

TALK TO ME

How Voice Computing Will Transform the
Way We Live, Work, and Think

让机器能听会说、能理解会思考，用人工智能建设美好世界！

——科大讯飞董事长刘庆峰

智能语音时代

商业竞争、技术创新与虚拟永生

[美] 詹姆斯·弗拉霍斯 (James Vlahos) 著 苑东明 胡伟松 译

解密苹果、亚马逊、谷歌、脸书、微软、
科大讯飞等商业巨头的智能语音布局与商业角逐

决胜未来的智能语音核心技术、商业思维和未来趋势



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

TALK TO ME

How Voice Computing Will Transform the
Way We Live, Work, and Think

智能语音时代

商业竞争、技术创新与虚拟永生

[美] 詹姆斯·弗拉霍斯 (James Vlahos) 著

苑东明 胡伟松 译

Talk to Me: How Voice Computing Will Transform the Way We Live, Work, and Think

by James Vlahos

Copyright © 2019 by James Vlahos

Simplified Chinese translation copyright © 2019

by Publishing House of Electronics Industry

Published by arrangement with author c/o Levine Greenberg Rostan Literary Agency

through Bardon-Chinese Media Agency

All rights reserved.

本书中文简体版专有翻译出版权由 James Vlahos 授予电子工业出版社。未经许可，不得以任何手段和形式复制或抄袭本书内容。

版权贸易合同登记号 图字: 01-2019-1623

图书在版编目 (CIP) 数据

智能语音时代：商业竞争、技术创新与虚拟永生 / (美) 詹姆斯·弗拉霍斯 (James Vlahos) 著；苑东明，胡伟松译. —北京：电子工业出版社，2019.6

书名原文: Talk to Me: How Voice Computing Will

Transform the Way We Live, Work, and Think

ISBN 978-7-121-36208-8

I. ①智… II. ①詹… ②苑… ③胡… III. ①人工智能—应用—语音识别

IV. ①TP18②TN912.34

中国版本图书馆 CIP 数据核字 (2019) 第 057398 号

出版统筹：刘声峰

策划编辑：黄 菲

责任编辑：黄 菲

文字编辑：王欣怡

特约编辑：刘广钦

印 刷：北京盛通印刷股份有限公司

装 订：北京盛通印刷股份有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：720×1 000 1/16 印张：26.5 字数：353 千字

版 次：2019 年 6 月第 1 版

印 次：2019 年 6 月第 1 次印刷

定 价：98.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888，88258888。

质量投诉请发邮件至 zltts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：1024004410 (QQ)。

献给我的父亲约翰，因为他没来得及看到本书完稿。献给
我的妻子，因为她见证了本书的整个写作历程。

出版前言

作为一个见证者，我非常庆幸自己亲历了波澜壮阔的互联网时代、大数据时代、人工智能时代，深切感受到了接踵而至的技术浪潮是如何快捷地改变我们的工作和生活的，如果未来有闲暇时间追根溯源，那一定是一件非常美好的事情。

作为一个出版人，我也非常高兴能够遇到今天正式呈献给各位读者的这样一部优秀的科技人文作品，对于这本深入浅出、充满趣味又可能引领一个新的时代到来的科普著作，能够先睹为快，实在是一件令人赏心悦目的事情。

关于智能语音时代，我们大部分人已经多多少少有了一些直接或间接的感受。比如我们手机上的 Siri 或者其他语音软件，这曾是苹果手机最令人瞩目的新功能之一；又比如我们经常在用的语音导航软件里传出的林志玲的“娃娃音”，还有微软小冰展现自己诗歌“别才”的诗集《阳光失了玻璃窗》；当然，还有电子鸡、旅行青蛙这样风行一时的电子宠物。

对我来说最新的例子是，我在 2019 年春节期间购买了一台小米智

能音箱（“小爱同学”）作为礼物送给父母。“小爱同学”的乖巧能干，可着实把他们惊呆了。“小爱同学”为他们做的第一件事是播放花鼓戏《浏阳河》，这是现代技术与古老文化碰撞出的新奇体验。

与我们直接的生活经验不同，被称为美国鬼才科普作家的本书作者詹姆斯·弗拉霍斯（James Vlahos），给我们带来了语音技术这样一个蔚为大观的科技发展趋势的深入洞察。作为智能语音领域的开山之作，作者确实出手不凡，他把语音技术、应用与产业的讨论引向了难得的高度，非常清晰地给我们展现了智能语音时代的新场景。

