科技工具人

当ChatGPT的广东话“讲唔正”：AI 年代，弱势语言是否注定被边缘化？

在AI 半吊子的广东话背后，是语言传承与社会资源分配的角力。

图：Mantha Mok / 端传媒

你听过 ChatGPT 说广东话么？

如果你是普通话母语者，恭喜你瞬间收获“精通粤语”成就。反而是会说广东话的人，这时可能要一头雾水了－－ChatGPT 自带奇特口音，像外地人在努力说广东话。

2023年9月的一次更新中， ChatGPT第一次拥有了“说”的能力；2024年5月13日，最新一代模型 GPT-4o 发布，虽然新版的语音功能尚未正式面世，只存在于 demo 中，但从去年的更新中，已经可以一窥 ChatGPT多语言语音对话的能力。

而很多人也发现了，ChatGPT 讲广东话口音浓重，虽然语气自然，像真人一样，但那个“真人”肯定不是广东话母语者。

0:00 / 3:52

为了查证这一点，探寻背后的原因，我们展开了粤语语音软件的对比测试：受测者有 ChatGPT Voice、苹果 Siri、百度文心一言，以及 suno.ai。其中，前三者均为语音助手，suno.ai 则是近期红热极一时的人工智能音乐生成平台。它们都具备根据提示词用粤语或近似粤语来生成回应的能力。

就词汇发音而言，Siri 和文心一言都发音正确，但回答比较机械和死板，其余两位选手则有不同程度的发音错误。很多时候，错误之处都是在用倾向普通话的方式来发音，比如“影”粤语应作“jing2”，变成了普通话“ying”；“亮晶晶”应作“zing1”，却读成“jing”。

0:00 / 0:06

苹果Siri讲广东话绕口令

“高楼大厦”的“高”被 ChatGPT 发为“gao”，而实际应为粤拼“gou1”。土生土长的广东人 Frank 也指出，这是一个非母语者中常见的发音错误，还常被本地人拿来开玩笑——因为“gao”是指涉性器官的广东话脏话。ChatGPT每次发音表现都会略有不同，“高楼大厦”的“厦”有时能发为正确的“haa6”，有时又错读为“xia”，一个广东话中不存在，近似普通话中“厦”的发音。

0:00 / 0:16

ChatGPT用广东话介绍端传媒

语法上，生成的文本明显更偏书面，只偶尔夹杂口语化表达。遣词造句也时常会突然切换为普通话的模式，脱口而出“买东西”（广东话：买嘢），“用粤语来给你介绍一下香港啦”（广东话：用粤语同你介绍下香港啦）等不符广东话惯用口语语法的句子。

0:00 / 0:07

百度文心一言用广东话介绍香港

suno.ai 在创作广东话饶舌歌词时，也写出类似“街坊边个仿得到，香港嘅特色真正靓妙”的，语义不明的歌词；我们把这句拿给ChatGPT 评价，它指出“这句似乎是普通话的直译，或者是普通话混合广东话的句法（syntax）”。

0:00 / 1:22

Suno.ai 创作的广东话饶舌歌曲

作为对比，我们也发现，在它们尝试使用普通话时，这些差错基本都不会出现。当然，同是广东话，广州、香港、澳门都有不同的口音与用语差别；被视为粤语“标准”的西关口音，与香港的常用广东白话就非常不一样。 但ChatGPT的广东话，最多只能说是“唔咸唔淡”（指不熟练，半吊子）的普通话母语者会有的口音。

这是怎么一回事？ChatGPT是不会广东话吗？但它没有直接表示不支持，而是对它展开了一番想象，而这种想象明确建立在一种更强势，更有官方背书的语言之上。这会不会成为一个问题？

语言学家兼人类学家沙皮尔（Edward Sapir）认为，口语影响着人们与世界互动的方式。当一种语言无法在人工智能时代声张自己，这意味着什么？对于广东话的样貌，我们会逐渐与AI共享同样的想象么？

没有“资源”的语言

翻阅 OpenAI 公开的信息，去年ChatGPT推出的语音模式展现的对话能力，实则由三个主要部分组成：首先由开源的语音识别系统 Whisper 将口语转为文本——再由 ChatGPT 文字对话模型生成文字回复——最后由一个文本转语音模型（Text-To-Speech，以下简称 TTS）来生成音频，并对发音方式进行微调。

