

The background features a vibrant blue digital landscape. A central figure, a person wearing a headset, is shown in profile, interacting with a large, glowing interface. Surrounding the figure are several floating panels and icons: a yellow padlock, a bar chart with a red line graph, a brain with neural connections, and a fingerprint scanner. The overall aesthetic is clean, modern, and high-tech, representing artificial intelligence and data processing.

中国 AI 语音识别市场研究报告

报告提供的任何内容 (包括但不限于数据、文字、图表、图像等) 均系弗若斯特沙利文公司独有的高度机密性文件 (在报告中另行标明出处者除外)。未经弗若斯特沙利文公司事先书面许可, 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, 弗若斯特沙利文公司保留采取法律措施, 追究相关人员责任的权利。弗若斯特沙利文开展的所有商业活动均使用“弗若斯特沙利文”或“Frost & Sullivan”的商号、商标, 弗若斯特沙利文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弗若斯特沙利文开展商业活动。

A Frost & Sullivan Research Report

目录

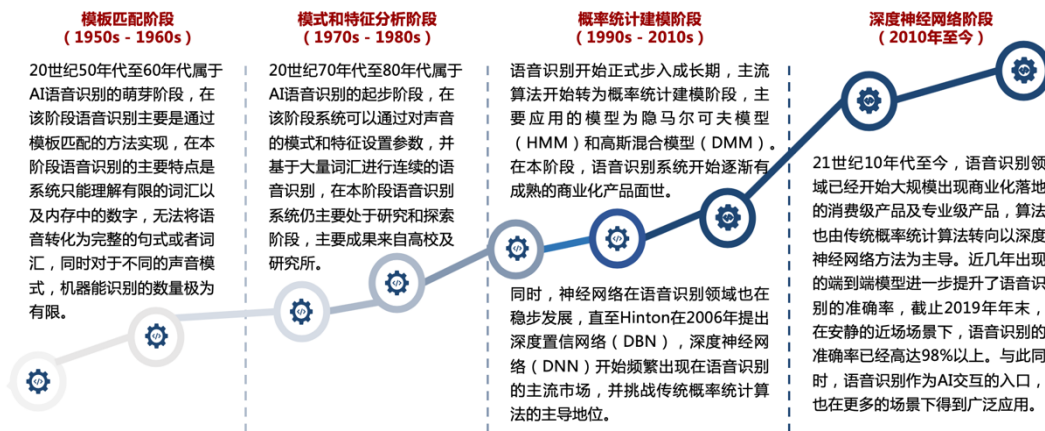
1 AI 语音识别定义与解读.....	2
2 中国 AI 语音识别市场概览.....	3
2.1 AI 语音识别产业链分析.....	3
2.2 中国 AI 语音识别市场驱动因素.....	5
2.2.1 需求端：下游需求增长，AI 语音识别市场空间稳步提高.....	5
2.2.2 技术端：算力、算法、大数据升级，AI 语音识别准确率持续提升.....	6
2.2.3 政策端：人工智能上升至国家战略地位，AI 语音识别行业加快布局和落地.....	7
2.3 中国 AI 语音识别市场趋势洞察.....	7
2.3.1 云计算渐发展，商业化前景更广阔.....	7
2.3.2 多技术协同发展，语音交互更生动.....	8
2.3.3 语音技术渐开放，普惠生态更繁荣.....	10
2.4 中国 AI 语音识别市场关键成功因素分析.....	10
3 中国 AI 语音识别市场竞争格局分析.....	12
3.1 沙利文企业增长评价数据来源和研究主体.....	12
3.2 中国 AI 语音识别市场企业增长评价结果及分析.....	12
3.3 沙利文评价模型的设计.....	16
4 研究方法和研究范围介绍.....	18
4.1 研究方法.....	18
4.2 研究范围.....	18

1 AI 语音识别定义与解读

语音识别是人机交互的入口，是指机器/程序接收、解释声音，或理解和执行口头命令的能力。在智能时代，越来越多的场景在设计个性化的交互界面时，采用以对话为主的交互形式。一个完整的对话交互是由“听懂——理解——回答”三个步骤完成的闭环，其中，“听懂”需要语音识别 (Automatic Speech Recognition, ASR) 技术；“理解”需要自然语言处理 (Natural Language Processing, NLP) 技术；“回答”需要语音合成 (Text To Speech, TTS) 技术。三个步骤环环相扣，相辅相成。语音识别技术是对话交互的开端，是保证对话交互高效准确进行的基础。

语音识别技术自 20 世纪 50 年代开始步入萌芽阶段，发展至今，主流算法模型已经经历了四个阶段：包括模板匹配阶段、模式和特征分析阶段、概率统计建模阶段和现在主流的神经网络阶段。目前，语音识别主流厂商主要使用端到端算法，在理想实验环境下语音识别准确率可高达 98% 以上。

图 1-1: AI 语音识别发展历程



来源: fsTEAM 软件采编, 沙利文研究院绘制

2 中国 AI 语音识别市场概览

2.1 AI 语音识别产业链分析

中国 AI 语音识别市场参与者众多，主要分为上游、中游、下游。

图 2-1: AI 语音识别产业链分析



来源：沙利文研究院绘制

- 上游：底层技术提供强力支撑，云计算助推 AI 语音应用普及**---语音识别解码过程中包含了声学模型和语言模型的识别建模和模型训练两个部分。在运行过程中训练数据量和计算量需求极大，传统的 CPU 或者单一处理器几乎无法快速单独完成一个完整的模型训练过程，主要原因在于 CPU 内部仅含少量逻辑单元，且指令执行是逐一进行的串行计算，使用该架构进行语音识别运算的处理时间过长，无法满足海量数据计算的实时性需求。因此，能提供海量数据处理、存储以及高性能运算能力的云计算技术成为语音识别行业的应用热点。目前，主流语音识别公司的模型训练和语音识别基本都在云端采用 GPU 并行架构或异构计算方案进行。