詹姆斯·弗拉霍斯是一位长期追踪研究语音技术和语音人工智能领域的专题记者，30多年来，他零距离地见证和细致地观察了这一领域的研究进展，曾与这一领域中的许多杰出人物直接对话，本书中的许多观点就是来自他对第一手访谈资料的提炼。

更难能可贵的是，本书作者还是这一领域的深度涉猎者和亲身参与者。他从十几岁起就对人与机器的对话产生浓厚兴趣，并曾尝试用BASIC语言开发了自己的文本对话游戏（虽然只能运行寥寥几步），他还开发出了以自己父亲为蓝本的聊天机器人，让自己的父亲实现了“虚拟永生”。

因此，本书不管是观察与思考的深度和广度，还是研究资料的广泛和翔实等，各个方面都令人钦佩，也让人感到“解渴”和过瘾。读这样一本书，从实用的角度说，花费的时间有更高的性价比，它能让我们真切地感受到一个新的产业生态的发展趋势，启发我们从经济、社会和文化等角度来思考语音技术和语音人工智能已经或即将带来的影响。

翻开这本书，你首先就会被作者这样的话语所吸引：每十年左右，

人与技术的互动方式就会有一个根本性的转变。数十亿美元的财富会“恭候”那些定义了新的时代范式的公司，而落伍者将破产倒闭。在计算机的大型机时代，IBM 是主宰者；微软公司是桌面时代的王者；谷歌公司靠索引领了互联网时代；苹果公司和脸书公司则在移动互联网时代一飞冲天。

最近的一次范式转移正在进行中。

最新的平台之战已经打响。

最新的技术颠覆正在发生，无论是其规模还是其重要性，都可能是世人前所未见的。

我们正在迈入智能语音时代。

语音正在变成影响现实的通用遥控器，成为几乎能控制任何一种技术装置的手段。语音能够让我们指挥各种数字产品助理——“行政助理”“门房”“主妇”“管家”“顾问”“保姆”“图书管理员”“演艺人员”等。语音打破了世界上一些最有价值的公司的商业模式，为新的应用创造了机会。语音把对人工智能的控制权交给了用户。很久之前科幻作品就预言过这样的关系模式，在这样的关系模式中，拟人化的人工智能成为我们的“助手”“看门人”“预言者”“朋友”。

作者这样言之凿凿，显然并非空穴来风。了解作者的思考逻辑，把握这样一个大趋势，并以此指导我们的工作和生活，必然有利于我们做出更多正确的选择。书中提到，当 Siri 在 2010 年刚刚被开发出来时，先知先觉的苹果公司前 CEO 乔布斯曾经连续 17 天每天给开发者之一的吉特劳斯打电话，有时甚至深更半夜也打，终于把 Siri 收入苹果公司囊中。

远见从来都是人类最宝贵的品质之一，作为智能语音领域的第一本书，这也可以视为一本“远见之书”。

除了具有经济性含义的远见，本书还非常重视这项新技术对人类精神和感性世界的影响，甚至作者也把自己和自己的家庭带入了与语音人工智能的互动过程中，这大大增强了本书的故事性。本书有文采、有温度、有趣味，展卷在手，没有同类书的枯燥和沉重，反而有一种引人入胜、不忍释卷之感。

正如作者在书中所言：智能语音时代的到来是人类历史的转折，因为运用语音是我们人类这个物种的特质——这一能力把我们和其他物种区分开来。人类的内部意识的中心不在肺部的空气里，也不在血管里的血液中，而是在大脑的语言区里。语言调整着我们的关系，它能塑造思想、表达感受、沟通需求；它能发起变革、挽救生命、激起爱恨情仇；它把我们所知道的一切记录下来。

不管语言是由人说出来还是由机器说出来的，尤其是当“你应我答”的模式出现，在人与人之间、人与机器之间，交谈就绝不只是一种纯粹依靠逻辑展开的过程。语言永远不是脱离内容的外壳，人都会被语言影响或打动。作者在书中讨论的种种事例和情境，都让我们领悟到人和机器之间的语言交流对我们的情感世界带来的影响和改变。未来，我们与无处不在的机器构成的世界，将是一个前所未有的更加丰富多彩的感性世界。在云时代，“只要简单地加上一个麦克风和一个 Wi-Fi 芯片，任何装置都能实现语音驱动。从浴室的水龙头到孩子玩的布娃娃，任何装置都能利用分布在全球的几千台计算机所提供的计算能力。”这几乎意味着“万物能言”的童话世界真的实现了。

