

2023年中国语音合成声学（TTS）行业词条报告

作者：唐文卿

行业：头豹分类/综合及概念/元宇宙/元宇宙底层技术/人工智能技术

头豹分类/信息传输、软件和信息技术服务业/软件和信息技术服务业/信息系统集成和物联网技术服务港股分类法/信息科技/软件服务

关键词：TTS、语音合成技术、人工智能

摘要：近年来，智能语音行业多家企业在上市程序上折戟沉沙。2月19日，上交所发布公告称，云知声和其保荐人中金公司于2021年2月18日向上交所提交了撤回IPO申请文件。根据规定，上交所决定终止云知声的发行上市审核。智能语音在上市程序上的表现不禁让市场疑惑，作为AI领域的重要组成部分，智能语音为何屡屡受挫，长期亏损？是技术原因还是市场导致？本文主要关注于智能语音中合成语音（TTS）部分的市场现状，了解TTS技术的现状，产业链情况，分析了TTS的技术应用场景以及未来的发展难点和趋势。

1. 语音合成（TTS）行业定义

语音合成又称文语转换（Text to Speech,TTS）技术，是语音处理领域的一个重要研究方向，旨在让机器生成自然动听的人类语音。语音合成技术既可以单独适用于不同的场景之内，也可以作为局部环节嵌入到语音交互的整体方案之中。语音合成的应用可以划分为单向语音输出类和交互类，单一使用单向语音输出类或交互类的场景并不多见，导航技术、阅读、配音、语音播报等场景下，单向语音输出应用比例较大，智能客服、智能机器人、泛娱乐产业、教育等场景下，交互类语音合成使用较多。

2. 语音合成（TTS）行业分类

语音合成技术内部分为前端和后端。前端主要负责文本的语言解析和处理，其处理内容主要包括语种、分词、词性预测、多音字处理、韵律预测、情感等。把文本上的发音的这些信息进行预测出来之后，将信息发送给TTS的后端系统，后台声学系统融合这些信息之后，将内容转化为语音。后端声学系统发展历程较长，从第一代的语音拼接合成，到第二代的语音参数合成，到第三代端到端合成，后端声学系统的智能化程度逐步增加，训练素材需要标记的详细程度和难度也在逐步减弱。

≡ 类型名称	≡ 类型说明
前端语言分析	负责文本的语言解析和处理，其处理内容主要包括语种、分词、词性预测、多音字处理、韵律预测、情感等。
后端声学系统	通过前端语言分析把文本上的发音的这些信息进行预测出来之后，将信息发送给TTS的后端系统，后台声学系统融合这些信息之后，将内容转化为语音。

3. 语音合成（TTS）行业特征

语音合成的产业链上中游融合是一体化趋势，其中中游有较多新企业出现去填补个性化TTS的空缺，但个性化TTS的制作成本昂贵，因此导致较多企业尚未实现盈利。一旦产品研发成功便能实现高达60%的毛利率，所以TTS行业发展较为成熟，进入难度较大。

一体化趋势明显
TTS产业链的上中游呈现一体化的融合趋势
语音合成企业需要向上游购买训练数据或服务，以达成人工智能模型的实现。部分数据标记企业为了使语音能够可持续发展而不受到产业自身限制，也开始拓展自己的业务边界，开拓了自有TTS的产品线。

研发投入较大
个性化TTS的制作成本相对高昂，前期研发投入较多导致部分TTS企业尚未实现盈利
个性化TTS在通用TTS的基础上更加注重音色和情感，情感和声音的表现力比通用型TTS更强，这类声音主要应用于对声音质量较高的教育，长音频、直播以及影视游戏配音等场景当中。现阶段，互联网大企业涉足个性化TTS服务这一业务较少，创新型中小企业踊跃出现。个性化TTS的制作成本相对高昂，TTS中游利润率保持在60-80%之间，但前期研发投入较多导致部分TTS企业尚未实现盈利。

进入难度大
TTS行业发展成熟，新进入者难以生存
TTS行业的发展可追溯到1990s，到现阶段，技术发展较为成熟。原有竞争者已经占领了结构性优势，保持其原有地位。除非新进入者能够弯道超车，或实现产品差异化，才有可能生存下来。但鉴于TTS的发展已经较为成熟，因此在技术的突破范围有限。这造成了新进入者难以发展的局面。