- **中游：语音技术持续升级，生态圈建立赋能产业**---语音识别的中游主要为将语音识别技术实现商业化落地的硬件及软件服务供应商。根据终端消费者类型，语音识别的中游厂商主要可以分为消费级市场和专业级市场，其中消费级市场中的主要语音识别产品包括消费级智能硬件、智能音箱及语音输入法等，专业级市场的语音识别产品则主要以行业解决方案(以项目制交付的软硬件产品及服务)和平台化技术输出 (SDK 或 API 形式的智能语音开放平台) 两种形式呈现，其中更为垂直落地的解决方案形式在目前专业级商业化收入市场中占比更高。目前，智能语音开放平台在智能语音市场中收入占比较小的主要原因是，以阿里、百度及



科大讯飞为首的各大厂商为加速 AI 语音技术对下游应用场景的渗透，采用多种优惠甚至免费形式向开发者提供语音识别服务，希望将语音识别技术应用在更多软件及场景中，与广大开发者携手建立一个完整的 AI 产业生态圈。

- **下游分析：行业应用多样化，一站式服务需求广**---语音识别作为 AI 交互的重要入口，在人工智能领域属于最重要和发展最为成熟的技术之一，目前已经以多种商业化形式广泛应用于下游市场。从应用领域来看，目前消费级市场主要应用于智能硬件、智能家居、智慧教育、车载系统等领域，专业级市场主要应用于医疗、公检法、教育、客服、语音审核等领域。广泛的应用领域也就意味着更加多元化的使用场景，然而目前的语音识别技术对于使用场景具有较强的限制性。尽管快速更新迭代的神经网络结构已经将安静环境下的近场语音识别的错误率降低至 3% 以下，但现实环境中多数应用场景无法满足理想的环境条件，因此在进行语音识别时需要同时考虑到各种噪声、信道等因素。为使语音识别技术在更广泛的使用场景下保持良好的表现，AI 语音厂商需要提供硬件与软件协同的一站式服务，并根据用户实际痛点进行针对性优化，从而有效提升在多元下游场景下语音识别的渗透率。

2.2 中国 AI 语音识别市场驱动因素

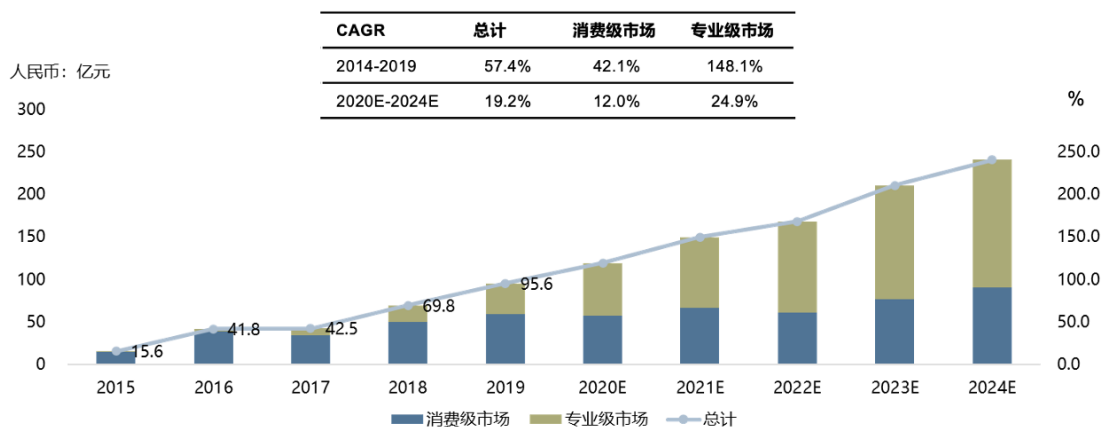
2.2.1 需求端：下游需求增长，AI 语音识别市场空间稳步提高

在过去五年间，中国 AI 语音的需求最先在消费级市场爆发，主要得益于互联网及智能硬件设备厂商加大语音识别的投入经费，以及厂商为提前占据市场推行的智能音箱硬件补贴。目前，消费级产品及服务主要包括智能音箱、智能车载和智能硬件及消费级互联网增值服务。然而，目前包括直接面向消费者的产品及服务在内，语音识别的相关应用及使用场景仍具有局限性。未来，在消费级产品供应商和开发者共同构建产业生态圈的过程中，语音识别技术将更好地与其他语音交互技术及软件功能融合，为消费者提供更优质的体验，未来 AI 语音识别市场将迎来广阔的发展空间。

对于专业级市场而言，主要的产品形式包括智能语音开放平台和行业解决方案，下游应用领域目前主要包括数字化水平相对较高的智慧医疗、智慧教育、企业客服、司法政务、金融领域等。AI 语音识别作为人机交互的重要入口之一，除了在语音识别的领域表现出色外，也要能更好地与其他智能语音技术（包括语义理解、远场语音识别、唤醒目标检测、全双工交互、个性化识别技术等）进行融合，从而综合提升真实场景中的用户体验。近年来 AI 语音识别专业级市场的快速增长主要原因除了深度神经网络算法为语音识别带来的准确率大幅提升外，更重要的是其他智能语音和 AI 技术的发展带来了更广阔的应用场景，预计未来专业级市场的商业化需求将得到进一步释放。



