

精品报告：
从语音合成（TTS）的发展探寻智
能语音行业的现状与未来

High-quality Report:
From The Development of Speech
Synthesis (TTS) Explore The Current
Situation and Future of The Intelligent
Voice Industry

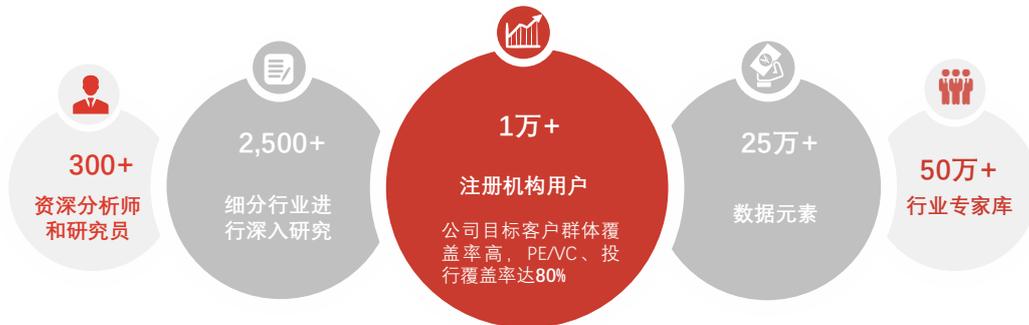
プレミアムレポート:音声合成(tts)の発展か
ら、スマート音声業界の現状と未来を探る

报告标签：语音交互、智能语音、科大讯飞

报告提供的任何内容（包括但不限于数据、文字、图表、图像等）均系头豹研究院独有的高度机密性文件（在报告中另行标明出处者除外）。未经头豹研究院事先书面许可，任何人不得以任何方式擅自复制、再造、传播、出版、引用、改编、汇编本报告内容，若有违反上述约定的行为发生，头豹研究院保留采取法律措施、追究相关人员责任的权利。头豹研究院开展的所有商业活动均使用“头豹研究院”或“头豹”的商号、商标，头豹研究院无任何前述名称之外的其他分支机构，也未授权或聘用其他任何第三方代表头豹研究院开展商业活动。

头豹研究院简介

- ◆ 头豹研究院是中国大陆地区首家**B2B模式人工智能技术的互联网商业咨询平台**，已形成集**行业研究、政企咨询、产业规划、会展会议**行业服务等业务为一体的一站式行业服务体系，整合多方资源，致力于为用户提供最专业、最完整、最省时的行业和企业数据库服务，帮助用户实现知识共建，产权共享
- ◆ 公司致力于以优质商业资源共享为基础，利用**大数据、区块链和人工智能**等技术，围绕**产业焦点、热点问题**，基于**丰富案例和海量数据**，通过开放合作的研究平台，汇集各界智慧，推动产业健康、有序、可持续发展



四大核心服务：

企业服务

为企业提供**定制化报告服务、管理咨询、战略调整**等服务

云研究院服务

提供**行业分析师外派驻场服务**，平台数据库、报告库及内部研究团队提供技术支持服务

行业排名、展会宣传

行业峰会策划、**奖项评选**、行业白皮书等服务

园区规划、产业规划

地方**产业规划**，**园区企业孵化服务**

报告阅读渠道



图说



表说



专家说



数说

头豹科技创新网 —— www.leadleo.com PC端阅读全行业、千本研报

头豹小程序 —— 微信小程序搜索“头豹”、手机扫上方二维码阅读研报

添加右侧头豹研究院分析师微信，邀您进入行研报告分享交流微信群



详情咨询



客服电话

400-072-5588



南京

杨先生：13120628075

唐先生：18014813521



上海

王先生：13611634866

李女士：13061967127



深圳

李先生：18916233114

李女士：18049912451

概览摘要

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。

■ 语音合成模型已经发展到第三代

语音合成模型经过长时间的发展，由最初的基于拼接合成，到参数合成，逐渐达到了现阶段感情充沛的基于端到端合成，最新一代端到端合成降低了对语言学知识的要求，可批量实现多语种的合成系统，语音自然程度高。

■ TTS技术主要分为两种

TTS语音主要分为通用TTS和个性化TTS，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，个性化TTS则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。

■ TTS技术助力长音频行业的发展

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。

■ 语音合成行业格局稳定，未来注重商用落地

语音合成行业发展较成熟，原有格局难以打破，新进入者只能通过弯道超车或差异化竞争打破僵局，用户的需求将会变得多样，这对语音合成行业也是巨大挑战，在未来，语音合成技术也将逐步打开商用化道路，应用场景将逐渐细化和专业。

目录

◆ 名词解释	-----	07
◆ 语音合成的定义与结构	-----	08
◆ 语音合成声学系统的发展	-----	09
◆ 语音合成的主观评价标准	-----	10
◆ 语音合成的客观评价标准	-----	11
◆ 语音合成产品质量分析评价	-----	12
◆ 语音合成产业链分析	-----	13
• 上游——数据产品及服务	-----	14
• 中游——语音合成行业	-----	16
• 下游——TTS的主要应用领域	-----	17
• 下游——TTS在导航领域的应用	-----	18
• 下游——TTS在长音频领域的应用	-----	19
• 下游——TTS在泛娱乐领域的应用	-----	20
• 下游——TTS在教育领域的应用	-----	21
◆ 语音合成的应用风险	-----	22
◆ 语音合成行业政策分析	-----	23
◆ 语音合成的发展难点与未来	-----	24
◆ 语音合成行业企业分析——科大讯飞	-----	25
◆ 语音合成行业企业分析——云知声	-----	27
◆ 语音合成行业企业对比——毛利率	-----	30
◆ 语音合成行业企业对比——销售费用	-----	31
◆ 中国语音合成优质企业——倒映有声	-----	32
◆ 海外语音合成优质企业——Sonantic	-----	33
◆ 方法论	-----	34
◆ 法律声明	-----	35

Contents

◆	Terms	07
◆	Definition and structure of TTS	08
◆	Development of TTS acoustic systems	09
◆	Subjective Evaluation Criteria for TTS	10
◆	Objective Evaluation Criteria for TTS	11
◆	Quality analysis and evaluation of speech synthesis products	12
◆	TTS Industry Chain	13
◆	• Upstream - Data Products and Services	14
◆	• Midstream – TTS Industry	16
◆	• Downstream -- TTS Applications in Navigation	17
◆	• Downstream - TTS Applications in The Long Audio Field	18
◆	• Downstream -- TTS Applications in The Field of Pan-Entertainment	19
◆	• Downstream -- Application of TTS in The Field of Education	20
◆	Application Risks of Speech Synthesis	21
◆	Speech Synthesis Industry Policy Analysis	22
◆	Development Difficulties and Future of TTS	23
◆	Enterprise Analysis of Speech Synthesis Industry -- IFlytek	24
◆	Enterprise Analysis of Speech Synthesis Industry - Cloud Voice	25
◆	Speech Synthesis Industry Enterprise Comparison - Gross Margin	27
◆	Speech Synthesis Industry Enterprise Comparison - Selling Expenses	30
◆	High-quality Enterprise of China TTS - DaoYingYouSheng	31
◆	High quality Enterprise of Overseas TTS -- Sonantic	32
◆	Methodology	33
◆	Legal Statement	34

名词解释

- ◆ **算法**：解题方案的准确而完整的描述，是一系列解决问题的清晰指令，算法代表着用系统的方法描述解决问题的策略机制
- ◆ **算力**：计算能力
- ◆ **人工智能 (AI)**：Artificial Intelligence，是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门技术科学
- ◆ **机器学习 (ML)**：Machine Learning，是一门专门研究计算机模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身性能的学科
- ◆ **深度学习 (DL)**：Deep Learning，机器学习研究中的一个新的领域，其动机在于建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如图像、声音和文本
- ◆ **深度神经网络 (DNN)**：Deep Neural Network，是一个多层的人工神经网络模型，包含了：输入层，隐藏层和输出层
- ◆ **人工神经网络 (ANN)**：Artificial Neural Networks，是一种应用类似于大脑神经突触联接的结构进行信息处理的数学模型。在工程与学术界也常直接简称为“神经网络”或“类神经网络”
- ◆ **智能语音**：实现人机语言的通信，包括语音识别技术 (ASR) 和语音合成技术 (TTS)
- ◆ **语音识别 (ASR)**：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
- ◆ **自然语言/自然语言处理 (NLP)**：Natural Language Processing,是计算机科学领域与人工智能领域中的一个重要方向，研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法
- ◆ **隐马尔可夫模型 (HMM)**：Hidden Markov Model，它用来描述一个含有隐含未知参数的马尔可夫过程

■ 语音合成的定义与结构

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

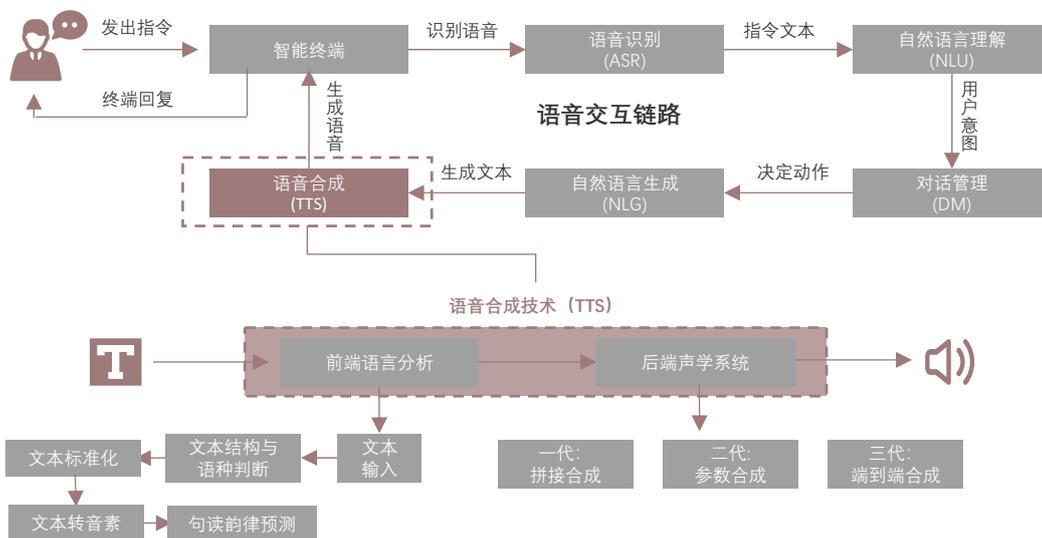
■ 语音合成让人工智能开口说话

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，也可以作为尾部环节嵌入到语音交互的整体方案之中。

■ 语音合成技术内部分为前端和后端

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都预测出来之后，将信息送给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。

