出，包括职工薪酬、数据服务费、折旧与摊销等，各项支出均按照所属项目在“研发费用-数据库产品开发”中进行明细核算。基础研发支出为公司研究开发训练数据开发生产所需的算法、技术以及工具、平台等的各项支出，包括职工薪酬、语言研究支出、折旧与摊销等，各项支出在“研发费用-基础研发支出”中进行明细核算。当期实际发生的研发费用各项支出均已计入当年/期的利润表“研发费用”科目予以反映。

研发费用的相关费用和人工成本的归集和分配的方法如下：

（一）职工薪酬

参与训练数据产品开发的数据业务中心人员按照工时积分制将职工薪酬归集到各训练数据产品开发的项目中。每月业务人员按其实际工作情况将工时积分填写到相应训练数据产品开发项目上，每人每月填写至所有项目（包括定制服务项目和训练数据产品开发项目等）的工时积分共计为 8 分/日*月工作日。各个项目应分摊的人工成本的计算

方式为：各业务人员填写在该项目的积分*该人员所属级别的人工成本单价=该项目的人工支出。

研发中心的人员按照各自的基础研发项目安排，将其薪酬直接、全部归集至基础研发支出。

（二）数据服务费

数据服务费与训练数据产品开发相关的原料数据采集与标注服务的采购支出。业务人员按项目实际发生情况进行支出申请和报销。在进行支出申请和报销时，必须填报项目编号、涉及环节、采购量、采购单价、报销金额、供应商名称等关键信息，保证发生额直接归集到单个项目上，不得与其他项目存在交叉。

（三）语言研究支出

语言研究支出与基础研发相关，主要是根据研发项目的需求聘请外部语言学家协助进行语言学基础研究所发生的费用。其发生额直接归集至基础研发支出。

（四）折旧与摊销

对于训练数据产品开发项目中使用的设备和软件的折旧和摊销费用，按各项目领用该设备、使用该软件时间的长短，由系统自动计算在各个项目上的分摊金额。计算方式为：项目折旧与摊销支出=项目领用天数/365 天*设备全年折旧额或无形资产全年摊销额。对于基础研发活动中使用的设备和软件，将其折旧和摊销费用直接归集至基础研发支出。

（五）差旅费等其他费用

对于训练数据产品开发项目上发生差旅费等其他费用，按项目实际发生情况报销，在报销时，必须选择项目编号及费用类别，费用发生额直接归集到单个项目上，不得与其他项目存在交叉。对于基础研发活动中发生的差旅费等其他费用则直接归集至基础研发支出。

报告期内，发行人的研发费用在项目间进行了适当的归集和分摊，并且全部计入当期成本费用。根据《企业会计准则第 6 号——无形资产》的规定，企业内部研究开发项目研究阶段的支出，应当于发生时计入当期损益，企业内部研究开发项目开发阶段的支出，在满足条件的情况下才能确认为无形资产。基于谨慎性原则，发行人将报告期内的

研发支出全部费用化并计入当期损益，相关会计政策和软件行业的同行业公司会计政策一致。

三、针对职工薪酬、数据服务费、折旧与摊销等训练数据定制服务、训练数据库产品开发均会发生的支出类型，不存在成本和研发费用划分不当的情形

发行人的训练数据定制服务、训练数据库产品开发均为项目制管理，每个项目立项之初，即需在业务管理平台中选择项目类型，每个项目只能选择一个类型，选择训练数据定制服务的项目，支出归集入“制造费用”，并按前述原则结转计入“主营业务成本”；选择训练数据库产品开发的项目，支出归集入“研发费用”。因此不会出现训练数据定制服务和训练数据库产品开发共享一个项目的情形，相应的，两者的支出不会出现划分不当的情形。

(2) 计入主营业务成本的人员部门、员工数量及其变动和人均薪酬，相关人员的具体职责以及在相关项目中的具体作用，人均薪酬与同行业可比公司的对比情况。

报告期内，发行人主营业务成本中的薪酬情况如下：

类型	项目	单位	2020年1-3月	2019年	2018年	2017年
与主营业务成本相关	薪酬总额	万元	82.18	1,185.95	971.87	537.23
	员工数量	人	3.63	40.41	35.02	26.52
	人均薪酬	万元/人	22.64	29.35	27.75	20.26

注 1：计入主营业务成本中的员工数量按照训练数据定制服务项目在当期结转确认的成本金额中所对应的数据中心员工的项目工时占全年总工时的比例加总计算得出。

注 2：人均薪酬反映当期确认成本的项目对应的数据中心员工，在项目周期覆盖的年度内的年化薪酬按项目工时进行加权平均计算后的年化薪酬。

报告期内，发行人计入主营业务成本核算的人员部门为数据业务中心。

数据业务中心的主要职责为负责训练数据定制服务项目以及训练数据库产品开发项目的需求分析、方案确定、预算制定、项目实施与监控、项目验收及项目决算等项目全流程的管控工作。数据业务中心按照业务领域下设语音识别采集业务部、语音识别转写业务部、语音合成业务部、计算机视觉业务部及自然语言处理业务部。各业务部门具体负责其相关业务领域下的项目管控工作。

数据业务中心的员工由于同时负责训练数据库产品开发项目和训练数据定制服务项目，因此在计算报告期各年/期间员工数量时，按照训练数据定制服务项目在当期结转确认的成本金额中所对应的数据中心员工的项目工时占全年总工时的比例加总计算得

出,同时由于 2020 年 1-3 月验收决算的训练数据定制服务的项目较少,因此 2020 年 1-3 月计入主营业务成本的员工人数较低。

报告期内,发行人计入主营业务成本中的员工人均薪酬分别为 20.26 万元、27.75 万元、29.35 万元和 22.64 万元,除 2020 年 1-3 月外,员工数量、人均薪酬均呈现逐年增长的态势,这与公司收入和业务规模的增长态势较为吻合。

2020 年 1-3 月的计入主营业务成本的人均薪酬为 22.64 万元,与 2019 年度相比金额较低,主要是由于 2020 年验收的训练数据定制服务项目以语音转写等短期项目为主,人工成本主要发生在 2020 年年初,而 2020 年 1-3 月的职工薪酬中不含年终奖,因此导致 2020 年 1-3 月计入主营业务成本的人均薪酬较低。

报告期内,同行业可比公司年报未直接披露计入主营业务成本的职工薪酬和员工人数,因此计算同行业可比公司的整体的人均薪酬。将发行人计入主营业务成本的人均薪酬与其进行对比,结果如下:

单位:万元

公司名称	2019 年	2018 年	2017 年
荣科科技	22.26	19.04	18.68
彩讯股份	15.89	16.66	16.52
天玑科技	20.95	20.62	19.42
四方精创	18.70	18.25	15.31
恒生电子	28.78	26.82	24.56
用友网络	26.12	27.25	23.75
万兴科技	32.27	31.56	30.80
泛微网络	23.46	20.97	21.66
平均值	23.55	22.65	21.34
海天瑞声	29.35	27.75	20.26

注 1: 以上同行业可比公司年报未直接披露计入主营业务成本的职工薪酬和员工人数,因此按照年报中披露的职工薪酬总额除以披露的总员工人数计算同行业整体的人均薪酬。

注 2: 2020 年 1-3 月无公开可比公司薪酬数据。

如上表所示,报告期内发行人计入主营业务成本中的人均薪酬,与同行业可比公司整体的人均薪酬相比,2017 年基本持平,2018 年和 2019 年较高,主要是由于公司处于软件和信息技术服务业中成长性较高的人工智能板块,业务人员获得的平均薪酬逐年增长且高于所属行业整体平均薪酬水平,较为合理。

中介机构核查事项

(1) 核查程序

保荐机构、申报会计师执行的主要程序如下：

1、询问相关人员以了解公司的训练数据定制服务成本和研发费用相关的业务流程和会计核算方法；

2、了解和评价与训练数据定制服务成本和研发费用核算、归集、结转相关的关键财务报告内部控制的设计和运行有效性；

3、获取报告期内发行人制造费用和研发费用明细，检查计算准确性；并与明细账、总账及财务报表合计数核对一致；

4、复核训练数据定制服务成本倒轧表，了解成本构成，并分析其合理性；

5、对于训练数据定制服务成本和训练数据产品开发支出，获取报告期内业务管理平台导出的训练数据定制服务成本和训练数据产品开发的项目明细及其支出明细，检查项目的各项支出是否与对应项目相关；

6、对于基础研发支出，获取报告期内基础研发项目的支出明细台账，检查计入基础研发的各项支出是否与基础研发项目相关，同时检查相关支出归集为基础研发支出的支持性证据；

7、分析训练数据定制服务成本和研发费用的构成，检查是否存在异常或变动幅度较大的情况，分析其合理性；

8、对训练数据定制服务成本和研发费用的各项费用明细进行测试，包括抽样检查各项费用支出的支持性依据，并执行费用支出的截止性测试；

9、获取发行人员工花名册及公司组织架构图，了解计入主营业务成本的人员部门构成、人数、薪资构成等要素，与计入主营业务成本的人员进行核对；

10、对主营业务成本中的职工薪酬与员工人数进行对比分析，计算人均薪酬，并分析在报告期内的波动是否合理；

11、查阅同行业上市公司公开信息，将公司计入主营业务成本的人均薪酬与同行业上市公司人均薪酬相比，分析是否合理；

12、根据员工花名册、具体岗位职责、薪酬明细表等情况，对职工薪酬的计提和成本费用分配情况进行了检查；

13、选取工资支付明细与工资计提情况和支付的银行流水进行核对。

(2) 核查结论

经核查，保荐机构、申报会计师认为：

1、训练数据定制服务的成本构成及各类别成本构成变动原因合理；

2、训练数据定制服务成本和研发费用的核算流程和结转方法，相关费用和人工的归集和分配方式符合企业会计准则的规定；

3、计入主营业务成本的人员的人均薪酬逐年增长且高于所属行业整体平均薪酬水平，较为合理。

15. 关于毛利率

报告期各期，发行人综合毛利率为 72.26%、64.77%和 70.25%，毛率先减后增，主要是训练数据资源定制服务的毛利率波动所致。

请发行人补充披露：（1）语音识别采集毛利率下降的原因，语音合成毛利率显著高于其他两类的的原因；（2）计算机视觉毛利率逐年下降的原因；（3）自然语言毛利率波动较大的原因。

请发行人说明：结合主要项目的收入和成本构成，说明报告期各期主要训练数据资源定制服务前五大客户的毛利率情况及变动原因。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人补充披露事项

(1) 语音识别采集毛利率下降的原因，语音合成毛利率显著高于其他两类的的原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 /（三）毛利及毛利率分析/2、毛利率情况分析”中进行了补充披露：

语音识别采集的毛利率 2018 年和 2019 年较 2017 年有所下降，2020 年 1-3 月有所上升。2017 年毛利率较高的主要原因是由于行业下游各大公司在当年普遍加大投入，

开发符合自身定位的智能音箱产品，产生了较多基于特殊设备的数据需求；除智能音箱外，部分客户还针对其机器人、车机等专有设备进行语音交互技术的研发，需要发行人人为其提供专有设备数据。上述需求定制化程度较高，毛利率普遍高于一般水平，该等业务的毛利率水平如下：

特定业务	毛利率	收入占比
智能音箱定制设备	59.20%	25.90%
机器人定制设备	48.79%	16.57%
车机定制设备	53.57%	7.74%
扣除特殊项目后的语音识别采集业务	42.88%	49.78%

扣除该等特殊项目影响后，语音识别采集 2017 年毛利率为 42.88%，与 2018 年和 2019 年毛利率接近。2018 年开始随着下游语音识别产品及应用的广泛落地，及市场竞争者的增多，定制化程度高成为普遍需求，对公司毛利率不再有显著提升的效果。

2020 年 1-3 月，公司承接的语音识别采集项目主要为唤醒词类项目，该类项目无需后期处理或后期处理较简单，较为节省成本，因此毛利率偏高。

语音合成业务涉及到高素质发音人及音素、韵律、词性等多种标注类型，相对复杂度较高，所需技术处理环节较多，因此毛利率相对高于语音识别采集业务和语音识别转写业务。语音合成的毛利率在 2017-2019 年较为稳定，2020 年 1-3 月明显上升，一方面是由于 2020 年 1-3 月期间较短，且该类业务规模较小，容易受到个别特殊项目的影响，另一方面是因为公司在技术驱动下引入标注工具使得标注效率进一步提升，节省成本提高了毛利率水平。

(2) 计算机视觉毛利率逐年下降的原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（三）毛利及毛利率分析/2、毛利率情况分析”中进行了补充披露：

由于发行人在报告期内逐步调整计算机视觉类业务价格策略，适度降低毛利空间，以获取更多与客户合作的机会，从而在计算机视觉领域树立良好的口碑和品牌形象，为未来的深入拓展奠定基础，因此该类业务的毛利率在 2017-2019 年呈现逐年下降趋势。

计算机视觉领域细分类别报告期内的收入占比及毛利率情况如下：

细分类别	2020年1-3月		2019年度		2018年度		2017年度	
	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比	毛利率	收入占比
手写体	67.18%	54.79%	48.19%	38.51%	40.02%	8.13%	57.92%	58.38%
光学字符识别	60.16%	30.85%	53.02%	21.87%	57.39%	18.51%	42.73%	7.79%
行车相关视频	86.43%	0.52%	36.26%	19.57%	41.30%	33.97%	24.53%	0.49%

就细分业务类别具体而言，手写体业务在2018年和2019年主要为中英文滑行输入法项目，相较于手写数据采集，滑行输入相对简单，因此发行人降低了报价，使毛利率较2017年有所下降；行车相关视频业务在2019年部分客户的订单竞争激烈，发行人通过较低报价争取业务机会；光学字符识别在2017年收入贡献较小，2018年和2019年毛利率基本稳定。

2020年1-3月，发行人计算机视觉类训练数据定制服务主要集中在手写体及光学字符识别项目，其中手写体涉及多国家、多地区繁体中文手写体采集项目，对被采集人年龄、性别、学历、设备使用经验等均有要求，且为华语项目，发行人议价能力较强；光学字符识别涉及主要为实况图片采集项目，客户指定采集设备与图片类别，采集难度大；后期需质检图片多项内容信息，质检流程复杂。上述两类项目由于较为复杂、要求较高，且发行人在行业内具备一定的竞争优势，因此毛利率较高，提升了计算机视觉类训练数据定制服务的整体毛利率。

(3) 自然语言毛利率波动较大的原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 / (三) 毛利及毛利率分析/2、毛利率情况分析”中进行了补充披露：

由于自然语言类业务包含的细分项目类型较多，如语义理解、特定领域文本采集、发音词典、机器翻译、相关性标注等众多类型，同时自然语言类训练数据定制服务在报告期内整体业务规模较小，而项目个数则较多，因此毛利率受单个项目需求差异化特征影响，在报告期内呈现上下波动的情况。

自然语言类业务中，发音词典业务是毛利率较高的业务，该类发音词典项目主要解决发音与文本对应关系与规则的问题，需要较强的语音语言学研究基础，由于行业内能够处理该类语言数据的企业较少，而发行人经过多年积累，拥有多语种数据处理能力，较高的资源稀缺性使得该类项目毛利率较高，进而提升了自然语言类业务整体毛利率水平。2019年，随着外语种拓展的整体趋势，公司词典业务需求增加显著，且

拓展到亚洲及欧洲的小语种，相继完成了印尼语、塔加路语、僧伽罗语、阿塞拜疆语等亚洲小语种，及巴西葡萄牙语、阿根廷西班牙语等欧洲语种的词典数据开发和销售，该类语种稀缺性高，市场同类竞品较少，因此毛利率较高。

上述外语种及方言词典业务的毛利率水平如下：

特定业务	年度	毛利率	收入占比
外语种及方言词典业务	2019 年度	89.59%	34.03%
	2018 年度	65.12%	3.07%
	2017 年度	58.48%	14.82%
扣除特殊项目后的自然语言类业务	2019 年度	44.62%	65.97%
	2018 年度	38.15%	96.93%
	2017 年度	40.70%	85.18%

由上表可以看出，特殊词典业务毛利率随着项目具体执行情况具有较大的波动性。扣除该等特殊项目影响后，自然语言类业务毛利率水平在 2017-2019 年波动较小。

2020 年 1-3 月，公司承接了部分复杂语种的特定领域文本采集项目，该类语种稀缺性高，发行人业务具有较强的市场竞争力，因此毛利率较高。

发行人说明事项

结合主要项目的收入和成本构成，说明报告期各期主要训练数据资源定制服务前五大客户的毛利率情况及变动原因

报告期内，发行人训练数据资源定制服务前五大客户的毛利率情况如下：

客户	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
前五大客户合计	57.53%	56.45%	48.21%	49.21%

发行人训练数据资源定制服务前五大客户的毛利率在 2017 年和 2018 年较为接近，2019 年和 2020 年 1-3 月明显上升，主要有以下两方面原因：

第一，前五大客户中某客户 2019 年收入在前五大客户收入占比中明显提升，由于该客户主要涉及语音合成业务，毛利率较高，因此提升了前五大客户整体毛利率；

第二，前五大客户中某客户 2020 年 1-3 月收入在前五大客户收入占比中明显提升，由于该客户主要涉及毛利率较高的计算机视觉领域手写体项目，因此提升了前五大客户整体毛利率。

申报会计师核查事项

(1) 核查程序

除上述第 13 题有关收入以及第 14 题有关成本的核查程序以外，申报会计师对发行人毛利率进行核查，执行的主要程序如下：

1、获取发行人报告期内收入成本明细表，对各主要类别产品和服务的销售收入、销售成本、产品结构、毛利率等明细及其变化进行分析；查阅并取得发行人与主要客户签订的合同，分析发行人训练数据定制服务前五大客户的销售收入、毛利率及变动趋势；

2、访谈发行人销售部门负责人，了解发行人主要客户需求、行业竞争因素、销售价格调整情况等，将了解的情况与收入成本明细表及分析情况进行比对；

3、对报告期内的主要客户进行实地走访或视频访谈，了解其和发行人之间的合作历史、合作模式、业务往来的发展情况。

(2) 核查结论

经核查，申报会计师认为：

1、发行人语音识别采集毛利率下降、语音合成毛利率显著高于其他两类、计算机视觉毛利率逐年下降、自然语言毛利率波动较大，具备合理原因；

2、报告期内，发行人主要训练数据定制服务前五大客户的毛利率变动存在合理原因。

16. 关于期间费用

16.1 关于销售费用

报告期各期，公司销售费用分别为 741.50 万元、874.48 万元和 925.28 万元，销售费用率分别为 6.23%、4.54%和 3.89%。

请发行人补充披露：结合销售人员的主要职责，分析定制服务项目数量和训练数据产品销售数量与人员数量的匹配关系。

请发行人说明：报告期各期计入销售费用的人员部门构成、员工人数、人均薪酬，相关人员的绩效考核指标以及完成情况，人均薪酬是否与同行业可比公司存在显著差异，是否与营业收入增长匹配。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人补充披露事项

结合销售人员的主要职责，分析定制服务项目数量和训练数据产品销售数量与人员数量的匹配关系。

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 /（四）期间费用分析/1、销售费用分析”中进行了补充披露：

基于发行人所属行业专业化强、高速发展、客户定制服务需求种类日益丰富等特征，发行人对销售人员素质和能力要求较高，其主要职责为结合行业发展趋势、客户需求及公司特点，为客户提供高附加值的顾问式销售服务，提升公司在 AI 基础数据服务领域的口碑。具体包括：熟悉市场趋势、客户需求、发行人的产品与服务的技术特点等，通过行业会议、展会宣传、网络信息宣传等多种途径与客户初步建立合作关系；基于已有经验帮助客户细化需求，并将需求及时、准确的反馈给公司管理和技术人员，协助制定基本解决方案；对整体合同的工期、内容和报价等与客户进行协商、完成合作协议的签署；持续跟进和维护客户长期合作关系等。

报告期内，训练数据定制服务、训练数据产品等相关的合同数量与销售人员数量的情况如下表所示：

各期签订合同数量	2020年1-3月	2019年	2018年	2017年
合同数量	119	390	341	303
销售人员数量	14.63	12.18	11.25	8.42
销售人员人均合同数量	8	32	30	36

注：销售人员数量是按照当年/期营销中心各月份员工的在职情况计算的平均值。

报告期内，发行人的合同数量呈逐年上升的趋势，同时发行人的销售人员平均人数也逐年上涨，趋势相匹配，销售人员年人均合同数量基本维持在 30-36 个之间。

发行人说明事项

报告期各期计入销售费用的人员部门构成、员工人数、人均薪酬，相关人员的绩效考核指标以及完成情况，人均薪酬是否与同行业可比公司存在显著差异，是否与营业收入增长匹配。

一、报告期各期计入销售费用的人员部门构成、员工人数、人均薪酬

报告期内，发行人将营销中心部门员工的薪酬归集到销售费用的职工薪酬中。销售费用中的职工薪酬的构成明细如下：

单位：万元

项目	2020年1-3月		2019年度		2018年度		2017年度	
	金额	占比	金额	占比	金额	占比	金额	占比
工资	88.67	60.81%	293.28	35.30%	279.74	36.46%	184.60	27.65%
奖金	24.66	16.91%	403.93	48.62%	375.97	49.00%	415.10	62.17%
社保费用	19.52	13.39%	88.27	10.63%	79.32	10.34%	64.19	9.61%
小计	132.85	91.11%	785.48	94.55%	735.03	95.80%	663.89	99.43%
劳务费	12.96	8.89%	45.25	5.45%	32.20	4.20%	3.82	0.57%
合计	145.81	100.00%	830.73	100.00%	767.23	100.00%	667.71	100.00%

报告期内，销售费用中的职工薪酬（不含劳务费）、销售人员的员工人数及人均薪酬情况如下：

类型	项目	单位	2020年1-3月	2019年	2018年	2017年
与销售费用相关	职工薪酬（不含劳务费）	万元	132.85	785.48	735.03	663.89
	员工人数	人	14.63	12.18	11.25	8.42
	人均薪酬	万元/人	9.08	64.49	65.34	78.85
	剔除奖金后人均薪酬	万元/人	7.39	31.33	31.92	29.55

注 1：职工薪酬为销售费用中的职工薪酬剔除劳务费后的金额；

注 2：员工人数为按照当年营销中心各月份员工的在职情况计算的平均值。

报告期内，发行人销售人员平均人数分别为 8.42 人、11.25 人、12.18 人及 14.63 人，逐年递增，但总体而言人数较少。基于发行人的顾问式营销模式，比较依赖于营销人员的素质和能力，而不是数量，因此公司营销团队人员数量相对较少，并匹配了较高的人均薪酬水平。

销售人员薪酬主要有基本工资和销售奖金组成，剔除销售奖金后，2017 年至 2019 年销售人员人均薪酬分别为 29.55 万元、31.92 万元、31.33 万元，总体呈稳定增长趋势。

总人均薪酬受到奖金变动影响呈现一定的波动，报告期内，销售人员的奖金发放情况主要是由其绩效考核指标的完成情况决定的。具体分析详见下述“二、相关人员的绩效考核指标以及完成情况”。

二、相关人员的绩效考核指标以及完成情况

发行人对销售人员的绩效考核指标主要为营业收入及营业毛利两个指标，并且按照指标完成情况发放销售奖金。销售人员的奖金主要由基础业绩奖金和超额业绩奖金构成。基础业绩奖金主要根据业绩考核指标的完成情况确定和发放，同时 2019 年起增加了客户及领域拓展力度的考核指标（包括新客户数量及收入贡献率、新领域数量及收入贡献率等）；超额业绩奖金则根据实际业绩与预算的完成程度进行评估，如发生显著超额完成预算的情况，参考超额比例发放超额业绩奖金。

报告期内销售人员的绩效考核指标的完成情况如下：

2017 年至 2019 年，发行人的营业收入和营业毛利两项绩效考核指标均逐年增长，因此销售人员的基础业绩奖金也随之增长。另一方面，2019 年新增的客户及领域拓展力度的考核指标由于为新的执行要求，实现程度有限，因此拉低了 2019 年基础奖金的增幅。

超额业绩考核方面，在 2017 年由于通过有效的市场推广，发行人的训练数据产品销售实际业绩大幅超出预算业绩，发行人在 2017 年发放了较高的超额业绩奖金；在 2017 年业绩实现基础上，2018 年及 2019 年发行人制定了较高的业绩预算，但未出现显著超额完成的情况，2018 年和 2019 年未发生超额业绩奖金，因此造成了发行人 2017 年奖金总额高于 2018 年和 2019 年的奖金总额，同时也导致 2018 年和 2019 年销售人员人均薪酬较 2017 年下降。2020 年一季度的薪酬相比较而言较低，主要是因为未包含年终奖金造成的。

2017 年至 2019 年，发行人绩效考核指标的完成情况与奖金情况具体如下：

单位：万元

项目	2020 年 1-3 月		2019 年		2018 年		2017 年
	金额	同比增幅	金额	同比增幅	金额	同比增幅	金额
销售人员奖金合计	24.66	21.12%	403.93	7.44%	375.97	-9.43%	415.10
其中：基础业绩奖金	24.66	21.12%	403.93	7.44%	375.97	58.50%	237.20
超额业绩奖金	-		-		-	-100.00%	177.90

主要考核指标:							
营业收入	3,126.09	0.06%	23,755.81	23.31%	19,265.77	61.80%	11,907.09
营业毛利	2,589.22	32.38%	16,688.29	33.73%	12,478.99	45.04%	8,603.62
客户及领域拓展力度	有		有		无		无

三、人均薪酬是否与同行业可比公司存在显著差异，是否与营业收入增长匹配

报告期内，销售人员人均薪酬与同行业可比公司比较如下：

单位：万元

公司名称	2019年	2018年	2017年
荣科科技	36.63	28.73	23.12
彩讯股份	37.61	32.25	35.07
天玑科技	23.05	23.08	22.38
四方精创	21.42	16.76	21.33
恒生电子	31.92	29.27	23.60
用友网络	39.62	42.19	34.05
万兴科技	24.32	25.10	26.50
泛微网络	19.65	16.45	18.39
平均值	29.28	26.73	25.56
海天瑞声	64.49	65.34	78.85

注 1：同行业公司销售人员人均薪酬的计算方法，为其年报中披露的销售费用中职工薪酬总额除以销售人员数量；
注 2：个别同行业公司某些年度销售人员数量无法准确取得，因此未进行列示，仅就可取得的数据计算了同行业公司销售人员人均薪酬平均值。

注 3：2020年1季度无公开可比公司薪酬数据

2017年至2019年发行人销售人员人均薪酬分别为78.85万元、65.34万元和64.49万元，同期可比公司销售人员人均薪酬平均值为25.56万元、26.73万元和29.28万元，发行人人均薪酬高于所属行业平均水平，主要是由于发行人处于软件和信息技术服务业中成长性和专业性较高的人工智能板块，并且推行和实施顾问式营销模式，对营销人员的专业性要求较高。虽然发行人的销售人员数量较少，但是均具有本科或研究生学历，平均从业年限10年左右，发行人给予了销售人员具备行业竞争力的基本薪酬水平。基于上述原因，发行人销售人员的人均薪酬高于所属行业平均水平较为合理。