4. 语音合成（TTS）发展历程

语音识别技术最早可以追溯到上个世纪50年代的第一个口头数字机器Audrey，大规模的语音识别研究是在进入了70年代以后，在小词汇量、孤立词的识别方面取得了实质性的进展。进入80年代以后，在研究思路上发生了重大变化，提出了将神经网络技术引入语音识别问题的技术思路。进入90年代以后，在语音识别技术的应用及产品化方面出现了很大的进展。2010年后，机器学习的算法提升和计算机性能进步，基于RNNs神经网络模型，使得语音识别达到更好的精度与准确度，从而在应用领域提供了更多的可能性，带动了相关产品的面世。

开始时间：1950 结束时间：1980 阶段：萌芽期

行业动态：贝尔研究所Davis等人研究成功了世界上第一个能识别10个英文数字发音的实验系统，能够理解口头数字的机器Audrey。1960年英国的Denes等人研究成功了第一个计算机语音识别系统。

开始时间：1980 结束时间：2010 阶段：启动期

行业动态：研究的重点逐渐转向大词汇量、非特定人连续语音识别。在研究思路上也发生了重大变化，即由传统的基于模板匹配的技术思路开始转向基于统计模型（HMM）的技术思路。此外，再次提出了将神经网络技术引入语音识别问题的技术思路。同时，在语音识别技术的应用及产品化方面出现了很大的进展。

开始时间：2011 结束时间：2016 阶段：变革发展期

行业动态：2010年，机器学习算法和计算机性能的进步带来了更有效的训练深层神经网络（DNN）的方法。因此，语音识别系统开始使用DNNs，具体地说，是使用一种DNNs的特殊变体，即循环神经网络（RNNs）。此后，基于RNNs的模型表现出比传统模型更好的精度和性能。2016年的语音识别准确率达到了90%。

开始时间：2016 结束时间：2022 阶段：落地应用期

行业动态：端到端的语音识别开始广泛运用，针对远场的语音识别和唤醒得到进一步发展，全双工语音交互开始出现，此阶段的语音识别准确率可达98%以上，且能根据实际应用痛点针对性优化。

5. 语音合成（TTS）产业链分析

语音合成的产业链上中游呈现**一定的一体化趋势**，尤其是上中游融合趋势较为明显，中游竞争格局已初步显现，下游应用场景多样，**商业化落地实现盈利成为重点**。为了满足企业自建的语音合成训练的要求，部分大型互联网企业建立自有语音标准公司。同时，语音合成产业链上游由于产业自身存在一定局限，因此，产业链上中游一体化或将成为重要发展趋势。随着5G时代的来临，通讯业逐渐发展，社会整体需要快速高效的信息传播方式，而语音相比于文字的信息传播速度更高。城市的扩大，人口密度的增加也为用户带来大量碎片化时间，语音内容受到社会欢迎，因此，语音合成也找到了用武之地。

≡ 上游环节	≡ 上游说明	≡ 上游参与方
数据产品及服务	人工智能数据产品及服务隶属于人工智能产业链的基础层，是自主研发人工智能技术的企业与机构必不可少的基础生产要素，其数量多寡和质量高低将会直接影响到人工智能产业链内企业的研发周期、产品性能和可扩展性。语音合成企业需要向上游购买训练数据或服务，以达成人工智能模型的实现。	北京百度网讯科技有限公司、阿里巴巴（中国）网络技术有限公司、北京海天瑞声科技股份有限公司、北京爱数智科技有限公司、信赛信息技术有限公司（北京）有限公司、深圳莫比诺客股份有限公司、腾讯云计算（北京）有限责任公司