TTS在语音交互场景里的应用

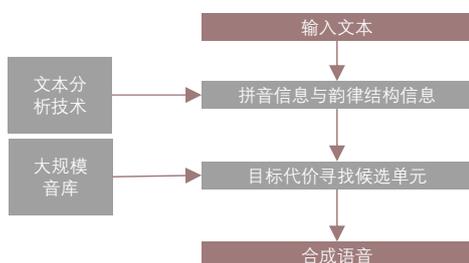


来源：头豹研究院编辑整理

■ 语音合成声学系统的发展

语音合成模型经过长时间的发展，由最初的基于拼接合成，到参数合成，逐渐达到了现阶段感情充沛的基于端到端合成，最新一代端到端合成降低了对语言学知识的要求，可批量实现多语种的合成系统，语音自然程度高

第一代：基于拼接合成

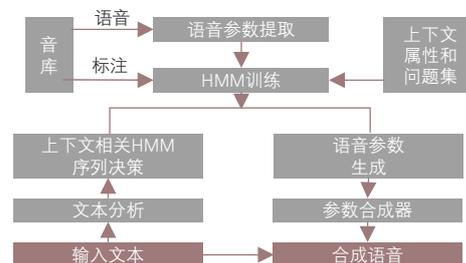


拼接合成：为了更好的拟合人声，拼接合成技术需要一个大规模的真人音库，音库内容按照音素和不同特征进行标注，合成时根据语言学特征寻找符合的音素，拼接起来完成合成

优点：效果接近真人，运算量低

缺点：拼接而成，语音不连贯，依赖音库，需要人工介入制定很多挑选规则和参数，制作成本高

第二代：基于参数合成

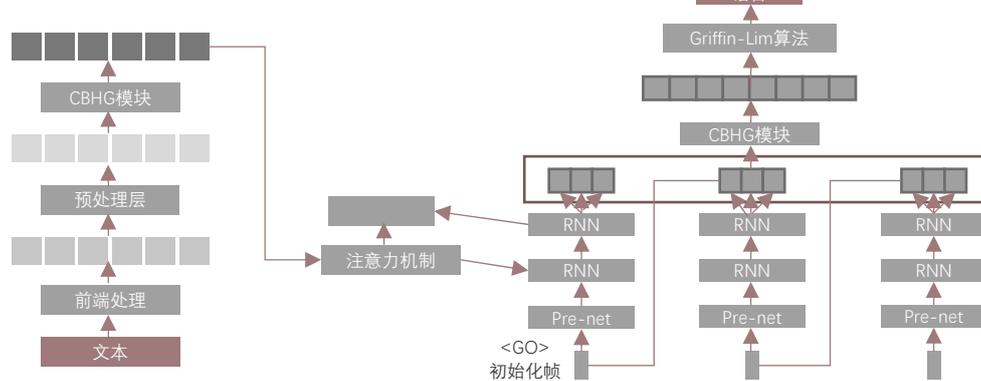


参数合成：通过深度学习构建文本特征和音库之间的映射关系，构建参数合成模型，当输入一个语言学特征时，基于神经网络给出声频特征，后通过声码器合成语音波形

优点：音库数量要求不多，合成语音连接平稳，质量高

缺点：对声码器依赖程度高，同时，由于传统参数系统建模时存在信息损失，限制了合成语音表现力的进一步提升

第三代：基于端到端（以Tacotron为例）



端到端合成：端到端语音合成一定程度上解决了拼接合成和参数合成存在的部分缺陷。端到端合成系统直接输入文本或者注音字符，通过文本或者文本特征和语音直接建模，跳过声码器阶段，减少了对声码器的依赖，弱化前端概念

优点：降低对语言学知识的要求，可方便的在不同语种上复制，批量实现几十种甚至更多语种的合成系统，语音自然程度高

缺点：运算量大，无法人工调优，实时性低

来源：头豹研究院编辑整理

■ 语音合成的主观评价标准

语音合成的主观效果指标分为Mos值测试和ABX测试两种，MOS值主要关注于语音本身的质量。而ABX则更注重通过参照对比，用来评价合成语音的适宜性和效果

语音合成MOS值测试指标

Mos评分	评级	音质	流畅度	正确性	自然度	分词与停顿	音色
5		广播级别	高	高	很自然，已无法分辨合成声音与自然人声	高	优秀
4.5	优	普通人对话水平	高	高	整体完整，没有明显的异常的韵律起伏	高	较优秀
4		一到两个音节模糊	无明显卡顿	无明显错误	无严重韵律错误	无明显错误	良好
3.5	良	偶尔有几个音节不清晰	比较流畅	错误较少	韵律起伏较为正常	较为正常	较好
3	中	有一些音节不清晰	不太流畅	有容易察觉的语言错误	有一些不太正常的韵律起伏	有一定错误	一般
2	差	有一些词不太清晰	不流畅	难以理解	基本没有韵律起伏	错误较多	差
1	劣	很不清晰	无流畅可言	基本无法理解	基本没有韵律起伏	错误较多	明显机器音

■ MOS值测试

语音合成发展历史悠久，技术较为成熟，已形成产品广泛应用于商业。作为产品，语音合成技术需要一套量化的评价标准用以评判合成语音的质量。现阶段，业内较为认可的主观评价标准为Mos (Mean Opinion Score) 值测试。Mos测试值评价会邀请业内专家，从音质、流畅程度、正确性、自然度、分词与停顿、音色6个方面，根据专家主观意见，对合成语音进行打分。再根据打分的平均值得到合成语音的最终评分。

MOS值能够有效合理的评价合成语音质量，其从使用者的感官角度出发，既评测了合成语音的声音质量，也对合成语音的拟人化、自然度做出了评价，是行业内较为通用的评价机制。但MOS值主要依靠人为打分，具有一定的主观性，同时，MOS值的打分需要许多领域内的人员进行打分，这需要昂贵的人力资源和时间开销，因此自动打分MOS的系统被提出，比如MOSNET, AutoNET等等，一定程度上替代了全人工MOS值测试。

■ ABX测评

ABX测评主要是对合成效果进行对比性测试。主要方式是通过相同文本与音色来对比不同的TTS系统的语音合成效果，ABX测评人为主观判断较多，这种方式更注重通过参照对比，来评价合成语音的适宜性和效果，有一定的优越性。

来源：头豹研究院编辑整理

■ 语音合成的客观评价标准

语音合成的客观性能指标主要包括实时率(RTF)、首包响应时间和并发数，实时率和首包响应时间注重与用户体验，而并发数关注的重点则是技术层面和生产成本

实时率(RTF) 与 首包响应时间

$$RTF = \frac{\text{文字合成所需时长}}{\text{文字合成出的音频时长}}$$

- TTS所对应的合成时间保持在500ms内较为理想

实时率是评估合成语音的参数之一，要求语音合成系统在接收到指令后快速以最快的速度合成语音，力求尽可能缩短用户的等待时间，提高用户体验。一般情况而言，语音交互期望的整体时延在2秒左右，理想情况在1.5秒，整体时延包括前端ASR、NLP等所花费的时间，因此，TTS所对应的合成时间保持在500ms内较为理想。

- 非流式合成适合语音输出，流式合成适合语音交互

语音合成分为非流式合成和流式合成，两者在合成语音实时性上关注的重点也不同。非流式合成方式一次性输入文字，一次性输出语音，注重语音合成系统的整体运算速度，该TTS方案不适合做语音交互。而流式合成、可以基于文本预测进行分词断句，分段传回合成的音频，这种语音合成方式主要关注其首包响应时间，首包响应时间越短，用户就会越快取得回应，用户等待和空白的的时间减少，不会因为等待回应而失去耐心，因此整体体验感较好，是更适合作为语音交互的TTS方案。

- 语音交互场景下离线TTS为较好选择

当前的语音合成系统分为云端TTS和离线TTS。云端TTS主要配套端到端或者多层神经网络的算法，语音输出质量更高，算力更强，但因此实时性更差，不适宜语音交互。而随着离线TTS算力以及算法得逐步更新，发展情况较好，一些参数化合成的TTS方案已经能够以10字/100ms左右的延时，合成质量也达到一定水平，适合于合成语音的交互类场景。