基于这样的前景，作者指出：当聊天机器人同时作为工具和准生命

进入我们的生活时，它们模糊了人与机器人的界限，模糊了隐私、自主权和亲密感的界限，还模糊了人际关系与数字关系、现实与虚拟、生与死的界限。

可以想象当这些界线模糊之后，在我们的生活中将会发生多少故事。这些故事肯定不会按照单一的模式进行，必定会有更多“人机情未了”式的故事演绎。

除了上述简单提到的精彩内容，还值得一提的是本书中充满浓厚的中国元素，从另外一个侧面拉近了中国读者与这一话题的距离。

作者在书中用很大篇幅讨论了亚马逊公司主办的亚历克莎奖竞赛，他这样介绍在比赛中拔得头筹的华盛顿大学团队，“这种方法是由该团队 28 岁的学生领袖郝方提出的。郝方来自中国宜春市，他活力四射、性格开朗。他和他的团队成员希望让他们的聊天机器人的评审用户也能感到快乐。”正是这位郝方同学带领的团队所开发出的聊天机器人创造了交谈长度 20 分钟的记录。

“当华盛顿大学团队的成员上台后，普拉萨德把那份令人满意的奖品发给了他们——一张金额达 50 万美元的巨额奖券式支票。郝方大笑着拿过支票，对着镜头竖起了大拇指。”

此情此景也让我们为这位郝方同学高兴。

在由 10 万个问题组成的斯坦福问答数据集测试中，真人平均能答对 82% 的问题。微软公司、阿里巴巴公司在 2018 年 1 月公布，它们所开发的系统得分和普通人得分一样高，这成了当时的头条新闻。

另外，还有在微软公司负责 Zo 聊天机器人项目的王颖，以及大家所熟悉的微信，都是书中屡屡提及的对象。这些中国元素让我们看到，

我们与这一项划时代科技突破的关系从来没有像今天这样接近过。这令我们感到自豪，也让我们与本书的主题产生了千丝万缕的关系。更何况，我们的人工智能领域的标杆企业——科大讯飞，经过在智能语音领域的勇敢探索，已经成为全球智能语音产业的主力军和技术领先者。

作为一项具有重大颠覆性的技术，语音技术和语音人工智能带来的影响是非常深远的，我们难以给出一个简单判断。作者对此的认识非常深刻，他指出：“从鱼钩到火星探测器，我们一直在制造工具。虽然我们制造出了很多对我们有用的东西，但它们在更深层次上都不像我们。即使是类人机器人，它们能做的也只是笨拙地移动。使用语言是人类这个物种真正与众不同的地方。语言把我们连接起来。因此，教机器掌握语言不同于通过编程让它们学会进行衍生品交易、做手术、进行海底航行或其他事情。我们正在“共享”人类的核心特征。”

我们应该看到，“就像历史上的其他给人带来便利的新技术一样，人工智能也可能会让我们付出新的代价。我们可能在智力活动上变得更加消极，我们将更少自主地寻找答案。寻找答案是一种激发好奇心、激发思考的过程。有了人工智能，答案会来找我们。与打开水龙头放水相比，从井里费力地打水明显过时了，而费力地寻找答案也正在变得过时。”

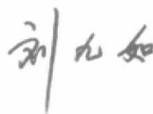
这显然可以视为其消极的一面，但人类从未因为其消极的一面而排斥过任何一项能够带来巨大便利的新技术。

因此，作者又向我们指出：如果应对得当，语音技术有可能成为我们发明的最有感情的技术。认为人工智能只能是冷冰冰的算法的观点是错误的。我们可以将最好的价值观和同理心注入其中。我们可以让它变得聪明、令人愉快、精灵古怪，并且善解人意。有了语音技术，我们最

终可以制造出不那么陌生、更像人类的机器。

未来已来，一场智能语音科技大秀的帷幕正在拉开。随着 5G 时代的到来，包括语音技术在内的人工智能技术，一定会让世界更美好。

本书在出版过程中，得到了工业和信息化部信息化和软件服务业司副司长董大健先生，科大讯飞董事长刘庆峰先生，以及北京市科学技术协会、科大讯飞的大力支持，特此致谢。我们相信，本书的出版发行，能够更好地助力我国语音智能产业的发展。我们期待，各位打开这本书，能更加全面地把握语音技术与人工智能的发展态势，激发起创新创业的强烈愿望。让我们积极迎接智能语音时代到来！

A handwritten signature in black ink, reading '刘九如' (Liu Jiaru). The characters are written in a cursive, flowing style.