图 2-2: 中国 AI 语音识别市场商用收入规模, 2015-2024 年预测



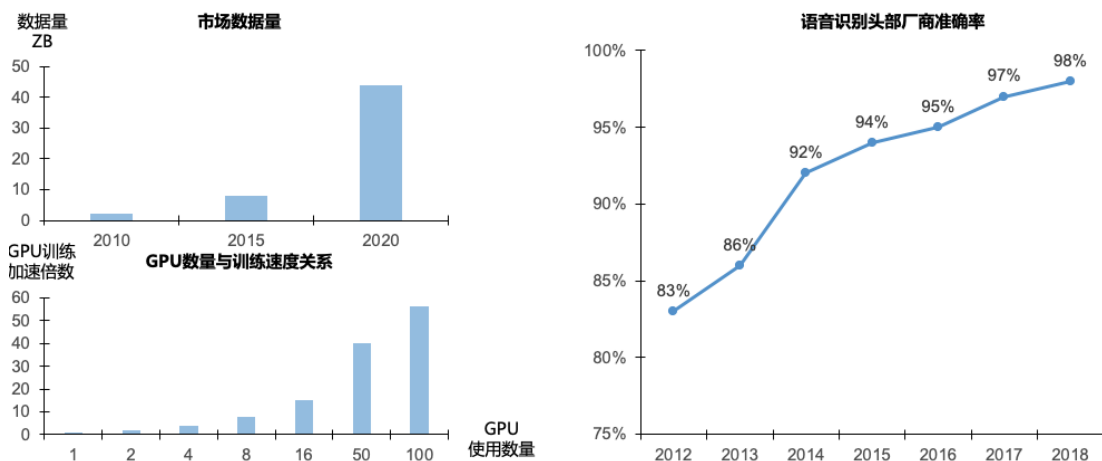
* 统计对象包括: 1) 专业级市场: 智能语音行业解决方案、智能语音开放平台等; 2) 消费级市场: 与语音识别直接相关的硬件设备, 如智能音箱及相应消费级软件和服务如个性化教与学平台、语音输入法、智慧考试等。以上商业化收入仅包括智能语音直接相关收入, 硬件收入及其他技术相关收入不纳入本市场规模。

来源: 沙利文研究院绘制

2.2.2 技术端: 算力、算法、大数据升级, AI 语音识别准确率持续提升

在过去 5-10 年, AI 语音识别技术的快速商业化的主要原因在于技术端的快速发展, 如计算能力的提升、算法框架的优化和大数据的升级等。

图 2-3: 中国 AI 语音识别市场技术发展情况



来源: 沙利文研究院绘制

- 从计算能力来看, 芯片处理能力的大幅提升、GPU 的大量应用、云服务的普及还有硬件价格的快速下降共同为人工智能计算能力的提升提供了重要支撑;
- 从算法框架来看, 目前主流语音识别模型已经以深度神经网络为主导, 神经网络的出现及普及为语音识别准确率的提升起到了重要作用;

- 从大数据来看,更加贴近真实使用场景的语料库也为语音识别技术提供了更加有效的训练素材,从而大幅提升了 AI 语音识别产品及服务的使用体验。以上底层技术的升级,为语音识别技术的准确率提升及商用渗透提供了强大的市场驱动力。

2.2.3 政策端：人工智能上升至国家战略地位，AI 语音识别行业加快布局和落地

人工智能发展水平一定程度上体现了各国最高的科技水平。考虑到人工智能发展对于国家经济发展的重要性,中国政府已针对人工智能行业颁布了多项国家层面的发展政策,自 2017 年以来人工智能行业已经连续三年被写入《全国政府工作报告》内。具体支持政策包括项目发展基金、人才引进政策及其他国家扶持政策。目前,语音识别技术属于中国 AI 领域中最为成熟落地的技术之一,在国家政策的强力扶持下,预计未来能够加速在垂直行业的渗透和布局。

同时,在《中国制造 2025》的大背景和智能经济新形态下,各省市响应中央号召,截至 2019 年上半年,已有 30 多个省市发布人工智能相关规划或专项政策,以人工智能为技术手段,发挥当地产业集群优势,促进产学研融合及协同发展。