并发数

$$\text{单核并发数} = \frac{N\text{段文字合成出的音频时长}}{N\text{段文字合成所需时长}}$$

- 实现一核多线能有效减少成本花费

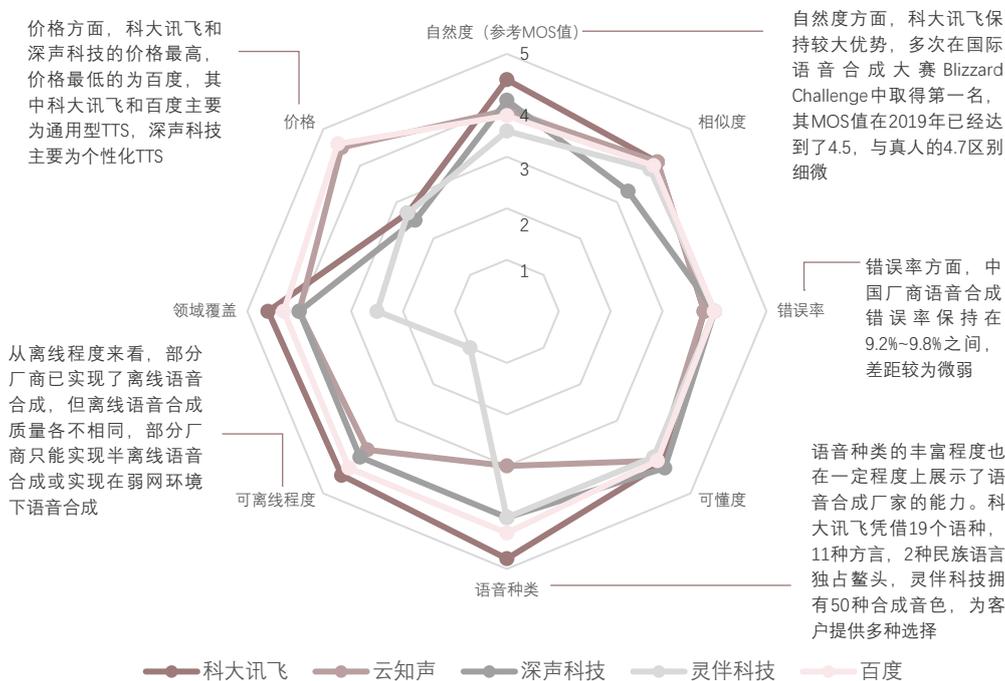
为了资源的利用率最大化，应用时保持TTS方案的RTF接近1或等于1即可，当RTF远小于1则可实现一核多线。线数即单线并发数，单线并发数越多，则需要的核数越少，并发数量的增加将减少服务器的成本费用。比如当用户需要200线程的语音合成并发数，如果实现一核5线，就需要64核CPU服务器实现，如果能达到一核10线，则24核的CPU服务器即可满足客户需求，为客户节省了成本。

来源：头豹研究院编辑整理

■ 语音合成产品质量分析评价

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

中国语音合成厂商产品质量雷达图，2021年



科大讯飞：行业内领军企业，在语音合成乃至智能语音行业均属于龙头企业，其技术水平和产品质量均较好，不足在于产品价格较高



百度：整体技术和产品质量均较好，以通用TTS为主，价格是多家中最低的，应用领域较为广泛，定价多元，市场认可度较高



云知声：产品性能稍逊色于科大讯飞和百度，但依旧受到市场认可，其产品的性价比较高，但可选择的语音和语种较少

来源：各企业官网，Blizzard Challenge官网，头豹研究院编辑整理

■ 语音合成产业链分析

语音合成的产业链上中下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，商用化落地实现盈利成为重点

上游：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需的基础生产要素，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。
- 部分数据标记企业为了使业务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。
- 截止到2020年12月，中国数据标准企业数量达到705家，数据标记企业30%左右分布于北京，其次为上海和成都。行业内不少企业成立于中国人工智能行业兴起的2015年之后，数据标记乃至数据产品服务行业整体较为新兴。

数据标记准确性，2021年数据标记准确率

60%~90%

中游：语音合成行业



- 通用TTS发音相对的机械化，声韵比较平淡，但是优点在于通用TTS口音清晰，听者可以准确获取声音信息，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。由于他对音色，情感上没有过多要求，因此其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，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，创新性中小企业踊跃出现。基于个性化TTS的复杂程度，相比通用型TTS，个性化TTS的制作成本相对高昂。

利润率（不包括研发成本）



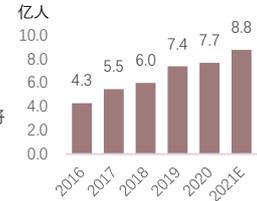
- TTS中游利润率保持在60~80%之间，但前期研发投入较多导致部分TTS企业尚未实现盈利。

下游：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话术较为固定，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剧，有声书等的制作周期，能够更快节奏的，及时的将长音频作品呈现到用户眼前。

中国长音频用户规模，2016-2025E



来源：各企业官网，头豹研究院编辑整理

■ 语音合成产业链上游——数据产品及服务（1）

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需的基础生产要素，实现机器的深度学习需要大数据和大算力的支持，这为数据产品及服务行业带来了巨大的市场

■ 数据是人工智能达成的基础

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需的基础生产要素，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。

■ 数据服务随AI行业成长而发展

随着人工智能的发展，机器深度学习技术取得了快速提高，而实现机器的深度学习需要大数据和大算力的支持。这为数据产品及服务行业带来了巨大的市场。

截止到2020年12月，中国数据标准企业数量达到705家，数据标记企业30%左右分布于北京，其次为上海和成都。行业内不少企业成立于中国人工智能行业兴起的2015年之后，其中2017、2018年成立的企业占到行业整体企业数量的80%，数据标记乃至数据产品服务行业整体较为新兴。

但整体来说，投资界对行业投资热情较高，从2015-2020年，数据标注服务行业的投资数量达到33起，其中龙头企业海天瑞声与2020年登陆科创板，行业发展较为红火。数据显示，到2019年，数据标注行业市场规模已超过25亿元人民币，在未来5年，数据标记及服务行业的年复合增长率将超过15%。市场的快速发展有利于推动行业的进步，同时，有利于中游企业挑选合适的数据服务供应商。

■ 行业发展缺乏核心技术，较容易复制，可替代性强

数据标记行业技术性不强，但对规模的要求较高，需要人力进行准确地识别和转写，对人力的依赖程度较大，属于劳动密集型行业。同时，由于自然人工作的局限性，数据标注的准确度质量只能保持在70%-90%之间，难以达到100%。数据标注行业整体缺乏核心技术，商业模式较容易复制，随着新进入者越来越多，行业内的竞争会进一步加剧。



来源：海天瑞声招股说明书，头豹研究院编辑整理

■ 语音合成产业链上游——数据产品及服务（2）

为了满足企业自我的语音合成训练的要求，部分大型互联网企业建立自由语音标准公司。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

■ 大厂设立自有数据标注公司

数据标记行业内的公司主要分为两种，一种为内需驱动型，为的是满足自有公司的需求，中国的一些互联网企业，如百度、阿里巴巴、腾讯等都建立了自有的数据标记众包平台。这类公司一般已经拥有完善的供应商体系和供应商评价机制，标注工具和标注管理方式已经实现标准化管理，因此，对标准供应商在技术实力、质量、及时性和多样化上均有较高要求。

■ TTS企业后向一体化，产业链出现融合趋势

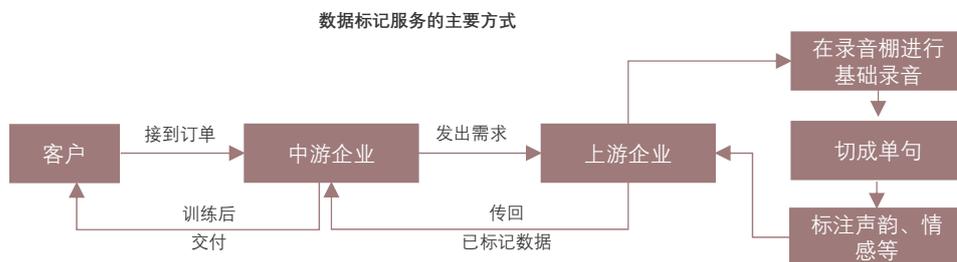
另一种为技术驱动型的，可能本身企业是TTS企业或数据标记工具生产企业，但是在自有产品的生产过程中，也加入了数据标记服务的生产行列。这类企业会根据自己本身的优势，例如TTS企业更加理解客户需求、数据标记工具生产企业借助自己的标记工具提高标记效率等，以此取得行业竞争优势。同时，部分数据标记企业为了使业务能够可持续发展而不受到产业自身限制，也开始拓展自己的业务边界，开拓了自有TTS的产品线。例如作为数据标记龙头企业的海天瑞声开展GPS业务。总体来看，TTS产业链的上中游呈现一体化的融合趋势。

■ 数据标记对人工依赖程度较大，暂时无法被人工智能取代

通过人工智能标记数据在应用层面已经实现，这种标记方式主要是借助之前用人工的方式标记过类似的大量的数据，专门训练一个模型去模拟人去实现标记操作，这种方式在一定程度上可以减轻数据标记的人工工作量。但其并不能实现完全自动化。原因在于实际上人工智能训练数据的错误率较高，需要人工进行校正，因此人工智能只能作为数据标记的一个辅助工具，而不能有效替代人工在数据标记行业的作用。

■ 数据标记行业将朝着高质量，专业化发展

现阶段，数据标记的准确性在约60%-70%，高质量的数据标记准确性在约90%，在未来，标注质量将成为竞争的核心因素，在算法模型的不断优化，应用场景的不断多样的趋势下，人工智能训练所需要的数据标记质量和精度将会不断提高，因此保持高准确度将会是数据标注企业的行业核心竞争力。同时，随着应用场景的多元化，涉及较为专业的细分领域，如医疗、教育等数据标记企业也将会突出重围。



来源：头豹研究院编辑整理

■ 语音合成产业链中游——语音合成行业

TTS语音主要分为通用TTS和个性化TTS，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，个性化TTS则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

TTS的主要类型

TTS类型	通用TTS	个性化TTS
流畅程度	高	高
机械声程度	高	低
清晰度	高	高
情感	低	高
制作成本	相对较低	相对较高
应用场景	导航、语音播报、智能客服和大多数语音交互场景	教育，书籍阅读、直播、影视配音、游戏配音等
主要厂商	BAT、科大讯飞、思必驰	倒映有声，标贝科技

■ TTS语音主要分为通用TTS和个性化TTS，使用场景各不相同

通用TTS发音相对的机械化，声韵比较平淡，但是优点在于通用TTS口齿清晰，听者可以准确获取声音信息，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。由于他对音色，情感上没有过多要求，因此其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，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，创新性中小企业踊跃出现。基于个性化TTS的复杂程度，相比通用型TTS，个性化TTS的制作成本相对高昂。