电子工业出版社总编辑

智能语音，开启万物互联时代的大门，让 AI 闪耀人性光芒

一位被诊断为患有晚期肺癌的父亲，在生命末期，为孩子留下了 91970 个单词的口述。孩子打造了一台爸爸机器人，让父亲在声音的世界里“永生”——这个孩子就是本书的作者。

这个令人动容的故事让我们感受到语音的温度和科技的温暖。

在中国，科技也在创造着这样的温暖。2018 年年初，在全球首部利用人工智能配音的纪录片《创新中国》中，我们合成了中央电视台已故配音大师李易的声音，用技术向艺术致敬。在首映式上，李易老师的弟子们集体起立、热泪盈眶。

语音，是人类呱呱坠地后最早使用的沟通方式，也是现代人际交往最基本的方式，更是未来人机交互最重要的方式。人工智能跌宕起伏发展 60 多年，智能语音是发展到今天最为成熟、也是最重要的板块之一。“最近的一次范式转移正在进行中。”作者在书中提到，这次转移正是关于智能语音的。

语音，开启万物互联时代的大门。

在互联网发展的下半场，我们将进入万物互联的新时代。随着越来越多的设备在无屏、移动、远场状态下被使用，作为人类最自然、最便捷的沟通方式，语音将会成为所有设备至关重要的入口。未来，我们将迎来以语音交互为主、键盘触摸为辅的全新的人机交互时代，人和机器之间的沟通，可能完全是基于自然语言的，你不需要去学习如何使用机器，只要对机器说出你的需求即可。

比如在导航软件中，你能听到各种明星的合成声音，可以用他们的声音为你指路；在电视上，你能看到虚拟主播播报的多语种新闻，与真人相比不仅相似度高，而且 24 小时无休；在居家生活中，你能通过语音控制音乐、灯光、温度，实现智慧家居；甚至在医院里、社区里，你能用语音调动机器人帮你办理事项，节省时间……人工智能已经在为我们的日常生活服务，智能生活的大门正缓缓打开。

语音，让时代更具人性温度。

智能语音是通向万物互联时代的必经之路，它的存在让交互方式拥有无限的可能，也让这个时代更具人性的温度。

20 世纪 90 年代，我在就读于中国科学技术大学时被选进人机语音通信实验室，研究“如何让机器像人一样开口说话”。那时，团队的一个梦想是研发一台能自动翻译的电话，即使交流时语言不通，通过人工智能技术也能让我们无障碍地交流；20 多年后的今天，我们自主研发的翻译机已经支持中文与 50 种语言的实时翻译，每个月总共为全球提供超过 5000 万次服务。智能语音让被地域、文化等因素隔离的人们也能无障碍地沟通。

此外，我们通过技术在听障和视障人群间搭起沟通的桥梁，让听障群体通过语音识别技术“看得见”声音，让视障群体通过语音合成技术“听得见”文字。2017 年我们发布了“三生有幸”公益计划，目前已有几十万残障人士受益。语音转写、语音朗读为他们获取信息带来了极大便捷。我们希望，随着语音技术的使用与发展，未来每个人都将因 AI 而能。

语音，在万物互联时代技术门槛将会更高。

在以语音为主、键盘触摸为辅的万物互联时代，人们对语音交互提出了更高的技术期待与需求。今天，虽然在安静、发音标准的情况下，中文的语音识别准确率已经可以达到 98%，英文的语音识别准确率可以达到 95%，但在有方言、噪音、口音和远场的情况下，距离语音识别高准确率或许还有很长一段路要走。