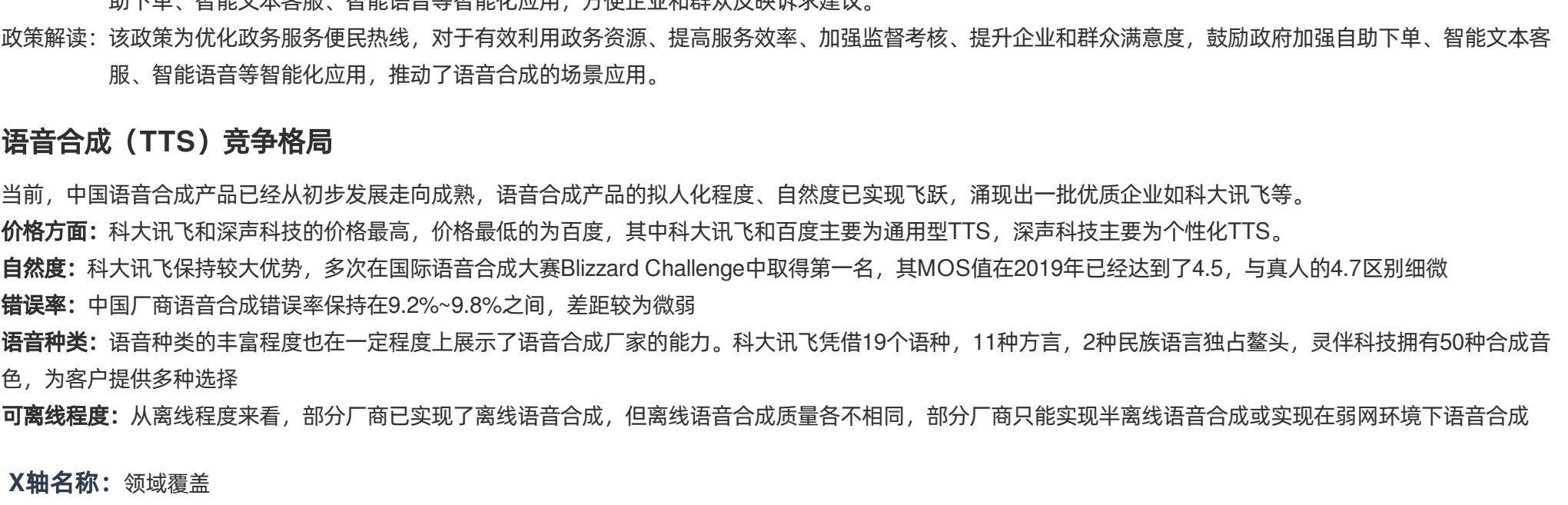
≡ 中游环节	≡ 中游说明	≡ 中游参与方
语音合成行业	语音合成主要分为 通用TTS发育和个性化TTS发育 。通用TTS发育相对的机械化，但其优点在于通用TTS口音清晰。因此，通用TTS适用于导航、语音播报、智能客服和大多数语音交互场景， 制作成本整体而言较为低廉 。个性化TTS在通用TTS的基础上更加注重音色和情感，这类声音主要应用于对声音质量较高的教育，长音频、直播以及影视游戏配音等场景当中。现阶段， 互联网大企业涉足个性化TTS服务这一业务较少，创新型中小企业踊跃出现 。相比通用型TTS， 个性化TTS的制作成本相对高昂。TTS中游利润保持在60-80%，但前期研发投入较大，部分企业仍未盈利 。	云知声智能科技股份有限公司、杭州倒映有声科技有限公司、思必驰科技股份有限公司、微软（中国）有限公司、科大讯飞股份有限公司、标贝（北京）科技有限公司、北京百度网讯科技有限公司、阿里巴巴（中国）网络技术有限公司

≡ 下游环节	≡ 下游说明	≡ 下游参与方
应用领域	个性化TTS最早落地于 导航领域产品 ，其原因主要在于导航语音的语术较为固定，只有约20%变量的部分是用TTS合成出来的，基于个性化TTS的早期水平并不成熟，因此，个性化TTS的IP价值首先体现在了导航领域。 长音频是个性化TTS发展的有利平台 ，TTS技术能够大大减少将文字转为语音的时间和财务成本，极大的压缩了长音频广播制，有声书等的制作周期，能够更快节省的，及时的将长音频作品呈现到用户眼前。	高德软件有限公司、中文在线数字出版集团股份有限公司、上海喜马拉雅科技股份有限公司、掌阅科技股份有限公司、北京聚力教育科技有限公司、北京捷派教育咨询有限公司