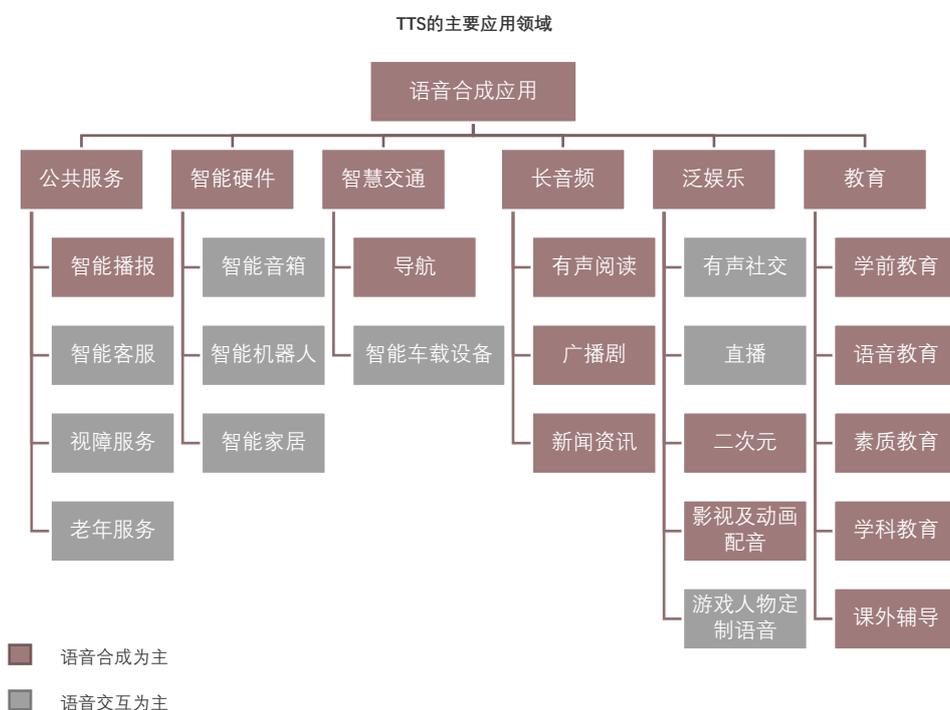
■ TTS产品前期投入较多，产品成型后利润较高

作为人工智能工具之一，TTS产品均有一个从模型搭建到深度训练学习到人工智能成型的过程，这其中需要大量的学习数据作为支持，而且要反复纠错，优化TTS的整体算法以达到理想的输出效果。因此，TTS产品在前期需要投入大量的研发成本，包括设备、人工、网络等，研发过程持续较长时间。一旦算法模型训练成熟，TTS产品得以落地，实现商用化，其运营成本较低，主要成本在于人工维护和服务器等，整体利润率较高，超过60%，市场前景良好。

来源：头豹研究院编辑整理

■ 语音合成产业链下游——TTS的主要应用领域

随着5G时代的来临，通讯业逐渐发展，社会整体需要快捷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



■ TTS应用领域非常广泛，一般分为单项应用或嵌入语音交互链

语音合成乃至整个智能语音行业，如何实现商业化落地乃至盈利成为了企业以及投资者关注的重点。语音合成技术可单独使用，也可作为语音交互的重要输出环节，因此，在智能语音应用越来越充分的背景下，语音合成技术应用非常广泛，涵盖公共服务、智能硬件、智慧交通、长音频、泛娱乐、教育等行业。

■ 5G时代和碎片化时间增加带动语音合成发展，助力高效率社会

随着5G时代的来临，通讯业逐渐发展，社会整体需要快捷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。

来源：头豹研究院编辑整理

■ 语音合成产业链下游——TTS在导航领域的应用

导航领域是TTS常用的、普及程度较高的应用场景之一，导航领域对TTS的主要要求为音质清晰，音色较为温和平静，断句清晰，精度高差错小，个性化TTS最早落地于导航领域产品，现如今个性化程度进一步深入

■ 整体以非交互TTS为核心

导航领域是TTS常用的、普及程度较高的应用场景之一。导航领域的TTS数据量较大，语音内容覆盖范围较大，又应用于交通领域，因此导航领域对TTS的主要要求为音质清晰，音色较为温和平静，断句清晰，精度高差错小。导航领域的相关产品，语音部分由交互语音和非交互TTS组成，其中，交互语音的功能占比较小，因此产品整体以非交互TTS为核心。

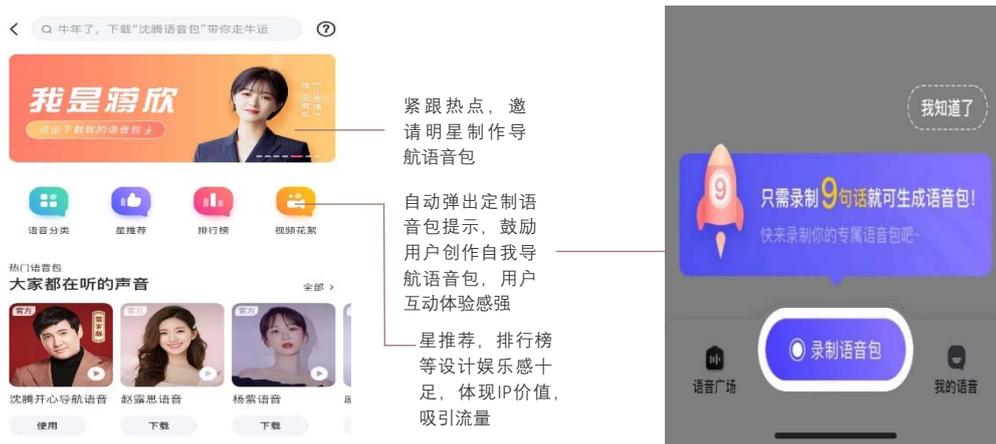
■ 个性化TTS最早落地于导航领域产品

最早的导航领域产品多数使用通用型TTS帮助用户实现听音导航，提高了产品使用的便捷性和驾驶的安全性。随着导航领域产品的逐步成熟，个性化TTS已经成为了差异化竞争的有力工具。

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话术较为固定，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。

以百度地图为例，最早，百度地图通过推出岳云鹏、郭德纲、李佳琦等广受欢迎的明星语音来作为导航语音，吸引用户群体下载和使用软件。随着语音库的不断扩大，百度地图将不同声音特点，适用于不同季节的语音进行归类，还建立语音排行榜，让用户找到自己最喜欢的声音。现阶段，百度推出“9句话定制自己的声音导航”让每个人拥有属于自己的导航语音，使百度地图的导航语音实现进一步个性化。

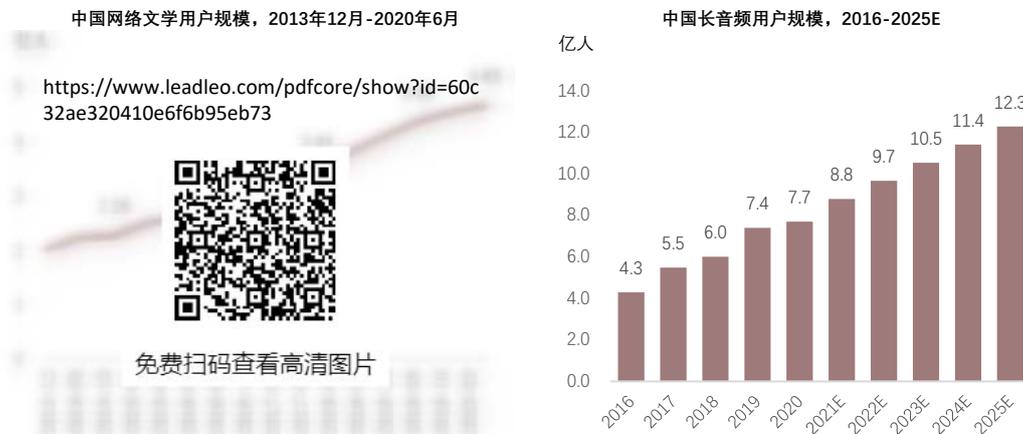
导航产品的语音包设置



来源：百度地图，头豹研究院编辑整理

■ 语音合成产业链下游——TTS在长音频领域的应用

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



■ 长音频市场未来发展潜力巨大

长音频主要是指有声阅读，广播剧，声音广播等产品。据QuestMobile的调研，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。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猛增加三至四倍，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。长音频市场发展潜力巨大。

■ 长音频发展为个性化TTS开设平台

长音频是个性化TTS发展的有利平台。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，将广播剧，有声书的制作时间从漫长的几个月时间缩短到几天甚至几个小时，极大的压缩了长音频的制作周期，能够更快节奏的，及时的将长音频作品呈现到用户眼前。

个性化TTS现阶段的语音质量已经发展到了较高的水平，已经完全可以满足市场对于长音频的质量需求，同时个性化TTS的操作灵活，声音效果多样，限制少，能为长音频作品带来个性化的听觉体验。

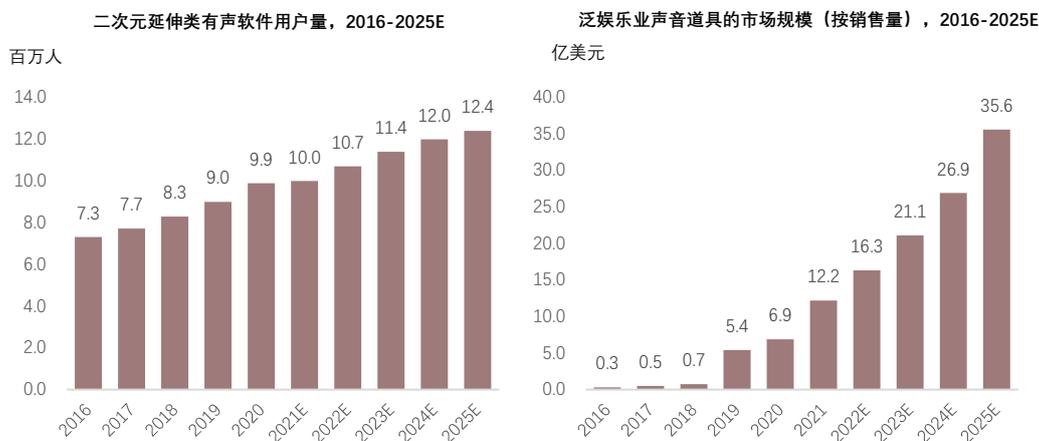
■ 明星IP助力长音频市场

明星IP已经成为长音频的特色之一，例如喜马拉雅利用明星IP进行演播，从而吸引用户流量。而个性化TTS已经可以达到相当高的演播水平，能够直接代替明星录音，使长音频的明星IP打造更加快捷，助力明星IP的商业落地。