2017年至2019年发行人的营业收入逐年增长，分别为11,907.09万元、19,265.77万元、23,755.81万元，销售人员的奖金金额在2017年至2019年分别为415.10万元、375.97万元和403.93万元，扣除2017年发放的超额业绩奖金177.90万元后，2017年至2019年基础业绩奖金金额分别为237.20万元、375.97万元、403.93万元，保持逐年

增长，而且除 2019 年增幅低于营业收入的增幅之外，其他年度的基础业绩奖金增幅与营业收入增幅基本匹配。2019 年度基础业绩奖金增幅低于营业收入增幅，主要是由于 2019 年发行人新增客户及领域拓展力度的考核指标，由于当年该新目标的完成情况有限，因此拉低了 2019 年基础业绩奖金的增幅。另外，超额业绩奖金主要与是否超额完成当年业绩预算目标有关，由于 2017 年发行人实现超额完成业绩预算目标，因此发放了超额业绩奖金，而 2018 年和 2019 年发行人制定了较高的业绩预算，但未出现显著超额完成的情况，因此 2018 年和 2019 年未发放超额业绩奖金。整体来看，发行人的销售人员奖金金额是与营业收入增长相匹配的。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

1、获取发行人薪酬管理制度、绩效考核政策，并访谈人力资源部门相关人员，了解人事薪酬循环相关流程，并执行了穿行测试和控制测试程序；

2、获取发行人员工花名册及公司组织架构图，了解销售费用的人员部门构成、人数、薪资构成等要素，与计入销售费用的人员进行核对；对销售费用中的职工薪酬实施实质性分析程序；分析人员薪酬总额、人均薪酬及其构成在报告期的波动；检查发行人计入销售费用的职工薪酬变动趋势与营业收入的变动趋势的一致性；

3、查阅同行业上市公司公开信息，将公司的销售人员人均薪酬与同行业上市公司销售人员人均薪酬相比，分析是否合理；

4、根据员工花名册、具体岗位职责、薪酬明细表等情况，对职工薪酬的计提和成本费用分配情况进行了检查；

5、选取工资支付明细与工资计提情况和支付的银行流水进行核对。

(2) 核查结论

经核查，申报会计师认为：

1、发行人训练数据定制服务项目数量和训练数据产品销售数量与销售人员数量相匹配；

2、发行人报告期各期销售费用中的职工薪酬变动情况与营业收入增长相匹配，销售人员人均薪酬与同行业公司差异符合发行人的实际情况。

16.2 关于管理费用

报告期各期，公司管理费用分别为 1,866.24 万元、2,180.65 万元和 3,155.96 万元，管理费用率分别为 15.67%、11.32%和 13.28%，主要为职工薪酬、租赁费和中介机构费等。

请发行人补充披露：职工薪酬大幅上升的原因。

请发行人说明：（1）报告期各期计入管理费用的部门构成、员工人数及变动、人均薪酬，相关人员的绩效考核政策及执行情况，人均薪酬是否与同行业可比公司存在显著差异。

请申报会计师核查并发表明确意见。

答复：

发行人补充披露事项

报告期内职工薪酬大幅上升的原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（四）期间费用分析/2、管理费用分析”中进行了补充披露：

报告期内，发行人管理费用中职工薪酬主要包括发行人管理及行政人员、财务人员薪酬，及董事会会费、劳务费等管理活动相关人员薪酬。董事会会费、劳务费不涉及发行人全职员工，且在职工薪酬中占比较低，对职工薪酬变动不构成显著影响。报告期内职工薪酬大幅上升主要受发行人管理及行政人员、财务人员薪酬总额增长拉动，具体情况如下：

项目	单位	2020年1-3月	2019年	2018年	2017年
薪酬总额（管理及行政人员、财务人员）	万元	283.72	1,587.97	1,205.39	935.79
员工人数	人	36.67	37.22	31.82	29.72
人均薪酬	万元/人	7.74	42.66	37.88	31.49

注：员工人数，按照当年计入管理费用核算薪酬的员工按照各年/期间的工时扣除其在服务项目和研发项目中所填列工时后的数量占总工时的比例加总计算。

报告期内，发行人管理费用中管理及行政人员、财务人员的职工薪酬金额分别为935.79万元、1,205.39万元、1,587.97万元和283.72万元，管理费用相关的平均人数分别为29.72人、31.82人、37.22人和36.67人。除了2020年1-3月由于未包含年度奖金因此职工薪酬较低之外，2017-2019年逐年大幅增长，主要是由于一方面随着公司业务规模的扩张、对内部管理水平要求的提升和上市筹备事宜，管理及行政人员和财务人员数量增加，另一方面管理人员结构中的中层比例上升，且高级管理人员薪酬随公司业绩规模扩大而增长，管理人员平均薪酬有所提高。

发行人说明事项

报告期各期计入管理费用的部门构成、员工人数及变动、人均薪酬，相关人员的绩效考核政策及执行情况，人均薪酬是否与同行业可比公司存在显著差异。

一、报告期各期计入管理费用的部门构成

报告期内，发行人主要将董事、监事及高级管理人员、董事会办公室、总裁办公室、财务部、审计部、人力资源部、行政部、采购部的薪酬归集到管理费用的职工薪酬中。

二、报告期各期计入管理费用的员工人数及变动及人均薪酬

报告期内，计入管理费用中的职工薪酬构成明细如下：

单位：万元

项目	2020年1-3月		2019年度		2018年度		2017年度	
	金额	占比	金额	占比	金额	占比	金额	占比
管理及行政人员、财务人员薪酬总额	283.72	93.86%	1,587.97	92.46%	1,205.39	91.51%	935.79	93.94%
劳务费	18.57	6.14%	98.67	5.74%	81.00	6.15%	60.18	6.04%
董事会会费	-	-	30.88	1.80%	30.83	2.34%	0.17	0.02%
合计	302.29	100.00%	1,717.52	100.00%	1,317.22	100.00%	996.14	100.00%

报告期内，公司管理费用相关的平均人数分别为29.72人、31.82人、37.22人及36.67人，人数逐年增加，与公司稳步提升内部经营及管理水平的举措相匹配。

报告期内，管理人员中管理及行政人员、财务人员的薪酬总额和人均薪酬情况如下：

类型	项目	单位	2020年1-3月	2019年	2018年	2017年
管理及行	薪酬总额	万元	283.72	1,587.97	1,205.39	935.79

政人员、 财务人员	员工人数	人	36.67	37.22	31.82	29.72
	人均薪酬	万元/人	7.74	42.66	37.88	31.49

注1：薪酬总额为管理费用中的职工薪酬，不包括劳务费和董事会会费；

注2：员工人数，按照当年计入管理费用核算薪酬的员工按照各年/期间的工时扣除其在服务项目和研发项目中所填列工时后的数量占总工时的比例加总计算。

报告期内，发行人管理费用的人均薪酬分别为 31.49 万元、37.88 万元、42.66 万元及 7.74 万元，除 2020 年一季度因不含年终奖导致人均薪酬均偏低以外，报告期内管理费用相关的人均薪酬呈平稳上升趋势。这主要是由于一方面，随着公司业务规模的扩张、对内部管理水平要求的提升和上市筹备事宜，管理及行政人员和财务人员数量增加；另一方面，管理人员结构中的中层比例上升，且高级管理人员薪酬随公司业绩规模扩大而增长，管理人员平均薪酬有所提高。

三、相关人员的绩效考核政策及执行情况

发行人对管理相关人员的绩效考核指标由各岗位的岗位职责及工作目标决定，主要受员工职级、出勤、日常综合表现情况等影响。其中，分为高级管理人员和普通人员进行分别考核：

高级管理人员的考核机制根据发行人《高级管理人员薪酬与考核管理办法》执行，根据高级管理人员分管领域工作计划完成情况发放绩效奖金，根据公司整体业绩情况发放效益奖金，董事会薪酬与考核委员会作为考核方，并将考核结果和奖金金额报董事会审批通过。2017-2019 年，高级管理人员的考核均按该制度执行。

普通管理人员按照发行人相关制度规定，每季度结束后两周内，以及最后一个季度考核结束后两周内分别进行季度考核和年度考核，考核指标根据不同部门、不同岗位职责确定，考核流程为员工自评后提交直属领导，直属领导根据被考核人的实际表现进行打分，提交部分分管总审核，经人力资源部及总裁审批后，提交人力资源部备案，作为奖金发放的依据。2017-2019 年，普通管理人员均按发行人相关制度执行。

四、人均薪酬与同行业可比公司的比较

报告期内，管理人员人均薪酬与同行业可比公司比较如下：

单位：万元

公司名称	2019年	2018年	2017年
荣科科技	19.63	21.25	24.90
彩讯股份	17.25	19.05	22.62

天玑科技	15.99	18.28	14.39
四方精创	19.31	15.24	12.23
恒生电子	44.81	36.67	35.71
用友网络	28.05	28.70	26.36
万兴科技	30.38	37.78	30.27
泛微网络	69.77	60.34	54.88
平均值	30.65	29.66	27.67
海天瑞声	42.66	37.88	31.49

注 1：同行业公司管理人员人均薪酬的计算方法，为其年报中披露的管理费用中职工薪酬总额除以管理人员数量；
注 2：个别同行业公司某些年度管理人员数量无法准确取得，因此未进行列示，仅就可取得的数据计算了同行业公司管理人员人均薪酬平均值。

注3：2020年1季度无公开可比公司薪酬数据

2017 年至 2019 年发行人管理人员人均薪酬分别为 31.49 万元、37.88 万元和 42.66 万元，同期软件和信息技术服务业管理人员的人均薪酬分别为 27.67 万元、29.66 万元和 30.65 万元。发行人人均薪酬高于所属行业平均水平，主要是由于公司处于软件和信息技术服务业中成长性和专业性较高的人工智能板块，因此平均薪酬高于所属行业平均水平较为合理。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

1、获取发行人薪酬管理制度、绩效考核政策，并访谈人力资源部门相关人员，了解人事薪酬循环相关流程，并执行了穿行测试和控制测试程序；

2、获取发行人员工花名册及公司组织架构图，了解管理费用的人员部门构成、人数、薪资构成等要素，与计入管理费用的人员进行核对；对管理费用中的职工薪酬实施实质性分析程序；分析人员薪酬总额、人均薪酬及其构成在报告期的波动。

3、查阅同行业上市公司公开信息，将公司的管理人员人均薪酬与同行业上市公司管理人员人均薪酬相比，分析是否合理。

4、根据员工花名册、具体岗位职责、薪酬明细表等情况，对职工薪酬的计提和成本费用分配情况进行了检查；

5、选取工资支付明细与工资计提情况和支付的银行流水进行核对；

(2) 核查结论

经核查，申报会计师认为：

- 1、发行人管理费用中职工薪酬大幅上升具备合理原因，符合发行人的实际情况；
- 2、发行人管理人员人均薪酬与同行业公司的差异符合发行人的实际情况。

16.3 关于研发费用

报告期各期，公司研发费用分别为 2,527.99 万元、2,734.53 万元和 4,169.61 万元，研发费用率分别为 21.23%、14.19%和 17.55%。研发费用主要由数据服务费和职工薪酬构成，合计占研发费用的比例分别为 90.09%、85.25%和 88.83%。

请发行人补充披露：（1）训练数据产品开发和基础研发的费用构成并分析变动原因；（2）报告期各期训练数据产品开发的开展情况，2019 年度支出大幅上升的原因；（3）报告期各期训练数据产品的库存量、新增量和销售量，是否存在将定制服务产生的数据直接或修改后用于训练数据产品的情形。

请发行人说明：（1）报告期各期计入研发费用的部门构成、人员数量及其变动情况,并结合人均薪酬变化分析职工薪酬逐年大幅上升的原因，人均薪酬与同行业可比公司之间是否存在显著差异，发行人如何区分训练数据产品开发和基础研发的人员薪酬支出；（2）报告期发行人向税务机关申请研发费用加计扣除优惠政策的研究费用金额与发行人实际发生的研发费用金额之间的差异并逐项定量分析原因。

请申报会计师对上述事项进行核查并发表明确意见，进一步说明：（1）对报告期内发行人的研发投入归集是否准确、相关数据来源及计算是否合规的核查过程及核查结论；（2）对发行人研发相关内控制度是否健全且被有效执行核查过程及核查结论。

答复：

发行人补充披露事项

(1) 训练数据产品开发和基础研发的费用构成并分析变动原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析 /（四）期间费用分析/3、研发费用分析”中进行了补充披露：

报告期内，发行人的研发费用主要由训练数据产品开发支出及基础研究支出两部分构成，训练数据产品开发支出主要为开发训练数据产品相关的人工成本、数据服务费及其他相关费用支出。基础研究支出主要为研究开发训练数据所需的算法、技术以及工具、平台等的人工成本和其他费用支出。报告期内这两类支出的金额及占比如下：

单位：万元

分类	2020年1-3月		2019年度		2018年度		2017年度	
	金额	占比	金额	占比	金额	占比	金额	占比
训练数据产品开发支出	336.18	48.05%	2,217.81	53.19%	1,264.60	46.25%	1,455.04	57.56%
基础研究支出	363.46	51.95%	1,951.79	46.81%	1,469.94	53.75%	1,072.95	42.44%
合计	699.64	100.00%	4,169.61	100.00%	2,734.53	100.00%	2,527.99	100.00%

一、训练数据产品开发支出

报告期内训练数据产品开发支出的费用构成、金额、占比情况如下：

单位：万元

类别	2020年1-3月		2019年		2018年		2017年	
	金额	占比	金额	占比	金额	占比	金额	占比
职工薪酬	50.47	15.01%	320.01	14.43%	225.37	17.82%	204.40	14.05%
数据服务费	276.41	82.22%	1,844.97	83.19%	957.04	75.68%	1,196.38	82.22%
折旧与摊销	4.40	1.31%	19.88	0.90%	15.89	1.26%	25.95	1.78%
房租费	0.84	0.25%	7.70	0.35%	8.69	0.69%	5.35	0.37%
其他	4.07	1.21%	25.24	1.14%	57.60	4.55%	22.96	1.58%
合计	336.18	100.00%	2,217.81	100.00%	1,264.60	100.00%	1,455.04	100.00%

2017至2019年随着发行人加大训练数据产品开发的力度，巩固在自有知识产权训练数据产品规模性和多样性上的优势，训练数据产品开发总支出及各项费用构成总体呈上升趋势。其中，2018年训练数据产品开发支出各项费用构成出现一定波动，原因如下：（1）2018年数据服务费的金额及占比均有所降低，职工薪酬占比提高，主要是由于2018年发行人在拥有的资源和产能的基础上优先满足客户增加的定制化需求，训练数据产品开发规模的增速有所放缓，导致发行人2018年度投入训练数据产品的开发支出有所减少，发生的数据服务费也随之降低，而由于投入的研发职工薪酬受人均薪酬增长的影响并未同比减少，因此也导致了2018年职工薪酬占比较高；（2）2018年其他支出的占比略高，主要是由于2018年发行人同乐乐创新签订抵债协议，发行人及

附属子公司以人民币 40 万元购买乐乐创新中文女声语音合成数据库的永久使用权和授予第三方使用的权利及相关收益权，发行人应支付的人民币 40 万元直接冲抵乐乐创新所欠公司债务，乐乐创新向公司开具相应发票，公司将不含税金额人民币 37.74 万元计入训练数据产品开发支出的其他支出，因此造成了 2018 年其他支出占比较高。

二、基础研发支出

报告期内基础研发支出的费用构成、金额、占比情况如下：

单位：万元

类别	2020 年 1-3 月		2019 年		2018 年		2017 年	
	金额	占比	金额	占比	金额	占比	金额	占比
职工薪酬	258.16	71.03%	1,538.93	78.85%	1,148.87	78.16%	876.78	81.72%
语言研究	27.20	7.48%	182.04	9.33%	175.46	11.94%	166.77	15.54%
折旧与摊销	30.66	8.44%	108.25	5.55%	53.96	3.67%	14.62	1.36%
房租费	18.03	4.96%	66.63	3.41%	66.63	4.53%	-	0.00%
其他	29.41	8.09%	55.94	2.86%	25.02	1.70%	14.78	1.38%
合计	363.46	100.00%	1,951.79	100.00%	1,469.94	100.00%	1,072.95	100.00%

报告期内，基础研发支出各项费用构成的变化及原因如下：（1）2020 年 1-3 月职工薪酬占比出现明显降低，主要是由于 2020 年 1-3 月职工薪酬中不含年终奖，因此占比降低；（2）语言研究支出占比在报告期内逐年降低，主要是由于随着公司自身在语言学研究方面的积累加强，对外采购语言学专家服务方面的支出在各年间基本稳定，未发生显著增长；（3）折旧与摊销占比在报告期内有所增加，主要是由于各年均有新购入的房产或设备用于从事研发活动，因此导致分摊的折旧费用逐年增加；（4）房租费从 2018 年开始发生，之后保持稳定，主要是由于公司自 2018 年开始为研发部门设立专属办公场地并将房租费用在研发费用中进行分摊；（5）其他支出占比在报告期内逐年增长，主要是由于 2019 年专利申请和系统平台升级相关的支出均有增加，因此导致其他支出占比逐年增长。

（2）报告期各期训练数据产品开发的开展情况，2019 年度支出大幅上升的原因

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（四）期间费用分析/3、研发费用分析”中进行了补充披露：

报告期内，公司持续研发训练数据产品，丰富自身训练数据库资源。报告期内各期处于开发过程中的训练数据产品的个数如下：

项目	2020年1-3月	2019年	2018年	2017年
训练数据产品开发支出（万元）	336.18	2,217.81	1,264.60	1,455.04
当期进行开发的训练数据产品个数（注）	42	149	112	112
单位训练数据产品开发支出（万元/个）	8.00	14.88	11.29	12.99

注：当期进行开发的训练数据产品个数为当期有研发支出的训练数据产品个数。

2019年训练数据产品开发支出大幅上升，从2018年的1,264.60万元增长到2,217.81万元，增幅为75%，主要是由于以下两方面原因：一方面是由于下游客户对外语种训练数据产品的需求有所增长，公司准确把握市场趋势，推进了训练数据产品储备，2019年进行开发的训练数据产品个数明显增长，从112个增长至149个，增幅达33%；另一方面是由于2019年训练数据产品的增长主要是来源于外语种智能语音训练数据产品，这类产品需要由母语发音人进行录音，且标注难度较高，相关数据服务费用支出较一般产品更高，因此2019年的单位训练数据产品开发支出也有明显增长，进一步导致了训练数据产品开发支出的大幅上升。

2020年1-3月，进行开发的训练数据产品个数和单位训练数据产品开发支出均有所下降，主要是由于受2020年年初疫情的影响，训练数据产品开发活动减少所致。

（3）报告期各期训练数据产品的库存量、新增量和销售量，是否存在将定制服务产生的数据直接或修改后用于训练数据产品的情形

以下内容已在招股说明书“第六节 业务与技术/三、销售情况和主要客户/（一）报告期内销售情况/3、训练数据产品销售情况”中进行了补充披露：

报告期内，发行人各期间训练数据产品的库存量、新增量和销售量如下：

单位：个数/次数

类别	2020年1-3月	2019年	2018年	2017年
期末 / 年末库存量（个数）	641	619	516	465
新增量（个数）	22	103	51	87
销售量（个数）	65	293	183	166
销售量（授权次数）	73	414	265	263

报告期内，发行人不存在将定制服务产生的数据直接或修改后用于训练数据产品的情形。训练数据定制服务为发行人根据客户的个性化要求，为客户量体裁衣地提供定制化训练数据集的设计及开发服务，训练数据定制服务所形成的训练数据所有权在销售时已经转移至客户，公司不再拥有该等训练数据的所有权，并且根据公司内部相关业务制度，由专人在有效期截止日清除训练数据定制服务已交付的业务数据，有效期需要根据合同约定确定，合同中无明确约定的，有效期为 6 个月。发行人严格执行上述规定，不存在将定制服务形成的训练数据直接或修改后用于训练数据产品的情形。

发行人说明事项

(1) 报告期各期计入研发费用的部门构成、人员数量及其变动情况，并结合人均薪酬变化分析职工薪酬逐年大幅上升的原因，人均薪酬与同行业可比公司之间是否存在显著差异，发行人如何区分训练数据产品开发和基础研发的人员薪酬支出

一、报告期内计入研发费用的部门构成、人员数量及其变动情况：

部门	2020年1-3月	2019年	2018年	2017年
研发中心人员数量	29.28	26.98	25.29	20.71
语言研究部人员数量	3.85	3.21	2.75	1.63
质量保证部人员数量	2.67	2.42	2.00	1.58
数据业务中心参与研发活动的人员数量（注）	12.33	12.07	8.75	10.93
合计	48.13	44.68	38.79	34.86

注：数据业务中心参与研发活动的员工人数根据其在工时表中填列的训练数据产品开发工时数占该员工当年/期总工时数的比例加总。

报告期内，发行人投入研发活动的研发人员数量分别为 34.86 人、38.79 人、44.68 人及 48.13 人，呈逐年上升趋势，主要由于发行人持续增加和完善研发和技术人员数量和结构，提升研发和技术人员能力与水平，支撑公司技术、平台及工具迭代和拓展的研发需要，以及持续增长的训练数据产品的开发需求，以保持公司领先的技术优势及市场地位。

二、报告期内，发行人研发费用中的薪酬总额和人均薪酬情况如下：

类型	项目	单位	2020年1-3月	2019年	2018年	2017年
与研发费用相	薪酬总额	万元	308.63	1,858.94	1,374.24	1,081.18
	员工数量	人	48.13	44.68	38.79	34.86

关	人均薪酬	万元/人	6.41	41.61	35.43	31.02
---	------	------	------	-------	-------	-------

报告期内，发行人研发费用中薪酬成本金额分别为 1,081.18 万元、1,374.24 万元、1,858.94 万元及 308.63 万元，2017 年至 2019 年呈持续增长趋势，是由员工数量与人均薪酬双双逐年增长驱动的。员工数量的变动情况及变动原因参见上述“一、报告期内计入研发费用的部门构成、人员数量及其变动情况”。2017 年至 2019 年的研发费用人均薪酬分别为 31.02 万元、35.43 万元、41.61 万元及 6.41 万元，除 2020 年 1-3 月因不含年终奖导致薪酬总额和人均薪酬均偏低以外，报告期内研发相关的薪酬总额和人均薪酬呈平稳上升趋势，主要是由于发行人基于整体经营情况、技术研发成果、行业内平均市场薪酬水平等因素的考虑，适当增加研发人员薪酬水平所致。

三、报告期内，发行人研发人员人均薪酬与同行业比较如下：

单位：万元

名称	2019 年	2018 年	2017 年
荣科科技	7.96	8.71	4.41
彩讯股份	6.57	6.35	6.72
天玑科技	13.01	11.91	11.31
四方精创	7.49	9.66	8.46
恒生电子	25.59	24.18	22.72
用友网络	25.57	22.65	19.53
万兴科技	44.14	35.73	29.22
泛微网络	22.69	21.87	21.70
平均值	19.13	17.63	15.51
海天瑞声	41.61	35.43	31.02

注 1：同行业公司研发人员人均薪酬的计算方式，是年报中披露的研发费用中职工薪酬除以研发人员数量；

注2：2020年1-3月无公开可比公司薪酬数据。

如上表所示，发行人研发人员的人均工资高于所属行业平均水平，这主要是由于发行人处于软件和信息技术服务行业中成长性和专业性较高的人工智能板块，因此研发人员的平均薪酬高于所属行业平均水平较为合理。

四、发行人如何区分训练数据产品开发和基础研发的人员薪酬支出：

训练数据产品开发的人员主要为数据业务中心人员，该等人员薪酬按照工时积分制归集到各个项目，参与训练数据产品开发的人员每月按其实际工作情况填写工时积分到训练数据产品开发项目，每人每月工时积分总量为 8 分/日*月工作日。各个项目应分摊

的人工成本的计算方式为：各业务人员填写在该项目的积分*该人员所属级别的人工成本单价=该项目的人工支出。

基础研发人员主要为研发中心、语言研究部、质量保证部人员，按照各自的基础研发项目安排，将其薪酬全部直接归集至基础研发支出。

因此发行人通过上述部门区分、核算流程和制度安排，可以确保将人员薪酬支出在训练数据产品开发支出与基础研发支出之间进行合理和准确的区分，不存在无法区分或混同的情况。

(2) 报告期发行人向税务机关申请研发费用加计扣除优惠政策的研发费用金额与发行人实际发生的研发费用金额之间的差异并逐项定量分析原因

报告期内，发行人研发费用与加计扣除金额的差异情况如下：

单位：万元

项目	公式	2020年1-3月	2019年	2018年	2017年
申请加计扣除的研发费用金额	A	332.61	1,826.71	1,342.80	955.56
实际发生的研发费用金额	B	699.64	4,169.61	2,734.53	2,527.99
差异金额	C=B-A	367.03	2,342.90	1,391.73	1,572.43
原因分析：					
(1) 训练数据产品开发费用		336.18	2,217.81	1,264.60	1,455.04
(2) 内部交易合并抵消影响		-4.22	-14.04	-47.17	-
(3) 合并范围内亏损主体研发费用未申请加计扣除		-	-	74.17	117.39
(4) 其他不得计入加计扣除的费用		35.07	139.13	100.13	-
(1) - (4) 项原因合计		367.03	2,342.90	1,391.73	1,572.43

1、训练数据产品开发支出

根据《财政部关于企业加强研发费用财务管理的若干意见》（财企〔2007〕194号）的规定，企业研发费用（即原“技术开发费”）指企业在产品、技术、材料、工艺、标准的研究、开发过程中发生的各项费用。因此发行人数据库产品开发费用属于在产品开发过程中发生的费用，属于研发费用的范畴。

但是根据《国家税务总局关于研发费用税前加计扣除归集范围有关问题的公告》(国家税务总局公告 2017 年第 40 号)规定,企业研发活动直接形成产品或作为组成部分形成的产品对外销售的,研发费用中对应的材料费用不得加计扣除。发行人的训练数据产品是企业研发活动直接形成产品对外销售的,因此相关支出按照上述税法规定不得进行加计扣除。

2、内部交易合并抵消影响

实际发生的研发费用按抵消发行人合并范围内母子公司之间交易后的金额列示,加计扣除金额按照各个主体直接加总计算列示,合并抵消是造成研发费用与加计扣除金额存在差异的原因之一。

3、合并范围内亏损主体研发费用未申请加计扣除

发行人全资子公司中瑞智 2017 年、2018 年未实现盈利,应纳税所得额为负数。与此同时,中瑞智研发费用金额不大,申请加计扣除对税负影响较小,故未申请享受研发费用加计扣除。2019 年和 2020 年 1-3 月,中瑞智研发活动增加,申请进行研发费用加计扣除。

4、其他不得计入加计扣除的费用

单位:万元

项目	2020 年 1-3 月	2019 年	2018 年	2017 年
房屋租赁费	28.26	102.72	66.63	-
工会经费和职工教育经费	6.81	36.41	24.06	-
委托研发费用的 20%部分	-	-	9.44	-
小计	35.07	139.13	100.13	-