6. 语音合成（TTS）长音频行业规模

中国用户在网络音频上的MAU已经从2019年的342分钟快速增长至2020年的600分钟，**长音频市场未来发展潜力巨大**，长音频作品大量都基于文字脚本而来，TTS技术能够大大减少将文字转为语音的时间和财务成本，也为长音频作品带来创新体验。根据腾讯音乐2020年Q2到Q3两个季度的财报，公司的长音频授权作品数量同比均迅速增加三至四倍，内容涉及文学作品、广播剧和脱口秀等各种类型。来自Q3财报的数据显示，长音频用户数进一步提升，其MAU渗透率从去年同期的4.7%提高至11.7%，其中长音频IP授权数量增至去年同期的四倍，持续提升在长音频方面的平均用户时长。

中国长音频用户规模



头豹研究院测算

头豹研究院

7. 语音合成（TTS）政策梳理

政策名称：《国务院关于积极推进“互联网+”行动的指导意见》 颁布主体：国务院 生效日期：2015-07 影响：7 政策性质：鼓励性政策

政策内容：进一步推进计算机视觉、智能语音处理、生物特征识别、自然语言理解、智能决策控制以及新型人机交互等关键技术的研发和产业化，推动人工智能在智能产品、工业制造。

政策解读：此政策推动依托互联网平台提供人工智能公共创新服务，加快人工智能核心技术突破，促进人工智能在智能家居、智能终端、智能汽车、机器人等领域的推广应用，培育若干引领全球人工智能发展的骨干企业和创新团队，形成创新活跃、开放合作、协同发展的产业生态。

政策名称：《关于促进“互联网+医疗健康”发展的意见》 颁布主体：国务院 生效日期：2018-04 影响：6 政策性质：鼓励性政策

政策内容：研发基于人工智能的临床诊疗决策支持系统，开展智能医学影像识别、病理分型和多学科会诊以及多种医疗健康场景下的智能语音技术应用，提高医疗服务效率。

政策解读：政策鼓励通过多种医疗健康场景下的智能语音技术应用以提高医疗服务效率，此政策推动了智能语音在医疗行业的运用。

政策名称：《最高人民法院关于深化人民法院司法体制综合配套改革的意见》 颁布主体：最高人民法院 生效日期：2019-02 影响：8 政策性质：鼓励性政策

政策内容：全面提升语音识别技术在庭审语言同步转录中的应用效能，建成全国法院智能语音云平台，实现全国法院语音识别的模型共享和统一管理。

政策解读：此政策为智能语音的运用场景提出了建成全国法院智能语音云平台，这对智能语音行业提供了应用场景以及行业的发展方向，也对智能语音对政务上的业务效率提升表示了肯定。

政策名称：《鼓励外商投资产业目录(2020年版)》 颁布主体：发改委 生效日期：2020-12 影响：6 政策性质：鼓励性政策

政策内容：鼓励智能语音、量子通信等设备研发、制造

政策解读：此政策鼓励外商进入中国进行智能语音行业的投资，对智能语音行业看好，并在政策上给予相应支持，促进该行业的未来发展以及带动相关的产业。

政策名称：《关于进一步优化地方政务服务便民热线的指导意见》 颁布主体：国务院办公厅 生效日期：2021-01 影响：8 政策性质：鼓励性政策

政策内容：各地区要做好热线接通能力保障建设，提供与需求相适应的人工服务，同时拓展互联网渠道，丰富受理方式，满足企业和群众个性化、多样化需求。加强自助下单、智能文本客服、智能语音等智能化应用，方便企业和群众反映诉求建议。

政策解读：该政策为优化政务服务便民热线，对于有效利用政务资源、提高服务效率、加强监督考核、提升企业和群众满意度，鼓励政府加强自助下单、智能文本客服、智能语音等智能化应用，推动了语音合成的场景应用。

8. 语音合成（TTS）竞争格局

当前，中国语音合成产品已经从初步发展走向成熟，语音合成产品的拟人化程度、自然度已实现飞跃，涌现出一批优质企业如科大讯飞等。

价格方面：科大讯飞和深声科技的价格最高，价格最低的为百度，其中科大讯飞和百度主要为通用型TTS，深声科技主要为个性化TTS。

自然度：科大讯飞保持较大优势，多次在国际语音合成大赛Blizzard Challenge中取得第一名，其MOS值在2019年已经达到了4.5，与真人的4.7区别细微

