

奋进新征程 建功新时代·北京劳动者之歌

“小爱同学!”“我在!”从2017年9月问世,不论是小米智能音箱,还是小米电视、手机,“能听会道”的“小爱同学”月活跃用户数量已达到1.15亿,成为世界上最忙碌的语音助手之一。细心的用户会发现,“小爱同学”潜移默化中一直在成长、进步,2023年全国工人先锋号、小米科技有限责任公司小米技术委人工智能实验室语音组(以下简称“语音团队”)就是扶持它成长的强大“技术天团”,让它越来越智能,越来越有温度。

让中国语音技术领跑世界

——记2023年全国工人先锋号、小米科技有限责任公司小米技术委人工智能实验室语音组

□本报记者 孙艳/文 彭程/摄

赋予世界最大的手机 AIoT平台以“耳朵和嘴巴”

2017年1月,小米科技有限责任公司(以下简称:小米公司)成立人工智能实验室,下设声学语音技术等多个研究方向。声学语音技术总监王育军是语音组的“元老”,“团队成立初期仅4人——两名正式员工和两名实习生,我们把这段经历戏称为‘444岁月’。”王育军回忆,就是这仅有的4个人,在4个月的时间内,经过40多次繁复的实验尝试,不仅上线了小米第一代语音识别系统,还公开了一篇小米在电视语音交互中“端到端”语音识别的研究成果。该研究成果展示了当年中文“端到端”语音识别技术最好的识别率。

从此语音团队一路披荆斩棘,除了小米电视,还成功把智能语音交互技术应用到了小米手机、智能音箱系列。由于自研的语音引擎不但速度快,识别率更是超越了微软、Nuance等知名公司提供的引擎,小米公司决定全面使用自研语音识别技术。有了这项技术,小米公司的智能产品就像长了耳朵一样,能够倾听用户的指令和诉说。

王育军介绍,为了与用户进行远距离语音交互,语音团队把研发目光转向了中远距离拾音。小米公司研发的设备中,无论是电视等大型家用设备,或是音箱、手机这样的小型或便携设备都拥有多颗麦克风,语音团队利用这些麦克风组成的阵列设计了利用声音波束指向目标说话人的算法以及利用声源分离提升目标说话人声音质量的算法。

在此期间,小米公司率先建设了中国人工智能产业联盟首批合作实验室。专家们在实验室里进行专业的测量与标定,同时在中国南、北方几十个不同户型的住宅中进行数据采样,最终结合回声消除、房间混响抑制、增益控制和特定噪声压制等算法,形成了小米特有的中远距离,即60厘米到5米的清晰拾音算法。

有了倾听用户语音的“耳朵”,团队开始为小米智能设备打造会说话的“嘴巴”——语音合成的算法能力。王育军介绍,2017年的6月,语音团队开始跟踪最前沿的端到端语音合成技术,经过14个月反复的论文研读、试验以及评测,语音团队追平了国际最先进的语音合成技术水平。

拥有在小米手机、音箱和电视上应用人工智能“听说”的技

术能力之后,团队开始把自研语音技术全面应用于小米手表、耳机、平板电脑、手环、翻译机、大家电、儿童机器人、机器狗等50多个品类的智能产品中。

为世界最忙的语音助手 打造全栈语音技术

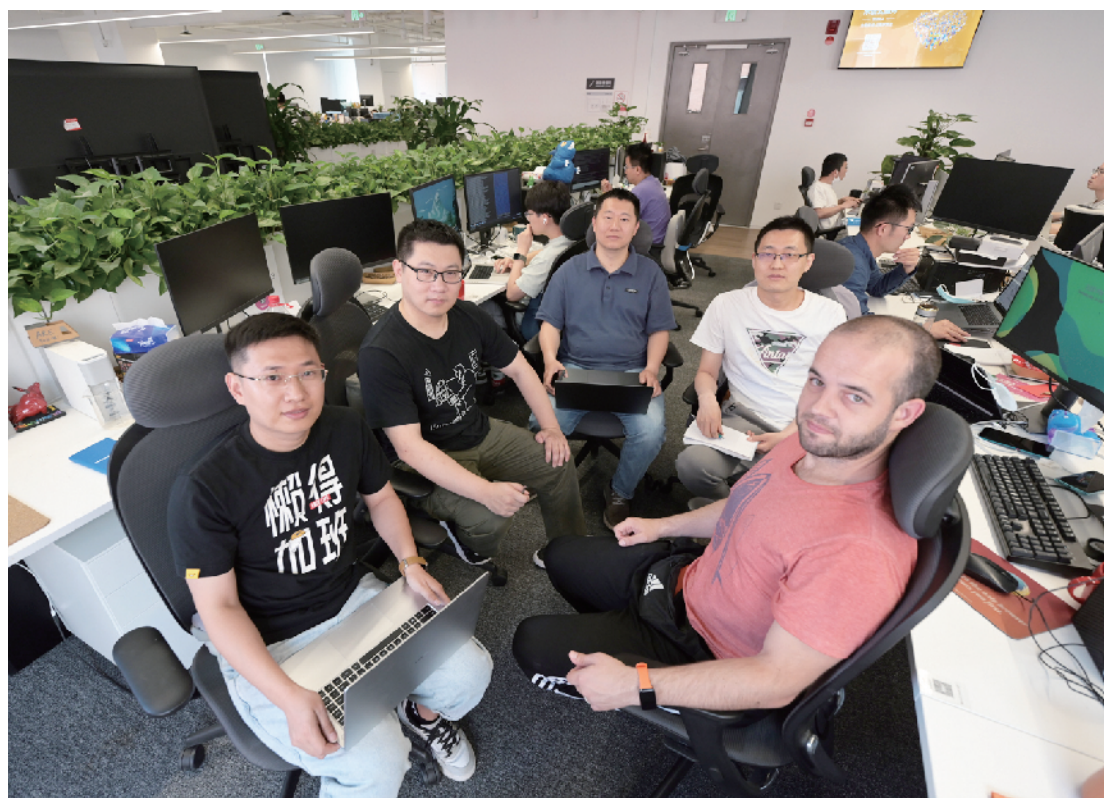
2019年,小米公司提出了“手机×AIoT”战略。为了持续提升这个大生态上的声音交互体验,小米公司引入了首席语音科学家Daniel Povey博士,他是全世界开发者使用最多的开源语音识别社区Kaldi的创始人,被业界称为“Kaldi之父”。Povey博士加入小米后,语音团队又持续邀请了几十位业内顶尖的声学语音算法专家和研发人员加入,一起在世界上最肥沃的声音技术土壤上耕耘。