注 1: 发行人 2018 年开始为研发部门设立专属办公场地并开始将房屋租赁费分摊计入研发成本,而且自 2018 年开始将工会经费、职工教育经费在研发费用中单独进行核算。

注 2: 不得计入加计扣除的房屋租赁费按照各个主体直接加总计算列示,基础研发费用构成中的房租费按抵消发行人合并范围内母子公司之间交易后的金额列示,合并抵消是造成两项房租费存在差异的原因。自 2019 年发行人开始用自有房产为中瑞智的研发活动提供房屋租赁服务,中瑞智支付的租赁费用在合并财务报表层面被抵消,2019 年抵消金额为 36.08 万元、2020 年 1-3 月抵消金额为 10.23 万元。

根据《财政部 国家税务总局 科技部关于完善研究开发费用税前加计扣除政策的通知》(财税〔2015〕119 号)、《国家税务总局关于企业研究开发费用税前加计扣除政策有关问题的公告》(国家税务总局公告 2015 年第 97 号)和《国家税务总局关于研发费用税前加计扣除归集范围有关问题的公告》(国家税务总局公告 2017 年第 40 号)

规定，可以享受加计扣除的研发费用范围包括：直接从事研发活动人员的人工费用，研发活动直接投入费用，用于研发活动的折旧费用，用于研发活动的无形资产摊销费用，新产品设计试验费用，与研发活动直接相关的其他费用。

报告期内，发行人发生的与研发活动相关的房屋租赁费、工会经费和职工教育经费等不属于上述规定列举的加计扣除范围，因此不满足加计扣除条件，未申请加计扣除。

另根据财税〔2015〕119号文件规定，企业委托外部机构或个人进行研发活动所发生的费用，按照费用实际发生额的80%计入委托方研发费用并计算加计扣除。2018年，发行人发生委托研发费用47.17万元，按照80%计算加计扣除金额，剩余20%部分即9.44万元因不符合加计扣除条件，未申请加计扣除。

申报会计师核查事项

（1）核查程序

申报会计师执行的主要程序如下：

- 1、访谈公司管理层和相关业务人员，了解公司主要的研发成果和研发投入情况；
- 2、了解和评价与研发活动相关的关键财务报告内部控制的设计和运行有效性；
- 3、获取报告期内发行人研发费用明细，检查计算准确性；并与明细账、总账及财务报表合计数核对一致；
- 4、对于训练数据产品开发支出，获取报告期内业务管理平台导出的数据库开发项目明细及其支出明细，检查计入数据库开发项目的各项支出是否与该数据库产品开发相关，同时抽样检查相关支出归集为训练数据产品开发支出的支持性依据，检查发行人是否存在将定制服务产生的数据直接或修改后用于训练数据产品的情形；
- 5、对于基础研发支出，获取报告期内基础研发项目的支出明细台账，检查计入基础研发的各项支出是否与基础研发项目相关，同时抽样检查相关支出归集为基础研发支出的支持性证据；
- 6、分析研发费用的构成，检查是否存在异常或变动幅度较大的情况，分析其合理性；
- 7、对研发费用进行实质性测试，包括检查所选样本的支持性文件及人力成本、折旧摊销的合理性测试及费用截止性测试；

8、获取报告期各期训练数据产品的库存量、新增量和销售量的明细，同各期训练数据产品开发项目明细进行比对，检查有无异常；

9、获取发行人员工花名册及公司组织架构图，了解研发费用的人员部门构成、人数、薪资构成等要素，与计入研发费用的人员进行核对；

10、对研发费用中的职工薪酬实施实质性分析程序；分析人员薪酬总额、人均薪酬及其构成在报告期的波动；

11、查阅同行业上市公司公开信息，将公司的研发人员人均薪酬与同行业上市公司管理人员人均薪酬相比，分析是否合理；

12、根据员工花名册、具体岗位职责、薪酬明细表等情况，对职工薪酬的计提和成本费用分配情况进行了检查；

13、选取工资支付明细与工资计提情况和支付的银行流水进行核对；

14、查阅发行人报告期各年度所得税汇算清缴报告、《2017年-2019年度企业研究开发费用加计扣除专项鉴证报告》，与发行人实际发生的研发费用逐项进行比较，并明确具体差异原因；

(2) 核查结论

经核查，申报会计师认为：

1、报告期内，发行人训练数据产品开发和基础研发的费用构成稳定，变动原因合理，符合发行人的实际情况；

2、报告期内，发行人训练数据产品开发的开展情况与实际情况相符，2019年度训练数据产品开发支出大幅增长具备合理原因；

3、报告期内，发行人训练数据产品的库存量、新增量和销售量与发行人研发活动开展的实际情况相符，发行人不存在将定制服务产生的数据直接或修改后用于训练数据产品的情形；

4、报告期内，发行人研发费用的人员部门、员工数量和人均薪酬及变动情况与发行人的经营情况相符，职工薪酬逐年大幅上升具备合理原因；发行人研发人员人均薪酬略高于同行业公司，符合发行人的业务情况；

5、报告期内，报告期发行人向税务机关申请研发费用加计扣除优惠政策的研究费用金额与发行人实际发生的研发费用金额之间的差异主要为会计处理和税务部门认定口径的差异，差异原因合理，符合发行人的实际情况。

申报会计师说明事项

(1) 对报告期内发行人的研发投入归集是否准确、相关数据来源及计算是否合规的核查过程及核查结论

一、核查程序

针对上述核查部分相关事项，申报会计师执行了以下核查程序：

- 1、了解和评价与研发活动相关的关键财务报告内部控制的设计和运行有效性；
- 2、获取报告期内发行人研发费用明细，检查计算准确性；并与明细账、总账及财务报表合计数核对一致；
- 3、对于训练数据产品开发支出，获取报告期内业务管理平台导出的数据库开发项目明细及其支出明细，检查计入数据库开发项目的各项支出是否与该数据库产品开发相关，同时抽样检查相关支出归集为训练数据产品开发支出的支持性依据；
- 4、对于基础研发支出，获取报告期内基础研发项目的支出明细台账，检查计入基础研发的各项支出是否与基础研发项目相关，同时抽样检查相关支出归集为基础研发支出的支持性证据；
- 5、分析研发费用的构成，检查是否存在异常或变动幅度较大的情况，分析其合理性；
- 6、对研发费用进行实质性测试，包括检查所选样本的支持性文件及人力成本、折旧摊销的合理性测试及费用截止性测试；

二、核查结论

经核查，申报会计师认为：

报告期内发行人的研发投入归集准确、相关数据来源及计算过程合规。

(2) 对发行人研发相关内控制度是否健全且被有效执行核查过程及核查结论。

一、发行人研发相关的内控制度

发行人针对研发活动制定了《海天瑞声研究与开发管理办法》和相关与研发活动相关的内控制度，包括《海天瑞声全面预算管理办法》、《海天瑞声项目管理办法》、《海天瑞声采购管理制度》等，对研发项目从研发计划制定与审批、研发预算建立、研发项目立项、可行性研究、研发费用管理、项目人员安排、实施方案的制定与审批、项目进度管理、项目测试、成果交付等环节进行了明确规范。具体来讲，发行人与研发活动相关的内控制度包含以下几方面的内容：

1、建立研发项目的跟踪管理系统，有效监控、记录各研发项目的进展情况

针对训练数据产品开发项目，发行人通过业务管理平台进行线上管理，实时跟踪项目进度、记录项目支出；对基础研发项目，发行人以专项研发项目组组长为责任主体，研发部门负责人对项目进度进行监督检查。项目组组长负责编写工作周报，以邮件方式发送给研发部门负责人汇报，定期与研发部门负责人沟通、汇报项目进度情况。

2、建立与研发项目相对应的人财物管理机制

发行人的研发活动均要求按项目进行管理，每个项目均指定了项目参与人员，采购数据服务、聘请外部专家、使用固定资产、占用租赁房屋时均需要按照项目口径进行，人财物管理机制与研发费用严格对应。

3、明确研发支出开支范围和标准，严格按照研发开支用途、性质据实列支研发支出，建立研发支出审批程序

报告期内，发行人在研发管理制度中明确了研发支出开支范围和标准，将研发费用管理纳入了公司全面预算体系，研发项目涉及费用支出时，均按照公司统一的财务支出管理办法执行，总体遵循“先预算，后动支”的管理原则，严格控制预算进度与时间进度或业务计划达成进度的匹配性，同时执行严格的层级审批制度，确保研发支出开支范围和标准有效执行，确保严格按研发开支用途、性质列支研发支出。若涉及重大研发项目或截至本项目本年度累计项目支出大于年度研发预算，还需经总经理办公会审批。

综上所述，发行人已建立与研发相关的内部控制制度，能有效监控、记录各研发项目的进展情况。

二、核查程序

针对发行人与研发相关的内控制度，申报会计师执行了下述核查程序：

1、了解发行人建立的与研发活动相关的内部控制制度，对研发流程及其控制活动的情况执行穿行测试；

2、对研发流程中识别的关键控制点执行控制有效性测试，选择样本检查关键控制点是否得到有效执行；

3、对研发流程涉及的业务管理系统平台及其相关控制活动，引入信息系统专家执行与信息系统相关的控制测试。

三、核查结论

经核查，申报会计师认为：

发行人与研发相关的内控制度健全且有效执行。

17. 关于税务事项

报告期各期，发行人享受税收优惠的金额分别为 1,129.34 万元、2,080.29 万元和 2,097.82 万元，占当期利润总额的比重分别 28.22%、28.84%和 22.96%，占比较高。

请发行人说明：（1）逐项对照说明发行人是否符合国家规划布局内重点软件企业的认定条件，相关认定到期后是否会对公司持续经营及业绩产生重大影响；（2）结合软件产品增值税优惠政策的相关规定，说明发行人销售软件产品的具体名称和销售收入金额以及是否符合条件，合并报表中各类别营业收入与增值税申报表申报的软件收入、增值税即征即退额之间的匹配关系，进项税额和销项税额与采购和销售的匹配关系，逐项分析差异原因及合理性；（3）“收到的税费返还”“支付的各项税费”与报表科目的勾稽关系。

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

答复：

发行人说明事项

(1) 逐项对照说明发行人是否符合国家规划布局内重点软件企业的认定条件，相关认定到期后是否会对公司持续经营及业绩产生重大影响

一、国家规划布局内重点软件企业的认定

根据《财政部、国家税务总局关于进一步鼓励软件产业和集成电路产业发展企业所得税政策的通知》（财税[2012]27号），国家规划布局内的重点软件企业，如当年未享受免税优惠的，可减按10%的税率征收企业所得税。

根据《财政部、国家税务总局、发展改革委、工业和信息化部关于软件和集成电路产业企业所得税优惠政策有关问题的通知》（财税〔2016〕49号），享受财税〔2012〕27号文件规定的税收优惠政策的软件、集成电路企业，每年汇算清缴时应按照《关于发布修订后的〈企业所得税优惠政策事项办理办法〉的公告》（国家税务总局公告2018年第23号）规定向税务机关备案。

《财政部、国家税务总局、发展改革委、工业和信息化部关于软件和集成电路产业企业所得税优惠政策有关问题的通知》（财税〔2016〕49号）以及《国家发展改革委、工业和信息化部、财政部、税务总局关于印发国家规划布局内重点软件和集成电路设计领域的通知》（发改高技[2016]1056号）规定的国家规划布局内重点软件企业的认定条件，以及发行人对照符合认定条件的具体情况如下：

认定条件	发行人（母公司）情况	是否符合
在中国境内（不包括港、澳、台地区）依法注册的居民企业	注册地址为北京市海淀区	是
汇算清缴年度具有劳动合同关系且具有大学专科以上学历的职工人数占企业月平均职工总人数的比例不低于40%，其中研究开发人员占企业月平均职工总数的比例不低于20%	报告期内，具有劳动合同关系且具有大学专科以上学历的职工人数占企业月平均职工总人数的比例为100%，其中研发人员占企业月平均职工总数比例在报告期内分别为：34.04%，33.33%，32.17%及31.20%。	是
拥有核心关键技术，并以此为基础开展经营活动，且汇算清缴年度研究开发费用总额占企业销售（营业）收入总额的比例不低于6%；其中，企业在中国境内发生的研究开发费用金额占研究开发费用总额的比例不低于60%	目前拥有20项已授权专利（19项发明专利授权、1项实用新型专利授权）、94项软件著作权，核心技术覆盖基础研究、平台工具、训练数据生产三个层次，应用于训练数据生产的设计、采集、加工、质检全流程之中，构成了公司生产经营的基础，公司通过核心技术的成果转化形成了基于核心技术的训练数据相关产品和服务；	是

	报告期内，符合条件的研究开发费用总额占企业销售（营业）收入总额的比例分别为8.03%、7.07%、6.92%及9.07%；其中，在中国境内发生的研究开发费用金额占研究开发费用总额的比例均为100%。	
汇算清缴年度软件产品销售（营业）收入占企业收入总额的比例不低于50%（嵌入式软件产品和信息系统集成产品开发销售（营业）收入占企业收入总额的比例不低于40%），其中：软件产品自主开发销售（营业）收入占企业收入总额的比例不低于40%（嵌入式软件产品和信息系统集成产品开发销售（营业）收入占企业收入总额的比例不低于30%）	报告期内，软件产品销售（营业）收入（包括软件产品销售和软件开发服务）占企业收入总额的比例均接近100%，其中：软件产品和服务自主开发销售（营业）收入占企业收入总额的比例接近100%。	是
主营业务拥有自主知识产权	目前拥有20项已授权专利（19项发明专利授权、1项实用新型专利授权）、94项软件著作权，多年积累的核心技术使得发行人自成立至今一直从事训练数据产品和服务的开发与销售，能够更大规模、更有效率、更加精准地生产训练数据，提高了训练数据对于客户AI算法模型的改善、优化效果。	是
具有与软件开发相适应软硬件设施等开发环境（如合法的开发工具等）	拥有研究开发专用场地，及与软件开发相关的服务器、PC机、交换机、路由器等硬件设备，及拥有录音系统软件、音视频处理软件、数据标注平台等多个软件开发相关的软件设施。	是
汇算清缴年度未发生重大安全、重大质量事故或严重环境违法行为	截至目前，未发生重大安全、重大质量事故或严重环境违法行为。	是
在国家规定的重点软件领域内，汇算清缴年度软件产品销售（营业）收入不低于5000万元，应纳税所得额不低于250万元，研究开发人员占企业月平均职工总数的比例不低于25%，企业在中国境内发生的研究开发费用金额占研究开发费用总额的比例不低于70%	报告期在国家规定的重点软件领域内软件产品销售（营业）收入为分别为9,961.05万元、15,915.86万元、19,330.34万元及2,234.09万元。应纳税所得额分别为3,830.65万元、6,675.31万元、8,103.59万元及1,465.29万元。研究开发人员占企业月平均职工总数的比例分别为34.04%、33.33%、32.17%及31.20%，企业在中国境内发生的研究开发费用金额占研究开发费用总额的比例均为100%。	是
国家规定的重点软件领域：（八）语言文字信息处理软件：汉语和少数民族语言相关文字编辑处理、语音识别/合成、机器翻译软件	智能语音及机器翻译领域的产品和服务属于该项重点软件领域。	是
选择领域的销售（营业）收入占本企业软件产品开发销售（营业）收入的比例不低于20%	报告期内，属于上述领域的销售（营业）收入占软件产品开发销售（营业）收入的比例分别为83.67%、83.80%、82.38%及71.38%。	是

综上，发行人在报告期内符合国家规划布局内重点软件企业认定的相关条件，且发行人已按相关规定办理了 2017 年度、2018 年度和 2019 年度的重点软件企业备案认定。

二、相关认定到期后若不能被持续认定对公司的影响

报告期内，发行人符合高新技术企业及国家规划布局内的重点软件企业的相关认定，按照相关税收优惠政策，可享有 15%或 10%的所得税优惠税率，发行人报告期内实际执行重点软件企业 10%的所得税优惠税率。相应的税收优惠对经营成果的影响情况分别分析如下：

单位：万元

项目	2020 年 1-3 月	2019 年度	2018 年度	2017 年度
高新技术企业所得税税率优惠影响金额（15%优惠税率相较 25%的法定税率）	151.33	864.41	693.46	391.72
重点软件企业所得税税率优惠额外影响金额（10%优惠税率相较 15%的高新技术企业税率）	75.67	432.20	538.26	-
利润总额	1,558.97	9,135.95	7,212.51	3,994.18
高新技术企业所得税税率优惠影响金额占当期利润总额的比例	9.71%	9.46%	9.61%	9.81%
重点软件企业所得税税率优惠额外影响金额占当期利润总额的比例	4.85%	4.73%	7.46%	-
企业所得税税收优惠影响占当期利润总额的比例合计	14.56%	14.19%	17.07%	9.81%

注：发行人于 2017 年享受高新技术企业的企业所得税优惠税率 15%，并按照该税率进行了 2017 年所得税的纳税申报。2018 年 10 月发行人获得 2017 年国家规划布局内的重点软件企业认定，可以对 2017 年的所得税享受 10%的所得税税率优惠政策，因此发行人按 10%的优惠税率重新申报了 2017 年度的企业所得税，上表中 2018 年度重点软件企业所得税税率优惠额外影响金额包括当年收到退回的 2017 年度多缴纳的企业所得税 191.53 万元的影响。

报告期内，公司享受的企业所得税税率优惠影响金额占当期公司利润总额的比例分别为 9.81%、17.07%、14.19%和 14.56%，公司经营成果对该税收优惠不存在重大依赖。其中重点软件企业所得税税率优惠额外影响金额占当期公司利润总额的比例分别为 0%、7.46%、4.73%和 4.85%，占比较小。重点软件企业相关认定到期后若不能被持续认定，公司仍可享受高新技术企业 15%的所得税优惠税率，因此预期不会对公司的生产经营产生重大影响。

发行人相关认定到期后若不能被持续认定，相关风险已于招股说明书“第四节 风险因素/十一、税收优惠不确定的风险”进行了披露：

“报告期内，公司享受了增值税即征即退、高新技术企业和国家规划布局内的重点软件企业的企业所得税税率优惠、研发费用加计扣除等一系列税收优惠，如果未来公司

所享受的税收优惠政策出现不利变化，或者公司不再符合税收优惠的条件，则公司将无法持续享受税收优惠或存在所享受的税收优惠减少的可能性。并进而导致公司未来经营业绩和现金流水平受到不利影响。”

(2) 结合软件产品增值税优惠政策的相关规定，说明发行人销售软件产品的具体名称和销售收入金额以及是否符合条件，合并报表中各类别营业收入与增值税申报表申报的软件收入、增值税即征即退额之间的匹配关系，进项税额和销项税额与采购和销售的匹配关系，逐项分析差异原因及合理性

一、增值税即征即退的优惠政策

根据《财政部 国家税务总局关于软件产品增值税政策的通知》财税〔2011〕100号规定：增值税一般纳税人销售其自行开发生产的软件产品，按 17% 税率征收增值税后，对其增值税实际税负超过 3% 的部分实行即征即退政策。满足下列条件的软件产品，经主管税务机关审核批准，可以享受本通知规定的增值税政策：1) 取得省级软件产业主管部门认可的软件检测机构出具的检测证明材料；2) 取得软件产业主管部门颁发的《软件产品登记证书》或著作权行政管理部门颁发的《计算机软件著作权登记证书》。

二、发行人适用即征即退优惠政策的软件销售收入

发行人 2012-2018 年完成了识别语音库软件 V1.0.0.0、语音合成数据库软件 V1.0.0.0、文本数据库软件 V1.0.0.0、电子书记员系统服务软件 V1.0.0、对话语音库软件 V1.0.0.0、词典数据库软件 V1.0.0.0、图形图像数据库软件 V1.0.0.0、智能音频采集系统软件 V1.1.4、视频数据库软件 V1.0.1、智能会议系统软件 V1.0.1、ASIORecord 多通道录音软件 V3.4 共 11 个软件的优惠备案，均已获得相关部门的检测证明并取得计算机软件著作权登记证书。该类训练数据产品及其他软件销售享受财税〔2011〕100号规定的优惠政策，在销售合同及销售发票中注明软件著作权编号的训练数据产品及其他软件销售收入可享受增值税即征即退优惠政策。

发行人申请并拥有软件著作权的上述软件是针对各类型训练数据产品的结构设计、数据管理与索引、数据在物理磁盘上的存储、定制化展示和导出等功能，以适配客户的不同算法和使用环境，并提高训练数据产品的易用性。由于同一类型训练数据产品具体数据内容和数量的变化不影响软件的整体框架及功能，因此同一软件著作权的软件可匹

配其对应类型的不同训练数据产品。例如，软件著作权识别语音库软件 V1.0.0.0 可以用于发行人所有语音识别数据库产品。

报告期内，公司增值税申报表申报的上述软件产品销售收入金额如下：

单位：万元

软件名称	批号	2020年1-3月	2019年	2018年	2017年
识别语音库软件 V1.0.0.0	海国税软字 [2015] 20151116091815 号	2,374.35	4,833.84	4,412.17	3,722.13
语音合成数据库软件 V1.0.0.0	海国税软字 [2015] 20151116091816 号	-	774.90	290.02	627.30
文本数据库软件 V1.0.0.0	海国税软字 [2015] 20151116091817 号	-	322.53	40.91	132.85
电子书记员系统服务软件 V1.0.0	税软字 [2018]201803280	78.23	108.53	123.94	-
对话语音库软件 V1.0.0.0	海国税软字 [2015] 20151116091814 号	-	88.58	-	-
词典数据库软件 V1.0.0.0	海国税软字 [2015] 20151116091818 号	-	53.97	85.15	117.69
图形图像数据库软件 V1.0.0.0	税软字 [2016]201611090	-	48.88	22.82	25.50
智能音频采集系统软件 V1.1.4	税软字 [2018]201803280	-	-	51.52	-
合计		2,452.58	6,231.24	5,026.53	4,625.47

上述软件产品销售收入符合增值税即征即退优惠政策的相关条件。

三、合并报表中各类别营业收入与增值税申报表申报的软件收入、增值税即征即退额之间的匹配关系

报告期内，发行人软件产品销售收入按照产品类别分别计入训练数据产品和训练数据相关的应用服务。

对于训练数据产品和训练数据相关的应用服务销售，主管税务机关认为，销售合同条款中是否包含软件著作权编号是判断该类业务适用税率选择的重要标志：

1、包含软件著作权编号的，属于销售软件产品，适用 17%（2018 年 5 月 1 日起变更为 16%，2019 年 4 月 1 日起变更为 13%）的销项税率，在增值税申报表中列示在 17%/16%/13% 税率的销售收入中；

2、未包含软件著作权编号的，属于提供特许权使用服务，适用 6% 的销项税率，在增值税申报表中列示在 6% 税率的销售收入中。

发行人销售合同是否包含软件著作权编号，是根据与客户双方协商的结果确定的。通常销售合同条款中会记载软件著作权编号，针对上述销售合同发行人按销售软件产品

进行增值税销项税纳税处理；但部分客户因其自身原因，如部分大型科技公司不同意在其制式化的销售合同条款中加入软件著作权编号，针对上述销售合同发行人按提供特许权使用服务进行增值税销项税纳税处理。

以上两种不同的处理方式只针对增值税销项税额的纳税处理，不影响发行人训练数据产品和训练数据相关的应用服务销售的业务实质，在收入确认等财务处理上不存在区别。

报告期内，合并报表中符合即征即退优惠政策的各类别营业收入与增值税申报表申报的软件收入具体情况如下：

单位：万元

类别		2020年1-3月	2019年	2018年	2017年
合并报表中享受即征即退优惠政策的软件收入	训练数据产品收入	1,854.23	7,501.50	4,477.32	4,848.76
	训练数据相关的应用服务收入	-	162.30	239.26	-
	合计	1,854.23	7,663.80	4,716.58	4,848.76
调节过程	加：当期开票尚未验收	1,210.87	676.26	503.43	278.78
	加：当期开票前期验收	1,484.32	292.01	370.06	105.51
	减：当期验收前期开票	661.03	499.04	277.94	238.46
	减：当期验收尚未开票	1,435.81	1,901.80	285.60	369.12
增值税申报表申报的软件收入		2,452.58	6,231.24	5,026.53	4,625.47

注：上述合并报表中的软件收入只包括训练数据产品和训练数据相关的应用服务收入中包含软件著作权编号的，适用销售软件产品税率为17%（2018年5月1日起变更为16%，2019年4月1日起变更为13%）的收入。

合并报表中的软件收入与增值税申报表中的软件收入存在差异的原因系合并报表中的软件收入金额为当年/期已验收确认的符合即征即退税收优惠条件的收入，增值税申报表中的软件收入为当年/期开具发票并按享受即征即退的软件收入进行增值税申报的销售额，两者存在会计确认收入时点与开具发票时点的时间性差异。

报告期内报表中享受即征即退优惠政策的软件收入与增值税纳税申报表中的软件收入总体差异为747.55万元，差异较小，基本匹配。

报告期内，公司增值税申报表中的软件收入和即征即退税额的匹配关系如下：

单位：万元

项目	适用税率	计算公式	报告期合计	2020年1-3月	2019年度	2018年度	2017年度
增值税申报表申	17%	①	7,006.67	-	-	2,381.20	4,625.47

报的软件收入	16%	②	3,814.48	-	1,169.15	2,645.34	-
	13%	③	7,514.67	2,452.58	5,062.09	-	-
	小计		18,335.82	2,452.58	6,231.24	5,026.53	4,625.47
名义税率	17%	④=①*17%	1,191.13	-	-	404.80	786.33
	16%	⑤=②*16%	610.32	-	187.06	423.25	-
	13%	⑥=③*13%	976.91	318.84	658.07	-	-
	小计		2,778.36	318.84	845.14	828.06	786.33
实际税负	3%	⑦= (①+②+③) *3%	550.07	73.58	186.94	150.80	138.76
即征即退收入对应的进项税额		⑧	34.30	-	10.52	2.42	21.35
按 3% 税负计算应退税额		⑨=④+⑤+⑥-⑦-⑧	2,193.99	245.26	647.68	674.84	626.21
实际收到即征即退税款		⑩	2,174.03	157.56	638.49	747.86	630.12
差额		⑪=⑨-⑩	19.95	87.69	9.19	-73.02	-3.91

注：销售合同中注明软件著作权编号，按照 17% 税率（2018 年 5 月 1 日起变更为 16%，2019 年 4 月 1 日起变更为 13%）缴纳增值税的销售收入可享受增值税即征即退优惠政策。

报告期内，发行人实际收到的增值税即征即退额分别为 630.12 万元、747.86 万元、638.49 万元和 157.56 万元，与按照按 3% 税负计算的应退税额之间的差异主要是由于实际退税有一定的滞后性，从而导致年度之间产生差异，报告期内总体差异为 19.95 万元，差异较小。公司于实际收到即征即退税款时计入当期“其他收益”进行核算。