也就是说，对话内容仍然是由 ChatGPT3.5 的本体生成的，其训练集为网络上已经存在的大量文本，而非语音资料。

在这点上，广东话存在显著的劣势，因为它很大程度上存在于口语而非书写中。官方层面，粤语区使用的书面语为源自北方汉语的标准书面中文，它更接近普通话而非粤语；而书面粤语，也就是符合粤语口语的语法与词汇习惯的书写系统，又称粤文，则主要出现于非正式的场合，比如网络论坛中。

这种使用时常不遵循统一的规则。“大约有 30% 广东话的字，我也不知道该怎么写。” Frank 就表示，人们在网络聊天时遇到不会写的字，常常也只是在中文拼音键盘上找个发音近似的字打上去。例如广东话中的“乱噏廿四”（lyun6 up1 jaa6 sei3；意即胡说八道），就常被写成“乱up廿四”。虽然彼此之间大多能理解，但这进一步让现存的粤语文本变得杂乱且标准不一。

大语言模型的出现让人们理解了训练集对于人工智能的重要性，以及其可能带有的偏见。但实际上，在生成式 AI 出现之前，不同语言之间的数据资源差距就已经造成了鸿沟。大多数自然语言处理系统都是用高资源语言设计和测试的，在全球所有活跃语言中，只有 20 种被认为是“高资源”语言，比如英语、西班牙语、普通话、法语、德语、阿拉伯语、日语、韩语。

全球约有20种高资源语言

「资源」包括语言数据、教育和研究材料、技术支持，使用和传播渠道

全球母语和非母语使用者的数目

极高资源语言		高资源语言	
英语	15 亿	阿拉伯语（包括所有方言）	3.1 亿
		日语	1.2 亿
		德语	1.3 亿
		西班牙语	5.5 亿
		普通话	11 亿
中高资源语言			
法语	2.8 亿	土耳其语	8800 万
俄语	2.6 亿	波斯语（法尔西语）	7900 万
葡萄牙语	2.7 亿	瑞典语	1000 万
印地语	6 亿	波兰语	4500 万
意大利语	6800 万	印度尼西亚语	2 亿
韩语	8200 万	越南语	8500 万
荷兰语	2400 万	希伯来语	900 万
部分被广泛使用的低资源语言			
广东话 / 粤语			8500 万
泰米尔语			7900 万
菲律宾语 / 他加禄语			2900 万
爪哇语			8200 万
高棉语			1600 万

注：从技术上而言，如一种语言缺乏大型单语或平行语料库和/或人工制作的语言资源，不足以构建统计自然语言处理（NLP） 应用程序，这种语言就被视为低资源语言。图中部分语言应归属的类别存有争议。

资料来源：Center for Democracy & Technology；端传媒综合整理

端傳媒 Init

而拥有 8500 万使用者的广东话，在自然语言处理（NLP）中则时常被视为是一种低资源语言。作为深度学习的起点，维基百科的英文内容压缩后大小为 15.6GB，繁简混合版压缩后为 1.7GB，粤版压缩后仅有 52MB，与近 33 倍的差距。

同样地，现存最大的公开语音数据集 Common Voice 中， Chinese (China) 的语音数据有 1232 小时，Chinese (Hong Kong) 为 141 小时，Cantonese 为 198 小时。

语料缺失会深刻影响到机器的自然语言处理表现。2018 年的一份研究发现，如果语料库中的平行句子少于 13K，机器翻译就无法实现合理的翻译结果。这也影响到机器“听写”的表现。ChatGPT Voice 采用的开源 Whisper 语音识别模型（V2 版本）性能测试，粤语字符错误率要明显高于普通话。

模型的文本表现显示出粤文的资源不足，而决定我们听感的发音和语调又是如何出错的呢？

机器是怎么学会说话的？

人类很早就萌生出让机器说话的念头，最早可以追溯到 17 世纪，早期的尝试包括使用风琴或风箱等，机械地将空气泵入模拟胸腔、声带和口腔结构的复杂装置。这一思路随后被一名叫费伯（Joseph Faber）的发明家纳用，打造了一个身着土耳其服饰的说话假人——但当时人们都不理解这有什么意义。