图 2-4：国家及地方相关政策及影响

政策名称	颁布日期	颁布主体	主要内容及影响	地方	政策名称	主要内容及影响
《国家新一代人工智能创新发展试验区建设工作指南》	2019.08	科技部	到2023年,布局建设20个左右试验区,对京津冀、粤港澳大湾区等 重大区域发展战略进行布局	北京	《北京市加快科技创新培育人工智能产业的指导意见》	加强人工智能产业发展基础和人工智能融合应用 强化人工智能基础理论研究,培育人工智能新兴产业,攻克核心技术,打造人工智能产业集群
《2019年政府工作报告》	2019.03	国务院	深化大数据、人工智能等研发应用	上海	《关于本市推动新一代人工智能发展的实施意见》	强化前沿基础研究布局,攻关关键共性技术,如芯片、传感器等,打造人工智能产业集群 拓展人工智能融合应用场景,营造人工智能多元创新生态
《新一代人工智能产业创新发展重点任务工作方案》	2018.11	工信部	确定了 神经网络芯片、开源开放平台等17个揭榜方向	深圳	《深圳市新一代人工智能发展行动计划(2019-2023年)》	通过产业发展与技术创新深度融合,发展核心零部件 围绕人工智能产业数据开放、资源互通的核心需求,建设人工智能创新服务平台,推进产业应用
《科技创新2030-新一代人工智能“重大项目2018年度项目申请报告”	2018.10	科技部	基于新一代人工智能的基础理论、重大需求的关键共性技术、新型感知与智能芯片等 3个技术方向启动16个研究任务,安排国拨经费概算8.7亿元	杭州	《杭州市建设国家新一代人工智能创新发展试验区的若干政策(征求意见稿)》	支持高校和科研机构人工智能领域的科研布局,构建科研与产业资源的协同 对基础研究平台建设、示范应用场景项目等设立 多项具体资金补助
《高等学校人工智能创新行动计划》	2018.04	教育部	构建人工智能领域多层次教育体系 ,到2020年建立50家人工智能学院、研究院或交叉研究中心	其他举例	《浙江省促进新一代人工智能发展规划》 《广东省新一代人工智能发展规划(2018-2030年)》	例如粤浙均重视加快发展人工智能软件产业,致力于成为 中国领先的新一代人工智能核心技术引领区、产业发展示范区和创新型产业集群
《促进新一代人工智能产业发展三年行动计划(2018-2020年)》	2017.12	工信部	以三年为期限明确多项任务的具体指标,对 产业链关键环节的产品和应用领域做了细致的量化			
《新一代人工智能发展规划》	2017.07	国务院	以 2020年、2025年和2030年为时间节点对人工智能提出三步走战略			
《2017年政府工作报告》	2017.03	国务院	“人工智能”首次被写入全国政府工作报告, 国家层面促进 AI 产业发展			

来源：沙利文研究院绘制

2.3 中国 AI 语音识别市场趋势洞察

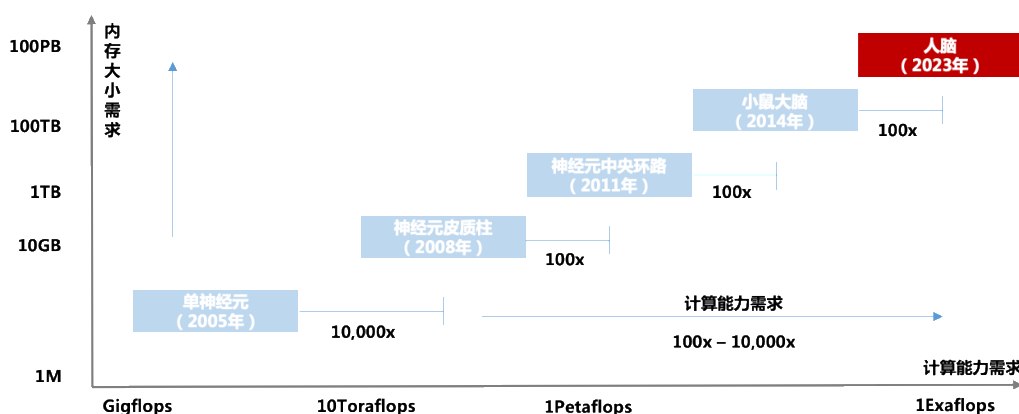
2.3.1 云计算渐发展，商业化前景更广阔

AI 语音识别发展至今,主流算法模型已经从模板匹配阶段转变为深度神经网络阶段。在深度神经网络算法下,考虑到训练过程中大量数据的使用,计算量巨大,对于应用企业而言,采用本地计算方式的算力门槛过高。而在当下的智能时代,日渐普及的云计算环境提升

AI 语音识别运算效率的同时也降低了企业的进入门槛,因此大大促进了 AI 语音的技术发展。语音识别终端把采集到的语音片段进行模数转换后,进行传送和决策,然后通过通信网络将语音数据上传至云端进行语音识别,反馈结果至语音识别终端。在此过程中,云计算可以完成语音数据库和语言数据库的训练,最高效输出反馈结果,促进 AI 语音技术的准确率提高。

基于云计算的发展,部分头部厂商也在逐渐推出基于云上的语音产品,商业化落地的步伐正在加快。在个别应用场景领域中,基于云计算的 AI 语音技术应用市场销售规模已近乎领先于头部的基于传统硬件厂商所服务的市场规模,大量的独立软件开发商 (ISV) 趋于与云上语音技术厂商达成合作关系,从而在低成本的情况下在云开放平台上获取最前沿的云上智能语音技术和行业语音解决方案。例如,现已有超过 5 万家语音客户与阿里云智能语音达成合作,覆盖多行业场景,包括中国移动、中央电视台、招商银行在内的传统行业的大型企业。其中,在电话客服行业,与阿里达成合作的独立软件开发商 (ISV) 头部 8 家用户年销售额接近 6 亿元人民币,在法院语音识别市场的联盟商 2019 年也达到年销售额 1.6 亿元人民币。基于云计算的 AI 语音技术能够满足 ISV 在录音文件识别、实时语音识别、一句话识别、语音自学习平台、短文本语音合成、长文本语音合成、语音唤醒、声纹识别、语音模组和语音交互 SDK 等方面的技术需求,从而支撑他们实现和拓展更多的应用场景和渠道的发展需求。