以 2018 国际语音识别比赛 CHiME-5 为例，它是世界上最权威的语音识别比赛，考察在噪声和远场环境下的语音识别效果。但是让人意想不到的是，比赛主办方用最新的算法和深度学习模型做了参考系统，在测试中语音识别错误率竟高达 81.14%，可以说是“史上最难语音识别任务”。科大讯飞虽然在这次比赛中取得全部四个项目的第一名，将错误率降低了 35 个百分点，但是距离高准确率仍有不小的差距。可以看到，在万物互联时代，语音识别技术还有非常大的提升空间，语音识别的门槛不是降低了，而是提高了。

20 年前，我和实验室的同学们一同创立科大讯飞，就是认定了智能语音巨大的潜力和广阔的前景，它会让人机信息沟通无障碍。今天，看到这本《智能语音时代》，我非常高兴。作者对语音技术的发展趋势有着深刻洞察，从 Siri 诞生到谷歌助理、亚历克莎的规模化应用，以翔实

的资料、细致的文笔讲述智能语音时代的到来及其可能带来的影响。这本书不仅是对智能语音的科普，更让读者对智能语音未来的发展有了更多的了解。

“他山之石，可以攻玉。”人工智能正在成为全球化发展的关键力量，中国的语音技术和产业也必将在其中发挥更大力量。相信本书的出版，会让更多人重新认识神秘且熟悉的语音世界，让我们一起携手，让世界聆听我们的声音，让沟通从 AI 开始。

刘庆峰

科大讯飞董事长

译者序

因为幸运地托庇于一家优秀的企业，因为有家庭这个稳定的大后方，我四十岁后的生活，显得波澜不惊，也因为有稳定的预期而变得无忧、无惧。

这是不可否认的幸福生活。

我为此而深深感恩。因为这样的生活能够让我以一种从容的心态去超越生活，而不必以剑拔弩张的姿态去与生活争斗，更不必“赋到沧桑句便工”。

与电子工业出版社（以下简称电子社）的相遇和相知与我而言就是这样一种从容而幸福的超越，是在不知不觉中，漂流到了一处未曾意料过的“桃花源”，自己的生命也因此有意无意之间变得更加丰盈起来。

第一次接触电子社的书是在 1988 年，那时我正读大学二年级，从此便对这家出版社有了印象。

成为电子社的译者则始于 2015 年翻译《学会学习》一书，从此，电子社成了与我的生命有最多交集的文化机构。四年过去了，本书已经是我为电子社翻译的第 11 本书，在这四年间，这 11 本书成为我这

段生命航程中虽不耀眼，但足以让我感到小小满足的一份成绩。

这 11 本书的翻译是在工作之余完成的，它们不是我生活内容的主体，也不是我发力死磕的对象，一切似乎都是很自然地生发出来。赶工的辛苦自然是有的，译完一本书的那个瞬间所体会到的轻松畅快也沉淀在记忆中，但让我印象更深的是那种一本书译完之后大约十几天到一个月就会产生的虚空感，仿佛一切已经归零，又该继续“战斗”了。这个时候，当电子社的刘声峰老师、黄菲老师问“有本书愿不愿意翻译”时，我真仿佛如闻“纶音”，肾上腺素会陡然升高，对生活的意义似有了更明显的感知。

毫不夸张地说，与电子社合作的翻译事业，在无意中丰富甚至改变了我的人生。这固然不是什么了不起的大事，但作为一个普通人，我们的人生本就平淡无奇。电子社的 11 本书，加上为中国人民大学出版社翻译的 9 本书，把我这四年的闲暇时光填充得满满当当，有力地提高了我生命的密度，驱走了许多可能是庸人自扰的无聊。人生的陀螺旋转得更顺畅、更自信，生活也在运动中达到了更理想的平衡。四年来，当生命和时间像流水一般逝去，在一片琐碎的生活汪洋中，还分布着这样一些属于真诚努力和用心探讨的“岛屿”，这让我深感幸运。

还要说说电子社的刘声峰老师、孙学瑛老师和黄菲老师，其中只是与刘声峰老师有过一面之交，但感觉与各位老师都神交已久。他们的豪爽与真诚，质朴与平易，让我产生了要与电子社风雨同舟的亲切感和使命感，推动我突破理性的界线，夸张地想以对历史负责的态度，对一本译作视若己出、尽心用情。

最后说一下这本《智能语音时代》。译罢本书我有一种如饮醇醪的感觉，感觉十分幸运。在我的阅读范围之内，在我国，无论是对一个产