直到家用电器愈加普及，让机器说话的主意，才引发了更多人的兴趣。

毕竟对绝大多数人来说，用编码进行交流并不自然，也有相当一部分残障人群因此被隔绝在技术之外。

1939 年的世博会上，贝尔实验室工程师达德利（Homer Dudley）发明的语音合成器 Voder 向人类发出了最早的“机器之声”。

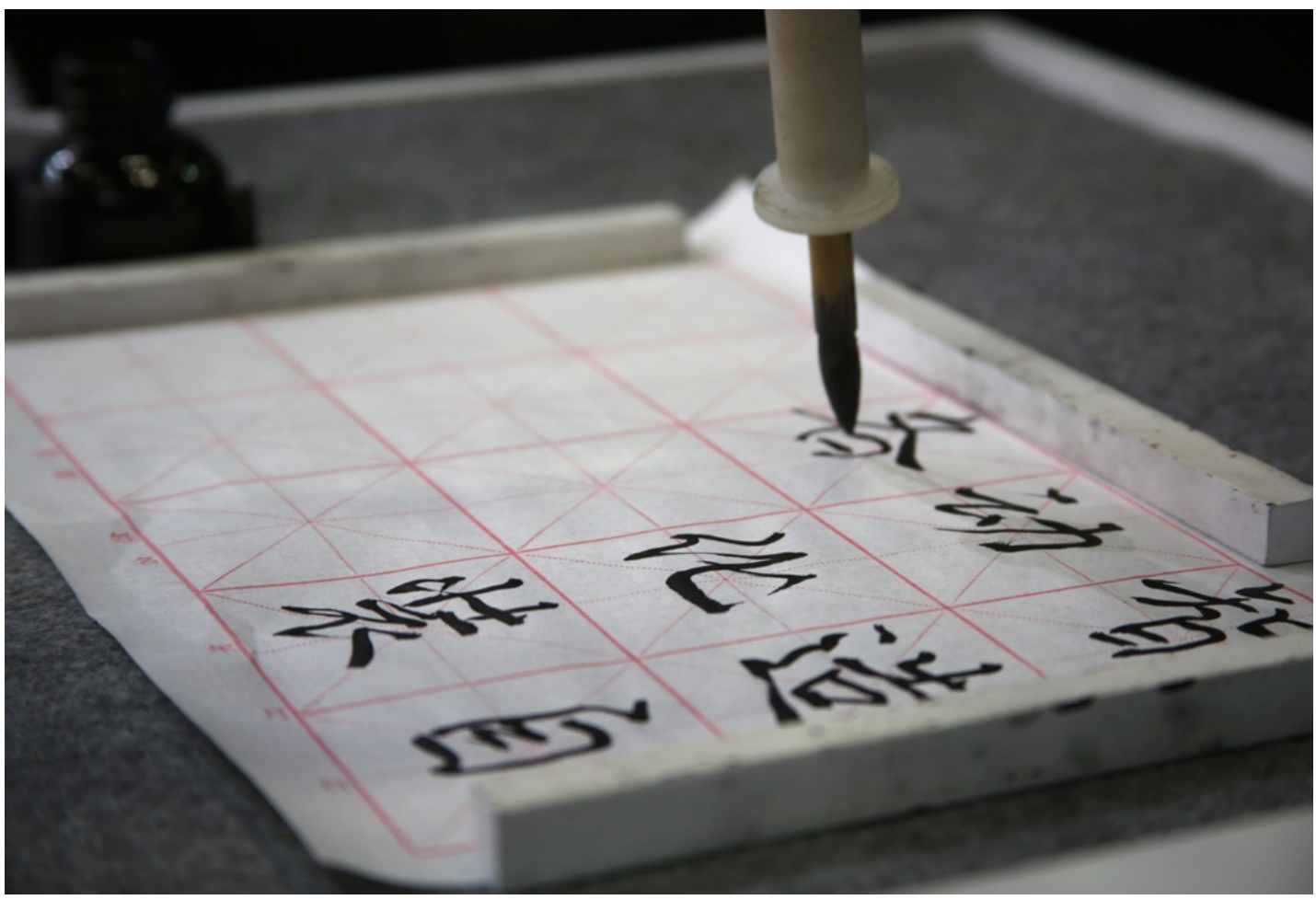
1939 年的世博会上，贝尔实验室工程师达德利（Homer Dudley）发明的语音合成器 Voder 向人类发出了最早的“机器之声”。对比现今机械学习的“神秘”，Voder的原理简单易明，而且场观众都能看到：一名女性操作员坐在一台玩具钢琴一样的机器前，通过熟练控制 10 个按键，来产生近似于声带摩擦的发音效果。操控员还可以踩下脚踏板，改变音高，模拟更欢快或是更沉重的语气。一旁，一名主持者不断让观众提出新的词语，以证明 Voder 的声音并非预先录制。