错误率：中国厂商语音合成错误率保持在9.2%-9.8%之间，差距较为微弱

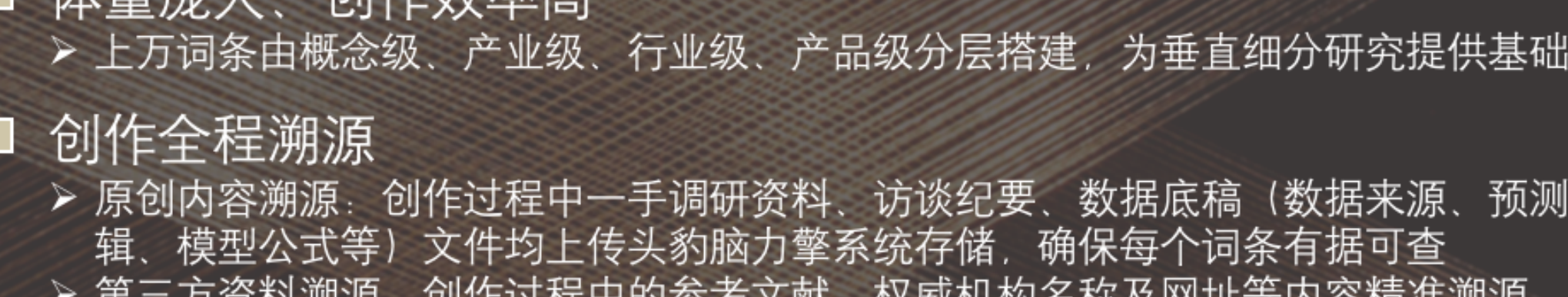
语音种类：语音种类的丰富程度也在一定程度上展示了语音合成厂家的能力。科大讯飞凭借19个语种，11种方言，2种民族语言独占鳌头，灵伴科技拥有50种合成音色，为客户提供更多选择

可离线路度：从离线路度来看，部分厂商已实现了离线路语音合成，但离线路语音合成质量各不相同，部分厂商只能实现半离线路语音合成或实现在弱网环境下语音合成

X轴名称：领域覆盖	Y轴名称：价格	α轴名称：语音种类	β轴名称：自然度
-----------	---------	-----------	----------

≡ 竞争参与方	↗ 横轴	↑ 纵轴	≡ 气泡大小	≡ 气泡色深	≡ 得分依据	⊙ 气泡显示	⊙ 名称显示
科大讯飞股份有限公司	4.6	2.7	4.8	4.5		☑	☑
云知声智能科技股份有限公司	4	4.5	3	3.9		☑	☑
广州深声科技有限公司	4	2.5	4	4.1		☑	☑
杭州灵伴科技有限公司	2.5	2.7	4	3.5		☑	☑
北京百度网讯科技有限公司	4.3	4.6	4.3	3.8		☑	☑

气泡大小表示：语音种类；气泡色深表示：自然度



x轴为企业的领域覆盖度，y轴为企业产品的价格指数，气泡的大小为企业产品支持的语音种类，气泡颜色的深度为企业产品的自然度。

上市公司速览

≡ 股票代码	≡ 上市公司	⊕ 总市值	⊕ 营收规模	同比增长(%)	毛利率(%)
002230	安徽科大讯飞信息科技股份有限公司	82978400000	350,598.50万元	40.17	40.53
688100	威胜信息技术股份有限公司	10970000000	94,843.20万元	11.47	36.94
300229	拓尔思信息技术股份有限公司	7813000000	42,494.86万元	8.53	67.95
09988	阿里巴巴集团控股有限公司	37640.74亿	853,062,000,000.00	18.9300	--
00700	腾讯控股有限公司	3493000000000	560,118,000,000.00	16.1900	43.91

9. 语音合成声学（TTS）代表企业分析

杭州倒映有声科技有限公司

企业状态	存续	注册资本	123.4568万人民币
企业总部	杭州市	行业	软件和信息技术服务业
法人	肖翔	统一社会信用代码	91330110MA2GMLLM6E
企业类型	有限责任公司(自然人投资或控股)	成立时间	2019-05-28
品牌名称	杭州倒映有声科技有限公司		
经营范围	服务：基础软件服务，应用软件服务，计算机系统集成，计算机系统集成、互联网信息服务，产品设计，网页设计，电脑动画设计，个人形象设计，市场营销策划，企业形象策划，技术进出口（法律、行政法规禁止经营的项目除外，法律、行政法规限制经营的项目取得许可后方可经营）（依法须经批准的项目，经相关部门批准后方可开展经营活动）		