2-5: 神经网络模拟对于内存大小和计算能力的需求关系



来源: 沙利文研究院绘制

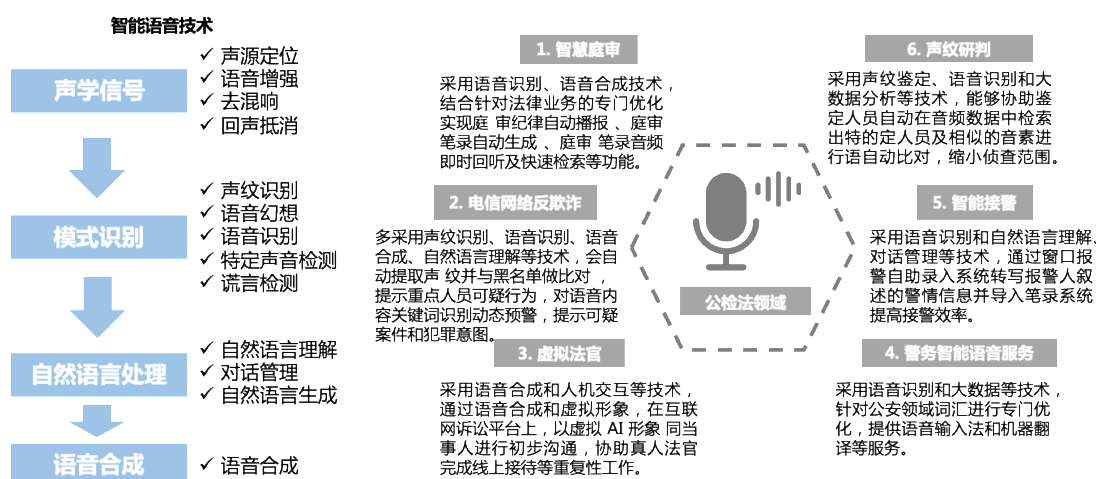
2.3.2 多技术协同发展, 语音交互更生动

语音识别属于人工智能中的感知智能,其核心功能是将物理世界的信息转化成可供计算

机处理的信息,为后续的认知智能提供基础。因此,语音识别作为人工智能的重要感知入口,除了实现本身的单一功能外,还可以与其他 AI 技术进行深度集成,应用于更广泛的生活场景中。通过前端语音交互提供入口,后端互联网提供服务,多种技术协同发展的形式,不仅为单一的技术赋能,同时也能推动 AI 语音相关产业创新,有利于未来新兴产业的崛起。如服务机器人、智能客服等新兴产业在 AI 语音识别的技术推动下正在快速发展。

技术融合已成为当下的趋势,只有将多种技术充分结合,才能为用户带来更多价值。以公检法领域为例,通过融合声学信号、模式识别、自然语言处理、语音合成等技术,可以实现智慧庭审、电信网络反欺诈、虚拟法官、声纹研判、智能接警、警务智能语音服务等功能,为公检参与者提供全面高效的服务。

2-6: 语音识别在公检法领域的技术应用和落地



来源: 沙利文研究院绘制

2.3.3 语音技术渐开放，普惠生态更繁荣

未来将会是一个普惠 AI 的时代，技术开放将会是 AI 时代非常重要的特点。尤其在当下竞争激烈的环境下，AI 语音识别厂商技术的差距已经不太明显，若能开放语音技术能力，赋能行业伙伴，不仅可以快速抓住用户，占领市场，还能促进整体生态繁荣发展，普惠 AI 应用市场。

智能硬件的“智能”，体现在强大的感知能力、机器学习、自然语言理解等，而这些功能十分依赖于大数据以及云计算技术的支撑。目前，这些数据和计算资源基本掌握在大型 AI 语音识别厂商手中，给 AI 生态的发展无形中增设了障碍。此外，随着云计算、语音识别等技术的发展，AI 语音将会渗透到各行各业中，但各行业都有其独特的属性，很难有一套通用的 AI 语音技术适用于所有的行业。为适应多元化的行业应用，提高开发和应用效率，应渐渐将定制化模型的能力开放，使开发者在前期模型训练阶段不拘于行业属性，授之以渔，定制行业特属的算法模型，真正做到普惠生态。

目前，AI 语音厂商正在逐渐开始构建这样的普惠生态。AI 语音厂商基于可靠的技术基础赋能行业升级，将软件和硬件结合，提供芯片端到语音算法、平台的一站式服务，并通过平台化的方式开放其智能语音算法能力，赋能其他合作伙伴高效开发针对性产品，为整个 AI 生态提供更为普惠的服务。打造 AI 普惠生态系统，需与 AI 行业独立软件开发商达成合作，为其落地更多的应用场景降低技术使用门槛，合作提供更贴近需求、量身定制的解决方案。如，百度宣布语音技术接口永久免费开放，提供语音识别、语音合成、语音唤醒多平台 SDK。阿里云的语音交互技术服务平台聚焦语音的核心能力，在基于云的开放式平台为中下游 ISV 提供包括了语音原子能力、开箱即用的行业模型、和自学习平台的一站式服务，紧靠行业伙伴，和客户一起打造更贴近需求的产品。

2.4 中国 AI 语音识别市场关键成功因素分析

(1) 强劲的技术支撑

AI 语音从技术突破开始，到应用、终端和场景的不断突破，再回归到技术，渐成业内共识。语义识别的加入、知识图谱的构建和技术的快速迭代，也为智能语音技术进入 3% 红线、