四、进项税额和销项税额与采购和销售的匹配关系，逐项分析差异原因及合理性；

（一）发行人进项税额与采购的匹配关系

报告期内，发行人进项税额与采购商品或劳务金额之间的勾稽关系如下表所示：

单位：万元

采购类型	进项税率	2020年1-3月		2019年		2018年		2017年	
		采购金额	进项税额	采购金额	进项税额	采购金额	进项税额	采购金额	进项税额
数据服务费	6%	1,016.38	60.98	5,349.19	320.95	6,273.79	376.43	3,223.96	193.44
	3%-小规模纳税人	45.62	1.37	471.36	14.14	290.53	8.72	373.88	11.22
	0%-无增值税专票	106.63	-	279.40	-	788.33	-	459.47	-
办公房租	5%	155.54	7.78	625.49	31.27	487.31	24.37	335.93	16.80
	0%-无增值税专票	54.51	-	69.17	-	57.86	-	79.55	-
资产采购	17%	-	-	-	-	142.34	24.20	55.33	9.41
	16%	-	-	60.82	9.73	310.92	49.75	-	-
	13%	14.51	1.89	188.10	24.45	-	-	-	-
	6%	-	-	38.20	2.29	99.06	5.94	-	-
	0%-无增值税专票	3.25	-	-	-	0.58	-	148.39	-
岗位服务费	6%	36.40	2.18	177.60	10.66	145.26	8.72	-	-
	0%-无增值税专票	22.34	-	107.29	-	111.36	-	226.95	-
不动产	4%-税务局代开	-	-	-	-	2,008.57	80.87	-	-
其他采购	—	215.10	26.58	1,523.78	108.19	955.01	27.80	748.16	-

进项税转出	—	-	-	-	-1.00	-	-4.10	-	-0.36
增值税加计抵扣	—	-	7.93	-	40.75	-	-	-	-
合计	—	1,670.28	108.71	8,890.41	561.44	11,670.92	602.70	5,651.62	230.50

注 1：2018 年公司购置建筑物支出 2,102.20 万元，其中房产原值 2,008.57 万元适用 4% 税率；中介费（不含税）33.38 万元适用 6% 税率，包括在上表其他采购额中；剩余部分 60.25 万元为契税及印花税。

注 2：资产采购中定制化无形资产的适用税率为 6%，其他无形资产及有形动产的适用税率为 17%/16%/13%。

整体来看，公司采购业务支出与增值税进项税额勾稽关系是合理的。

（二）发行人销项税额与销售产品或提供劳务的匹配关系

报告期内，发行人的销项税额与销售产品或提供劳务金额之间的勾稽关系如下表所示：

单位：万元

销售业务	适用税率	2020 年 1-3 月		2019 年		2018 年		2017 年	
		销售金额	销项税额	销售金额	销项税额	销售金额	销项税额	销售金额	销项税额
训练数据产品	17%	-	-	-	-	2,270.12	385.92	4,848.76	824.29
	16%	-	-	1,969.94	315.19	2,207.20	353.15	-	-
	13%	1,854.23	241.05	5,531.56	719.10	-	-	-	-
	6%	134.11	8.05	1,674.97	100.50	2,124.34	127.46	640.55	38.43
训练数据定制服务	6%	1,141.52	68.49	14,272.43	856.35	12,369.55	742.17	6,297.23	377.83
训练数据相关的应用服务	17%	-	-	-	-	-	-	-	-
	16%	-	-	-	-	309.38	49.50	-	-
	13%	-	-	264.69	34.41	-	-	-	-
	6%	-	-	98.73	5.92	41.51	2.49	120.55	7.23

其他业务	9%	10.23	0.92	36.08	3.25	-	-	-	-
内部交易抵消	—	-14.00	—	-92.59	—	-56.33	—	-	—
其他因素	—	—	143.94	—	-13.79	—	118.58	—	51.00
合计	—	3,126.09	462.45	23,755.81	2,020.93	19,265.77	1,779.28	11,907.09	1,298.78

注：发行人2019年适用13%税率及2018年适用16%税率的训练数据相关的应用服务收入中分别包括102.39万元及70.12 万元涉及软件使用权与硬件配套销售，针对该业务适用税率应为13%或16%，但不符合《财政部国家税务总局关于软件产品增值税政策的通知》（财税[2011]100号）关于即征即退条件的规定，无法享受销售软件产品增值税即征即退政策，因此在本问题“（2）/三、发行人适用即征即退优惠政策的软件销售收入”中未作为软件销售收入列示。

1、发行人主要收入类型及适用的销项税率

公司主要包括三大类主要业务，分别是训练数据定制服务、训练数据产品和训练数据相关的应用服务。

训练数据定制服务：适用税率 6%。

训练数据产品：其中包含软件著作权编号的，适用销售软件产品税率为 17%（2018 年 5 月 1 日起变更为 16%，2019 年 4 月 1 日起变更为 13%）；未包含软件著作权编号的，适用提供特许权使用服务税率 6%。

训练数据相关的应用服务：其中应用软件的授权使用，以及配套硬件的销售（如有）适用税率 17%（2018 年 5 月 1 日起变更为 16%，2019 年 4 月 1 日起变更为 13%）；提供应用软件的定制开发服务适用税率 6%。

2、其他因素说明

报告期内，按照销售业务类型和适用税率比例计算的销项税额与公司实际承担的销项税额存在一定的差额，具体影响金额详见上表“其他因素”所示。主要是由于公司严格按照税法规定，将会计确认销售收入时点、开具增值税发票时点、收到预收款时点三者中的最早时点作为增值税纳税义务的发生时点。因此会计确认收入时点和增值税纳税义务发生时点产生差异。差异原因分析通常有两种情形：

第一种情形：部分项目执行过程中收到客户支付的预收款，会计上尚未确认收入，发行人根据三个时点孰早原则在收到预收款时申报销项税，早于会计确认收入时点；

第二种情形：部分客户在支付预付款前，先要求发行人开具发票，因此开具发票时点既早于预收款时点，也早于会计确认收入时点。

对于上述两种情形，即预收款项影响和只开票税金影响，进一步核对与“其他因素影响金额”之间的勾稽关系：

其他因素影响金额=（期末预收账款产生的销项税-期初预收账款产生的销项税）+（期末开票销项税-期初开票销项税）

单位：万元

类型	适用税率	6%		13%		16%		17%		合计		序号	计算公式	销项税额
		预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税			
预收账款产生销项税	项目期间(时点)	预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税	预收金额	计提销项税		预收账款对当期销项税额的影响	
	2016-12-31/2016年	1,068.80	64.13	-	-	-	-	231.26	39.31	1,300.06	103.44	①	—	—
	2017-12-31/2017年	2,350.75	141.05	-	-	-	-	234.35	39.84	2,585.10	180.89	②	a=②-①	77.45
	2018-12-31/2018年	3,178.68	190.72	-	-	618.59	98.97	53.98	9.18	3,851.25	298.87	③	b=③-②	117.98
	2019-12-31/2019年	1,742.91	104.57	1,285.56	167.12	-	-	-	-	3,028.47	271.70	④	c=④-③	-27.17
	2020-03-31/2020年第一季度	1,009.11	60.55	2,678.60	348.22	-	-	-	-	3,687.71	408.76	⑤	d=⑤-④	137.07
开票销项税	项目期间(时点)	开票金额	计提销项税	开票金额	计提销项税	开票金额	计提销项税	开票金额	计提销项税	开票金额	计提销项税		只开票销项税对当期销项税影响	
	2016-12-31/2016年	893.44	53.61	-	-	-	-	8.75	1.49	902.19	55.10	⑥	—	—
	2017-12-31/2017年	348.55	20.91	-	-	-	-	45.55	7.74	394.10	28.65	⑦	e=⑦-⑥	-26.45
	2018-12-31/2018年	61.25	3.68	-	-	158.06	25.29	1.62	0.28	220.93	29.25	⑧	f=⑧-⑦	0.60
	2019-12-31/2019年	123.37	7.40	270.96	35.22	-	-	-	-	394.33	42.63	⑨	g=⑨-⑧	13.39
	2020-03-31/2020年第一季度	97.54	5.85	335.74	43.65	-	-	-	-	433.27	49.50	⑩	h=⑩-⑨	6.87
其他影响因素合计	2017年	—	—	—	—	—	—	—	—	—	—	—	i=a+e	51.00
	2018年	—	—	—	—	—	—	—	—	—	—	—	j=b+f	118.58
	2019年	—	—	—	—	—	—	—	—	—	—	—	k=c+g	-13.79
	2020年第一季度	—	—	—	—	—	—	—	—	—	—	—	l=d+h	143.94

上表中所列示的预收账款金额与财务报表列示的预收款项/合同负债金额之间的核对关系列示如下：

单位：万元

项目	2020年3月31日	2019年12月31日	2018年12月31日	2017年12月31日
计提销项税的预收金额	3,687.71	3,028.47	3,851.25	2,585.10
按照应收/预收净额列报的抵消金额	-418.85	-547.81	-443.56	-349.15
财务报表列示金额	3,268.86	2,480.66	3,407.69	2,235.95

(3) “收到的税费返还” “支付的各项税费”与报表科目的勾稽关系

报告期内发行人现金流量表中“收到的税费返还”“支付的各项税费”与其他报表科目的勾稽关系如下表所示：

一、收到的税费返还

单位：万元

项目	2020年1-3月	2019年度	2018年度	2017年度
所得税费用-按税法及相关规定计算的当年所得税（收到的多缴企业所得税）	-	-	1.02	35.03
所得税费用-调整以前期间所得税的影响	-	-	191.53	-
其他流动资产-预缴企业所得税的减少	-	114.60	-	-
应交税费-企业所得税的（减少）/增加（预缴所得税退回的影响）	-	-1.09	57.43	-
应交税费-增值税的增加（预缴增值税退回的影响）	-	-	15.34	-
合计	-	113.51	265.32	35.03
收到的税费返还	-	113.51	265.32	35.03
差异	-	-	-	-

二、支付的各项税费

单位：万元

项目	2020年1-3月	2019年度	2018年度	2017年度
所得税费用-按税法及相关规定计算的当年所得税（不含代扣代缴境外所得税及收到的多缴企业所得税）	146.53	810.36	666.76	621.38
其他流动资产-预缴企业所	-	-	114.60	-

得税的增加				
应交税费-企业所得税的减少/(增加)	252.92	-389.75	141.93	104.46
应交税费-增值税的减少	364.98	1,340.80	1,192.90	1,182.86
税金及附加(包括管理费用中核算的税费)	44.16	215.34	141.27	166.17
应交税费-税金及附加的减少/(增加)	2.28	-15.80	5.98	10.98
合计	810.87	1,960.95	2,263.44	2,085.85
支付的各项税费	810.87	1,960.95	2,263.44	2,085.85
差异	-	-	-	-

中介机构核查意见

(1) 核查程序

保荐机构、申报会计师执行的主要程序如下：

1、与发行人财务部门相关人员访谈，了解发行人相关税务流程、核算方法及核算过程，并执行了穿行测试和控制测试程序；

2、获取发行人软件产品开发销售收入/研究开发费用专项审计报告、人员花名册、企业所得税申报表等重点软件企业认定的相关材料，及税务局优惠备案回执，逐项对照检查发行人是否符合重软企业的认定条件；

3、获取发行人享受增值税优惠政策的软件产品清单，及相应的计算机软件著作权登记证书、增值税优惠备案回执等相关文件，检查是否符合增值税优惠政策的相关规定；

4、了解发行人软件销售业务类型，获取软件收入明细表、了解软件收入情况、收入确认时点等，与增值税申报表申报的软件收入之间是否存在差异；

5、了解发行人软件增值税即征即退情况，包括退税基础及计算依据；获取公司申报增值税即征即退的相关资料，并与发行人账面记录核对；检查银行收款相关单据，确认软件增值税退税金额已确实收到且金额正确；

6、获取并检查发行人报告期内增值税纳税申报表，检查、核实增值税申报情况是否与账面情况一致，并将销售收入与销项税额、采购与进项税额进行勾稽检查；

7、检查了发行人报告期内“收到的税费返还”及“支付的各项税费”与对应的资产负债表项目、利润表项目的勾稽关系。

(2) 核查结论

经核查，保荐机构、申报会计师认为：

1、报告期内，发行人符合国家规划布局内重点软件企业认定的相关条件，相关认定到期后若不能被持续认定，不会对公司持续经营及业绩产生重大影响，发行人已在招股说明书中充分披露相关风险；

2、发行人享受增值税即征即退优惠的软件产品符合增值税优惠政策的相关规定；发行人增值税即征即退金额与软件收入匹配关系合理；进项税额和销项税额与采购和销售的匹配关系合理；

3、发行人“收到的税费返还”“支付的各项税费”与对应的资产负债表项目、利润表项目勾稽相符。

18. 关于货币资金

报告期各期末，公司货币资金分别为 1,301.73 万元、4,669.16 万元和 16,325.52 万元，公司货币资金主要为银行存款，报告期内发行人存在购买银行理财产品的情形。

请发行人说明：（1）结合报告期各期购买理财产品的具体情况，分析相关会计处理是否符合企业会计准则的相关规定，与报表相关科目的勾稽关系，理财产品购买履行的相关决策程序，发行人建立的与货币资金相关的内部控制及执行情况；（2）结合目前大额货币资金的使用计划，说明发行人需要进一步补充流动资金的原因及合理性。

请保荐机构核查并发表明确意见，请申报会计师对说明事项（1）核查并发表明确意见。

答复：

发行人说明事项

(1) 结合报告期各期购买理财产品的具体情况，分析相关会计处理是否符合企业会计准则的相关规定，与报表相关科目的勾稽关系，理财产品购买履行的相关决策程序，发行人建立的与货币资金相关的内部控制及执行情况

一、报告期内公司购买理财产品的情况

2017 年度购买理财产品情况：

单位：万元

银行名称	理财产品名称	理财产品类型	期初金额	本期购入	本期赎回	期末未收利息	期末金额	利息收入	年化收益率	是否保本
浦发银行	利多多对公结构性存款公司固定持有期	保证收益型	-	6,000.00	3,000.00	-	3,000.00	30.78	4.15%-4.25%	是
平安银行	平安银行卓越计划滚动型人民币公司理财产品	非固定期限保本浮动收益型	-	7,500.00	3,000.00	-	4,500.00	100.40	3.50%-4.35%	是
合计			-	13,500.00	6,000.00	-	7,500.00	131.18		

2018 年度购买理财产品情况：

单位：万元

银行名称	理财产品名称	理财产品类型	期初金额	本期购入	本期赎回	期末未收利息	期末金额	利息收入	年化收益率	是否保本
浦发银行	利多多对公结构性存款公司固定持有期	保证收益型	3,000.00	10,500.00	12,500.00	-	1,000.00	140.14	3.75%-4.55%	是

平安银行	平安银行卓越计划滚动型人民币公司理财产品	非固定期限保本浮动收益型	4,500.00	-	1,700.00	-	2,800.00	111.68	3.50%-4.35%	是
	平安银行对公结构性存款（挂钩利率）开放型人民币产品	固定期限保本浮动收益型	-	13,500.00	9,000.00	-	4,500.00	86.62	3.10%-4.70%	是
合计			7,500.00	24,000.00	23,200.00	-	8,300.00	338.43		

2019年度购买理财产品情况：

单位：万元

银行名称	理财产品名称	理财产品类型	期初金额	本期购入	本期赎回	期末未收利息	期末金额	利息收入	年化收益率	是否保本
浦发银行	利多多对公结构性存款公司固定持有期	保证收益型	1,000.00	-	1,000.00	-	-	3.54	3.75%	是
浦发银行	利多多对公结构性存款（挂钩金融衍生品）	固定期限保本浮动收益型	-	5,000.00	-	-	5,000.00	-	3.75%	是
平安银行	平安银行卓越计划滚动型人民币公司理财产品	非固定期限保本浮动收益型	2,800.00	-	2,800.00	-	-	45.67	3.50%-4.35%	是
	平安银行对公结构性存款（挂钩利率）开放型人民币产品	固定期限保本浮动收益型	4,500.00	44,100.00	38,000.00	1.69	10,601.69	273.85	3.08%-4.15%	是

中信银行	中信银行共赢利率人民币结构性存款产品	固定期限保本浮动收益型	-	19,950.00	19,950.00	-	-	132.61	3.45%-4.12%	是
合计			8,300.00	69,050.00	61,750.00	1.69	15,601.69	455.67		

2020年1-3月购买理财产品情况：

单位：万元

银行名称	理财产品名称	理财产品类型	期初金额	本期购入	本期赎回	期末未收利息	期末金额	利息收入	年化收益率	是否保本
浦发银行	利多多对公结构性存款（挂钩金融衍生品）	固定期限保本浮动收益型	5,000.00	-	-	46.75	5,046.75	-	3.75%	是
平安银行	平安银行对公结构性存款（挂钩利率）开放型人民币产品	固定期限保本浮动收益型	10,600.00（注）	500.00	10,600.00	0.95	500.95	29.77	3.20%-3.73%	是
中信银行	中信银行共赢利率人民币结构性存款产品	固定期限保本浮动收益型	-	21,400.00	-	179.59	21,579.59	-	3.55%-4.20%	是
合计			15,600.00	21,900.00	10,600.00	227.29	27,127.29	29.77		

注：平安银行对公结构性存款（挂钩利率）开放型人民币产品 2020 年年初余额与 2019 年末的期末余额差异人民币 1.69 万元为截至 2019 年末公司未收的理财产品利息。

二、理财产品的会计处理是否符合相关会计准则的规定

报告期内，发行人购买的理财产品主要类型包括以下三种类型：保证收益型，非固定期限保本浮动收益型，及固定期限保本浮动收益型三类，具体如下：

1、保证收益型（如利多多对公结构性存款公司固定持有期），银行确保本金和约定收益，投资到期日一次性返还产品存款本金和支付约定的产品收益。

2、非固定期限保本浮动收益型（如平安银行卓越计划滚动型人民币公司理财产品），银行对产品的收益不提供保证承诺，发行人可在第二个理财期及之后的每个理财期的产品开放期提出申购与赎回申请，若发行人在产品开放日不提出赎回申请，将视为发行人选择继续理财。

3、固定期限保本浮动收益型（如平安银行对公结构性存款（挂钩利率）、中信银行共赢利率人民币结构性存款以及利多多对公结构性存款（挂钩金融衍生品），银行对产品的收益不提供保证承诺，银行与发行人约定固定投资周期，每个投资周期内，投资者无提前终止该投资周期内结构性存款计划的权利。

财政部于 2017 年颁布了《企业会计准则第 22 号——金融工具确认和计量（修订）》、《企业会计准则第 23 号——金融资产转移（修订）》、《企业会计准则第 24 号——套期会计（修订）》及《企业会计准则第 37 号——金融工具列报（修订）》（统称“新金融工具准则”）。新金融工具准则修订了财政部于 2006 年颁布的《企业会计准则第 22 号——金融工具确认和计量》、《企业会计准则第 23 号——金融资产转移》和《企业会计准则第 24 号——套期保值》以及财政部于 2014 年修订的《企业会计准则第 37 号——金融工具列报》（统称“原金融工具准则”）。

发行人自 2017 年 1 月 1 日至 2018 年 12 月 31 日止执行原金融工具准则，自 2019 年 1 月 1 日起执行新金融工具准则。

根据原金融工具准则：

针对保证收益型理财产品，银行承诺本金和固定收益，表明可回收金额固定或可确定，且该理财产品再活跃市场中没有报价，故发行人将该理财产品归类为贷款和应收款项，发行人按实际利率法计提利息，并确认投资收益。

针对固定期限保本浮动收益型和非固定期限保本浮动收益型理财产品，由于投资收益的浮动性，理财产品不具有未来回收金额固定或可确定的性质，因此不能被划分为贷款及应收款项及持有到期投资，同时发行人购买理财产品的目的不是为了近期内出售，也不能被划分为交易性金融资产，因此该理财产品符合原金融工具准则下可供出售金融资产的规定，故发行人将该理财产品分类为可供出售金融资产核算，在实际收到利息时确认为投资收益。

上述三种理财产品的期限均在一年以内，故发行人将 2017 年末与 2018 年末未赎回的理财产品在资产负债表上列报为“其他流动资产”，理财产品取得的利息收入确认为投资收益符合企业会计准则的相关规定。

根据新金融工具准则：

企业应当根据其管理金融资产的业务模式和金融资产的合同现金流量特征，将金融资产划分为以摊余成本计量的金融资产；以公允价值计量且其变动计入其他综合收益的金融资产；以公允价值计量且其变动计入当期损益的金融资产。

针对上述保证收益型理财产品，银行承诺本金和固定收益，其合同现金流量特征与基本借贷安排一致，即相关金融资产在特定日期产生的合同现金流量仅为对本金和以未偿付本金为基础的利息的支付，同时发行人购买理财产品的目的不是为了近期内出售，而是以收取合同现金流量为目标，因此该理财产品可归类为以摊余成本计量的金融资产，发行人按实际利率法计提利息，在实际收到利息时确认为投资收益。

针对固定期限保本浮动收益型和非固定期限保本浮动收益型理财产品，其合同现金流量特征与基本借贷安排不一致，即相关金融资产在特定日期产生的合同现金流量不是对本金和以未偿付本金为基础的利息的支付，因此不能分类为以摊余成本计量的金融资产，故作为以公允价值计量且其变动计入当期损益的金融资产，在实际收到利息时确认为投资收益，理财产品在到期日前公允价值的变动计入公允价值变动收益。

公司于 2019 年末和 2020 年 1-3 月末持有的理财产品全部为固定期限保本浮动收益型理财产品，期限均在一年以内，故发行人于 2019 年末与 2020 年 1-3 月末将未赎回的理财产品在资产负债表上列报为“交易性金融资产”，符合企业会计准则的相关规定。

三、理财产品与报表科目的勾稽关系

各报告期期末，理财产品同报表科目的勾稽关系列示如下：

单位：万元

报表类型	财务报表科目 / 项目	项目	2020年1-3月	2019年度	2018年度	2017年度
资产负债表	其他流动资产	理财产品余额	-	-	8,300.00	7,500.00
	交易性金融资产		27,127.29	15,601.69	-	-
现金流量表	投资支付的现金	购买理财产品	21,900.00	69,050.00	24,000.00	13,500.00
	收回投资收到的现金	赎回理财产品	10,600.00	61,750.00	23,200.00	6,000.00
	取得投资收益收到的现金	赎回理财产品	31.46	453.98	338.43	131.18
利润表	投资收益	赎回理财产品	29.77	453.98	338.43	131.18
	投资收益	期末未收利息	-	1.69	-	-
	公允价值变动损益		227.29	-	-	-

四、公司购买银行理财产品需要履行的决策程序

根据公司的《海天瑞声对外投资管理办法》和《海天瑞声非日常经营交易事项决策制度》，公司在年度预算中制定投资理财计划及预期收益，若购买理财产品预算金额占公司最近一期经审计总资产的比例未达到 50%的，由总经理办公会议讨论通过后，提交董事会审批决定；超过 50%的，由董事会讨论通过后，提交股东大会审议决定。若会计年度内购买理财产品金额在预算内，在每次购买时无需再次提交董事会审批。财务部根据未来一周、月度、季度的现金流变动情况编制投资理财报告。投资理财报告中包含对理财产品的考察、分析，从而选出最适宜的理财产品，财务人员将编制的投资理财报告交由财务部门经理审批，部门经理审批完成后由财务负责人审批，财务负责人审批通过后方可购买相关理财产品。

五、发行人建立的与货币资金相关的内部控制及执行情况

报告期内，发行人公司建立了《海天瑞声货币资金管理制度》、《海天瑞声对外投资管理办法》、《海天瑞声财务支出管理办法》、《海天瑞声财务报销及借款管理办法》和《海天瑞声第三方电子账户支付管理办法》作为资金管理工作的制度依据，发行人对货币资金相关的内部控制制度如下：

关键控制点	控制活动
职责分工和职权分离制度	发行人实施钱账分管制度，出纳人员不得兼任稽核、会计档案保管和收入、支出、费用、债权债务账目的登记工作，发行人严禁将办理资金支付业务的相关印章和票据集中一人保管。
库存现金管理	发行人建立了健全严密的库存现金内部控制制度，配备专职出纳人员，办理现金收付和结算业务，发行人对库存现金有明确的开支范围规定，对库存现金支付的审批权限有明确的规定，库存现金的管理实现日清月结，出纳人员每日对现金进行盘点，编制盘点表，财务部主管每月

	末对现金盘点进行监盘。
银行账户的开立与取消	发行人将所有银行账户均纳入用友-资金管理平台统一管理，财务部门资金会计定期梳理银行账户，及时根据需要，提出增加或清理账户申请，上报财务负责人审批。
银行存款收支管理	发行人实行严格的资金支付业务审批程序，资金支付的申请需要经过部门负责人和财务经理审批，金额重大的需经财务负责人审批。
银行对账管理	非出纳人员根据银行对账单或银行对账电子文件，与财务系统中的银行日记账相核对，编制上月《银行余额调节表》，《银行余额调节表》由财务经理及时进行复核。对银行存款日记账和银行对账单的差异，非出纳人员分析是否为月末未达事项或记账错误，经财务负责人复核后，对于记账错误，在发现问题当月进行调整。
电商平台企业账户收支管理	发行人因采购需要开通电商账户企业平台，由行政部专员发起申请，电商平台企业账户须经财务部审批后，由财务部在电商平台注册企业账号。行政部采购专员负责企业账户的日常管理与使用，定期更换密码，当电商账户余额不足时，向财务部提出账户充值申请，并经财务部审核。

综上所述，发行人根据《企业内部控制基本规范》的规定，针对货币资金的管理建立了完善的内控制度，实行钱、账分离以及收支两条线的管理制度，严格按照授权审批办理资金业务，确保了公司资金在使用上符合合理性、效率性、安全性的原则，确保为公司发展提供充足的资金支持。

(2) 结合目前大额货币资金的使用计划，说明发行人需要进一步补充流动资金的原因及合理性

一、大额货币资金的使用计划及具体预算

截至2020年3月31日，发行人货币资金金额为5,587.67万元，主要使用计划及预算如下：

单位：万元

用途	金额
购买商品、接受劳务	2,000.00
支付职工薪酬	1,200.00
支付各项税费	800.00
支付其他与经营活动有关	600.00
购建固定资产、无形资产和其他长期资产	100.00
支付其他与筹资活动有关	300.00
合计	5,000.00

此外，截至 2020 年 3 月 31 日，发行人交易性金融资产余额为 27,127.29 万元，主要为发行人购买的理财产品。发行人后续拟相机决策，主要将前述资金运用于长期发展相关的各项支出：1、持续扩大产能，拓展境内外销售渠道，提供更为全面的训练数据定制服务，完善训练数据库产品体系，提升训练数据相关的应用服务市场竞争力；2、在保持基础训练数据领域的优势外，投入专项资金建设行业训练数据研发体系并建立市场销售渠道。

二、发行人需要进一步补充流动资金的原因及合理性

结合上述大额货币资金使用计划情况，发行人的货币资金主要运用于购买商品、接受劳务及支付职工薪酬、税费等各项经营活动相关支出以及长期发展相关的各项支出。本次发行拟募集 10,000 万元补充流动资金，主要原因是随着公司收入增长，预计会产生一定的流动资金投入需求，另外一方面是下游行业快速发展，要求公司持续关注市场需求变化，拓展核心技术优势，预计也会增加流动资金投入需求，具体说明如下：