透过当年的录音，《纽约时报》评价，Voder 的声音像“深海中传来外星人的问好”，又像个烂醉如泥的人囫囵吐字，难以理解。但在当时，这种技术已足以让人惊奇不已，这届世博会期间， Voder 吸引了全世界超 500 万人次前来参观。

早期智能机器人、外星生物的声音想象从这些装置中获取了诸多灵感。1961 年，贝尔实验室的科学家让 IBM 7094 唱起了 18 世纪的英国小曲 “Daisy Bell”。这是已知最早的由计算机合成声音演唱的歌曲。《2001：太空漫游》的作者克拉克曾去过贝尔实验室听 IBM 7094 唱 Daisy Bell，这本小说中，超级电脑 HAL 9000 最先学会的就是这首曲子。在电影版中，片末被初始化的 HAL 9000 意识混乱时，开始吟唱起 “Daisy Bell”，灵动拟人的声音逐渐退归于机械的低吼。

自此，语音合成经历了数十年的演进。而在 AI 时代的神经网络技术成熟前，串联（concatenative synthesis）和共振峰合成（formant synthesis）是最常见的方法——实际上如今常见的许多语音功能仍是通过这两种方法实现的，比如读屏。其中，共振峰合成在早期占据主导地位。它的发声原理与 Voder 的思路很相似，利用基频、清音、浊音等参数的控制结合，来生成无限量的声音。这带来了一个很大的优势，你能用它来产出任何语言：早在 1939 年，Voder 就能说法语了。

那么当然它当然也可以说广东话。2006 年，还在中山大学读计算机软件理论硕士的广州人黄冠能在计划毕业课题时，想到可以做一款适用于视障人士的 Linux 浏览器，过程中他接触到了 eSpeak，一款采用共振峰合成的开源语音合成器。由于在语言上的优势，eSpeak 出现后很快被投入实际应用，2010 年 Google 翻译开始为大量语言添加朗读功能，包括普通话、芬兰语、印度尼西亚语等，就是通过 eSpeak 实现的。



2015年11月24日，中国北京，一座机械臂在用毛笔写中文字。

黄冠能决定为 eSpeak 添加他的母语，也就是广东话的支持。但由于原理的局限，eSpeak 合成的发音有着明显的缝合感，“就像你学习中文，不是通过汉语拼音，而是英文的音标来拼读一样，效果就很像一个外国人学说汉语。”黄冠能表示。

因此他又做了 Ekho TTS。如今，这款语音合成器支持广东话、普通话，甚至是诏安客语、藏语、雅言、广东台山话等更为小众的语言。Ekho 采用的是串联的方法，更浅显的说法就是拼贴——预先录制人类发音，“说话”时将它们拼贴在一起。这样一来，单字发音会更加标准，而一些常用词汇如果被完整录

入，也会让听感更加自然。黄冠能整理了包含 5005 个音的广东话发音表，从头到尾录制完成需要 2 到 3 个小时。

深度学习的出现为这个领域带来了变革。基于深度学习算法的语音合成从大规模语音语料库中学习文本和语音特征之间的映射，而无需依赖事先设定的语言学规则和录制好的语音单元。这种技术让机器声音的自然程度向前迈进了一大步，很多时候效果已经与真人无异，且凭借十几秒的语音就克隆出一个人的音色与说话习惯—— ChatGPT 的 TTS 模块使用的便是这种技术。

相比于共振峰合成和串联技术，这类系统为语音合成省去了大量的前期人力成本，但也对文本和语音的配对资源提出了更高的要求。比如 Google 2017 年推出的端到端模型 Tacotron，就需要超过 10 小时的训练数据才能获得较好的语音质量。

为照顾到很多语言的资源稀缺，近年来，研究者提出了一种迁移学习的方法：先用高资源语言的数据集训练出一个通用模型，再将这些规律迁移到低资源语言的合成中。 一定程度上，这种迁移而来的规律仍然携带着原本数据集的特征——就像拥有第一母语的人去学习一门新语言时，会带入自身母语的语言知识。2019 年 Tacotron 团队就曾提出过一个模型，可以在不同语言之间克隆同一说话者的嗓音。在 demo 演示中，英语母语者在“说”普通话时，尽管发音标准，却带有十分明显的“外国人口音”。