甚至达到更高识别率带来可能。但由于 AI 语音发展时间从整个技术长河的角度而言并不长，从听清逐渐实现听懂，最后到满足用户随心所欲，还需要不断的训练、试验，以及技术迭代。对于语音识别厂商而言，如何在现实场景下将声学、语言学等多学科技术融合，实现技术迭代和算法提升，从而提供高准确率的语音识别服务是行业的重要成功因素之一。

(2) 充足的语料积累

如果算法是 AI 语音技术的引擎，那么数据就是燃料。算法需要庞大的基于真实场景的数据，并需要对数据进行相对精确的标注，例如在建立声纹识别训练库时，至少要保证性别比例分布为 50%±5%，并包含有不同年龄区间、不同地域口音等训练样本。因此，对于语音识别厂商而言，



数据一方面可以通过长时间的终端数据积累，另一方面，可通过平台的数据资源聚集，平台能够在短时间内完成大量的语音数据积累，通过不同行业用户在平台上的使用频度，能够精准捕获市场需求。

在垂直行业的真实场景下积累充足且及时的语音资料和文本资料，并针对以上资料加以标注及进行实时更新和迭代是优化用户体验和提升客户粘度的关键，也是语音识别行业的关键成功因素之一。

(3) 丰富的场景土壤

应用、终端和场景带来了大量应用数据，更为应用于语音识别的机器学习、深度学习带来了技术突破。语音识别技术的爆发是源于大数据，数据量越多，语音识别的算法准确性越高，语音识别的识别准确率相应越高。

此外，语音识别技术需要丰富的场景土壤来培养快速的复杂场景处理能力。在消费级用户需求方面，从语音识别技术发展的开端起，消费者对语音识别技术就建立了高预期——希望利用新技术提升生活体验，将高准确率的语音识别技术创造性地融入到日常场景中。在专业级用户需求方面，司法、医疗、教育、电信、交通等行业的政府及企业级用户需要语音识别系统在实际业务应用中表现出功能的可靠性和稳定性，因此，这些专业级用户在选择语音识别产品时会通过严格的招投标选择最具实力和行业经验的 AI 语音识别产品和服务供应商。面对日益提升的用户需求，AI 语音厂商需要积累丰富的场景经验，培养快速的复杂场景处理能力以赢得市场。此外，用户在选定语音识别供应商后倾向于长期保持稳定合作，这也成为 AI 语音厂商持续成功的因素之一。

3 中国 AI 语音识别市场竞争格局分析

沙利文通过深度访谈和调研市场领导者、参与者、用户及行业专家，查阅公开信息，制定了一套包含一系列评价指标和市场权重的评价体系，并基于该评价体系对中国 AI 语音主流厂商进行客观公正的评估，分析其在中国的 AI 语音行业中增长力情况。

3.1 沙利文企业增长评价数据来源和研究主体

本报告数据来源于技术指数、市场数据及 AI 语音识别领域开发人员、服务提供商、行业专家访谈等。报告的评价主体为中国 AI 语音识别主流厂商，按照其属性可分为 IT 及互联网厂商、语音技术厂商两类。

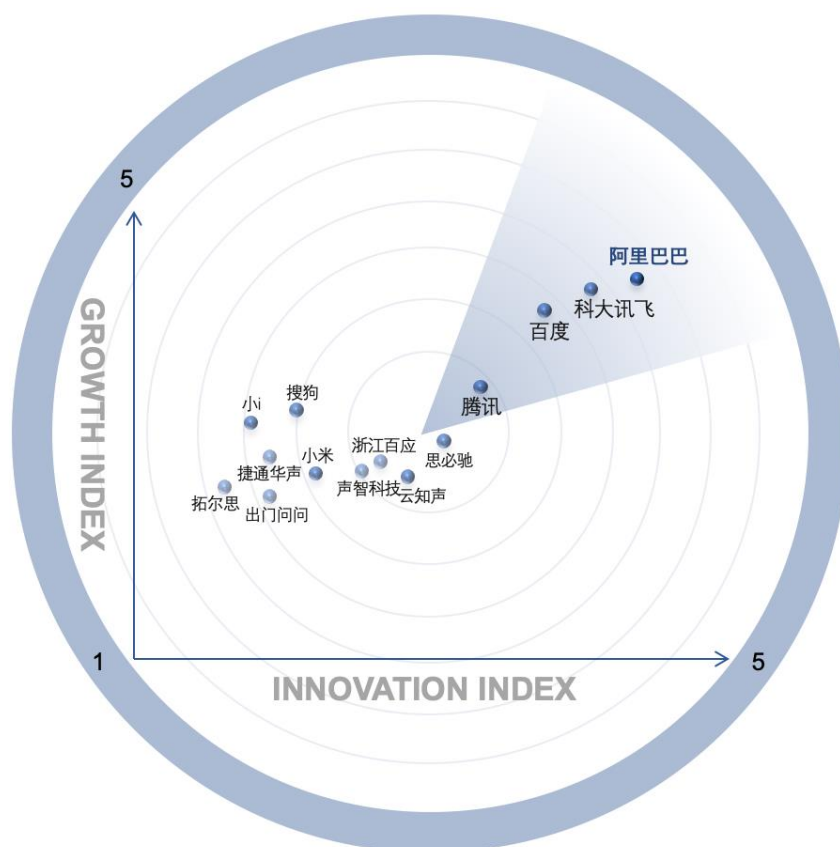
IT 及互联网厂商：IT 及互联网厂商包括百度、阿里巴巴、腾讯、搜狗、小米等厂商。凭借在互联网时代积累的大量资本，IT 及互联网厂商在众多科技领域成为了主要“玩家”。与依靠人工智能技术起家的 AI 垂直类企业不同，IT 及互联网厂商侧重以流量导向和满足庞大用户群体验和创新发展需求为导向去做 AI 开发，更多以流量和用户体验为导向，同时，会比传统的语音技术厂商更加注重创新实践。