1、随着收入增长产生的流动资金需求

随着公司未来收入的增长，会产生一定的流动资金投入，具体测算如下：

根据公司历史年度收入增长情况及业务经营特点，结合 2016 年-2019 年公司营业收入的平均增长率情况，预测公司 2020-2022 年的营业收入增长率为 40%。

金额单位：万元

项目	2016 年	2017 年	2018 年	2019 年
营业收入	8,422.86	11,907.09	19,265.77	23,755.81
增长率		41.37%	61.80%	23.31%
平均增长率				42.16%
复合增长率				41.29%
测算用增长率				40.00%

假设公司经营性流动资产、经营性流动负债主要项目与公司的销售收入呈一定比例。

根据上述基本假设，对未来三年公司新增流动资金需求测算如下：

金额单位：万元

项目	2016		2017		2018		2019		平均占比	2020年至2022年预计经营资产及经营负债数额			2022年期末预计数-2019年末实际数
	金额	占比	金额	占比	金额	占比	金额	占比		2020年(预计)	2021年(预计)	2022年(预计)	
营业收入	8,422.86	100.00%	11,907.09	100.00%	19,265.77	100.00%	23,755.81	100.00%	-	33,258.14	46,561.39	65,185.95	41,430.14
应收账款	1,684.24	20.00%	2,515.58	21.13%	2,938.96	15.25%	4,364.55	18.37%	18.69%	6,215.13	8,701.18	12,181.65	7,817.11
预付账款	64.10	0.76%	45.16	0.38%	83.57	0.43%	419.96	1.77%	0.84%	277.86	389.01	544.61	124.65
存货	852.54	10.12%	1,185.65	9.96%	2,063.20	10.71%	754.75	3.18%	8.49%	2,824.08	3,953.71	5,535.19	4,780.45
经营性流动资产合计	2,600.88	30.88%	3,746.39	31.46%	5,085.74	26.40%	5,539.26	23.32%	-	9,317.07	13,043.90	18,261.46	12,722.20
应付账款	43.42	0.52%	40.66	0.34%	193.46	1.00%	359.69	1.51%	0.84%	280.64	392.89	550.05	190.35
预收账款	877.17	10.41%	2,235.95	18.78%	3,407.69	17.69%	2,480.66	10.44%	14.33%	4,766.10	6,672.54	9,341.56	6,860.90
经营性流动负债合计	920.58	10.93%	2,276.61	19.12%	3,601.15	18.69%	2,840.35	11.96%	-	5,046.74	7,065.43	9,891.60	7,051.26
流动资金占用额(经营资产-经营负债)	1,680.30	19.95%	1,469.78	12.34%	1,484.59	7.71%	2,698.91	11.36%	-	4,270.34	5,978.47	8,369.86	5,670.95

注：上表测算不构成对发行人未来业绩的盈利预测。

由上表可知，未来三年发行人因业务规模扩张新增的流动资金需求预计为 5,670.95 万元。

2、下游行业快速发展、人才储备需要带来的流动资金投入需求

除前述业务扩张引致的流动资金投入需求外，下游行业快速发展、人才储备需要等也将带来流动资金投入需求，储备一定规模的流动资金是发行人实施未来发展战略规划的需要，原因及合理性具体如下：

人工智能行业发展迅速，下游客户对训练数据产品及服务的需求可能持续变化升级，发行人需要紧跟行业发展趋势，持续关注市场需求变化，着力挖掘和培养自身核心技术优势并拓展自身产品及服务结构，提升公司核心竞争力，同时，为弥补竞争劣势，公司须加强海外市场拓展，拓展更丰富的营销渠道。因此，公司需要准备充足的流动资金以满足未来可能持续新增的研发和市场拓展投入需求。

发行人业务涉及基础研究、工具平台、训练数据生产多层次核心技术，也需要持续招揽相应行业的优秀人才，为保持有竞争力的薪酬水平以留住现有人才并引进优秀行业人才，发行人需要充足的资金支持。

3、小结

综上，随着公司业务规模的持续扩张，发行人对流动资金的需求预计将持续扩大，公司需补充流动资金，以优化公司资本结构、增强公司资金实力、提升公司的抗风险实力，为公司业务的稳定增长提供财务保障；同时，下游行业的迅速发展、人才储备等都将催生的流动资金需求，公司补充流程流动资金具备必要性和合理性。

通过本次募集资金补充营运资金缺口，一方面将有利于增强公司的营运能力和市场竞争能力，另一方面还将一定程度上改善公司流动性指标，提高公司偿债能力，降低公司财务风险，使公司财务结构更为优化。

中介机构核查事项

针对说明事项（1）

保荐机构、申报会计师执行了下述核查程序：

1、对发行人管理层、财务主管人员进行访谈，获取发行人建立的与货币资金、购买理财产品相关的内部控制制度，对货币资金、购买理财产品相关的流程及其控制活动的情况执行穿行测试；

2、对货币资金、购买理财产品相关的流程中识别的关键控制点执行控制有效性测试，选择适当的样本，检查关键控制点是否得到有效执行；

3、获取发行人报告期内相关理财产品说明书、理财产品协议书、银行回单和相应凭证，核查理财合同金额、期限、预期收益率、风险等级、投资方向等信息与账面记录和披露信息是否相符，检查会计处理是否符合企业会计准则的规定；

4、获取报告期内理财产品购买和赎回交易记录及收益明细，测算理财收益计算是否准确；

5、检查理财产品的相关会计记录，并与财务报表的各个科目进行勾稽核对。

经核查，保荐机构、申报会计师认为：

1、发行人报告期内购买、赎回理财产品的相应的会计处理符合企业会计准则的规定，与报表相关科目勾稽相符；

2、发行人报告期内购买、赎回理财产品履行了必要的决策程序；

3、发行人建立的与货币资金和购买理财产品相关的内控制度健全且有效执行。

针对说明事项（2）

保荐机构访谈了发行人管理层、财务主管人员，取得了发行人关于大额货币资金的使用计划，了解发行人资金投入预期情况，结合发行人财务数据、未来发展计划等对发行人的流动资金缺口进行了测算。

经核查，保荐机构认为：

本次补充流动资金与发行人的收入增长趋势、下游行业发展带来的技术发展需求等相匹配，具备合理性。

19. 关于应收账款

报告期各期末，公司应收账款账面净额分别为 2,515.58 万元、2,938.96 万元和 4,364.55 万元，逐年上升。

请发行人补充披露：预期信用损失率的确认方法。

请发行人说明：（1）信用额度等级的确定标准，客户信用政策的审批制度、审批程序及执行情况，报告期内发行人对主要客户的信用结算政策及变动情况；（2）前五

大应收账款与前五大客户之间的匹配关系和差异原因；（3）期后回款的统计方法，是否存在将预付款项认定为应收账款回款的情形。

请申报会计师进行核查并发表明确意见。

答复：

发行人补充披露事项

（1）预期信用损失率的确认方法

发行人自 2019 年 1 月 1 日开始执行新金融工具准则。新金融工具准则以“预期信用损失”模型替代了原金融工具准则中的“已发生损失”模型。

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/四、报告期内主要采用的会计政策和会计估计/（八）金融工具/2、自 2019 年 1 月 1 日起执行新金融工具准则下的政策”中进行了补充披露：

发行人始终按照相当于整个存续期内预期信用损失的金额计量应收账款的减值准备，并以逾期天数与预期信用损失率对照表为基础计算其预期信用损失。预期信用损失率基于应收账款过去期间的实际信用损失经验计算，并考虑历史数据收集期间的经济状况、当前的经济状况与发行人所认为的预计存续期内的经济状况三者之间的差异。

第一步：确定客户群的构成

根据发行人的历史经验，发行人的客户群体比较稳定，主要为信誉良好的大型科技公司、研究机构等，客户群体发生信用损失的情况没有显著差异，因此在计算减值准备时未进一步区分不同的客户群体。

第二步：确定用于计算历史损失率的历史数据集合

资产负债表日，发行人选取截至资产负债表日为止的最近 3 个会计年度末应收账款账龄数据用以计算最新的历史损失率。对于截至 2019 年 12 月 31 日止的年度期间而言，发行人选择的是 2017 年至 2019 年这段期间。

报告期内，发行人最近三年末应收账款逾期账龄情况如下表所示：

单位：万元

项目	2019 年 12 月 31 日	2018 年 12 月 31 日	2017 年 12 月 31 日
未逾期	3,493.65	2,699.02	1,706.67

逾期1年以内(含1年)	1,107.37	338.59	884.23
逾期1至2年(含2年)	2.44	59.14	116.67
逾期2年以上	56.90	-	-
合计	4,660.36	3,096.75	2,707.57

第三步：计算应收账款迁徙率

迁徙率是指在一个时间段内没有收回而迁徙至下一个时间段的应收账款的比例。

发行人根据客户群最近3年期间的历史逾期情况，计算应收账款迁徙率。

发行人应收账款迁徙率计算过程如下：

单位：万元

项目	2019年12月31日	2018年12月31日	2017年12月31日	年末综合余额	注释	公式	迁徙率
未逾期	3,493.65	2,699.02	1,706.67	7,899.34	A	B/A	29.50%
逾期1年以内(含1年)	1,107.37	338.59	884.23	2,330.20	B	C/B	7.65%
逾期1至2年(含2年)	2.44	59.14	116.67	178.25	C	D/C	31.92%
逾期2年以上	56.90	-	-	56.90	D	D/D	100.00%
合计	4,660.36	3,096.75	2,707.57	10,464.69			

第四步：计算历史损失率，并结合当前状况以及对未来经济状况的预测，确定预期信用损失率

历史损失率代表在历史上每个时间段中最终不能收回的应收账款所占比例。发行根据历史数据经验判断认为逾期2年以上的应收账款确认无法收回，损失率为100%，在此基础上，发行人根据历史期间客户群体应收账款迁徙率计算得到各账龄段的历史损失率。同时管理层在历史损失率的基础上，基于当前经济状况以及对未来经济状况的预测，对计算出的历史损失率做出调整，确定各账龄段的预期信用损失率。发行人将预期信用损失率乘以应收账款余额来建立准备矩阵。

发行人预期信用损失率的计算过程如下：

单位：万元

项目	注释	迁徙率	历史损失率 计算过程	历史损失率	调整后的预期 信用损失率	同行业上市公司预 期信用损失率区间 (注)
未逾期	A	29.50%	A*B*C*D	0.72%	3.00%	0.01%-7.00%
逾期1年以内(含1年)	B	7.65%	B*C*D	2.44%	12.00%	1.71%-50.00%

逾期1至2年 (含2年)	C	31.92%	C*D	31.92%	50.00%	5.00%-100.00%
逾期2年以上	D	100.00%	D	100.00%	100.00%	5.01%-100.00%

注：同行业上市公司包括，包括荣科科技、彩讯股份、天玑科技、四方精创、恒生电子、用友网络、万兴科技、及泛微网络；对于上述同行业公司按照应收账款账龄划分预期信用损失率的情形，此处将应收账款账龄区间为1年以内的预期信用损失率作为未逾期和逾期1年以内（含1年）的预期信用损失率，将应收账款账龄区间为1-2年的预期信用损失率作为此处逾期1至2年（含2年）的预期信用损失率，将应收账款账龄区间为2年以上的预期信用损失率作为逾期2年以上的预期信用损失率。

发行人运用迁徙法计算历史损失率，基于当前经济状况以及对未来经济状况的预测，同时对比了同行业可比公司信息后，对计算出的历史损失率做出调整，并基于谨慎性原则采用调整后的预期信用损失率计算应收账款坏账准备。

发行人于每年末根据更新的信息重新计算应收账款的预期信用损失率；如果用于计算预期信用损失率的信息在年内出现重大变化，发行人会重新评估预期信用损失率，必要时进行调整以准确反应应收账款的预期信用损失。

发行人说明事项

(1) 信用额度等级的确定标准，客户信用政策的审批制度、审批程序及执行情况，报告期内发行人对主要客户的信用结算政策及变动情况

一、信用额度等级的确定标准

发行人获取客户信用资料，对客户进行信用调查，进行信用等级评定，根据客户的信用等级制定不同的授信策略，具体如下：

信用额度等级	信用额度	信用期限	抵押/担保
特级	1000万	不超过90个工作日	无
一级	500万-1000万	不超过90天	无
二级	200万-500万	不超过60天	无
三级	0万-200万	不超过45天	无
四级	现销，如赊销则需提供与信用额度相同的抵押、担保		

所有拟与发行人发生赊销交易的客户均需要进行信用等级评定。客户信用等级分为5个等级：特级、一级、二级、三级、四级。

(1) 特级为优秀客户，指规模较大，信誉良好，经营情况和财务状况均较好的客户；(2) 一级为良好客户，指有一定规模，与公司有长期业务往来，信誉较好的客户；(3) 二级为中等客户，指付款良好，经济实力较好，信誉较好，但与公司业务往来不

多的客户；（4）三级为一般客户，指经营情况和付款情况一般，与公司有一定业务往来的客户；（5）四级为较差客户，指有不良付款记录、信誉较差的客户。

二、客户信用政策的审批制度、审批程序及执行情况

根据发行人制定的《销售信用管理规定》，销售部负责赊销活动的对外沟通及资料收集，并负责主导信用申请流程及信用档案的管理。财务部负责对信用申请的审核与建议、信用执行的日常监督管理。财务总监负责审核一般的信用额度申请/变更、信用解冻申请和信用冻结审批；对于额度超过 1,000 万、账期超过 90 个工作日的信用额度申请，在财务总监审核批准后仍需由总经理办公会终批，另外总经理办公会也负责对特殊信用额度申请/变更进行审批。

所有客户的信用额度审批有效期均为一年，一年之后由销售部重新评估和申请授予客户的信用额度，由财务部对客户进行信用申请的审定与审核；一年之内如果客户发生重大变化，可以启动临时审定。

发行人在报告期内建立了完善的客户信用政策的审批制度和审批程序，并严格执行了相关制度规定。

三、报告期内发行人对主要客户的信用结算政策及变动情况

报告期内，发行人主要客户的信用政策与上述总体信用政策一致，信用额度在 200 万到 1,000 万之间，信用期限在 45 天到 90 个工作日之间，而且在报告期内未发生重大变动。

（2）前五大应收账款与前五大客户之间的匹配关系和差异原因

报告期内，发行人按照品牌统计的应收账款前五大客户及销售收入前五大客户情况具体如下：

单位：万元

2020 年 1-3 月			
应收账款前五大客户品牌	期末余额	销售收入前五大客户品牌	本期发生额
Google	609.99	某国际消费电子产品厂商	611.65
某国际消费电子产品厂商	511.54	Google	547.75
阿里巴巴	430.15	客户 1	272.22
腾讯	357.75	某大学	251.57

中国某大型科技公司	219.35	腾讯	236.71
合计	2,128.78	合计	1,919.90
2019 年			
应收账款前五大客户品牌	年末余额	销售收入前五大客户品牌	本年发生额
中国某大型科技公司	711.73	三星	4,762.06
腾讯	424.47	微软	2,876.20
阿里巴巴	423.70	阿里巴巴	2,208.26
微软	380.13	亚马逊	1,492.98
亚马逊	290.52	腾讯	1,459.31
合计	2,230.54	合计	12,798.80
2018 年			
应收账款前五大客户品牌	年末余额	销售收入前五大客户品牌	本年发生额
阿里巴巴	1,261.62	阿里巴巴	5,179.07
腾讯	454.27	三星	2,436.81
亚马逊	351.94	腾讯	1,574.16
百度	289.80	微软	1,299.73
Info-communications Media Development Authority (“IMDA”)	108.90	百度	888.29
合计	2,466.53	合计	11,378.06
2017 年			
应收账款前五大客户品牌	年末余额	销售收入前五大客户品牌	本年发生额
腾讯	836.63	腾讯	1,770.05
阿里巴巴	326.89	三星	1,553.08
乐乐创新	209.88	阿里巴巴	1,132.01
百度	202.76	微软	1,058.87
富士康	191.52	百度	585.29
合计	1,767.69	合计	6,099.29

报告期内，发行人应收账款与销售收入的前五大客户较为稳定，主要为人工智能产业链上的各类知名公司与机构，如阿里巴巴、腾讯、百度、微软、三星、亚马逊等。其中 2017 年，腾讯、阿里巴巴、百度既是应收账款前五大客户，又是销售收入前五大客户；2018 年度，阿里巴巴、腾讯和百度既是应收账款前五大客户，又是销售收入前五大客户；2019 年度，微软、阿里巴巴、腾讯和亚马逊既是应收账款前五大客户，又是

销售收入前五大客户；2020 年一季度，某国际消费电子产品厂商、Google 和腾讯既是应收账款前五大客户，又是销售收入前五大客户。

应收账款前五大客户与销售收入前五大客户之间存在差异的情况及原因分析：

1、三星和微软通常采取提前预付款项的结算模式，因此与其的往来款余额通常体现为发行人的预收账款余额，因此也导致在 2017 年、2018 年和 2019 年，三星和微软虽然是发行人销售收入的前五大客户，但并不一定是应收账款前五大客户。

2、乐乐创新不在 2017 年度销售收入前五大客户之列，但在 2017 年末是应收账款前五大客户，主要原因是由于发行人在 2016 年向乐乐创新销售训练数据产品，产生应收款 209.88 万元截至 2017 年末仍未结算，因此 2017 年末应收账款余额虽然较高，但是收入金额并不在 2017 年反映。富士康不在 2017 年销售收入前五大客户之列，但在 2017 年末是应收账款前五大客户，应收账款余额为 191.52 万元。2017 年与富士康的收入交易额为 233.85 万元，由于年末交付验收的收入尚未到付款结算时点，因此 2017 年末余额较高。

3、亚马逊与 IMDA 均为 2018 年度新发展的客户，不在 2018 年度销售收入前五大客户之列，但在 2018 年末进入应收账款前五大客户，原因主要是由于其项目均于 2018 年 12 月结束，并进行交付验收和确认收入，但截至 2018 年末尚未到付款结算时点，因此形成较大金额的应收账款。

4、中国某大型科技公司在 2019 年是应收账款前五大客户，但不是销售收入前五大客户。虽然中国某大型科技公司在 2019 年没有进入销售收入前五大客户，但仍然是销售收入的前十大客户，全年产生销售收入 1,019.94 万元，销售额较高，而且由于其信用期较长，收入产生的应收账款于 2019 年末尚未到付款结算时点，因此 2019 年末应收账款余额较高。

5、2020 年第一季度受新冠肺炎疫情影响，发行人在境内的业务受到一定的冲击，但由于新冠疫情对国际市场的影响相对滞后且发行人积极开拓海外市场，因此 2020 年第一季度发行人的销售收入前五大客户中增加了某大学和客户 1 两家海外公司，但这两家公司的应收账款均不在前五大之列，其中客户 1 期末应收账款为 184.26 万元，为 2020 年一季度末应收账款第六大品牌客户，某大学则是由于这类新客户的信用期较短，因此其交付验收的项目于 2020 年一季度末已达到付款结算时点，并且发行人已经收到回款。

同时阿里巴巴及中国某大型科技公司在 2020 年一季度末是应收账款前五大客户，但是由于应收账款余额中有部分对应的收入发生在 2019 年，因此不在 2020 年一季度销售收入前五大客户之列。

(3) 期后回款的统计方法，是否存在将预付款项认定为应收账款回款的情形

发行人期后回款的统计主要是依据发票信息进行核对和对应，发票可以直接核对至单个合同，避免了同一客户的不同合同的回款相互间混淆的情况。对于每个合同而言，发行人确认收入前收取的款项对应为客户预付款项，确认收入后收到的款项对应为客户应收账款的回款，因此能够避免将客户的预付款项认定为应收账款回款的情形。

在实际操作中，发行人的财务人员在单个合同下开具发票，并且对发票信息、客户名称、合同编号等信息进行核对并登记。开具发票后有专门的商务专员负责跟进，并积极与客户进行沟通。在收到应收账款的期后回款后，财务人员会与商务专员沟通并确认回款金额对应的发票信息、合同信息，并根据发票对应的合同信息记录入账。通过上述安排，发行人根据发票来进行期后回款的统计，避免了将预付款项认定为应收账款回款的情形。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

1、了解发行人销售管理相关的内部控制，关注主要客户的信用情况，应收账款账龄表以及逾期账龄表的编制及审核等，并测试关键财务报告内部控制的运行有效性；

2、取得发行人报告期各期末应收账款账龄明细表、逾期明细表、应收账款坏账准备计算表，通过分析实际坏账损失情况，评估预期损失模型相关参数、方法、假设的合理性；同时比较新旧金融工具准则下计提的应收账款坏账准备金额是否存在显著差异，比较与同行业公司应收账款坏账计提政策是否存在重大差异；

3、向主要客户函证应收账款余额，并对主要客户进行实地走访，回函及走访结果未见重大异常情形；

4、针对应收账款的期后回款的核查，申报会计师检查了销售合同编号与开具发票的对应关系、客户回款金额与发票的对应关系、应收账款与销售合同的对应关系。

(2) 核查结论

经核查，申报会计师认为：

1、发行人在新金融工具准则下，根据预期信用损失模型，应收账款的坏账准备计提准确、充分，相关会计处理符合《企业会计准则》的要求；

2、发行人所说明的主要客户信用政策、执行情况、信用期变更等信息与申报会计师在审计过程中了解到的情况一致；报告期内发行人对主要客户的信用结算政策与总体信用政策一致，报告期内未发生重大变动；

3、发行人期后回款逐一匹配至相应的应收账款，不存在将预付款项认定为应收账款回款的情形。

20. 关于存货

报告期各期末，发行人存货的账面价值分别为 1,185.65 万元、2,063.20 万元和 754.75 万元，为归集正在履行尚未完成验收的训练数据定制服务和数据资源相关的应用服务所发生的各项支出。

请发行人补充披露：各期末未完工项目的类别、数量、金额及对应的合同金额。

请发行人说明：（1）各期末未完工的项目数量，未完工服务成本的具体构成及其后结转成本确认收入的情况，结合合同约定和具体项目实施周期，说明是否存在长期未结转或期后提前终止的情况及原因，存货跌价准备计提是否充分；（2）报告期各期末主要未完工项目的客户名称、项目内容、成本结构、合同金额和预计总成本，分析各项目成本结构的差异原因，说明预计总成本的确认方法，与实际完工成本之间是否存在差异，是否存在应计提而未计提存货跌价准备的情形。

请申报会计师对上述事项进行核查并发表明确意见并说明对存货存在、计价和分摊等相关认定履行的审计程序。

答复：

发行人补充披露事项

各期末未完工项目的类别、数量、金额及对应的合同金额

以下内容已在招股说明书“第八节 财务会计信息与管理层分析/十、资产质量分析 /（一）流动资产分析/6、存货”中进行了补充披露：

单位：万元

类型	领域	2020年3月31日				
		项目数量	存货账面余额	存货跌价准备	存货账面净值	对应的合同金额
训练数据定制服务	智能语音	194	997.56	-8.72	988.84	5,167.19
	计算机视觉	44	57.15	-2.81	54.34	945.61
	自然语言	77	71.00	-	71.00	405.22
训练数据相关的应用服务		8	131.82	-1.90	129.91	1,539.77
合计		323	1,257.53	-13.44	1,244.08	8,057.78
类型	领域	2019年12月31日				
		项目数量	存货账面余额	存货跌价准备	存货账面净值	对应的合同金额
训练数据定制服务	智能语音	94	537.45	-24.35	513.10	3,729.47
	计算机视觉	20	50.43	-	50.43	763.46
	自然语言	32	83.95	-	83.95	319.55
训练数据相关的应用服务		5	109.17	-1.90	107.27	1,359.82
合计		151	780.99	-26.25	754.74	6,172.30
类型	领域	2018年12月31日				
		项目数量	存货账面余额	存货跌价准备	存货账面价值	对应的合同金额
训练数据定制服务	智能语音	141	1,653.62	-118.31	1,535.31	8,511.57
	计算机视觉	27	469.12	-5.37	463.75	1,251.35
	自然语言	46	210.01	-167.13	42.88	304.77
训练数据相关的应用服务		10	43.88	-22.63	21.25	1,479.49
合计		224	2,376.64	-313.43	2,063.21	11,547.19
类型	领域	2017年12月31日				
		项目数量	存货账面余额	存货跌价准备	存货账面价值	对应的合同金额
训练数据定制服务	智能语音	102	1,107.40	-29.41	1,077.99	5,745.95
	计算机视觉	18	35.17	-	35.17	166.63
	自然语言	50	68.39	-0.75	67.64	311.69
训练数据相关的应用服务		8	4.85	-	4.85	62.79

合计	178	1,215.81	-30.16	1,185.65	6,287.06
----	-----	----------	--------	----------	----------

发行人说明事项

(1) 各期末未完工的项目数量，未完工服务成本的具体构成及其后结转成本确认收入的情况，结合合同约定和具体项目实施周期，说明是否存在长期未结转或期后提前终止的情况及原因，存货跌价准备计提是否充分

一、各期末未完工的项目数量，未完工服务成本的具体构成及期后结转成本确认收入的情况

发行人各期末的未完工服务成本涉及训练数据定制服务和训练数据相关的应用服务两种业务类型。截至2017年12月31日、2018年12月31日、2019年12月31日及2020年3月31日，发行人未完工的项目数量分别为178个，224个、151个及323个，未完工项目成本的构成以及期后（截至2020年6月30日）结转成本、确认收入的情况详见下表：

单位：万元

类型	领域	2020年3月31日							期后结转成本，确认收入的项目		
		项目数量	数据服务费	职工薪酬	折旧摊销	差旅费	其他	未完工服务成本小计	项目数量	期后结转成本	期后确认收入
训练数据定制服务	智能语音	194	826.72	88.49	8.13	10.30	63.92	997.56	157	1,288.23	2,103.89
	计算机视觉	44	34.20	15.44	0.92	0.01	6.58	57.15	37	92.89	125.46
	自然语言	77	63.14	5.75	0.02	-	2.09	71.00	59	124.47	209.92
训练数据相关的应用服务		8	34.35	64.91	8.44	12.99	11.13	131.82	5	146	225.29
合计		323	958.42	174.58	17.51	23.30	83.71	1,257.53	258	1,651.59	2,664.56

注：2020年3月31日的未完工项目截至2020年6月30日尚有65个项目未结转成本和确认收入，项目均在执行周期内。

单位：万元

类型	领域	2019年12月31日							期后结转成本，确认收入的项目		
		项目数量	数据服务费	职工薪酬	折旧摊销	差旅费	其他	未完工服务成本小计	项目数量	期后结转成本	期后确认收入
训练数据定制服务	智能语音	94	371.24	55.43	6.15	10.28	94.35	537.45	77	1,011.78	1,613.48
	计算机视觉	20	31.19	13.13	0.96	1.64	3.52	50.43	15	243.85	573.15
	自然语言	32	63.60	16.49	0.07	0.12	3.66	83.95	29	128.98	241.86
训练数据相关的应用服务		5	32.89	46.85	15.00	12.67	1.76	109.17	2	57.32	66.04
合计		151	498.93	131.90	22.18	24.70	103.29	780.99	123	1,441.93	2,494.53