杭州倒映有声科技有限公司融资信息

融资时间	披露时间	投资企业	金额	轮次	投资比例	估值
2021-05-10	2021-05-10	中资在线	千万级人民币	Pre-A轮		
	2019-08-07	天使湾创投，凌晨资本	未披露	天使轮		

杭州倒映有声科技有限公司竞争优势

	倒映有声是一家致力于利用语音合成技术打造AI音频科技创新的高新技术企业。该公司研发的基于端对端的神经网络音频合成系统和解决方案，在音色模拟、情感展现、定制化服务、多语种等方面表现较好。2021年5月，倒映有声取得了来自中文在线的千万级Pre-A轮融资。
	倒映有声的技术优势主要存在于个性化TTS而非通用TTS，倒映有声的产品技术应用场景主要应用于有声书阅读、广播剧、游戏等泛娱乐领域以及教育等场景。这些场景对TTS的质量和情感有较高的要求，更能发挥倒映有声的差异化技术优势。倒映有声已经与超过5家版权方合作，上线了共10多部作品。
	倒映有声的产品主要分为三类：（1）第一类为有声合成系统，用户可以通过云端产品，自助或半自助的实现文字作品的有声化，该产品主要服务于版权方和融媒体中心；（2）第二类为AI主播定制，主要由自有AI主播授权生成和定制服务，客户可以去倒映有声的库里选择满足自己需求的AI主播声音，也可以根据自身需求定制属于自己的AI主播，倒映有声会通过联合语音合成的上下游为客户提供定制音频的全流程服务；（3）第三类产品的覆盖范围更加广泛，基于倒映有声的产品特点和技术优势，企业致力于布局TTS在泛娱乐行业的应用，其中包括对于文学的承包制作和IP运营，倒映有声不仅仅把产品的目光需锚于技术层面，而为TTS找到了商业落地的途径，通过这种方式打开语音合成应用的途径。这种将技术和商业应用相结合的方式有利于倒映有声抢占市场蓝海。
	中国市场上目前个性化语音合成的企业有限，倒映有声作为为数不多的企业之一，2021年5月，倒映有声取得了来自中文在线的千万级Pre-A轮融资，市场前景看好企业在行业中的发展。倒映有声进入行业时间早，把握市场的蓝海，占领结构性优势。倒映有声的产品专业化程度高，语音质量高，且专注于泛娱乐合成语音的打造，覆盖面专业，商业思维也较为活跃，在后疫情时代将会有较大的发展空间。

倒映有声官网

头豹“数字行研”——词条报告

■ 优质企业共建词条报告

■ 展示意见优势地位

■ 第三方数据机构应用合作招募

■ 头豹词条数据库流量赋能转化

■ 开通会员账号，查阅数据底稿

■ 市场规模、竞争格局工作底稿一览无余

详情咨询：400—072—5588

136—1163—4866

■ 体量庞大、创作效率高

➢ 上万词条由概念级、产业级、行业级、产品级分层搭建，为垂直细分研究提供基础

■ 创作全程溯源

➢ 原创内容溯源：创作过程中一手调研资料、访谈纪要、数据底稿（数据来源、预测逻辑、模型公式等）文件均上传头豹脑力擎系统存储，确保每个词条有据可查

➢ 第三方资料溯源：创作过程中的参考文献、权威机构名称及网址等内容精准溯源

➢ AI生成类内容溯源：AI生成的内容进行区分、标注

■ 科技赋能

➢ 脑力擎系统：词条数据库、写作指引及视频指南、溯源功能、写作助手、AI生成、专家访谈工具、数字资产确权等功能，实现数字行研

➢ 开源、扩展性：词条内涉及的公司名可与第三方企业库对接获取信息；脑力擎系统接口可与第三方对接，获取实时数据或输出数据

■ 方法论模型

➢ 词条基于头豹行企研究8-D方法论组成，概述+数据+分析相结合，内容清晰，数据量足，观点结论丰富

➢ 依托多年行研咨询经验，脑力擎Size3.0控件独创市场规模及竞争格局搭建及测算模型