《南华早报》上的一篇评论中指出，香港人用标准汉语写作，为了让所有讲中文的人都能理解自己的意思，必须使用现代标准汉语中的“他们”——“他们”，粤拼为“taa1 mun4”，是一个粤语口语几乎永远不会用的词；粤语中的意指“他们”的，是发音写法都截然不同的“佢哋”（keoi5 dei6）。

在一个解法处理普遍问题这一点上，最新的 GPT-4o 模型做得更加极致，OpenAI 介绍，他们端到端地训练了一个跨文本、视觉和音频的模型，所有输入输出都由这一通用的神经网路进行处理。该模型如何处理不同语言，这一点尚不明确，但看起来它在跨任务之间的通用性要比过去都更强。

一名老师在教授中文。摄: Lucy Nicholson/Reuters/达志影像

但广东话和普通话之间的互通时而会让问题更为复杂。

在语言学上，有“语言分层”或“双层语言”（diglossa）这一概念，指在特定社会中存在两种紧密联系的语言，一种具更高威望，通常为政府所用，另一种则常作为方言口头使用、或谓之白话。

在中国的语境中，普通话是最高层次的语言，用于正式书写、新闻播报、学校教育和政府事务。而各地方言，如粤语、闽南语（台语）、上海话等，则是低层次语言，主要用于家庭和地方社区的日常口头交流。

因此，在广东、香港和澳门便造成了这样的现象，粤语是大多数人的母语，用于日常口语交流，而正式的书面语言则通常是使用普通话的书面标准汉语。

两者之间有许多相似却实际不同，诸多如“他们”与“佢哋”这般的“不和谐对”，也反而可能导致从普通话到粤语的迁移变得更加困难和误会重重。

日渐边缘化的粤语

> “对于粤语未来的担忧绝非空穴来风。语言衰微发生的速度很快，可能在一、两个世代之内就式微，而一旦语言迈向衰亡，就很难力挽狂澜。”
> James Griffiths《请说国语》

至此，似乎可以认为，语音合成在粤语上的表现不佳是技术处理低资源语言时的能力所致。采用了深度学习算法的模型，在面对不熟悉的词语时，会生出声音的幻象。 但香港中文大学电子工程系教授 Tan Lee，在听过 ChatGPT 的语音表现后，给出了一点不同的意见。

油麻地戏院上演的一出粤剧。摄：林振东/端传媒

Tan Lee 自 1990 年代初开始致力于语音语言相关的研究，领导开发了一系列以粤语为核心的口语技术，并得到了广泛的应用。他在 2002 年与团队合作推出的粤语语音语料库 CU Corpora，是彼时世界同类数据库中最大的，包含两千多人的录音数据。苹果的第一代语音识别在内，许多公司和研究机构希望开发粤语功能时，都曾向他们购买这套资源。

在他看来，ChatGPT 的广东话语音表现“水平不是很好，主要是不稳定，声音的质量、发音的准确性整体都不是让人很满意”。但这种表现不佳并非源于技术局限。实际上，如今市面上许多具备广东话能力的语音生成产品，质量都要远高于此。以至于他对网络视频中 ChatGPT 的表现感到难以置信，一度以为是深度仿冒的赝品，“如果是做语音生成模型的，做成这样基本不能见人，等于自杀”。

以香港中文大学自身开发的系统为例，最先进的一批在语音效果上已经很难分辨是真人还是合成的声音。与普通话和英语等更强势的语言相比，AI广东话只在一些更个性化和生活化的场景中，情感表现会逊色一些，比如在父母与孩子的对话、心理咨询、工作面试的场景中，广东话会显得比较冰冷。

“但严格来讲，在技术上这并没有什么难度，关键在于社会资源的选择。” Tan Lee 表示。

相比于 20 年前，语音合成领域已经发生了翻天覆地的变化，CU Corpora 的数据量跟如今的数据库相比“可能还不到万分之一”。语音技术的商业化让数据成为了一种市场资源，只要愿意，数据公司随时可以提供大量的定制数据。而广东话作为口语化语言，文本与语音的平行数据缺少的问题，近年来随着语音识别技术的发展，也已经不再是一个问题。在当下，广东话作为“低资源语言”的说法，Tan Lee 认为已经不再准确。