注：2019年12月31日的未完工项目截至2020年6月30日尚有28个项目未结转成本和确认收入，项目均在执行周期内。

单位：万元

类型	领域	2018年12月31日							期后结转成本，确认收入的项目		
----	----	-------------	--	--	--	--	--	--	----------------	--	--

		项目数量	数据服务费	职工薪酬	折旧摊销	差旅费	其他	未完工服务成本小计	项目数量	期后结转成本	期后确认收入
训练数据定制服务	智能语音	141	1,254.42	220.88	4.98	12.47	160.88	1,653.62	138	3,018.73	6,953.39
	计算机视觉	27	366.23	50.90	8.81	10.67	32.51	469.12	27	652.65	1,126.72
	自然语言	46	196.57	10.48	0.00	1.21	1.75	210.01	37	79.99	165.90
训练数据相关的应用服务		10	9.84	28.73	0.54	2.01	2.76	43.88	5	23.30	37.97
合计		224	1,827.06	311.00	14.32	26.36	197.90	2,376.64	207	3,774.67	8,283.98

注：2018年12月31日的未完工项目截至2020年6月30日尚有17个项目未结转成本和确认收入，其中10个项目（智能语音领域2个项目，自然语言领域6个项目，训练数据相关的应用服务2个项目）为日本纽康株式会社的项目，由于项目难度高于预期，双方已于2019年取消该合同；4个项目（智能语音领域1个项目，自然语言领域3个项目）为淘宝（中国）软件有限公司的项目，由于发音人身体情况不佳，项目暂停，导致项目周期延长，截至2020年6月30日尚未完工；其余3个项目均为训练数据相关的应用服务，项目执行周期较长，截至2020年6月30日尚未完工。

单位：万元

类型	领域	2017年12月31日							期后结转成本，确认收入的项目		
		项目数量	数据服务费	职工薪酬	折旧摊销	差旅费	其他	未完工服务成本小计	项目数量	期后结转成本	期后确认收入
训练数据定制服务	智能语音	102	767.29	214.31	1.80	12.09	111.91	1,107.40	101	2,388.05	4,991.88
	计算机视觉	18	22.90	11.80	-	-	0.48	35.17	18	82.47	157.43
	自然语言	50	40.25	25.00	1.20	-	1.95	68.39	44	115.87	176.27
训练数据相关的应用服务		8	0.51	4.33	-	-	-	4.85	6	5.34	21.6
合计		178	830.95	255.45	3.00	12.09	114.33	1,215.81	169	2,591.73	5,347.18

注：2017年12月31日的未完工项目截至2020年6月30日尚有9个项目未结转成本和确认收入，均为日本纽康株式会社的项目，由于项目难度高于预期，双方已于2019年取消该合同。

二、结合合同约定和具体项目实施周期，说明是否存在长期未结转或期后提前终止的情况及原因，存货跌价准备计提是否充分

发行人从项目最初建立到最终验收，根据业务类别、项目内容和规模等因素，开发周期也有所不同。发行人训练数据定制服务和训练数据相关的应用服务项目开发周期通常不超过 1 年，部分项目比较复杂，项目难度较高，因此部分项目周期可达 1-2 年左右。按照项目所处的服务类型和业务领域区分，各类项目合同约定以及一般的开发周期如下：

服务类型	业务领域	合同关于项目实施周期的约定	一般项目实施周期
训练数据定制服务	智能语音	合同中直接对项目实施周期和里程碑日期做出约定，对于由于客户或被采集人等非发行人主观原因造成的延误，发行人可以免责顺延工期。	3 个月至 1 年（部分长周期、采集难度高以及标注体系复杂项目开发周期为 1-2 年左右）
	计算机视觉		1 至 6 个月
	自然语言		1 周至 6 个月
训练数据相关的应用服务	--		1 个月至 1 年（部分长周期、项目难度高的项目开发周期为 1-2 年左右）

根据上述一般项目实施周期，结合期末未完工项目数量与期后结转成本项目数量的对比情况可以看出，报告期末，公司各类未完工项目均处于正常开发周期内，不存在长期未结转的项目。

报告期内，公司提前终止的主要项目情况及原因列示如下：

单位：万元

合同	项目数量	客户名称	取消年份	合同业务	业务类型	项目支出金额	报告期内存货跌价准备	取消原因
合同 1	1	日本纽康株式会社	2019 年	法语女声智能语音定制数据库开发	训练数据定制服务-智能语音	85.33	85.33	客户在项目实施中后期对口内噪音指标、环境噪音标准、发音标准度等临时提出更高要求，而发音人因身体情况不佳、个人其他工

合同 2	1			美国英语男声智能语音定制数据库开发		89.03	89.03	作安排等，使得无法短时间内满足客户需求，因此取消该项目。
合同 3	2			法语及西班牙语文本正则化处理	训练数据相关的应用服务	20.91	20.91	首次处理该两种语言的文本正则，且客户采用高于商用标准的高难度测试集对效果进行测试，导致测试通过效果不理想，因此取消该项目。
合同 4	6			日本专有名词对泰语、越南语、印尼语、缅甸语、西班牙语、法语机器翻译及标注	训练数据定制服务-自然语言	165.15	165.15	试验性尝试小语种目标语言与非中英源语言搭配的特殊领域名词的翻译及标注项目，但拥有该领域能力的标注资源稀缺性超出预期，进而影响了工期。此外，在项目实施后期，客户临时提出了更有挑战性的超常规规则 and 标准，因此取消该项目。

上述项目自 2017 年开始启动，2019 年取消，2017 年发生支出人民币 6.31 万元，2018 年发生支出人民币 302.43 万元，2019 年发生支出人民币 46.76 万元。由于项目难度高于预期，发行人判断该合同预计能够全部收回收入的可能性较低，因此发行人预计了未来可能收回的收入金额，扣除必要支出后，同存货的账面余额进行了比对，将低于存货账面余额的部分计提存货跌价准备。发行人于 2017 年和 2018 年计提存货跌价准备人民币 289.78 万元，2019 年计提存货跌价准备人民币 65.72 万元。截至 2019 年 12 月 31 日，上述项目累计发生支出人民币 355.50 万元，累计计提存货跌价准备人民币 355.50 万元，因此发行人在报告期内已对该类提前终止的项目计提了充分的存货跌价准备。

(2) 报告期各期末主要未完工项目的客户名称、项目内容、成本结构、合同金额和预计总成本，分析各项目成本结构的差异原因，说明预计总成本的确认方法，与实际完工成本之间是否存在差异，是否存在应计提而未计提存货跌价准备的情形

1、报告期各期末前十大未完工项目的客户名称、项目内容以及成本结构如下：

单位：万元

序号	客户名称	项目内容	存货余额	成本结构									
				数据服务费		职工薪酬		折旧摊销		差旅费		其他	
				金额	占比	金额	占比	金额	占比	金额	占比	金额	占比
2020年1-3月	百度、阿里巴巴、微软等	语音识别转写、语音合成等	503.23	400.5	79.59%	76.88	15.28%	0.32	0.06%	9.94	1.98%	15.57	3.09%
2019年	阿里巴巴、百度、微软	语音识别转写、语音合成等	380.74	297.4	78.11%	69.97	18.38%	0.72	0.19%	12.64	3.32%	-	0.00%
2018年	某国际消费电子产品厂商、三星等	语音识别采集、语音合成、手写体采集等	1,023.51	753.19	73.59%	103.89	10.15%	11.36	1.11%	19.35	1.89%	135.73	13.26%
2017年	三星、阿里巴巴、腾讯	语音识别采集、语音合成	496.27	324.26	65.34%	102.54	20.66%	0.70	0.14%	2.20	0.44%	66.57	13.41%

2017年末，发行人前十大未完工项目的成本结构中数据服务费占比较低，职工薪酬和其他支出占比较高。数据服务费占比较低而职工薪酬占比较高主要是由于当年末前十大未完工项目中的部分智能语音项目尚未开始密集进行原料数据的采集和标注工作，而人工投入则在项目立项后持续发生，因此相对降低了数据服务费占比，提升了职工薪酬占比。其他支出占比较高主要是由于：（1）个别项目对录制环境有较高的要求，需要安静的环境和包含家居设备的真实家居场地，并需要在以下三个房间进行录音：客厅、厨房、卧室，因此导致发生的项目租赁支出较高，（2）个别项目录音需要租用客户指定中高端车型，包括奔驰、帕萨特、途观等，因此导致发生的项目租赁支出较高。

2018年末，发行人前十大未完工项目的成本结构中其他支出占比较高，主要是由于个别项目为国外采集项目，需要在当地租赁专业录音棚，因此导致发生的项目租赁支出较高。

2019年末和2020年3月末，发行人前十大未完工项目的成本结构与发行人整体营业成本结构基本一致。

2、报告期各期末前十大未完工项目的合同金额、预计总成本、实际完工成本及其比较差异如下：

单位：万元

期间	合同金额	预计总成本	截至目前已完工项目		预计总成本与实际完工成本的差异	
			预计总成本	实际完工成本	金额	比例
2020年1-3月	1,038.34	657.28	286.72	275.16	11.55	4.03%
2019年	755.42	449.31	206.3	191.76	14.54	7.05%
2018年	3,313.51	1,539.68	1,539.68	1,403.06	136.60	8.87%
2017年	1,444.50	693.52	693.52	701.88	-8.38	-1.21%

报告期各期末，发行人前十大未完工项目中，截至目前已完工项目的预计总成本与实际完工成本的差异较小，不存在显著异常。

发行人预计总成本是根据项目预算及项目执行情况调整后综合判断得出，判断过程由项目负责人与财务部共同进行，经其签字确认后作为会计档案保存。其中项目预算纳入公司整体预算管理体系，由项目负责人按照项目具体内容，对项目所需人员及每人投入工时等直接人工内容，采集人数、采集单价、标注工时、标注单价等数据服务采购内容，折旧与摊销、租赁等项目所需场地、设备内容，差旅、交通、办公等其他项目相关支出内容制定详细预算。项目预算需要经过业务部门、财务部等多层级审批通过后方可执行。

3、发行人计提存货跌价准备的情况

报告期各期末，发行人已经充分考虑了存货中归集的未完工项目是否存在跌价准备的情形，并按照企业会计准则的要求，计算存货中未完工项目的可变现净值，对于可变现净值低于成本的金额计提存货跌价准备。计算可变现净值的方法是以合同价格为基础，同时考虑预计成本以及销售费用和相关税费的影响。根据上述方法，发行人于各报告期分别计提存货跌价准备人民币 30.16 万元、284.03 万元、104.05 万元、11.54 万元，各报告期期末的存货的预计总成本金额较实际成本金额不存在重大差异，各报告期期末存货的账面价值能够合理反映存货的可变现净值，因此存货跌价准备计提充分，不存在应计提而未计提存货跌价准备的情形。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

1、了解及评估存货相关的内部控制的设计和执行情况、发行人存货归集及核算原则是否适当；

2、获取报告期各期末未完工项目明细，与存货的账面余额进行核对，并分析期末未完工项目的成本构成是否合理；

3、取得报告期各期末未完工项目期后结转收入确认成本的情况，检查是否存在长期未结转或提前终止的情况，通过检查相关销售合同，与项目人员进行访谈，了解项目运行情况，评价是否存在未充分计提存货跌价准备的情况；

4、取得管理层编制的存货跌价准备计算表，复核存货减值测试的过程，检查管理层在存货减值测试中使用的预计总成本，预计总收入金额的支持性依据，评估其合理性。

(2) 核查结论

经核查，申报会计师认为：

1、发行人补充的披露各期末未完工项目的类别、规模、数量、金额及对应的合同金额等信息符合发行人实际情况；

2、报告期各期末未完工的项目数量、未完工服务成本的具体构成及期后结转成本确认收入情况真实、准确；报告期内不存在长期未结转的项目；期后提前终止的项目具备合理原因；不存在应计提而未计提存货跌价准备的情形；

3、报告期各期末主要未完工项目的客户名称、项目内容、成本结构、合同金额和预计总成本等信息符合发行人实际情况；

4、报告期各期末主要未完工项目间的成本结构差异具备合理原因，主要未完工项目的预计总成本的确认方法合理，预计总成本与实际完工成本之间的差异原因不存在异常，主要未完工项目的存货跌价准备计提充分，不存在应计提而未计提存货跌价准备的情形。

申报会计师说明事项

针对存货存在、计价和分摊等相关认定，申报会计师执行的审计程序如下：

1、了解及评估与存货相关的内部控制的设计及执行有效性、发行人存货归集及核算原则是否符合会计准则规定；

2、获取并检查成本倒轧表，分析其合理性；

3、对制造费用执行实质性测试，选取样本，对数据服务费支出检查支持性依据，对人工成本、折旧摊销在项目之间的分摊进行合理性测试；

4、获取存货余额的项目明细表，同项目预算进行比较，分析存货构成的合理性；

5、分析存货账龄，对长账龄存货通过访谈及查看相关销售合同，分析其项目周期是否合理，评估是否存在减值；

6、对存货余额中的未完工项目，选取样本，进行存货监盘和函证，检查项目的状态，确认存货的真实性；

7、取得管理层编制的存货跌价准备计算表，复核存货减值测试的过程，检查管理层在存货减值测试中使用的预计总成本，预计总收入金额的支持性依据，评估其合理性；

8、取得存货期后实际对外销售时确认的收入金额和结转的成本金额，同存货跌价准备计算表中的预计收入金额和预计总成本进行比对，检查存货跌价准备的计提是否充分。

21. 关于预收账款

报告期各期末，公司预收款项的金额分别为 2,235.95 万元、3,407.69 万元和 2,480.66 万元，大幅波动。

请发行人说明：（1）结合在手订单和项目执行进度，分析预收款项变动原因以及期后收入的实现情况；（2）超过合同执行周期的预收账款的情况，合同未履行完毕的原因。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人说明事项

（1）结合在手订单和项目执行进度，分析预收款项变动原因以及期后收入的实现情况

单位：万元

年份/期间	预收款项/合同负债对应在手订单总额	增长率	预收款项/合同负债年/期末余额	增长率	期后确认收入金额
2020年1-3月	4,840.32	43%	3,268.86	32%	1,575.54
2019年度	3,374.52	-28%	2,480.66	-27%	864.37
2018年度	4,669.20	18%	3,407.69	52%	3,407.69
2017年度	3,952.94	-	2,235.95	-	2,235.95

注：期后确认收入截止日期为 2020 年 6 月 30 日。

发行人前五大预收款项/合同负债的具体构成情况如下：

单位：万元

客户名称	2020年3月31日 余额	占预收款项/合同负债期末余额 合计数的比例
微软（中国）有限公司	1,925.51	58.90%
NTT DOCOMO, INC.	242.50	7.42%
Amazon Web Services, Inc.	189.79	5.81%
Tegara Corporation	145.13	4.44%
北京三星通信技术研究有限公司	129.10	3.95%
合计	2,632.03	80.52%

单位：万元

客户名称	2019年12月31日 余额	占预收款项/合同负债年末余额 合计数的比例
微软（中国）有限公司	1,990.59	80.24%
北京大米未来科技有限公司	106.13	4.28%
北京三星通信技术研究有限公司	100.67	4.06%
客户 2	47.03	1.90%
国家计算机网络与信息安全管理中心	44.55	1.80%
合计	2,288.97	92.27%

单位：万元

客户名称	2018年12月31日 余额	占预收款项/合同负债年末余额 合计数的比例
微软（中国）有限公司	1,429.38	41.95%
SAMSUNG ELECTRONICS CO.,LTD.	707.82	20.77%
北京三星通信技术研究有限公司	480.59	14.10%
SAMSUNG R&D INSTITUTE INDIA—BANGALORE PRIVATE LIMITED	163.36	4.79%
SOUNDHOUND INC	131.54	3.86%
合计	2,912.69	85.47%

单位：万元

客户名称	2017年12月31日 余额	占预收款项/合同负债年末余额 合计数的比例
北京三星通信技术研究有限公司	655.88	29.33%
SAMSUNG ELECTRONICS CO.,LTD.	611.99	27.37%
微软（中国）有限公司	434.82	19.45%
杭州海康威视数字技术股份有限公司	57.50	2.57%

客户名称	2017年12月31日 余额	占预收款项/合同负债年末余额 合计数的比例
北京羽扇智信息科技有限公司	55.11	2.47%
合计	1,815.30	81.19%

发行人按照合同对预收款项或者应收账款进行核算，对于每个合同而言，发行人确认收入前收取的款项对应为该合同的预收款项，确认收入后收到的款项对应为该合同应收账款的回款，因此能够避免出现预收款项与应收账款混同的情形，此外在每个合同预收款项或应收账款核算中，发行人亦不会进行抵消处理。根据上表所示，2018年末预收款项余额较上年末有较大增长，同时在手订单金额也有所增加。预收款项的上涨主要是微软和三星两个客户产生的，这两个客户的结算模式以预付为主，2018年这两个客户均与发行人签订大额销售合同，根据合同约定采取向发行人提前预付账款的形式进行支付，且预付比例较高，因此2018年末的预收款项余额有所上升，增幅高于对应的在手订单金额；2019年末预收款项余额有所下降，同时对应的在手订单金额也基本同比例降低，主要是由于三星在2019年对项目进行了集中验收，使得其预收款项确认为收入，在手订单和预收账款余额均有较大的下降；2020年3月末预收款项余额较大增长，同时对应的在手订单金额也基本同比例增长，主要是由于发行人2020年积极开拓海外市场，多个海外客户订单规模较上年度有明显增长，使得预收款项余额也有所增长。

发行人预收款项通常在1年以内结转确认收入。截至2020年6月30日，2017年末及2018年末的预收款项在期后已全部确认收入，2019年末及2020年3月末的预收款项余额在期后确认收入的比例也达到了35%和48%。

(2) 超过合同执行周期的预收账款的情况，合同未履行完毕的原因。

发行人合同执行周期通常不超过1年，部分长周期、标注体系复杂的智能语音类项目周期为1-2年左右。报告期内无超过合同执行周期的预收账款的情况。

申报会计师核查事项

(1) 核查程序

申报会计师执行的主要程序如下：

1、获取了预收账款明细表，选取样本对期末预收账款实施了函证程序并查阅收款的银行回单，核对是否相符；

2、获取了期末预收款项对应的销售合同，检查与发行人披露的在手订单金额是否一致；

3、获取预收账款的账龄明细，检查长账龄预收账款及其合同，询问管理层合同未履行完毕的原因，并判断其合理性。

(2) 核查结论

经核查，申报会计师认为：

1、发行人报告期各期预收款项余额的变动情况与发行人订单承接及项目执行情况相匹配，期后收入实现情况不存在异常；

2、发行人报告期内不存在超过合同执行周期的预收账款的情况。

22. 关于现金流量表

报告期内，公司销售商品、提供劳务收到的现金占营业收入的比例分别为 114.29%、111.05%和 96.74%。

请发行人说明：“销售商品、提供劳务收到的现金”“购买商品、接受劳务支付的现金”“支付给职工以及为职工支付的现金”“支付的其他与经营活动有关的现金”“收到的其他与经营活动有关的现金”的具体明细以及与报表科目的勾稽关系。

请申报会计师对上述事项进行核查并发表明确意见。

答复：

发行人说明事项

一、现金流量表中“销售商品、提供劳务收到的现金”的具体明细以及与报表科目的勾稽关系

现金流量表中“销售商品、提供劳务收到的现金”的具体明细如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
----	-----------	-------	-------	-------

提供训练数据定制服务、销售训练数据产品以及提供训练数据相关的应用服务收到的现金	5,074.74	22,980.66	21,395.54	13,608.72
---	----------	-----------	-----------	-----------

现金流量表中“销售商品、提供劳务收到的现金”与报表科目的勾稽关系如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
营业收入	3,126.09	23,755.81	19,265.77	11,907.09
应收账款账面余额的减少/(增加)	1,062.69	-1,563.61	-389.18	-934.69
应收账款账面余额的核销	-	-	-64.88	-
应收账款非现金偿债	-	-	-145.00	-
合同资产账面余额的增加	-286.00	-	-	-
预收款项的(减少)/增加	-2,480.66	-927.03	1,171.74	1,358.78
合同负债的增加	3,268.86	-	-	-
应交税费 - 增值税 - 销项税	462.45	2,020.93	1,779.28	1,298.78
财务费用 - 净汇兑(亏损)/收益- 应收账款	-27.55	-131.13	-173.40	7.89
财务费用 - 其他财务费用 - 现金折扣	-9.09	-23.36	-10.55	-8.52
所得税费用 - 代扣代缴境外所得税费用	-42.05	-150.94	-38.24	-20.61
合计	5,074.74	22,980.66	21,395.54	13,608.72
销售商品、提供劳务收到的现金	5,074.74	22,980.66	21,395.54	13,608.72
差异	-	-	-	-

二、现金流量表中“购买商品、接受劳务支付的现金”的具体明细以及与报表科目的勾稽关系

现金流量表中“购买商品、接受劳务支付的现金”的具体明细如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
数据服务费	1,281.74	5,917.72	7,275.10	4,042.37
进项税(注1)	62.35	335.09	385.15	204.66
项目租赁费	18.59	274.06	345.08	159.36
房租费用(注2)	7.80	40.27	35.58	36.44
其他(注3)	-	101.22	140.58	2.86

合计	1,370.48	6,668.36	8,181.49	4,445.69
-----------	-----------------	-----------------	-----------------	-----------------

注 1: 此处进项税为当期采购数据服务费的进项税。

注 2: 此处房租费用为与当期主营业务直接相关的房租费用。

注 3: 其他主要为与当期主营业务直接相关的差旅费、交通费、燃料费等支出。

现金流量表中“购买商品、接受劳务支付的现金”与报表科目的勾稽关系如下:

单位: 万元

项目	2020 年 1-3 月	2019 年	2018 年	2017 年
营业成本	536.87	7,067.52	6,786.78	3,303.47
研发费用 - 训练数据产品开发支出	336.18	2,217.81	1,264.60	1,455.04
存货账面余额的增加/(减少)	476.53	-1,595.64	1,160.82	362.57
存货跌价准备的转销	-	372.75	-	0.70
存货、营业成本以及研发费用-训练数据产品开发支出中的职工薪酬、折旧与摊销等	-198.48	-1,546.93	-1,338.29	-865.81
应付账款 - 数据服务费的减少/(增加)	79.62	-147.04	-102.31	5.97
预付款项 - 数据服务费的增加/(减少)	77.41	-35.19	24.74	-20.91
应交税费 - 增值税 - 进项税(注)	62.35	335.09	385.15	204.66
合计	1,370.48	6,668.36	8,181.49	4,445.69
购买商品、接受劳务支付的现金	1,370.48	6,668.36	8,181.49	4,445.69
差异	-	-	-	-

注: 此处进项税为当期采购数据服务费的进项税。

三、现金流量表中“支付给职工以及为职工支付的现金”的具体明细以及与报表科目的勾稽关系

现金流量表中“支付给职工以及为职工支付的现金”的具体明细如下:

单位: 万元

项目	2020 年 1-3 月	2019 年	2018 年	2017 年
职工薪酬	2,095.57	5,120.24	4,358.71	3,207.55

现金流量表中“支付给职工以及为职工支付的现金”与报表科目的勾稽关系如下:

单位: 万元

项目	2020 年 1-3 月	2019 年	2018 年	2017 年
应付职工薪酬 - 当期支付金额	1,906.85	5,308.17	4,364.51	3,207.82

应交税费 - 个人所得税的减少/ (增加)	188.72	-187.92	-5.80	-0.28
合计	2,095.57	5,120.24	4,358.71	3,207.55
支付给职工以及为职工支付的现金	2,095.57	5,120.24	4,358.71	3,207.55
差异	-	-	-	-

四、现金流量表中“支付的其他与经营活动有关的现金”的具体明细以及与报表科目的勾稽关系

现金流量表中“支付的其他与经营活动有关的现金”的具体明细如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
房租费用（注1）	202.25	654.39	509.59	379.04
中介费用	42.88	465.40	155.09	215.86
进项税（注2）	32.80	129.88	60.89	16.80
市场宣传费	12.62	44.76	52.82	29.71
业务招待费	9.09	90.47	42.13	23.79
水电费	8.54	36.78	31.08	20.55
办公费	3.07	38.17	52.38	32.82
差旅费	2.38	76.43	67.74	41.51
财务手续费	2.63	9.70	7.89	8.63
邮寄费	2.40	8.76	14.24	9.47
押金及保证金	2.14	25.26	59.36	135.08
装修费	-	64.83	44.24	15.91
滞纳金支出	-	-	-	15.34
捐赠支出	-	15.00	-	-
其他（注3）	78.20	204.48	180.06	281.46
合计	399.00	1,864.31	1,277.51	1,225.97

注1：此处房租费用为当期的管理部门和人员使用的租赁房屋的当期房租费用，以及研发部门和人员使用的租赁房屋的当期房租费用。

注2：此处进项税为当期扣除数据服务费采购以外的与经营活动相关的支出的进项税。

注3：其他主要为当期信息系统咨询服务、专利申请服务费、通讯费、劳保费等支出。

现金流量表中“支付的其他与经营活动有关的现金”与报表科目的勾稽关系如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
销售费用（不含职工薪酬、折旧与摊销）	14.72	94.55	107.25	73.79
管理费用（不含职工薪酬、折旧与摊销）	240.53	1,310.17	774.51	799.59
研发费用-基础研发支出（不含职工薪酬、折旧与摊销）	47.44	163.36	123.68	14.78
财务费用 - 其他财务费用 - 手续费支出	2.63	9.70	7.89	8.63
营业外支出（不含非流动资产毁损报废损失）	-	15.00	3.42	15.35
预付款项 - 其他（不含资产采购款）的增加/（减少）	24.03	-16.15	13.67	1.97
其他应收款账面余额 - 经营相关的增加（注1）	1.85	194.55	189.11	254.17
其他流动资产 - 待抵扣及待认证进项税的增加/（减少）	6.98	-39.57	39.77	-
其他流动资产 - 待摊房屋租赁费的增加/（减少）	29.33	31.25	7.76	-44.01
应付账款 - 其他采购费的减少	-0.71	-19.19	-50.49	-3.22
应交税费 - 增值税 - 进项税（注2）	32.80	129.88	60.89	16.80
其他应付款的（增加）/减少	-0.60	-9.24	0.05	88.12
合计	399.00	1,864.31	1,277.51	1,225.97
支付的其他与经营活动有关的现金	399.00	1,864.31	1,277.51	1,225.97
差异	-	-	-	-

注 1：其他应收款账面余额 -经营相关的增加主要为其他应收款中押金保证金、员工备用金等与发行人日常经营活动相关的往来款的当期/年增加额。

注 2：此处进项税为当期扣除数据服务费采购以外的与经营活动相关的支出的进项税。

五、现金流量表中“收到的其他与经营活动有关的现金”的具体明细以及与报表科目的勾稽关系

现金流量表中“收到的其他与经营活动有关的现金”的具体明细如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
政府补助	157.56	803.69	762.46	680.73
利息收入	3.90	6.24	7.20	5.23
押金保证金	0.43	32.81	14.20	85.11
其他（注）	36.88	13.29	64.23	11.58

合计	198.77	856.03	848.09	782.65
-----------	---------------	---------------	---------------	---------------