来源：wind，头豹研究院编辑整理

■ 语音合成产业链下游——TTS在泛娱乐领域的应用

TTS在泛娱乐领域的应用主要涉及二次元、直播、游戏等行业，声音IP价值实际上有较高价值，但现阶段商业化较少，常常被市场所忽略



■ 为人物打造独特声音，使声音更贴合形象

TTS技术可广泛应用于泛娱乐配音行业，主要应用包括影视配音、游戏配音等。以游戏为例，良好的声感能够极大程度的提高用户的游戏体验，声音也会成为游戏人物形象的一部分，单纯的人工后期配音不仅需要花费大量时间遴选最优质的配音人，还需要考虑人为因素对人工配音的影响，例如人工配音需要休息时间、续集声音的连贯性等。因此，游戏配音将会对游戏的制作产生较大影响，不顺利的游戏配音甚至会推迟游戏的发售时间，影响企业经营。而TTS则只需要输入脚本，便可以快速输出声音，并快速调整，高效的制作出贴合人物形象的声音。

随着二次元ACG等亚文化在中国的兴起，中国的游戏市场逐渐被打开，用户群体越来越多，例如猫耳FM这种二次元延伸类产品的用户量从2016年的730万上升到2020年的1,000万人次。因此，为虚拟人物打造独特也将会是个性化TTS发展的沃土。随着初音未来，洛天依这种纯虚拟偶像的出现，个性化TTS可以为虚拟人物打造一整套标志性的声学系统，为虚拟人物在声音层面为用户留下深刻印象。

■ 声音道具吸引流量

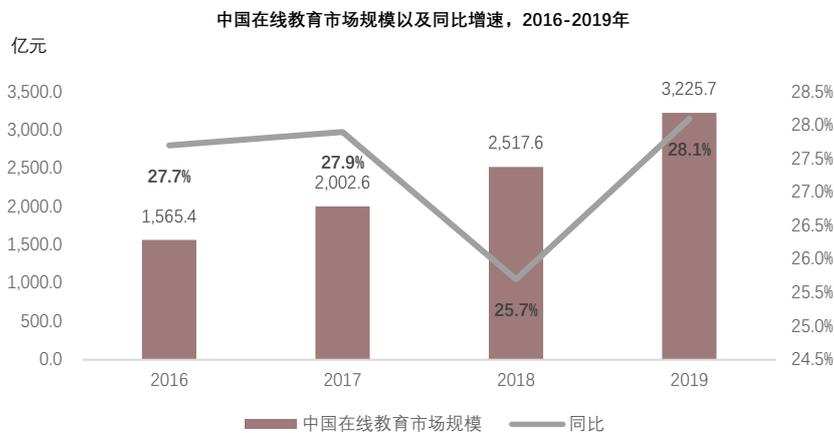
疫情影响和5G通信的完善间接推动了中国直播行业的加速发展，直播产品无论从博主人数到用户量人数都快速增长并且整体粘度较强。部分主播直播时会使用变声器增加趣味性，吸引流量。随着AI技术的逐渐普及，在视觉上，直播已经可以实现主播以虚拟形象出现，而TTS可以帮助主播实现虚拟声音的打造。

声音社交的出现也为TTS找到用武之地，令人愉悦或趣味性的声音道具作为产品特点，吸引年轻用户使用。

来源：头豹研究院编辑整理

■ 语音合成产业链下游——TTS在教育领域的应用

教育领域也是TTS应用的广阔蓝海，随着线上教育受到市场热捧，TTS也在课堂效率、教学方式，教学氛围等方面发挥出其应用优势



■ TTS对线上教育帮助较大

TTS技术在线下教育领域的应用较少，更多的应用于线上教育领域。从数据来看，中国在线教育需求呈增长趋势，市场规模从2016年的1,565.4亿元增长到2019年的3,225.7亿元，年增速保持在25%以上，增长迅速。在此背景下，TTS也迎来了其在教育市场上的应用机遇。TTS技术在教育领域的应用主要有：

- 能够将教师的语言即时转成文字、字幕等，让学生更好的抓住课堂信息，理解课堂内容，集中注意力，同时课后也较好巩固复习；
- 利用TTS技术，教师不需要再花费时间进行板书，提升课堂效率，帮助教师借助AI完成简单的语音读题、英文听力出题等辅助工作，将进一步释放老师的备课压力，让老师可以把更多精力投入到教学中；
- 可以联合AR和VR打造虚拟教师形象，使教学方式更加活泼，表现形式更加多元；
- 减少语言教育中听力练习录制所需要的时间，提升教学效率；
- 减少教师口音对学生的影响

■ TTS技术在学前教育作用明显

学前教育中，儿童接收的音频信息尤为关键。TTS技术将通过文字识别技术识别的文字转换为语音，实现学前儿童通过智能硬件收听故事的功能，一定程度上促进了学前儿童的阅读量。同时，情感丰富的高拟人化的智能语音能够帮助学前儿童建立语言情感。

来源：wind，头豹研究院编辑整理

■ 语音合成的应用风险

但随着语音合成不断创新突破，应用场景日益增多，其安全风险也将动态演进，将越发具有广泛化、场景化、融合化的特点，在此背景下，语音合成技术也产生了不同等级的应用风险

语音合成技术应用的主要风险

风险种类	具体情况	风险等级	影响等级
数据安全	个人声纹和个人肖像权一样，同属于受法律保护的个人自有权益，用户在使用语音合成软件时，个人声纹作为个人信息数据，可能面临泄露的风险	30%	30%
网络安全	声纹特征也同样被应用于加密、解锁领域。不法分子或可利用声纹复刻技术破解声音锁等等加密机制，窥探个人隐私或盗取个人财产	50%	70%
信息安全	语音合成可制作不良信息音频进行传播，制作虚假信息内容用以实施诈骗等非法活动。比如利用语音合成技术克隆亲友的声音对用户实行诈骗	60%	70%
社会治安	语音合成技术可被应用于制造不良导向舆论，利用有影响力的公众人物的声纹，制作虚假负面的音视频，制造社会混乱，对社会安定造成威胁	80%	80%
法律地位	语音合成技术所生产的语音产品在表演权、复制权、著作权上存在一定争议	20%	30%

■ 语音合成技术在信息安全和法律上存在风险，加强法律规范

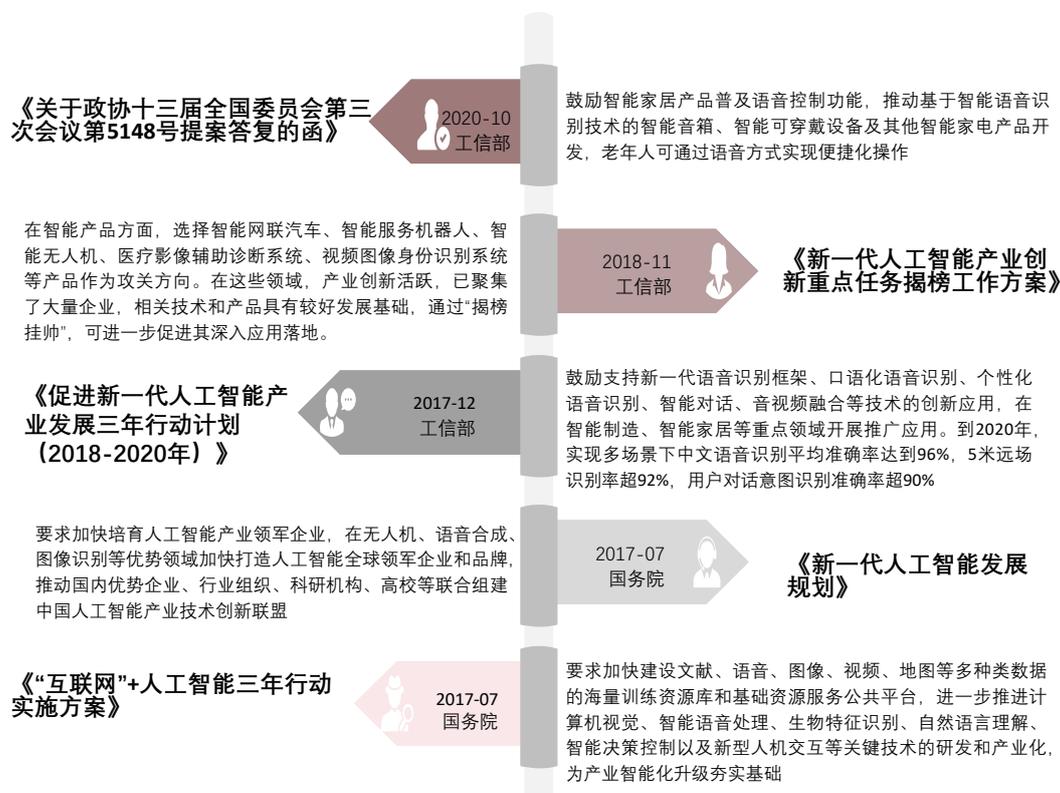
个人声纹是一种强生物特征，语音合成技术可以利用这种特征破解加密信息、骗取信任、混淆视听。一旦个人声纹信息遭到泄露，语音合成技术得到滥用，则会造成不同等级的安全风险，影响较为广泛。同时，作为人工智能最为成熟的板块，语音合成技术虽然已大量应用于社会生活，但在法律上的权利存在一定的争议。为减轻语音合成在应用而产生的风险，国家应建立法律法规加强对个人信息的保护；同时，严格审核语音合成行业的准入资质；另外，组织科学家及时研判隐含的技术风险，完善学术规范和监管框架，为科技创新发展提供先进的技术指导和良好的发展环境，是促进社会向智能化快速发展。