语音技术厂商：语音技术厂商也可以分为传统语音技术厂商和创业厂商。传统语音技术厂商包括科大讯飞、小 i 机器人、捷通华声等拥有自己的核心智能语音芯片以及语音相关软件系统的厂商，创业厂商包括云知声、思必驰、出门问问等专注于某些垂直领域（如汽车、家电）来推广自己的语音技术和产品的厂商。

3.2 中国 AI 语音识别市场企业增长评价结果及分析

通过以上评价和模型计算，可得到综合评价排名如下：

图 3-2: 中国 AI 语音识别主流厂商竞争力分析



注：圆点颜色深浅代表客户指数值大小，圆点颜色最深表示客户指数值越大。

来源：沙利文研究院绘制

基于本模型的评价，阿里巴巴、科大讯飞、百度、腾讯均处于高竞争力区间，且阿里巴巴在增长指数、创新指数和客户指数三个维度均表现突出。

● 阿里巴巴

阿里智能语音交互服务是业内领先的云原生语音服务平台，对阿里集团内服务于超过 99% 语音场景，对外提供各类云上语音产品。（1）在语音识别全产业链方面，阿里云上语音技术在大量数据积累、算法积累（独创的下一代端到端语音识别技术 SCAMA 和 SAN-M 算法，基于 CPU 服务器的高并发、实时/离线的语音识别能力）、工程积累（云原生 AI 技术、大规模弹性计算能力、支撑集团日均数亿请求服务能力）和阿里达摩研究院技术同步上云的成果加持下，极大提升了语音交互的准确率和性能。阿里智能语音交互技术已解锁并成熟布局于多个场景中，包括智能客服、

智能质检、法庭庭审实时记录、实时演讲字幕、访谈录音转写、声纹登录、设备端语音交互等场景，在政务、金融、物流、教育、电商、泛互联网、医疗、餐饮等多个领域均有应用案例以及大量的客户积累。(2) 在市场方面，目前，阿里智能语音分别在电话客服和法院语音识别的市场都占有领先的位置(全国电话客服领域最大的技术合作联盟；与法院行业龙头应用厂商全部达成合作，覆盖近万间线下法庭和超过 1.5 万间线上法庭)。阿里语音 AI 技术能在多个领域快速落地，占领市场并成为行业客户认知度前列的云上语音技术厂商，关键因素之一是与大量的独立软件开发厂商结成了阿里语音 AI 产业联盟。(3) 在企业经营和战略方面，阿里云在基建技术的布局的规划方面，将继续加大投入对云操作系统、服务器、芯片、网络等核心技术的研发，为语音 AI 与云的结合带来更具想象力的应用场景和价值空间。(4) 在创新方面，阿里重视对技术的研发投入和对行业的贡献，其智能语音核心技术能力是下一代端到端语音识别技术 SCAMA、SAN-M 技术和 DFSMN 技术，这几种技术都是业界首次在非科研领域的应用并大获成功，使得语音交互的准确率在高并发的情况下获得大幅度提高，同时 DFSMN 也已经面向全行业开源，为行业整体技术提升做贡献。另外阿里在业内率先推出了自学习平台改变了语音 AI 生产关系，即便行业实践者没有很多语音领域的专业知识，也能够用阿里云自学习产品，通过灌注入行业内的数据和知识，就能够获得所在行业不错的语音交互效果。阿里的智慧大屏解锁了多场景，阿里达摩院全球首创多模态语音交互方案，与语音、视觉、自然语言理解多 AI 技术融合，能够实现在强噪音环境下的免唤醒人机交互，具有突出的产品组合优势。(5) 在客户服务方面，阿里智能语音最为主要的商业策略就是为其合作伙伴提供语音原子能力、多个领域开箱即用模型和“自学习”平台，赋予客户产品“能说、会说、懂你”式的智能人机交互体验，在将阿里云智能语音领

先技术赋能于产业联盟企业的同时, 构建一个更为完整和繁荣的 AI 语音应用生态, 普惠市场。

- **科大讯飞**

科大讯飞是中国领先的智能语音技术提供商之一, 在语音识别等智能语音技术领域拥有国际先进的成果, 主要业务包含智能语音和语言技术研究、软件及芯片产品开发、语音信息服务及电子政务系统集成, 为企业、政府和个人提供智能应用。科大讯飞技术优势在同行业中出众, 拥有自主知识产权的智能语音技术, 已推出覆盖大型电信级应用到小型嵌入式应用, 以及满足不同应用环境的软硬件产品。科大讯飞在中文语音技术市场和语音合成产品市场都占有较大份额, 应用于电信、金融、电力、社保等主流行业。科大讯飞结合自己核心的智能语音技术和人工智能的研究, 已经实现 AI 产品化的布局。在 To B 端, 科大讯飞面向教育、政法、医疗、汽车和客服等多个行业提供相应的解决方案, 如智能汽车方案中的飞鱼车机服务, 是一款软硬数服一体化的智能车载设备, 将降噪、唤醒和识别等智能语音技术及人工智能技术应用于汽车, 让用户在复杂的行车环境中轻松获取娱乐、通讯、导航、车身控制等服务。在 To C 端, 科大讯飞提供面向消费者的家庭、汽车、翻译等消费级个人产品, 包括硬件产品如讯飞学习机, 讯飞翻译机, 智能办公本等, 以及软件产品如讯飞输入法, AI 电视助手等。