注：其他主要为当期收回的员工备用金、汇款退回等。

现金流量表中“收到的其他与经营活动有关的现金”与报表科目的勾稽关系如下：

单位：万元

项目	2020年1-3月	2019年	2018年	2017年
其他收益 - 政府补助	157.56	791.49	747.86	630.12
财务费用 - 存款及应收款项的利息收入（不含股东利息收入）	3.90	6.24	7.20	5.23
营业外收入 - 政府补助	-	12.20	14.60	50.61
营业外收入 - 其他	-	0.21	9.26	0.49
其他应收款账面余额 - 经营相关的减少（注）	37.31	45.89	69.17	96.20
合计	198.77	856.03	848.09	782.65
收到其他与经营活动有关的现金	198.77	856.03	848.09	782.65
差异	-	-	-	-

注：其他应收款账面余额 - 经营相关的减少主要为其他应收款中押金保证金、员工备用金等与发行人日常经营活动相关的往来款的当期/年减少额。

申报会计师核查事项

（1）核查程序

申报会计师执行的主要程序如下：

- 1、获取发行人编制现金流量表的底稿，核对其中各项目数据的计算过程及数据来源；
- 2、检查现金流量表中“销售商品、提供劳务收到的现金”“购买商品、接受劳务支付的现金”“支付给职工以及为职工支付的现金”“支付的其他与经营活动有关的现金”“收到的其他与经营活动有关的现金”等科目相关数据与资产负债表、利润表等科目的勾稽关系是否合理。

（2）核查结论

经核查，申报会计师认为：

报告期内，公司“销售商品、提供劳务收到的现金”“购买商品、接受劳务支付的现金”“支付给职工以及为职工支付的现金”“支付的其他与经营活动有关的现金”“收到

的其他与经营活动有关的现金”的具体明细与发行人的实际情况相符，与报表科目之间勾稽相符。

六、关于其他事项

23. 关于前次撤回和本次申报

2019年4月9日本所受理了发行人首次公开发行股票并在科创板上市的申请文件，2019年7月发行人和保荐机构分别提交了撤回申请。

请发行人说明：（1）与数据服务费采购相关的内部控制设计完善和整改过程以及有效运行的具体情况；（2）前后两次申报材料中披露的业务数据和财务信息的差异情况及修正原因。

请保荐机构说明：本次申报前保荐机构内核委员的投票情况，是否存在投反对票的情形，相关委员的反对原因、提出的具体问题、落实情况。

答复：

发行人说明事项

（1）与数据服务费采购相关的内部控制设计完善和整改过程以及有效运行的具体情况

一、报告期内，发行人与数据服务采购相关的内部控制设计合理且有效运行

（一）发行人根据《企业内部控制基本规范》的规定建立了完善的内控制度

报告期内，发行人建立了《海天瑞声采购管理制度》、《海天瑞声项目管理办法》、《海天瑞声供应商管理制度》、《海天瑞声合同管理制度》和《海天瑞声业务采购实施细则》等规章制度作为对外采购数据服务的制度依据。

针对数据服务采购业务环节涉及的控制点（年度总体预算、供应商选择和持续管理、采购协议签署、采购价格管理、立项及预算管理、采购申请与审批、下达采购需求、供应商提供服务、验收结算及费用报销、供应商付款、项目决算、业财核对），发行人均制定、采取了有效的控制措施/活动（具体控制点、控制活动情况请参见本问询函回复问题 10.1

“发行人说明事项/（1）发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效”相关回复），确保数据服务采购相关内控有效执行。

（二）发行人搭建了有效的内部信息系统，可确保采购信息完整、准确记载，数据服务采购量可复核，财务信息与实际采购业务一致

1、发行人搭建了有效的内部信息系统，可记载完整、准确的采购信息等业务数据。

2、发行人业务管理平台记录了完整、准确的采购信息，使得可以通过“以销定采”、“同类业务标产比可比”等业务逻辑对数据服务采购量进行复核、确认。同时，发行人在项目预算申请环节参考业务采购价格体系对预算采购价格进行复核、确认，在项目支出报销环节参考预算采购价格对实际采购价格进行复核、确认。

3、发行人财务系统中数据服务采购信息均基于业务管理平台产生，可保证实际采购业务与财务信息一致。

本部分详细说明及论述请参见本问询函回复问题 10.1 “发行人说明事项/（1）发行人采购数据采集和标注服务的人员管理、场地使用、进度管理和质量控制措施、验收标准、结算方式、结算对象和结算周期等情况，与数据服务采购相关内部控制的设计是否合理、执行是否有效”相关说明。

（三）小结

综上，报告期内发行人与数据服务费采购相关的内部控制设计合理且有效运行，发行人搭建了有效的内部信息系统，可确保采购信息完整、准确记载，数据服务采购量可复核，财务信息与实际采购业务一致。

二、发行人对原有内部控制架构、制度进行了审慎自查，针对可改进之处进行了整改、完善，整改完善后的内部控制设计有效运行，进一步提升了内部控制有效性

在原有内部控制架构、信息系统的基础上，发行人进行了审慎自查。经自查，发行人原有内部控制架构存在的可改进点分别为：1、报告期初未在生产系统（一体化数据处理平

台)中留存生产过程中的日志信息,使生产过程留痕程度不够;2、终端服务人员的身份信息、联系方式、签字确认等信息记录不全,导致内部控制中的复核工作未穿透至各终端人提供的劳务服务本身,且外部核查过程可能受到信息不全的限制。

针对上述可改进之处,发行人积极进行有针对性的整改提升,进一步提升内部控制完善性、有效性。具体情况如下:

改善措施	改善目的
1、自 2018 年开始保留生产系统中已记载的基本日志信息; 2、在原有生产系统中上线单独的日志管理系统,记录并保留更为全面的生产过程日志信息,如终端人的注册信息、各终端人的工作量的详细信息等,作为生产系统中已有数据的有效补充。	使得发行人训练数据生产过程充分留痕,可供追溯。
1、自 2019 年起在内部管理流程中加入终端人对提供服务量进行签字确认的环节; 2、在原有生产系统中上线终端人管理系统,完整记录劳务人员的个人信息、劳务量、确认信息等。	1、通过终端人对劳务量的直接签字确认,进一步确认采购量数据; 2、可通过终端人留存的个人信 息,将采购链条直接穿透至最终劳务提供者,并可供核查。

发行人于 2019 年初开始要求业务部门将质量验收合格的采集数据量、标注工作量录入表格/信息系统,并由终端劳务人员签字确认其提供的劳务工作量,使数据服务采购量在服务公司确认的基础上,可穿透至终端劳务人员确认,加强数据服务采购的外部证据留存。

另外,发行人于 2019 年下半年在原有生产系统基础上逐步上线了终端人管理系统和日志管理系统,在原有的基本日志信息基础上,获取更为详细的生产过程用户行为数据,使得整个数据生产过程更加可追溯、与财务数据可复核,从而提升了公司整体内部控制水平和经营管理效率。具体整改提升措施包括:

1、**新增终端人管理系统**,包括公司项目管理人员、服务公司、小外包人员、被采集人、标注人员等。保证所有人员的个人信息在系统内的完整、真实、可追溯。具体包括:

- (1) 建立人员信息数据库,提供标准的人员注册、登录入口;
- (2) 建立多种实名认证方式可供选择,包括身份证验证、实名电话验证等方式;
- (3) 终端个人有唯一的标识,各个工具中记录的日志信息均与终端个人对应。

2、**新增日志管理系统**,对于与经营业务相关的日志详细数据,做到集中处理、实时监控。系统可记录被采集人员参与采集的次数、采集事项发生时间、单次采集数据量、累

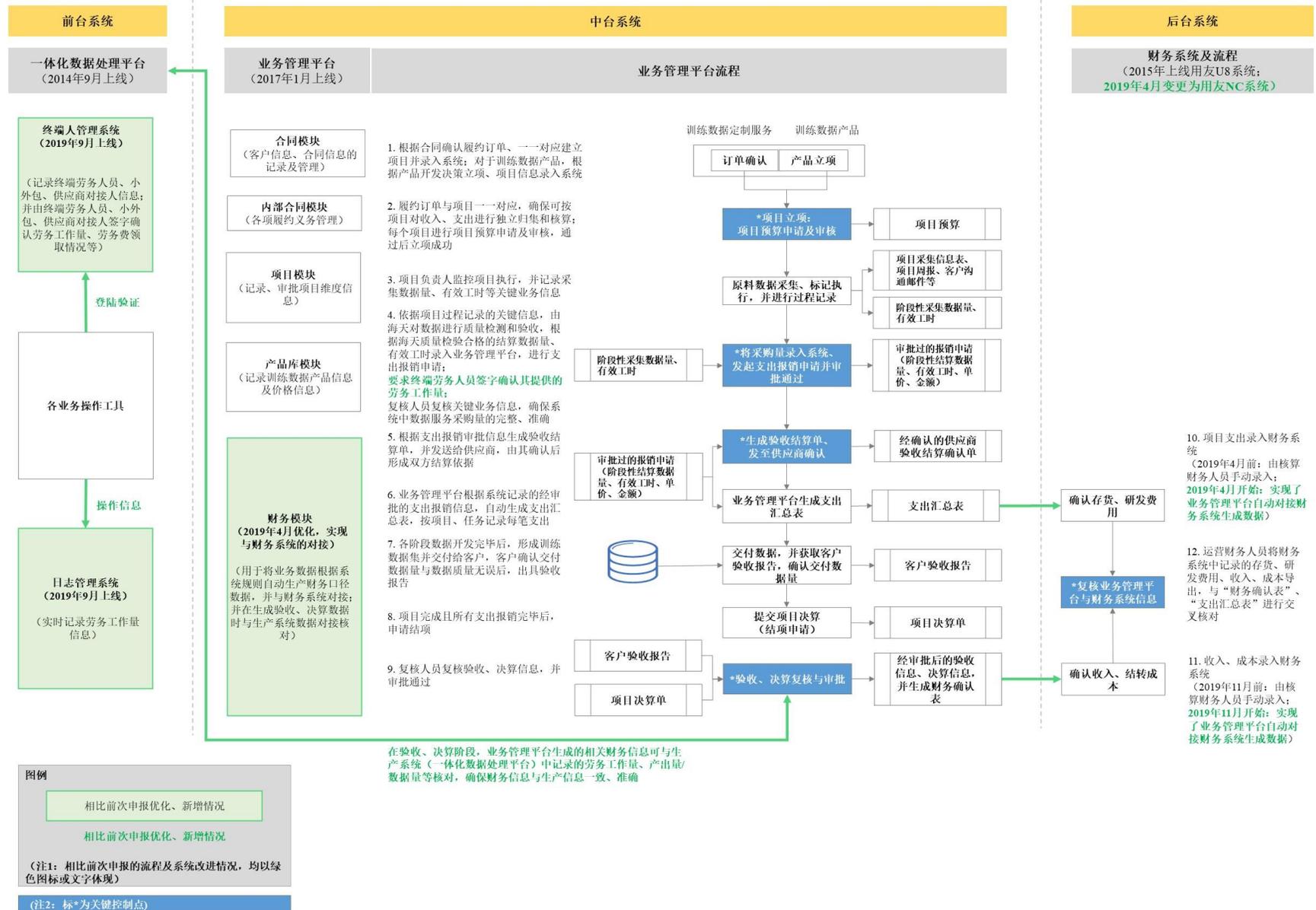
计采集数据量、测试标产比，各标注人员的操作工时、产出量等用户行为数据。具体包括：

（1）各采集和标注工具开发标准的日志记录模块，定义并执行统一的日志格式；所有原始日志均保存至相应工具的服务器端；

（2）建立日志收集服务，定期对分散在各服务器上的日志做定向整理、归并和收集，存储于特定的日志库；

3、优化了业务管理平台中的财务模块，实现收入、成本由业务管理平台信息向财务系统进行自动传输，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。

发行人本次申报时点内部信息系统整改提升情况如下图所示：



此外，保荐机构、申报会计师对发行人的信息系统进行了核查。经过信息系统核查，发行人会计师毕马威审计与保荐机构（聘请安永（中国）企业咨询有限公司提供 IT 审计专业服务支持）均证明发行人的信息系统总体控制是有效的，信息系统中记录保留的业务、财务数据真实、准确、完整。发行人 2019 年确认收入项目（不含训练数据产品重复销售）涉及终端劳务人员 5.14 万人，上述人员对其所提供劳务工作量的签字确认文件合计 80,761 份（存在部分终端人员参与多个采购项目的情形）。经保荐机构、申报会计师核查的终端劳务人员对工作量签字确认文件的项目收入占发行人总体收入（不含训练数据产品重复销售）的比例为 73.9%，发行人终端人签字确认其服务量的控制活动有效运行。

（2）前后两次申报材料中披露的业务数据和财务信息的差异情况及修正原因

发行人前后两次申报材料中披露的业务、财务数据差异情况主要是下述原因所致：

发行人根据《公开发行证券的公司信息披露内容与格式准则第 41 号——科创板公司招股说明书》的相关要求，自查完善了财务及非财务数据的选取和披露口径，细化完善业务、财务相关数据的内容、金额披露及分析，并全面检查复核了前次披露的业务数据情况，对发行人部分供应商采购金额、部分业务产销量等数据进行修正；发行人 2017、2018 年合并口径财务报表数据，调整了现金流量表科目金额，并根据 2019 年保荐代表人培训会议精神对其他应收款中押金和员工备用金补充计提了坏账准备。除此以外，发行人 2017、2018 年合并口径财务报表数据无其他调整事项。

具体差异情况及修正原因说明如下：

一、业务数据

（一）采购细分类别数据

1、差异情况

2018 年采购细分类别数据差异情况如下：

单位：万元

项目	本次申报	前次申报	差异
数据服务费	7,352.66	7,352.66	-
办公房租	545.17	545.17	-
资产采购	552.90	- (前次申报未列示此科目)	552.90

项目	本次申报	前次申报	差异
其他固定资产采购	- (本次申报未列示此科目)	408.00	-408.00
项目租赁费	345.08	345.08	-
岗位服务费	256.62	256.62	-
其他	- (本次申报未列示此科目)	863.40	-863.40
采购总额	9,052.42	9,770.93	-718.51

2017年采购细分类别数据差异情况如下：

单位：万元

项目	本次申报	前次申报	差异
数据服务费	4,057.31	4,057.31	-
办公房租	415.48	415.48	-
资产采购	203.72	- (前次申报未列示此科目)	203.72
其他固定资产采购	- (本次申报未列示此科目)	200.80	-200.80
项目租赁费	159.36	159.36	-
岗位服务费	226.95	226.95	-
其他	- (本次申报未列示此科目)	1,040.86	-1,040.86
采购总额	5,062.82	6,100.75	-1,037.93

2、修正原因

采购细分类别数据修正原因是：（1）列示口径调整：前次申报采购分类“其他固定资产采购”（仅列示固定资产采购相关金额）调整为本次申报采购分类“资产采购”（合并列示固定资产采购和无形资产采购）；（2）删除原采购总额中的“其他”细分类别金额，仅保留与业务生产经营直接相关的采购细分类别。

（二）主要供应商采购数据

1、差异情况

2018年主要供应商采购数据差异情况如下：

单位：万元

序号	供应商名称	本次申报	前次申报	差异
1	广州善世人力资源服务有限公司	3,324.46	3,297.57	26.89
	芜湖善慧人力资源服务有限公司			

序号	供应商名称	本次申报	前次申报	差异
	广州金社宝网络技术有限公司			
2	北京博睿驰信人力资源有限公司	2,078.46	1,930.67	147.79
	恩泽尔商贸有限公司			
3	北京冠华英才国际经济技术有限公司	905.61	905.61	-
4	李宏	410.29	410.29	-
5	TWIN STUDIOS	229.97	229.97	-
6	SAKURA JAPAN LLC	84.81	71.11	13.70

2017年主要供应商采购数据差异情况如下：

单位：万元

序号	供应商名称	本次申报	前次申报	差异
1	芜湖善慧人力资源服务有限公司	1,722.70	1,722.70	-
	广州善世人力资源服务有限公司			
2	北京博睿驰信人力资源有限公司	849.57	829.91	19.66
	恩泽尔商贸有限公司			
	北京零点起航营销策划有限公司			
3	北京冠华英才国际经济技术有限公司	662.12	539.82	122.30
4	李宏	218.55	218.55	-
5	SAKURA JAPAN LLC	150.69	189.68	-38.99

2、修正原因

主要供应商采购数据修正原因是：（1）2018年增加本次申报新识别的同一控制下供应商广州金社宝网络技术有限公司（曾用名：才星（广州）科技服务有限公司）；（2）北京博睿驰信人力资源有限公司前次申报遗漏“岗位服务费”采购金额，本次申报进行补充；（3）北京冠华英才国际经济技术有限公司前次申报遗漏“岗位服务费”采购金额，本次申报进行修正；（4）SAKURA JAPAN LLC 前次申报时误将少量2018年采购纳入该供应商2017年采购金额计算，同时存在少量统计错漏，本次申报进行修正。

（三）智能语音训练数据定制服务销量、单价数据

1、2018年数据差异情况

项目	本次申报	前次申报	差异
中文语音识别转写销量（小时）	87,856	88,196	-340

项目	本次申报	前次申报	差异
中文语音合成销量（小时）	766	784	-18
外语种语音合成销量（小时）	734	536	198

2、2018 年数据修正原因

语音识别转写销量修正原因是本次申报考虑母子公司的合并抵消，进行微调。语音合成销量修正原因是本次申报考虑分次验收的影响，进行微调。同时，根据销量的变动相应修正了对应的单价数据。

3、2017 年数据差异情况

项目	本次申报	前次申报	差异
中文语音识别采集销量（小时）	18,858	16,759	2,099
外语种语音识别采集销量（小时）	9,034	8,028	1,006
中文语音识别转写销量（小时）	46,362	41,646	4,716
外语种语音识别转写销量（小时）	2,123	1,887	236
中文语音合成销量（小时）	371	314	57
外语种语音合成销量（小时）	544	464	80

4、2017 年数据修正原因

上述销量修正原因是本次申报考虑分次验收的影响，进行微调，并修正前次申报销量数据的个别错漏。同时，根据销量的变动相应修正了对应的单价数据。

（四）智能语音训练数据产品产量数据

1、差异情况

项目	本次申报	前次申报	差异
2018 年度智能语音训练数据产品产量（小时）	19,917	19,815	102
2017 年度智能语音训练数据产品产量（小时）	20,920	14,368	6,552

2、修正原因

产量修正原因是本次申报与前次申报统计口径有所差异所致，前次申报统计口径为当年产成且形成收入的产量，本次申报统计口径为当年产成的产量。

二、财务信息

（一）其他应收款

1、差异情况及报表影响金额

单位：万元

项目	2018年12月31日	2017年12月31日
本次申报	172.84	150.09
前次申报	181.88	157.91
差异	-9.04	-7.82

上述差异较小，对发行人报表无实质性影响。

2、修正原因

本次申报根据 2019 年保荐代表人培训会议精神对其他应收款中押金和员工备用金补充计提了坏账准备，因此其他应收款账面价值减少。

(二) 现金流量表各科目

1、差异情况及报表影响金额

2018 年合并现金流量表差异情况如下：

单位：万元

项目	本次申报	前次申报	差异
一、经营活动产生的现金流量：			
销售商品、提供劳务收到的现金	21,395.54	21,424.93	-29.39
收到的税费返还	265.32	1,013.18	-747.86
收到其他与经营活动有关的现金	848.09	129.92	718.17
经营活动现金流入小计	22,508.94	22,568.03	-59.09
购买商品、接受劳务支付的现金	8,181.49	8,122.43	59.06
支付给职工以及为职工支付的现金	4,358.71	4,368.22	-9.51
支付的各项税费	2,263.44	2,285.09	-21.65
支付其他与经营活动有关的现金	1,277.51	1,348.14	-70.63
经营活动现金流出小计	16,081.15	16,123.89	-42.74
经营活动产生的现金流量净额	6,427.80	6,444.15	-16.35
二、投资活动产生的现金流量：			
收回投资收到的现金	23,200.00	18,700.00	4,500.00
处置固定资产、无形资产和其他长期资产收回的现金净额	11.39	0.63	10.76

项目	本次申报	前次申报	差异
取得投资收益收到的现金	338.43	338.43	-
收到其他与投资活动有关的现金	-	-	-
投资活动现金流入小计	23,549.82	19,039.06	4,510.76
购建固定资产、无形资产和其他长期资产支付的现金	2,629.37	2,626.09	3.28
投资支付的现金	24,000.00	19,500.00	4,500.00
支付其他与投资活动有关的现金	-	-	-
投资活动现金流出小计	26,629.37	22,126.09	4,503.28
投资活动产生的现金流量净额	-3,079.55	-3,087.03	7.48
三、筹资活动产生的现金流量：			
吸收投资收到的现金	-	-	-
筹资活动现金流入小计	-	-	-
分配股利、利润或偿付利息支付的现金	-	-	-
支付其他与筹资活动有关的现金	-	-	-
筹资活动现金流出小计	-	-	-
筹资活动产生的现金流量净额	-	-	-
四、汇率变动对现金及现金等价物的影响	19.18	10.31	8.87
五、现金及现金等价物净增加额/（净减少额）	3,367.43	3,367.43	-
加：期初现金及现金等价物余额	1,301.73	1,301.73	-
六、期末现金及现金等价物余额	4,669.16	4,669.16	-

2017 年合并现金流量表差异情况如下：

单位：万元

项目	本次申报	前次申报	差异
一、经营活动产生的现金流量：			
销售商品、提供劳务收到的现金	13,608.72	13,613.84	-5.12
收到的税费返还	35.03	672.93	-637.90
收到其他与经营活动有关的现金	782.65	219.78	562.87
经营活动现金流入小计	14,426.41	14,506.55	-80.14
购买商品、接受劳务支付的现金	4,445.69	4,553.60	-107.91
支付给职工以及为职工支付的	3,207.55	3,096.43	111.12

项目	本次申报	前次申报	差异
现金			
支付的各项税费	2,085.85	2,100.00	-14.15
支付其他与经营活动有关的现金	1,225.97	1,287.37	-61.40
经营活动现金流出小计	10,965.05	11,037.40	-72.35
经营活动产生的现金流量净额	3,461.36	3,469.16	-7.80
二、投资活动产生的现金流量：			
收回投资收到的现金	6,000.00	6,000.00	-
处置固定资产、无形资产和其他长期资产收回的现金净额	6.88	6.88	-
取得投资收益收到的现金	131.18	131.18	-
收到其他与投资活动有关的现金	39.61	39.61	-
投资活动现金流入小计	6,177.67	6,177.67	-
购建固定资产、无形资产和其他长期资产支付的现金	203.72	207.57	-3.85
投资支付的现金	13,500.00	13,500.00	-
支付其他与投资活动有关的现金	80.00	80.00	-
投资活动现金流出小计	13,783.72	13,787.57	-3.85
投资活动产生的现金流量净额	-7,606.05	-7,609.90	3.85
三、筹资活动产生的现金流量：			
吸收投资收到的现金	4,050.00	4,050.00	-
筹资活动现金流入小计	4,050.00	4,050.00	-
分配股利、利润或偿付利息支付的现金	500.00	500.00	-
支付其他与筹资活动有关的现金	-	-	-
筹资活动现金流出小计	500.00	500.00	-
筹资活动产生的现金流量净额	3,550.00	3,550.00	-
四、汇率变动对现金及现金等价物的影响	-17.87	-21.82	3.95
五、现金及现金等价物净增加额/（净减少额）	-612.56	-612.56	-
加：期初现金及现金等价物余额	1,914.29	1,914.29	-
六、期末现金及现金等价物余额	1,301.73	1,301.73	-

上述差异对发行人报表的影响如上表“差异”一列所示，经营活动现金流量净额、

投资活动现金流量净额、筹资活动现金流量净额差异较小，对发行人现金流量状况无实质影响。

2、修正原因

本次申报与前次申报的主要差异原因如下：

（1）“收到的税费返还”和“收到的其他与经营活动有关的现金”的差异

财政部于 2019 年颁布了《关于修订印发 2019 年度一般企业财务报表格式的通知》（财会 [2019] 6 号）和《关于修订印发合并财务报表格式（2019 版）的通知》（财会 [2019] 16 号），规定企业实际收到的政府补助，无论是与资产相关还是与收益相关，均在“收到其他与经营活动有关的现金”项目填列。发行人本次申报已根据财会 [2019] 6 号和财会 [2019] 16 号规定的财务报表格式编制整个报告期内财务报表，将 2017 年和 2018 年收到的增值税即征即退补贴款人民币 630.12 万元和 747.86 万元分别作为实际收到的政府补助款，从“收到的税费返还”调整到“收到的其他与经营活动有关的现金”列示。

（2）“收回投资收到的现金”和“投资支付的现金”的差异

发行人于 2018 年当年赎回理财产品收到现金人民币 4,500.00 万元，并于 2018 年当年继续购买该理财产品支付现金人民币 4,500.00 万元，前次申报中在“收回投资收到的现金”和“投资支付的现金”中并未包含上述金额，因此本次申报予以修正，在“收回投资收到的现金”和“投资支付的现金”中反映上述赎回理财产品和购买理财产品的金额。

（3）“购买商品、接受劳务支付的现金”和“支付给职工以及为职工支付的现金”的差异

前次申报发行人将 2017 年支付给语言专家的劳务费人民币 109.13 万元计入“购买商品、接受劳务支付的现金”列示，本次申报调整至“支付职工以及为职工支付的现金”列示。

本次申报与前次申报的其他差异较小，根据现金流量的实际情况对明细项目进行了调整。

（三）2017 年管理费用明细

1、差异情况及报表影响金额

2017 年管理费用明细项目中中介费、办公费、其他的差异情况如下：

单位：万元

项目	本次申报	前次申报	差异
中介费	231.88	215.86	16.02
办公费	32.82	91.86	-59.04
其他	70.80	27.78	43.02

上述差异较小，对发行人报表无实质性影响。

2、修正原因

本次申报根据管理费用性质的实际情况对明细项目进行了微调，将人才引进招聘费人民币 16.02 万元进行了明细项目的重分类，自“办公费”调整到“中介费”进行列示，将残保金和邮寄费等杂项支出共计人民币 43.02 万元进行了明细项目的重分类，自“办公费”调整到“其他”进行列示。

（四）主要税项缴纳情况

1、2018 年增值税缴纳

（1）差异情况及报表影响金额

单位：万元

项目	本次申报	前次申报	差异
期初未缴数	74.32	74.32	-
本期应缴数	1,176.59	1,160.01	16.58
本期实缴数	1,192.90	1,176.32	16.58
期末未缴数	58.01	58.01	-

上述差异对发行人报表无影响。

（2）修正原因

前次申报的“本期实缴数”遗漏发行人子公司中瑞智，本次申报进行了更正微调。

2、2018 年企业所得税缴纳

（1）差异情况及报表影响金额

单位：万元

项目	本次申报	前次申报	差异
期初未缴数	84.50	84.50	-
本期应缴数	474.24	724.22	-249.98
本期实缴数	673.34	923.32	-249.98
期末未缴数	-114.60	-114.60	-