来源：头豹研究院编辑整理

■ 语音合成行业政策分析

中国对于人工智能行业寄予较大期望，在政策上给与支持与鼓励。在政策上主要关注技术本身的研究与发展 and 各个细分的应用

语音合成及人工智能行业政策分析，2021



■ 中国政府大力扶持语音合成及智能语音行业

中国对于人工智能行业寄予较大期望，在政策上给与支持与鼓励。在政策上主要注重两个方向：（1）关注技术本身的研究与发展，提高科技创新能力，打造强有力的技术品牌；（2）积极将语音合成技术乃至整个人工智能投入各个应用细分，使之发挥出其应有实力。

来源：国务院，工信部，头豹研究院编辑整理

■ 语音合成的发展难点与未来

语音合成行业发展较成熟，原有格局难以打破，新进入者只能通过弯道超车或差异化竞争打破僵局，用户的需求将会变得多样，这对语音合成行业也是巨大挑战，在未来，语音合成技术也将逐步打开商用化道路，应用场景将逐渐细化和专业

► 发展难点

■ 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。但基于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。

■ 市场不买软件，只买硬件

市场对TTS软件的购买意向不高，但对于硬件捆绑软件接受程度较高，愿意购买带有TTS功能的硬件产品，如录音笔，智能音箱等。针对这样的市场形势，一方面，企业可以加强与硬件厂商的合作，推动TTS软件的销售；另一方面，可以更加便捷的方式，如手机软件等，推向市场。

■ 用户要求逐渐增高

随着通信技术和新媒体的发展，用户对于音频的质量要求将会越来越高。不清晰，错误较多的机械音较重的TTS将有可能逐渐被市场淘汰，用户对于TTS语音的清晰度、音色以及情感表现将会更加有更高的要求。这对于TTS的技术以及高质量的TTS的普及有很大的需求，这也意味着高质量TTS技术的成本控制成为TTS行业企业的发展难点。

► 未来

■ 市场对TTS的技术认识不足

实际上，市场上的用户群体对于TTS当前的认识依旧停留在机械化声音程度较高，智能化不足的阶段，而对于情感丰富，智能化程度较高，讲演水平较高，拟人化程度高的TTS认识不足，甚至会感到十分惊讶。

要实现TTS技术的商业化落地，从幕后走向前台，被大众所认知，是TTS行业的必由之路。一旦市场对于TTS技术的认知程度提高，无论是从C端还是B端，TTS的技术均能更好的得以实现。

■ TTS技术将应用于多元场景

TTS的应用场景将会更为多元，专业细分将会更加详细，例如科大讯飞和思必驰主推的医疗场景下的TTS应用。这对于TTS的精度要求将会提升。

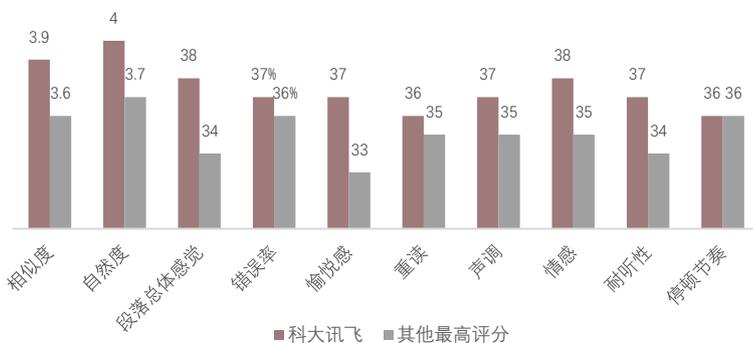
来源：头豹研究院编辑整理

■ 语音合成行业企业分析——科大讯飞（一）

科大讯飞是最早进入语音合成行业的中国企业之一，作为科大讯飞最成熟最专业的技术，科大讯飞的语音合成种类较多，又占领结构性优势，因此在语音课程行业市场占率高，地位稳固

- 科大讯飞股份有限公司成立于1999年，是亚太地区知名的智能语音和人工智能上市企业。自成立以来，长期从事语音及语言、自然语言理解、机器学习推理及自主学习等核心技术研究并保持了国际前沿技术水平；积极推动人工智能产品研发和行业应用落地，2008年，公司在深圳证券交易所挂牌上市（股票代码：002230）
- 科大讯飞在智能语音行业具有领先地位，在语音合成产品市场份额达到70%以上，在电信、金融、电力、社保等主流行业的份额更达80%以上，开发伙伴超过10000家，灵犀定制语音助手在同类产品中用户规模排名第一。
- 科大讯飞的语音合成能力保持世界领先水平，截止到2019年，科大讯飞已实现国际语音合成大赛Blizzard Challenge的14连贯，受到国际语音合成学术界的认可，在技术和产品质量上均有保障，并且未来发展势头良好，占据中国市场结构性优势。

科大讯飞在国际语音合成大赛Blizzard Challenge中的表现，2018年



- 科大讯飞是最早进入语音合成行业的中国企业之一，其于1993年进入市场，并以语音合成技术作为企业发展的起点，逐步拓展了语音识别、自然语言理解、机器学习推理及自主学习等方面的技术能力，最终成为中国人工智能的领军企业。作为科大讯飞最成熟最专业的技术，科大讯飞的语音合成种类较多，突破性技术较多，产品成熟质量较好，又占领结构性优势，因此在语音课程行业市场占率高，地位稳固。
- 同时，科大讯飞在toB、toC、toG 方面均有涉足，因此在产业层面，科大讯飞能够全方位覆盖市场需求，涉足范围颇为广泛，因此盈利能力十分可期。

来源：科大讯飞官网，头豹研究院编辑整理

■ 语音合成行业企业分析——科大讯飞（二）

科大讯飞在人工智能企业中较好的实现了盈利并且毛利率在行业内保持前列，二级市场表现也表现稳定而良好，2017、2019和2020年科大讯飞的每股价格稳步上升，每股未分配利润也保留充足，每股净资产较为充沛



■ 收入稳步上升，毛利润保持稳定

从2017-2020年，科大讯飞的营业收入保持稳步增长，四年增速达到了138%，而毛利率从2017年的51.4%下降到2020年的45.1%，净利率4年内保持在约8%~10%，稳定中略微下降。但总体来看，科大讯飞在人工智能企业中较好的实现了盈利并且毛利率在行业内保持前列。

根据年报显示，科大讯飞的流动比和速动比维持在1.2-1.5之间，企业流动性较好，库存周转较为灵活，流动性风险较小。



■ 二级市场表现良好

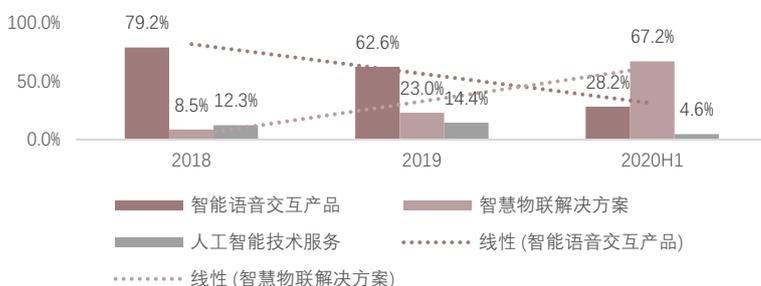
从2017-2020年，科大讯飞的二级市场表现也表现稳定而良好，除2018年股价略微下滑外，2017、2019和2020年科大讯飞的每股价格稳步上升，每股未分配利润也保留充足，每股净资产较为充沛，未来可期。

来源：科大讯飞企业年报，头豹研究院编辑整理

■ 语音合成行业企业分析——云知声（一）

2017年左右互联网头部企业为抢占智能语音硬件市场而大量采购智能语音模组，为云知声带来大量收入，但由于互联网头部企业议价能力太强，拉低智能硬件设备行业整体售价，在销量和售价的双重夹击之下，云知声主营业务收入增长后劲不足

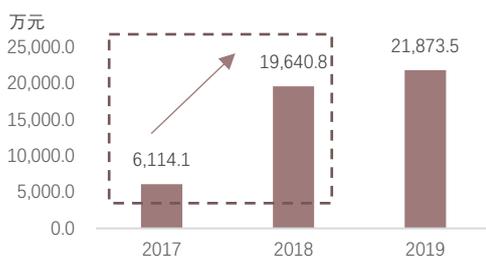
云知声主营业务收入结构占比



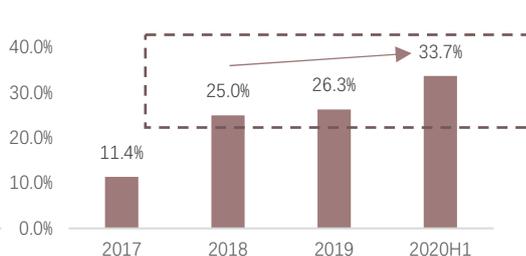
■ 产品收入结构从硬件走向软件

从2018-2020年H1，云知声的主要产品收入结构发生了较大改变，其硬件类产品智能语音交互产品，如智能语音模组和智能音箱，产品占比从2018年的79.2%下降到2020年上半年的28.2%，而服务类产品智慧物联解决方案占比则从2018年的8.5%上升到67.2%。出现该情况的原因在于2017年左右互联网头部企业为抢占智能语音硬件市场而大量采购智能语音模组，为云知声带来大量收入，但由于互联网头部企业议价能力太强，拉低智能硬件设备行业整体售价，在销量和售价的双重夹击之下，智能语音交互产品收入减少。但在2018-2020年H1，云知声和多家酒店、住宅、医院、社区等达成合作协议，为企业提供智慧物联解决方案，因此产品结构发生改变。

云知声主营业务收入，2017-2019



云知声毛利润，2017-2020H1



■ 产品收入结构从硬件走向软件

从2017-2020年H1，云知声的毛利润保持增长，从2017-2018的主营业务收入和毛利率的增长均依靠出售智能语音交互产品。随后云知声主要依靠业务拓展增加收入和毛利润，但整体涨幅不比从前，总体来看，云知声受行业宏观环境影响较大。