专家们把以语音识别为代表的一系列算法扩展为声音理解技术,内容包括语音识别、声音分析还原、阵列拾音、语音唤醒、声纹识别、通话降噪、主动降噪、声学测量、声音视觉跨模态感知理解、口语评测、语种识别和声音文本跨模态情绪识别;把以语音合成为代表的一系列技术扩展为声音生成技术,内容包括语音合成、声景合成、音乐生成、声音复刻、虚拟形象生成以及空间音频。从而形成了一个包含18个技术方向的完整声音交互技术栈,几乎囊括了所有民生领域中的声学语音应用。

截至2022年底,语音团队共98人,其中国内外博士20名、硕士75名。正是因为拥有了强大的声音“技术天团”的技术支撑,才让小米公司的“小爱同学”更加“能听会道”。截至目前,“小爱同学”月活跃用户数量达到1.15亿,成为世界上最忙碌的语音助手之一。小米手机日均收到2亿次语音识别请求、3亿次语音合成请求,日均通过远场拾音触发语音交互5200万次、唤醒语音识别5500万次、进行1.85亿次声纹识别。

挑战语音技术天花板, 领跑世界先进水平

语音技术最早起源于欧美,随着2011年移动互联网时代的到来,华人语音的研究者和团队逐渐踏入语音技术的聚光灯下,小米的语音团队成为其中的佼佼者。小米拥有世界上最大的智能设备生态网络,借此孕育了大量的语音交互需求。语音团队满足了这些真实的用户需求,为包括手机、耳机、手表、手环、笔记本电脑、大家电、机器人等



5000多款智能产品提供声学语音人机交互技术,并不断磨砺自己的人工智能算法。

语音团队不断挑战世界级的语音研究“天花板”,参加了语音技术领域内一系列挑战赛,并获得了6项国际声学语音挑战赛冠军和2项国内比赛冠军。同时,团队在语音研究领域持续创新,研发出超级拟人、语音成分分析与重建、AI作曲和编曲等技术,回馈工业界和学术界。其中,以“个性化情感化语音交互关键技术及产业化”获得北京市科技进步二等奖,以“小米智能语音技术在手机实时通信中的应用”获得深圳人工智能行业应用奖。截至2022年,语音团队发表的43篇论文为国际顶级语音会议收录。

一项项荣誉不仅见证了语音团队在和声音交互业务紧密相关的技术领域上实现了从“追随者”

到“领跑者”的蜕变,也为“中国智造”书写下浓墨重彩的一笔。

拥抱大模型时代, 科技赋能中国民生领域

为用户带来智能生活新体验的同时,小米始终关注智能设备无障碍建设。王育军特别介绍了语音声学团队为残障用户群体研发了一系列算法和应用。

据王育军介绍,语音团队为听障用户群体开发了“闻声技术”,不但可以让设备帮听障用户“看到”其他人说话,也可以帮他们“看见”周围环境中的声音,例如警报声、敲门声等;读屏技术可以让视障用户“看到”屏幕上的内容;“聆听”技术为构音困难用户提供了个性化的语音识别,让他们也可以通过语音和设备交互,这项技术被提名为2021年世界互联网领先科技成

果;“口语评测”技术为听障人群的语训提供帮助。作为语音合成技术在无障碍的典型应用案例,语音团队还开展了声音配型捐赠工作,赋予失去语言能力的用户属于自己的声音。

“做有温度的语音技术是我们团队共同的初心。如今,超大模型技术的涌现,已经标志着我们进入了人工智能2.5时代,小米将积极拥抱大模型时代,为中国民生领域再创技术佳绩。”王育军透露,小米已经集齐大模型的部分先决条件:跨模态编码的科研能力、个性化垂域模型与大模型的融合推理能力、设备端计算资源以及丰富的行业用户场景。“声学语音团队必将竭尽全力,参与大模型的研发,为中国AI声学语音技术创新做出贡献,持续在民生行业问鼎世界第一,用技术创新提升全民生活品质。”王育军说。