- **百度**

百度在 AI 开放平台上提供语音识别的技术, 共享 AI 领域的应用场景和解决方案。提供产品包括短语音识别、实时语音识别、远场语音识别、音频文件转写、语音自训练平台等, 应用于手机应用语音输入、机器人对话、语音内容分析和实时语音转

写等场景。百度的语音识别技术优势在于领先的语音识别技术，提供自助训练专属模型（使用业务场景文本语料，零代码自助训练语言模型，精准识别语音内容，有效提升业务领域识别准确率使用），可简单快速地接入（支持 API 及多种 SDK 接口，领先识别解码技术提升识别速度），提供稳定性保障，收取费用较低。

- **腾讯**

腾讯云语音识别目前主要应用于微信、王者荣耀、腾讯视频等内部业务使用，外部应用布局于录音质检、会议实时转写、法庭/审讯记录、语音输入法等场景；技术提供主要功能为开放实时语音识别、一句话识别和录音文件识别，支持云端和嵌入式两种服务模式。腾讯云语音识别的主要优势在于有大量的用户基础和数据积累可应用于丰富的行业场景，收费价格较低（购买预付费包后识别 1 小时语音不到 1 元），字准率高达 97%，支持多语言识别能力（中文普通话，粤语，英语和韩语），有领先的算法（基于创新网络结构 TLC-BLSTM，利用 ATTENTION 机制有效地对语音信号进行建模，通过 Teacher-Student 方式提升系统鲁棒性，对通用以及垂直领域下场景有领先业界的识别精度和效率，提供自主提升准确率的功能。

3.3 沙利文评价模型的设计

本次沙利文针对中国 AI 语音识别主流厂商进行增长力评价，旨在确定市场未来的增长机会，并评估能高效利用这些机会的公司。本次增长力评价主要评估两个关键指标：公司对持续创新的关注以及将创新转化为持续增长的能力，分别表现为创新指数 (Innovation Index) 和增长指数 (Growth Index)。此外，为体现公司的增长为客户带来的价值，同时将公司提供的客户服务价值--客户指数 (Customer Index) 作为辅助评价指标。

- **创新指数 (Innovation Index)**：公司对持续创新的关注。创新指数是衡量公司开发产品/服务/解决方案的能力的指标，这些产品/服务/解决方案的开发清楚地展示

了公司未来的趋势，能够发展和扩展以服务多个市场，并与客户的需求保持一致变化的能力。

- **增长指数 (Growth Index)**: 公司将创新转化为持续增长的能力。增长指数是对公司增长绩效和业绩记录的衡量, 以及其制定和执行完全一致的增长战略和愿景的能力; 健全的增长渠道系统; 有效的市场; 竞争策略; 和以用户为中心的销售和营销策略; 以及基于对未来的远见卓识和有效应对新挑战和机遇的能力, 持续制定新增长战略的能力。
- **客户指数 (Customer Index)**: 公司能将其创新及增长能力最终转化为产品及服务提供给客户的能力, 也即提供的客户服务价值。沙利文基于以上评价框架, 采用 360 度研究方法对中国 AI 语音识别主流厂商进行深入分析, 涵盖 AI 语音识别行业的整个价值链, 参考市场领导者、参与者、用户及行业专家意见, 获得 AI 语音市场中主流厂商的客观评价。

图 3-3: 中国 AI 语音识别市场企业增长评价体系

核心评价指数	评价指标分类	指标
增长指数	产业: 全产业链生态布局	上游-基础硬件和基础支撑能力
		中游-平台和框架, 基础核心技术
		下游-应用场景
		一站式服务的提供
	市场: 市场表现	市场份额
		收入增长
		合作伙伴赋能
	企业: 经营和战略	增长渠道丰富度及空间
		战略布局
创新指数		创新延展性
		研发投入
		产品组合
		创新趋势协同
客户服务指数		客户需求协同
		产品的普惠性

来源: 沙利文研究院绘制

4 研究方法和研究范围介绍

4.1 研究方法

沙利文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，从信息科技行业，大健康行业，新能源行业等领域着手，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。

- ✓ 弗若斯特沙利文本次研究于 2020 年 12 月完成。

4.2 研究范围

沙利文中国 AI 语音识别市场研究报告基于对语音识别的理解，从技术领域、行业领域、市场参与者等多个维度对中国 AI 语音识别市场进行分析，研究中国 AI 语音识别市场发展的驱动因素，主要趋势，以及进入壁垒和成功关键因素，并对中国 AI 语音识别主流厂商进行企业增长能力评价分析，为中国 AI 语音识别提供商及 AI 语音识别使用者提供参考。

- 行业领域范围：包括 AI 语音识别在中国不同行业领域中的应用，本报告将聚焦智慧法院、智能客服、智慧交通等领域进行深入探讨。

- 市场参与者范围：包括 AI 语音识别产业链中提供软硬件、产品和服务的 AI 语音识别提供商以及政府、企业、消费者、人工智能开发者等人工智能使用者。