也正是因此，在他看来，市面上机器的广东话表现反映的并非是技术的能力，而是市场与商业的考虑。“假设现在全中国一起学广东话，那肯定可以做起来；又比如，现在香港跟内地越来越融合，假设有一天教育政策变成，香港的中小学不能用广东话，只能说普通话，那就又会是另外一个故事了。”

“吃下什么便吐出什么”的深度学习展现出的口音，实际上是广东话在现实空间受到的挤压。

黄冠能女儿刚刚上广州的幼稚园中班，而从小只会说广东话的她，在上学一个月之后，就精通了普通话。如今，即便是与家人邻居的日常交流，她也更习惯用普通话，只有跟黄冠能还愿意说广东话，“因为她最想跟我一起玩，就要根据我的喜好来”。在他眼中，ChatGPT 的表现就很像是女儿如今说粤语时的样子，很多词汇想不起来怎么说，就用普通话来代替，或是通过普通话猜测它的发音。

这是广东话在广东地区长期不受重视，甚至从官方语境中被完全排除的结果。1981 年广东省人民政府的一份政府文件中写道，“推广普通话是一项政治任务”，尤其对于方言复杂，对内对外交往频繁的广东，“力争三、五年内大中城市一切公共场合都使用普通话；六年内各类学校基本普及普通话。”

2010年8月1日，中国广州的集会，数百抗议者走上广州街头，要求政府停止压制粤语。摄：Stringer/Reuters/达志影像

在广州成长的 Frank 对此也有很深的记忆，童年电视公共频道里播放的电影，外语片都没有中文配音，使用字幕，唯独粤语片一定会有普通话配音才会在电视上播放。 在此背景下，粤语日渐式微，使用者数量骤减，校园牵头“封杀粤语”，也引发了对粤语存亡以及与之相关的身份认同的激辩。2010 年，广州的网络与线下爆发大规模“撑粤语”行动。当年的报道中提及，人们将这场论战与法国小说《最后一课》中的场景相提并论，认为大半个世纪的文化激进主义使原本茂盛的语言枝干日益萎缩。 对于香港，广东话更是本地文化的关键载体，港片、港乐对外塑造了这里社会生活的面貌。

2014 年，教育局官网曾刊登一篇文章，文中称广东话为“不是法定语言的中国方言”，引发了激烈的争论，最终以教育局人员出面道歉收场。2023 年 8 月，香港捍卫粤语组织“港语学”宣布解散，创始人陈乐行在之后的采访中提及广东话在香港面临的现状：政府积极推动“普教中”，即用普通话教授中文科，但因市民关注，令政府“慢咗个步伐”。

这些都足见在香港人心目中广东话的重要性，但也显示出这个语言在本地面临的长期压力，没有官方身分的脆弱性以及政府与民间的持续角力。

网上粤语辞典 – 粤典。摄：卢翊铭/端传媒

不被代表的声音

语言的幻象不仅存在于粤语中。Reddit 论坛与 OpenAI 的讨论区，来自世界各地的用户都反映了 ChatGPT 在说非英语语言时存在类似表现：

“它的意大利语语音识别非常好，总是能听懂且表达流利，就像一个真人。但奇怪的是，它有英国口音，就像一个英国人在说意大利语。”

“本英国人表示，它有美国口音。我很讨厌这一点，所以我选择不用。”

“荷兰语也是，很烦人，仿佛它的发音是用英语音素训练出来的。”

语言学上，将口音定义为一种发音方式，每个人受到地理环境、社会阶层等因素影响，都或多或少会有发音选择上的差异，这常常体现在音调、重音或词汇选择上的不同。有趣的是，过去被广泛提及的一些口音，大多源于世界各地的人试图掌握英语时从母语中携带而来的习惯，比如印度口音、新加坡口音、爱尔兰口音——这反映了世界语言的多样性。但人工智能展现出的，则是主流语言对区域性语言的曲解和反向入侵。

技术放大了这种入侵。Statista 在今年二月的一份数据报告中着重点出，虽然全世界仅 4.6% 的人将英语作为母语，它却压倒性地占据网络文本的 58.8%，这意味着它在网络上具有比现实中更大的影响力。即便是将所有会说英语的人纳入，这 14.6 亿人也只占世界人口的不到 20%，也就是说世界上大约五分之四的人无法理解网络上发生的大部份事情。进一步来讲，他们也很难让精通英语的人工智能为自己工作。