上述差异对发行人报表无影响。

(2) 修正原因

前次申报的“本期实缴数”中加总金额计算有误，本次申报进行了更正。

3、2017 年企业所得税缴纳

(1) 差异情况及报表影响金额

单位：万元

项目	本次申报	前次申报	差异
期初未缴数	188.96	188.96	-
本期应缴数	586.35	656.41	-70.06
本期实缴数	690.81	760.87	-70.06
期末未缴数	84.50	84.50	-

(2) 修正原因

前次申报的“本期实缴数”中加总金额计算有误，本次申报进行了更正。

(五) 2017 年非经常性损益

1、差异情况及报表影响金额

单位：万元

项目	本次申报	前次申报	差异
非流动性资产处置损益	5.98	5.98	-
计入当期损益的政府补助	50.61	50.61	-
计入当期损益的对非金融企业收取的资金占用费	24.90	-	24.90
理财产品的投资收益	131.18	131.18	-
企业所得税退回	-	-	-
除上述各项之外的其他营业外收入和支出	-14.86	-14.86	-

小计	197.81	172.91	24.90
所得税影响额	-31.96	-28.22	-3.74
合计	165.85	144.69	21.16

上述差异对发行人报表无影响。

2、修正原因

本次申报根据股东借款利息的性质，计入非经常性损益。

保荐机构说明事项

本次申报前保荐机构内核委员的投票情况，是否存在投反对票的情形，相关委员的反对原因、提出的具体问题、落实情况

本次申报前，保荐机构内核委员在内核会议表决中全部投赞成票，不存在投反对票的情形。

24. 关于募集资金运用

招股说明书披露，截至2019年底，发行人资产总额为4.04亿元，发行人持有的货币资金和交易性金融资产分别为1.63亿和1.56亿元。发行人本次募集资金拟投向于自主研发数据产品扩建项目，拟投入募集资金7.6亿余元，远大于公司资产总额。

请发行人结合同行业公司资产收入情况、公司资金用途，说明公司募集资金总额的合理性，发行人是否具有足够的经营管理能力和资产运营能力。

答复：

发行人的募集资金投资项目及募集资金总额具备合理性，发行人有足够的经营管理能力和资产运营能力。具体说明如下：

一、发行人资产和收入规模较小，符合自身业务形态特征和行业特点；但发行人的经营管理能力、资产运营能力较强

发行人为科技企业，主营业务为训练数据的研发设计、生产及销售，当前仍处于成长期，暂未登陆A股资本市场平台，融资渠道有限，整体资产和收入规模尚相对较小。但报告期内，发行人展现出了较高的资产周转率、净资产收益率和较高的流动性，反映了发行人较强的经营管理能力和资产运营能力。

将发行人、发行人同行业²⁰上市/拟上市公司的资产、收入指标及盈利能力、运营能力指标与 A 股上市公司平均水平进行比较，可以得知：

1、信息和软件技术服务行业的科创类上市/拟上市公司普遍呈现出资产、收入规模相对较小、资产周转率、净资产收益率和流动比率等指标高于 A 股上市公司整体水平的特征。发行人资产和收入规模较小，符合自身业务形态特征和所在行业特点。

2、与其他软件和信息技术服务业科创板已上市/拟上市公司相比，发行人资产、收入规模低于平均水平，而资产周转率、净资产收益率、流动比率均高于平均水平，一定程度上体现出了发行人较高的运营能力和盈利能力。

二、发行人募集资金用途、测算合理，与发行人实际经营情况相适应，募集资金总额具备合理性

发行人本次募集资金运用围绕主营业务进行，在扣除发行费用后，将按轻重缓急顺序投入以下项目：

单位：万元

序号	项目名称	总投资金额	拟投入募集资金
1	自主研发数据产品扩建项目	24,921.18	24,921.18
2	一体化数据处理技术支撑平台升级项目	25,057.72	25,057.72
3	研发中心升级建设项目	16,614.98	16,614.98
4	补充流动资金	10,000.00	10,000.00
合计		76,593.88	76,593.88

1、发行人本次募集资金运用围绕主营业务进行，通过本次募集资金投资项目的实施，将全方位提升发行人现有研发、业务管理能力。

补充流动资金的原因及合理性分析请参见本问询函反馈第 18 题（2）相关回复；除补充流动资金外，发行人三个募集资金投资项目的实施目的、实施内容如下：

项目名称	对应现有业务和技术	实施目的	实施内容
自主研发数据产品扩建项目	训练数据产品	完善公司训练数据产品体系、优化训练数据产品储备结构，拓展公司训练数据产品覆盖场景和领域，加快对客户需求的	新开发 151 个自有知识产权的训练数据产品，包括 80 个智能语音训练数据产品、22 个计算机视觉训练数据产品和 49 个发音词典训

²⁰ 除发行人外，当前 A 股尚无从事训练数据相关业务的上市/拟上市公司，因此选取科创板信息和软件技术服务行业公司作为发行人的同行业公司进行对比。

项目名称	对应现有业务和技术	实施目的	实施内容
		响应速度	训练数据产品,将有效优化公司训练数据产品结构。
一体化数据处理技术支持平台升级项目	整体训练数据生产	在训练数据生产平台、工具的研发建设上继续增加投入,更好地满足日益增长的市场需求、提升产能和生产效率;进一步提升训练数据开发能力、开发效率,保障训练数据质量,加强训练数据的安全性。	在公司现有的一体化数据处理平台及其工具集合的基础上,扩充研发团队,提升生产工具、平台系统性开发实力;投入先进的开发设备和软件,推进测试实验室建设,升级建设数据处理工具及平台,为公司的训练数据生产提供支持。
研发中心升级建设项目	整体研发体系	增加公司的前瞻性技术储备,通过技术手段的运用提升产品服务质量	聚焦于语音识别、声纹识别、语种识别、语音合成、自然语言处理、计算机视觉识别、计算机视觉分割与处理7个方向的前沿技术,服务于公司语音识别、语音合成、自然语言处理、计算机视觉4个业务方向,提升完善从产品研发到服务客户全流程的各环节,保证公司发展的持续竞争力和领先优势。

可以看出,发行人各募集资金投资项目均是建立在现有业务和技术基础上的提升和扩展,与发行人现有业务和技术保持一致。

2、发行人募集资金投资项目构成测算合理。相关项目的投资支出构成情况如下:

序号	项目	自主研发数据产品 扩建项目		一体化数据处理技术 支撑平台升级项目		研发中心 升级建设项目	
		金额 (万元)	占比	金额 (万元)	占比	金额 (万元)	占比
1	场地费用	8,242.90	33.08%	8,379.20	33.44%	5,193.00	31.25%
1.1	场地购置费	7,674.00	30.79%	7,932.00	31.65%	5,100.00	30.70%
1.2	场地装修费	568.90	2.28%	447.20	1.78%	93.00	0.56%
2	硬件购置费用	2,092.25	8.40%	3,577.00	14.28%	2,549.00	15.34%
3	软件购置费用	100.00	0.40%	1,162.16	4.64%	1,351.70	8.14%
4	人员工资	3,399.50	13.64%	5,159.00	20.59%	7,195.50	43.31%
5	数据服务支出	7,946.40	31.89%	-	-	-	-
6	基本预备费	435.62	1.75%	365.55	1.46%	325.78	1.96%
7	铺底流动资金	2,704.51	10.85%	6,414.81	25.60%	-	-
	合计	24,921.18	100.00%	25,057.72	100.00%	16,614.98	100.00%

(1) 自主研发数据产品扩建项目投资支出主要由场地费用、数据费、技术人员工资构成,除场地费用支出为购置方式外,其他主要项目支出结构符合发行人现有训练数据产品业务类型的支出结构。

(2) 一体化数据处理技术支撑平台升级项目投资支出主要由场地费用、技术人员工资、硬件设备和铺底流动资金构成。其中硬件设备主要为 GPU 服务器、数据存储服务器和数据处理服务器等价值较高的硬件设备，故金额较大、占比较高。由于发行人平台升级及其效果实现需要一定周期，故需要较高比例的铺底流动资金保证项目的顺利实施。

(3) 研发中心升级建设项目投资支出主要由场地费用、技术人员工资、硬件和软件构成，与研发活动所需的智力投入、软硬件投入特征相一致。

由上述分析可得，发行人自主研发数据产品扩建项目、一体化数据处理技术支撑平台升级项目、研发中心升级建设项目支出构成具备合理性。发行人对各募集资金投资项目的合理性及具体支出进行了充分研究、审慎测算。各募集资金投资项目具体必要性、合理性及支出测算情况请参见招股说明书“第九节 募集资金运用与未来发展规划/二、募集资金运用情况”。

综上所述，发行人资产和收入规模较小符合自身业务形态特征和行业特点，发行人本次募集资金运用围绕主营业务进行，发行人各募集资金投资项目均是建立在现有业务和技术基础上的提升和扩展，与发行人现有业务和技术保持一致，投资支出构成符合相关业务和技术的特点和实际需求，募集资金项目投资额测算合理。发行人以除补充流动资金外的募集资金总额按轻重缓急投资于上述项目具备合理性。发行人募集资金用途、总额合理。

三、募集资金投资项目实施对发行人未来业绩的影响说明

本次募集资金金额及投资项目综合考虑了行业和市场状况、技术水平及发展趋势、场地、设备和人员等因素，发行人对其可行性进行了充分论证，具备合理性。但由于发行人为轻资产型公司，且处于成长阶段，公司资产和收入规模相对较小，而本次募集资金金额超过公司资产和收入规模，募投项目的实施将可能导致公司未来的业务经营、管理模式发生变化，也对公司各方面经营管理能力、资产运营能力均提出了更高的要求。如果未来行业或市场环境发生难以预期的不利变化，或由于发行人管理能力、资产运营能力不足等使得募投项目的按期实施及完全达产受到不利影响，则募集资金投资项目的经济效益可能无法如期实现。

同时，本次募集资金投资项目实施过程中将新增大量的固定资产、无形资产购置支出及研发投入，各年新增折旧摊销等费用金额较大。如果募投项目的经济效益不能如期实现，且发行人主营业务收入的增加不足以缓冲募投项目实施带来的折旧摊销等费用的增加，则公司利润将可能出现一定程度的下滑。

发行人已在招股说明书“重大事项提示/五、募集资金投资项目实施对发行人未来业绩的影响”对上述情形进行了充分披露及风险提示。

四、小结

与同行业公司相比，发行人资产收入规模较小，但运营能力、盈利能力较强；发行人募集资金用途合理、总额合理。发行人作为轻资产型公司，且处于成长阶段，资产、收入规模相对较小，本次募集资金投资项目投资额超过了公司的资产、收入规模，可能导致公司的业务经营和管理模式发生一定变化，也对发行人的经营管理等各方面提出了更高要求；同时募投项目的实施将使得发行人各年新增的折旧摊销等费用金额较大。发行人将妥善安排募集资金投资项目的建设节奏、持续提升主营业务营收水平，促进募投项目效益如期实现、减少其折旧摊销等费用支出对公司业绩带来的不利影响。

25. 关于风险因素

发行人部分风险因素中包含风险对策、竞争优势或类似表述。

请发行人结合公司实际情况作风险提示，提高风险因素披露的针对性和相关性，删除风险因素中包含的风险对策、发行人竞争优势及类似表述。

答复：

发行人已在招股说明书“重大事项提示/三、特别风险提示”和“第四节 风险因素”中对风险因素中包含的风险对策、发行人竞争优势及类似表述进行了删除，具体如下：

序号	风险因素	删除的风险对策、发行人竞争优势及类似表述
1	经营业绩波动风险	整体删除
2	市场竞争加剧的风险	“保持技术、产品和服务的领先性”
3	疫情风险	无
4	新业务拓展风险	“自 2005 年成立以来，发行人一直致力于为人工智能全产业链中的企业及科研机构提供训练数据产品和服务。”
5	国际贸易摩擦风险	修改更新为“中美贸易摩擦风险”
6	核心技术快速迭代风险	“发行人已将其核心技术体系运用于设计、采集、加工、质

		检等训练数据生产相关的各个环节,构成了发行人业务持续经营的技术基础,并在报告期内保障了发行人业务规模和经营业绩的成长性。”
7	核心技术人员流失风险	“虽然发行人已设立了较为完善的激励机制对核心技术人员进行激励”以及“报告期内,发行人的核心技术人员较为稳定,除2名核心技术人员分别为2017年新增、2018年新增外,其余核心技术人员均自报告期初即在公司任职且未发生变动。”
8	核心技术失密风险	“始终注重充分挖掘和培养自身的核心技术优势”以及“在多年的行业实践过程中,发行人自主研发了一体化数据处理平台,在基础研究、平台工具、训练数据开发三个维度均积累下核心技术。”
9	数据安全相关风险	“发行人高度重视数据安全问题,不断进行数据管理体系完善和信息安全管理体的建立和认证。”
10	训练数据产品无法实现授权销售的风险	无
11	采购成本上升风险	无
12	税收优惠不确定的风险	无
13	汇率波动风险	无
14	实际控制人控制不当的风险	“尽管公司通过相关制度安排尽可能避免实际控制人操纵公司的现象发生”
15	募集资金金额较大及募投项目实施的风险	整体删除,并在招股说明书“重大事项提示/五、募集资金投资项目实施对发行人未来业绩的影响”对该情形进行了充分披露及风险提示。
16	净资产收益率下降的风险	无
17	发行失败风险	整体删除
18	股价波动风险	整体删除

此外,发行人对风险因素的顺序和“疫情风险”、“新业务拓展风险”、“国际贸易摩擦风险”和“数据安全相关风险”的标题和内容表述进行了部分调整,进一步提高风险因素披露的针对性和相关性。

26. 关于业务数据来源

请发行人说明“业务与技术”章节引用数据来源的权威性,说明数据是否公开、是否专门为编写本次招股说明书而准备,发行人是否就获得此数据支付费用或提供帮助。

答复:

发行人说明事项

(1) 请发行人说明“业务与技术”章节引用数据来源的权威性，说明数据是否公开、是否专门为编写本次招股说明书而准备

招股说明书“业务与技术”章节中引用的数据除经审计的财务数据、发行人内部统计的业务数据外，其余主要为引用的外部行业相关数据。这些行业数据主要为行业相关数据及同行业可比公司相关数据，该等数据主要来自行业研究报告，部分来自于可比公司官方网站、政府网站、有关可比上市公司年报等公开发布数据。数据来源公开且具有权威性，不存在为编写本次招股说明书而准备的情形。

一、“业务与技术”章节引用数据情况

“业务与技术”章节引用数据情况如下：

序号	数据出现于“业务与技术”章节的摘录	数据来源报告名称	数据来源机构名称	数据获取渠道
1	“...约 1/3 的算法模型每月至少更新一次，约 1/4 的算法模型每日至少更新一次...”	《Notes From The AI Frontier Insights From Hundreds of Use Cases》	McKinsey Global Institute	公开渠道
2	“...作为推动人工智能技术进步的“三驾马车”，算法、数据和算力在过去的 5-10 年间不断优化、创新，AI 产业已全面进入机器学习时代，未来人工智能的发展路径将围绕着关键技术与实体经济的深度融合，全球人工智能的市场规模有望进一步扩大...”	《全球人工智能发展白皮书》	德勤 (Deloitte)	公开渠道
3	“...2019 年，全球人工智能产业规模达到 375 亿美元，预计 2023 年全球人工智能市场规模将达到 979 亿美元，年复合增长率为 27.11%...”	《Worldwide Spending on AI Systems Will Grow to Nearly \$35.8 Billion in 2019》	国际数据公司 (IDC)	公开渠道
4	《2018E-2022E 全球人工智能商业价值》	《Gartner Says Global AI Business Value to Reach \$1.2 Trillion in 2018》	高德纳咨询 (Gartner)	公开渠道
5	《2019 年中国与全球人工智能行业发展情况对比》	《2020 年中国 AI 基础数据服务行业研究报告》	艾瑞咨询	公开渠道
6	《2018-2023E 中国人工智能市场规模》	《2019-2020 中国人工智能计算力发展评估报告》	国际数据公司 (IDC)	公开渠道
7	《训练数据需求量》	《AI and Machine Learning Projects Are Obstructed by Data Issues》	Dimensional Research	公开渠道
8	《受访者遇到的与数据相关的难题及比例》	《AI and Machine Learning Projects Are Obstructed by Data Issues》	Dimensional Research	公开渠道
9	《中国 AI 基础数据服务行业市场规模与需求类型》	《2020 年中国 AI 基础数据服务行业研究报告》	艾瑞咨询	公开渠道
10	《中国 AI 基础数据服务行业主要市场参与者》	《2020 年中国 AI 基础数据服务行业研究报告》	艾瑞咨询	公开渠道

11	《全球数据量》及《中国数据量》	《IDC：2025年中国将拥有全球最大的数据圈》	国际数据公司（IDC）	公开渠道
12	“...揭榜工作主要涵盖智能产品、核心基础、智能制造关键技术装备和支撑体系四大方向，四大方向下共囊括17个细分领域，海天瑞声所属的细分领域为支撑体系下的行业训练资源库。揭榜工作将在每个细分领域择优遴选不超过5家揭榜单位，择优公布揭榜成功单位不超过3家...”	《新一代人工智能产业创新重点任务揭榜工作方案》解读	中华人民共和国工业和信息化部	公开渠道
13	“...在我国每年需要进行标注的语音数据超过200万小时...”	《2019年中国人工智能基础数据服务行业白皮书》	艾瑞咨询	公开渠道
14	《发行人与其同行业可比公司在经营情况、市场地位、技术实力、衡量核心竞争力的关键业务数据、指标等方面的比较情况》	各可比公司官方网站、中国国家知识产权局、澳大利亚知识产权局	各可比公司官方网站、中国国家知识产权局、澳大利亚知识产权局	公开渠道

二、数据来源所涉及机构的基本情况及其权威性

上述数据来源涉及机构的基本情况如下所述，均为知名研究机构、同行业可比公司、政府部门等，与发行人所处细分行业相关，数据公开且具备权威性。

1、McKinsey Global Institute

麦肯锡全球研究院（MGI）是成立于1926年的全球知名管理咨询公司麦肯锡（McKinsey & Company）的商业和经济研究机构，旨在为各领域的领导者提供基于管理和政策决策的事实和见解，其官方网站为（<https://www.mckinsey.com.cn/insights/mckinsey-global-institute/>）

麦肯锡全球研究院（MGI）的报告数据被上市公司招股说明书采用，如乐鑫信息科技(上海)股份有限公司、北京慧辰资道资讯股份有限公司招股说明书中均有采用麦肯锡全球研究院（MGI）数据。

2、德勤（Deloitte）

德勤（Deloitte）成立于1845年，在全球150多个国家拥有超过20万名专业服务人员，在中国拥有超过1500名的双语专业咨询专家与顾问，在“四大”中拥有规模最大最为丰富的管理咨询力量。以北京、上海、深圳、哈尔滨、成都及香港为六大人才集聚地，通过德勤中国的丰富服务网络，为整个大中华区的大中型国有企业、在华跨国公司、

高科技高成长的民营企业提供覆盖企业全价值链的高质量行业专精化的战略与运营管理咨询，人力资本咨询和信息技术服务，其官方网站为（<https://www2.deloitte.com/cn/zh.html>）。

德勤（Deloitte）的报告数据被上市公司招股说明书采用，如优刻得科技股份有限公司、江苏北人机器人系统股份有限公司、上海美迪西生物医药股份有限公司招股说明书中均有采用德勤（Deloitte）数据。

3、国际数据公司（IDC）

国际数据公司（IDC）是成立于 1964 年的国际数据集团（IDG）的子公司，致力于提供信息技术、电信和消费等行业的咨询服务，国际数据公司（IDC）目前共有一千多名分析师，业务范围涵盖全球一百多个国家，于 1982 年正式在中国设立分支机构，其官方网站为（www.idc.com）。

国际数据公司（IDC）的报告数据被上市公司招股说明书采用，如深圳震有科技股份有限公司、泽达易盛（天津）科技股份有限公司、上海韦尔半导体股份有限公司、深圳市超频三科技股份有限公司招股说明书中均有采用国际数据公司（IDC）数据。

4、高德纳咨询（Gartner）

高德纳咨询（Gartner）成立于 1979 年，是全球知名的信息技术研究和咨询机构，纽约证券交易所上市公司，总部位于美国，专注于 IT 产业的研究分析，致力于协助客户进行市场分析、技术选择、项目论证、投资决策，其官方网站为（www.gartner.com）。

高德纳咨询（Gartner）的报告数据被上市公司招股说明书采用，如山石网科通信技术股份有限公司、中科寒武纪科技股份有限公司、科大国盾量子技术股份有限公司、杭州迪普科技股份有限公司、鹏鼎控股（深圳）股份有限公司招股说明书中均有采用高德纳咨询（Gartner）数据。

5、艾瑞咨询

上海艾瑞市场咨询股份有限公司（简称“艾瑞咨询”），成立于 2002 年，是国内较早从事互联网研究的第三方研究机构，在国内互联网新经济领域拥有 18 年的研究和积累，自 2003 年开始研究中国互联网产业，已经累计出版超过 3000 份互联网研究报告，涵盖互联网、移动互联网、电子商务、人工智能等各个领域。艾瑞咨询专注新技术、新

经济领域 18 年，超过 200 位专家研究团队，研究范围广泛，累计发布上千份行业报告。近年来，艾瑞咨询在中国人工智能及 AI 基础数据领域发布了多篇研究报告，其研究报告具备权威性，其官方网站为（<https://www.iresearch.com.cn/>）。

艾瑞咨询的统计数据多次被上市公司招股说明书采用，如杭州当虹科技股份有限公司、杭州光云科技股份有限公司、北京金山办公软件股份有限公司均采用艾瑞咨询研究的公开数据。

6、Dimensional Research

Dimensional Research 是一家科技领域专业的实践型市场调查公司，已服务于 VMware、Zendesk、HP、DELL 等多家信息科技公司，其官方网站为（<https://dimensionalresearch.com/>）。

7、中华人民共和国工业和信息化部

中华人民共和国工业和信息化部是国务院直属部门，为政府部门官方信息披露平台，主要职责为：拟订实施行业规划、产业政策和标准；监测工业行业日常运行；推动重大技术装备发展和自主创新；管理通信业；指导推进信息化建设；协调维护国家信息安全等，其官方网站为（<http://www.miit.gov.cn/>）。

8、各可比公司官方网站

各可比公司官方网站为各可比公司官方的公开信息披露平台。

Appen 官方网站为（<https://appen.com/>）。

慧听科技官方网站为（<https://www.huitingtech.com/>）。

标贝科技官方网站为（<https://www.data-baker.com/>）。

9、中国国家知识产权局

中国国家知识产权局的中国及多国专利审查信息查询系统为确认相关专利状态的官方渠道，官方网站为（<http://cpquery.cnipa.gov.cn>）。

10、澳大利亚知识产权局

澳大利亚知识产权局是在澳大利亚境内管理专利授予的澳大利亚政府部门,包括执行专利授予前审查、对标准专利授予前的异议进行裁决等,官方网站为(<https://www.ipaustralia.gov.au/>)。

(2) 请发行人说明发行人是否就获得此数据支付费用或提供帮助

招股说明书“业务与技术”章节中引用的数据来源如上述表格所示,该等数据主要来自行业研究报告,部分来自于可比公司官方网站、政府网站、有关可比上市公司年报等公开发布数据。上述数据不存在专门为编写本次招股说明书而准备的情形,发行人并未就获得上述数据支付费用或提供帮助。

27. 关于媒体质疑

请保荐机构自查与发行人本次公开发行相关的媒体质疑情况,并就相关媒体质疑核查并发表明确意见。

答复:

保荐机构核查事项

海天瑞声本次申请公开发行于2020年2月28日在中国证监会北京监管局官网披露辅导备案情况,于2020年6月22日在上海证券交易所官网披露招股书,7月17日进入问询阶段。保荐机构通过网络搜索方式对媒体报道进行了全面搜索、查阅,核查媒体质疑情况。截至目前,针对公司本次申请首次公开发行股票并在科创板上市事宜,主要媒体报道摘录如下:

序号	时间	媒体	标题	主要关注内容
1	2020/3/4	第一财经	《财经早班车》:海天瑞声重启科创板 IPO	重启 IPO 简讯
2	2020/3/17	上海证券报	多个科创板“落榜考生”重启 IPO 征程	简介公司业务及客户情况;重启 IPO 情况
3	2020/6/22	上海证券报	海天瑞声科创板 IPO 获受理,拟募资 7.66 亿元	受理简讯;撤回首次申报材料后引进机构投资者情况;介绍公司基本情况、经营业绩、所获荣誉、未来发展战略及募投资金用途
4	2020/6/22	中国证券报	海天瑞声:重启科创板 IPO 获受理 公司拟募资 7.66 亿元	受理简讯;撤回首次申报材料后引进机构投资者情况;介绍公司基本情况、所获荣誉、未来发展战略及募投资金用途;科创板上市要求契合度
5	2020/6/23	上海证券报	海天瑞声二度冲刺 科创板再迎四“考生”	简述首次申报历程及撤回材料后引进机构投资者情况;本次募投资金用途;公司业务及客户情况
6	2020/6/23	科创板日报	海天瑞声二度冲击科创板:专利从无到有,又获两家知名机构入股	两次申报间核心技术与专利及股权结构变化情况

7	2020/7/20	资本邦	再下一城！海天瑞声科创板 IPO 申请近日已获上交所问询	问询简讯；公司基本情况、经营业绩
8	2020/8/12	集微网	【IPO 价值观】AI 训练数据提供商海天瑞声再度 IPO 业绩依赖前五大客户	公司基本情况；公司前五大客户占比情况
9	2020/8/15	集微网	【IPO 价值观】三年营收复合增长率 41%，海天瑞声五大挑战隐忧仍存；澜起上半年净利 6.02 亿；创业板注册制首批企业 8 月 24 日上市	公司所处行业情况；公司所处行业未来面临的挑战

保荐机构针对包含上述媒体报道在内的主要报道进行了全文查阅，上述媒体报道内容主要系对发行人已披露招股说明书等公开信息的简单摘录与评论。《上海证券报》、《中国证券报》等主流财经媒体陈述性报道公司科创板申报获得受理，少量媒体对比陈述了公司两次申报的前后变化。媒体报道整体呈现正面、中性的内容。

截至目前，媒体未对本次发行的发行申请文件的真实性、准确性及完整性提出质疑。

（1）核查程序

通过网络搜索方式对媒体报道进行了全面搜索、查阅，核查媒体质疑情况。

（2）核查结论

经核查，截至本回复出具日，媒体未对发行人公开发行上市信息披露的真实性、准确性、完整性提出质疑。

七、保荐机构在充分核查基础上的总体意见

对本回复材料中的发行人回复（包括补充披露和说明的事项），本保荐机构均已进行核查，确认并保证其真实、完整、准确。

（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申请文件审核问询函的回复》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审核问询函回复的全部内容，确认审核问询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，为华泰联合证券有限责任公司《关于北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市申请文件审核问询函的回复》之签章页）

保荐代表人： 张鹏 葛青
张 鹏 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审核问询函回复的全部内容，了解审核问询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