来源：云知声招股说明书，头豹研究院编辑整理

■ 语音合成行业企业分析——云知声（二）

云知声的业务模式尚未定型，同时，销售和研发消耗了大量成本，因此亏损严重，且随着业务模式的不断变化，商业模式的不断摸索，云知声的研发将会持续投入，亏损短时间内难以平衡

云知声营收与亏损对比，2017-2020H1



云知声研发费用占比，2017-2020H1

证券名称	2020H1	2019	2018	2017
科大讯飞	21.5%	16.3%	11.9%	10.9%
拓尔思	9.9%	9.5%	11.8%	11.8%
威胜信息	7.8%	7.9%	7.1%	8.2%
佳华科技	7.7%	9.1%	6.4%	6.6%
洪泉物联	18.3%	16.4%	16.1%	19.1%
同行均值	13.1%	11.7%	10.6%	10.9%
云知声	109.1%	117.8%	77.6%	163.6%

■ 入不敷出的财务局面

云知声连续4年的营业收入不及净亏损，净亏损额从2017年的17,721.7万元上涨到2019年的31,712.5万元，亏损增速达到78.9%，从2017-2021H1，云知声累计净亏损超过8亿元。盈利能力令人担忧。根据云知声招股说明书来看，云知声的业务模式尚未定型，同时，销售和研发消耗了大量成本，因此亏损严重。除2018年之外，从2017-2020H1，云知声的研发费用占营业收入的比重均超过100%，而同行均值均不超过15%，且随着业务模式的不断变化，商业模式的不断摸索，云知声的研发将会持续投入，亏损短时间内难以平衡。

来源：云知声招股说明书，头豹研究院编辑整理

■ 语音合成行业企业对比——毛利率

智能语音行业整体毛利润在40~50%之间，硬件产品由于其价格较为透明，产品溢价较少，因此，智能语音行业企业的毛利率与产品硬件比例有一定关系

智能语音行业毛利率对比，2017-2020H1

证券名称	2020H1	2019	2018	2017
科大讯飞	45.8%	46.0%	50.0%	51.4%
拓尔思	54.6%	61.8%	59.4%	62.6%
威胜信息	36.5%	34.1%	32.7%	32.6%
佳华科技	46.4%	51.5%	50.4%	51.2%
洪泉物联	48.5%	48.3%	41.0%	35.8%
云知声	31.7%	26.3%	24.9%	11.4%
同行均值	46.8%	48.3%	41.0%	48.1%

■ 软件毛利润高于硬件

智能语音行业整体毛利润在40~50%之间，其中主要出售硬件的威胜信息（硬件率在90~95%）的毛利率较同行业较低，而硬件率较低的拓尔思（硬件率15~25%）的整体毛利润率较高，均值达到约60%。

■ 行业最低毛利率在约30%

同行业中，毛利率表现最低的为云知声，2017-2020H1的毛利润率保持在约10-32%，低于行业平均水平，这主要是由于云知声相对来说收入规模较小，规模效应不明显。同时，占比最大的智慧物联解决方案中，硬件设备占比较大，由于硬件设备价格较为透明溢价较低，因此整体毛利率较低。

来源：云知声招股说明书，头豹研究院编辑整理

■ 语音合成行业企业对比——销售费用

从2017年到2019年，智能语音行业逐步被市场认可，因此在销售费用上逐渐减少，但由于新冠疫情的影响，企业需要扩大宣传，拓宽销售渠道，因此2020年H1，行业内部分企业销售费率一定程度增加。

智能语音行业销售费用占比对比，2017-2020H1

证券名称	2020H1	2019	2018	2017
科大讯飞	18.8%	17.7%	21.8%	20.4%
拓尔思	14.8%	17.1%	18.4%	18.1%
威胜信息	4.4%	5.1%	5.9%	6.5%
佳华科技	5.5%	4.5%	4.9%	4.2%
洪泉物联	6.0%	7.9%	6.2%	5.7%
云知声	37.6%	25.8%	14.8%	27.8%
同行均值	10.9%	10.8%	11.9%	11.3%

■ 行业平均销售费用占比保持在约10%

智能语音行业虽然为高新技术行业，但整体依旧以技术为导向，由于行业内竞争尚未发展到白热化程度，因此在销售费用率方面整体费率较低，平均销售费用约占整体营业收入的10%。行业内较低的销售费用甚至达到了4.2%。

■ 行业销售费用比例呈下降趋势，受疫情影响部分企业销售费用有所上涨

从2017年到2019年，智能语音行业逐步被市场认可，因此在销售费用上逐渐减少，2019年平均销售费率相比于2017年的11.3%下降0.5%，行业内各企业销售费率均有不同程度的减少。但由于新冠疫情的影响，企业需要扩大宣传，积极参加展会，拓宽销售渠道，因此2020年H1，行业内部分企业销售费率一定程度增加。

来源：云知声招股说明书，头豹研究院编辑整理

中国语音合成优质企业——倒映有声

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，其主要专注于个性化合成语音的打造以及和泛娱乐行业的商业应用，用高质量的技术手段打造用户体验感舒适的合成语音

- 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到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。倒映有声已与超过5家版权方合作，上线了共10多部作品。

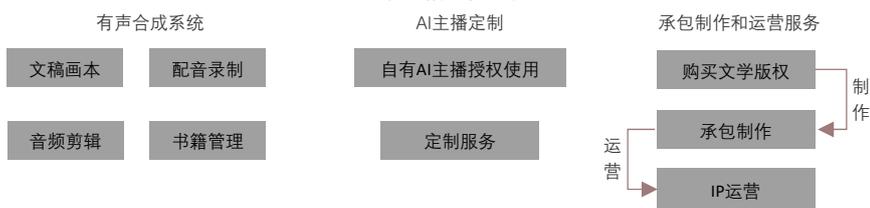
倒映有声的部分合作伙伴



- 企业产品线明晰，积极布局在泛娱乐行业的深度合作

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领域；（2）第二类为AI主播定制，主要分为自有AI主播授权使用和定制服务，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落脚于技术层面，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。

倒映有声的主要产品



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。



拟人化程度

高



产品覆盖面

专

企业分析

来源：倒映有声官网，头豹研究院编辑整理

■ 海外语音合成优质企业——Sonantic

Sonantic是一家英国的语音合成企业，其产品已经在泛娱乐行业投入使用，并在语音的精细化程度和应用的深度上相比于中国本土企业更胜一筹，中国合成语音在商业化应用上可以向海外企业寻找灵感，打造更适合中国市场的个性化语音合成产品

■ Sonantic主要将目光聚焦在泛娱乐行业

Sonantic是一家英国的语音合成企业，其主要为泛娱乐行业的从业者，特别是游戏和电影行业从业者，提供逼真又细腻的合成语音，让电影和游戏配音制作变得快捷而成本低廉。Sonantic已经与以黑曜石游戏公司为代表的10多家AAA游戏工作室建立了研发合作关系，已将产品投入商业应用。

■ 制作产品方便

Sonantic的产品设计操作简便，使用桌面应用程序无需任何代码即可创建出合成语音，工作方式较为灵活，这将大大促进合成语音产品的工作效率，制作人可以快速将脚本转换为音频，将配音冗长的制作时间从数月缩短至数分钟。

Sonantic的合成语音可完全控制用于调整声音演奏的参数，制作者可以通过调整软件的参数线条去调整的声音节奏，尝试不同的情绪，重新编写脚本，通过更改单词的音高来更改短语的变化，进一步甚至可以添加新建的单词和名称，并定义该词汇或名称应如何发音。同时，Sonantic可以简单的复制之前已经打造好的声音，方便游戏或电影配音的拓展以及续集制作。

■ 相比于国内市场的个性化语音合成产品，Sonantic产品的商业应用更加深入

中国市场上的定制化语音合成的产品主要应用于单一的音频输出，而Sonantic主要将产品和视觉效果结合起来。相比于单一的音频输出，与视觉效果相结合的合成语音需要更加准确的把握被配音角色的语气、节奏、神态和口型，这在语音质量和语音的艺术性上都提出了更高的要求。随着中国市场上语音合成技术的商用化需求进一步扩大，语音产品运用于视频行业势在必行，中国合成语音行业的参与者需保证自有产品的质量，同时产出更为多元化，更为丰富的特色合成语音。



Sonantic的优势在于将技术和商业化程度结合较好。高质量的语音帮助游戏和电影公司实现作品的声音需求，Sonantic自身的产品的的设计对于制作人简单易操作，复制性强，为作品的制作节约了时间成本以及费用，同时，Sonantic还听取客户意见不断改进产品。而电影和游戏作为承载平台，极大程度的发挥了合成语音的可定制化，灵活多变的优势，让合成语音拥有了绝佳的施展平台。