牛津英语字典。摄：Matthew Horwood/Getty Images

一些来自非洲的计算机科学家发现，ChatGPT 经常错解非洲语言，翻译很粗浅，对于祖鲁语（Zulu；班图语的一种，全球约有900万使用者），它的表现“好坏参半、令人捧腹”，对于提格雷尼亚语（Tigrinya；母语国主要为以色列和埃塞俄比亚，全球约有800万使用者）的提问，则只能得到乱码的回答。这一发现引发了他们的担忧：缺乏适用于非洲语言、可以识别非洲名称和地点的人工智能工具，会使非洲人民难以参与到全球经济体系，比如电子商务与物流中，难以获取信息并自动化生产过程，进而被阻挡在经济机会之外。

将某种语言作为“黄金标准”的训练方式，还会让人工智能在判别时有所偏差。史丹福大学 2023 年的一项研究发现，人工智能错误地将大量托福考试作文（非英语母语者的写作）标记为 AI 生成，对于英语母语学生的文章时却不会如此；另外一项研究则发现，在面对黑人说话者时，自动语音识别系统的错误率几乎是面对白人时的两倍，而且这些错误并非由语法，而是“语音、语音或韵律特征”，也就是“口音”引起。

让人更不安的是，在模拟庭审的实验中，面对非裔美式英语的使用者，大语言模型判处死刑比例要更高于说标准美式英语的人。

一些担忧的声音指出，如果不考虑底层技术的缺陷，只因便利就不假思索地讲现有的人工智能技术投入使用，将产生严重的后果。比如一些法庭转录已经开始使用自动语音识别，对于有口音或是不精通英语当事人的语音记录更可能产生偏差，而带来不利的判决。

更进一步思考，未来人们会不会为了被 AI 理解而放弃或改变自己的口音？现实中，全球化和社会经济发展的已经带来这样的改变。Frank 目前在北美读研究生，同班的加纳同学跟她分享过当下这个非洲国家的语言使用现状：书面文本基本上都使用英文，即便是私人的文本，比如书信也是如此。口语中则夹杂了大量英文单词，这导致即便是当地人，也逐渐开始忘记一些非洲母语词汇或表述方式。

在 Tan Lee 看来，如今人们正陷入对机器的一种痴迷。“因为机器现在做得好，我们就拼命地跟机器去说话”，这是一种本末倒置。“我们为什么说话？我们说话的目的不是为了转成文字，也不是让它生成回答。在现实世界，我们说话的目的是为了交流。”

他认为，技术发展方向应当是让人与人之间能沟通地更好，而非与电脑交流的更好。在这个前提下，“我们很容易想到很多有待解决的问题，比如有人听不到，可能因为耳聋，也可能离得太远，可能不懂这个语言，可能大人不会讲小孩的话，小孩不会讲大人话。”

如今有很多好玩的语言技术，但它们是否让我们沟通地更为顺畅？它在包容每个人的不同，还是让人们愈发与主流靠近呢？

ChatGPT 设计图片。摄：Beata Zawrzel/NurPhoto via Getty Images

当人们在庆祝 ChatGPT 带来的前沿突破，日常中的一些基础应用却仍并未从中受益。Tan Lee 至今仍能在机场广播中，听到合成语音发出错误的发音，“沟通的第一要点就是准确，但这都没有做到，这是不能接受的”。

几年前，因为个人精力有限，黄冠能停止了 Ekho 对安卓系统版本的维护，但停了一段时间，突然又有用户跑来希望他将其恢复。他才得知，如今安卓系统已经没有免费的粤语 TTS 可用了。

用当下的眼光看来，黄冠能开发的 Ekho 采用的已经是完全落伍的技术，但仍具有独特之处。作为本土的独立开发者，他在设计时带入了对于这个语言的切身经验。他记录的广东话包含了七个声调，其中第七个是香港语言学会提出的 Jyutping （粤拼）中不存在的一个发音。“‘烟’这个词在‘抽烟’和‘烟火’中，会发出不同的声调，也就是第一声和第七声。”

在整理发音字典时，他曾请教过 Jyutping 的研发者，得知随着时代变化，年轻一代的香港人不再分辨第一声与第七声的区别，这个音也因此逐渐消失了。但他仍选择将第七音纳入，这并非出于公认的标准，只是他个人的情感记忆，“土生土长的广州人是可以听出来的，现在使用还是非常普遍”。

只听到这个音，老广便能分辨，你是本地人还是外来的。

#生成式人工智能 #广东话 #人工智能