企业分析

来源：Sonantic官网，头豹研究院编辑整理

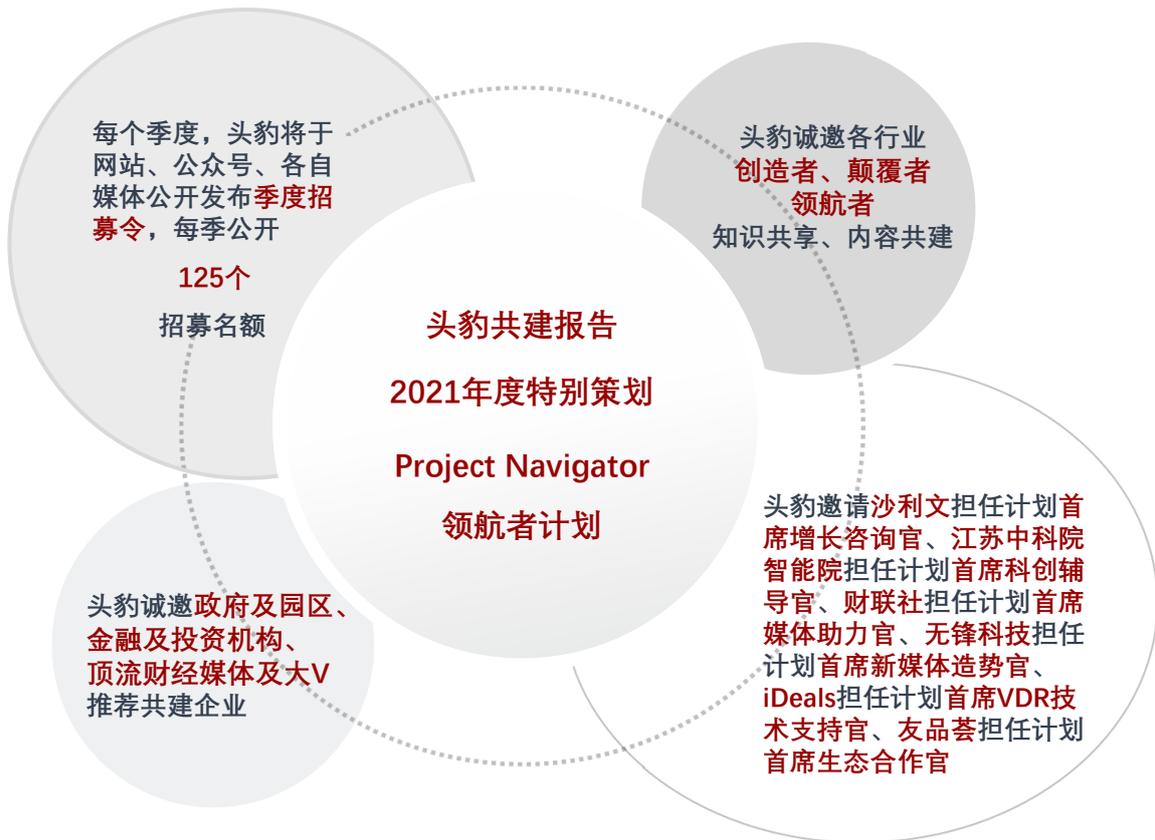
方法论

- ◆ 头豹研究院布局中国市场，深入研究10大行业，54个垂直行业的市场变化，已经积累了近50万行业研究样本，完成近10,000多个独立的研究咨询项目。
- ◆ 研究院依托中国活跃的经济环境，从语音交互、长音频市场、人工智能等领域着手，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ◆ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ◆ 研究院密切关注行业发展最新动向，报告内容及数据会随着行业发展、技术革新、竞争格局变化、政策法规颁布、市场调研深入，保持不断更新与优化。
- ◆ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。

法律声明

- ◆ 本报告著作权归头豹所有，未经书面许可，任何机构或个人不得以任何形式翻版、复刻、发表或引用。若征得头豹同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“头豹研究院”，且不得对本报告进行任何有悖原意的引用、删节或修改。
- ◆ 本报告分析师具有专业研究能力，保证报告数据均来自合法合规渠道，观点产出及数据分析基于分析师对行业的客观理解，本报告不受任何第三方授意或影响。
- ◆ 本报告所涉及的观点或信息仅供参考，不构成任何投资建议。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。在法律许可的情况下，头豹可能会为报告中提及的企业提供或争取提供投融资或咨询等相关服务。本报告所指的公司或投资标的的价值、价格及投资收入可升可跌。
- ◆ 本报告的部分信息来源于公开资料，头豹对该等信息的准确性、完整性或可靠性不做任何保证。本文所载的资料、意见及推测仅反映头豹于发布本报告当日的判断，过往报告中的描述不应作为日后的表现依据。在不同时期，头豹可发出与本文所载资料、意见及推测不一致的报告和文章。头豹不保证本报告所含信息保持在最新状态。同时，头豹对本报告所含信息可在不发出通知的情形下做出修改，读者应当自行关注相应的更新或修改。任何机构或个人应对其利用本报告的数据、分析、研究、部分或者全部内容所进行的一切活动负责并承担该等活动所导致的任何损失或伤害。

头豹 Project Navigator 领航者计划介绍

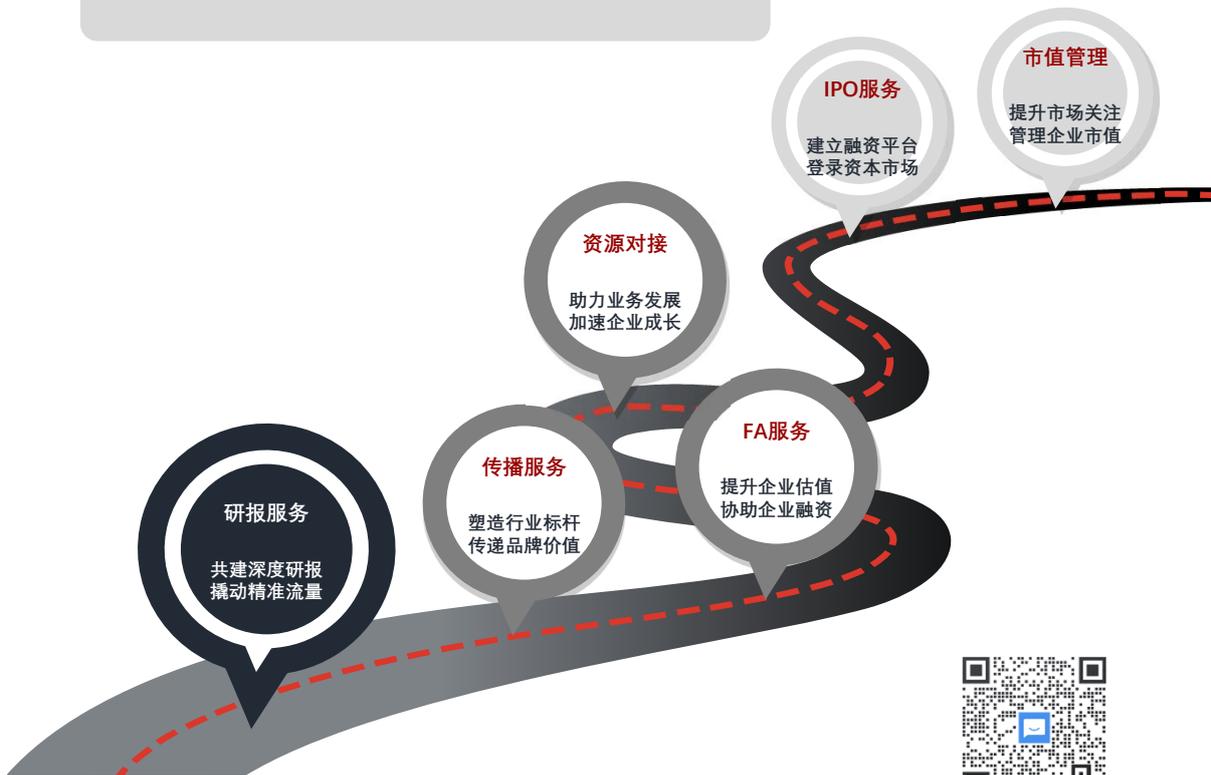


备注：活动解释权均归头豹所有，活动细则将根据实际情况作出调整。



头豹 Project Navigator 领航者计划与商业服务

- 头豹以**研报服务**为切入点，根据企业不同发展阶段的资本价值需求，以**传播服务、FA服务、资源对接、IPO服务、市值管理**为基础，提供适合的**商业管家服务解决方案**



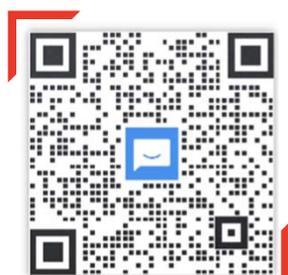
扫描上方二维码

联系客服报名加入

备注：活动解释权均归头豹所有，活动细则将根据实际情况作出调整。

读完报告有问题？

快，问头豹！你的智能随身专家



扫描二维码
即刻联系你的智能随身专家



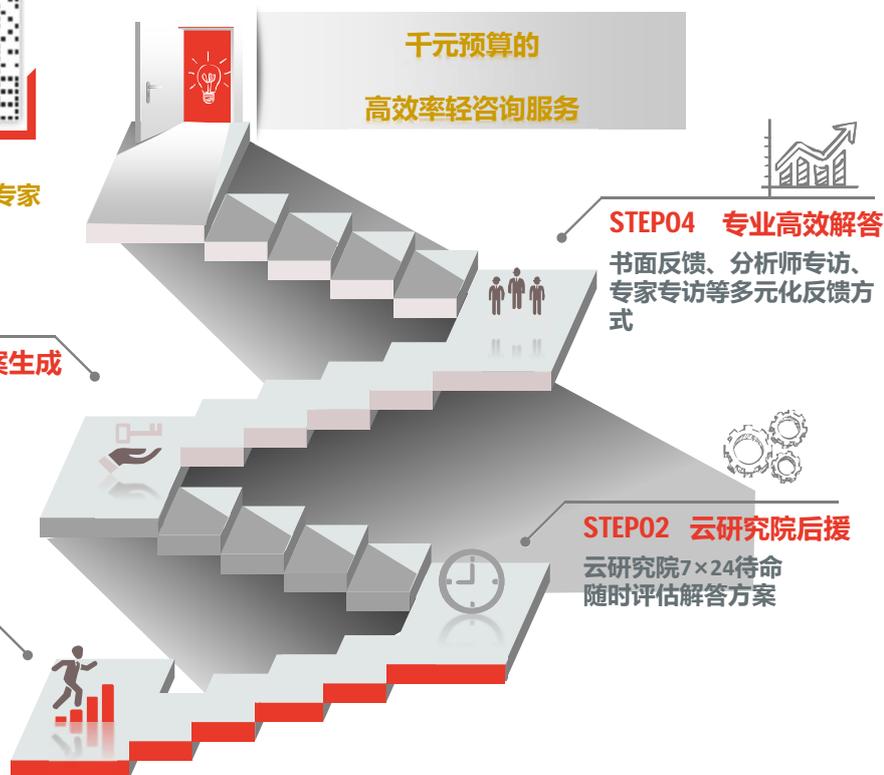
STEP03 解答方案生成

大数据×定制调研
迅速生成解答方案



STEP01 智能拆解提问

人工智能NLP技术
精准拆解用户提问



400-072-5588

www.leadleo.com
©2021 LeadLeo