

◀ 从技术探索到产业落地 ▶

探 秘

刘强东 AI 数字人

“大家好，好久不见，我是你们的老朋友东哥……”

4月16日下午6点18分，国人熟知的京东集团创始人、董事局主席刘强东以AI数字人“采销东哥”的形式空降京东超市、京东家居采销直播间，开启了直播首秀，成功引来众多网友围观。

直播间，“采销东哥”从容自若，大谈自己在运动、烹饪方面的心得，对带货物品的讲解，既专业又热情。独特的宿迁口音，不时搓手指的小动作，还有看手机回复的表情、动作，甚至还会提醒工作人员上链接，让很多观众非常怀疑在看的主播到底是真人还是数字人。

这是一场相当成功的直播。据京东超市数据，“采销东哥”上播30分钟，直播间观看人数破千万；直播时段用户平均停留时长达到日常均值的5.6倍；直播期间讲解的13款商品，整体订单量环比上周日均增长7.6倍；京东超市“百亿农补”货品，开播半小时成交额环比上周日均增长5.7倍。更重要的是，刘强东这次AI数字人直播，让外界对这个去年一度火爆却又很快沉静下去的行业，有了一次非常重要的重新认识和认可。

■新快报记者 郑志辉

产业
崛起

最快三天可诞生一个数字人主播

所谓数字人，简单来说就是运用数字技术创造出来的，与人类形象接近的数字化人物形象。数字人的创建是一项极为复杂的工作，需要经过建模、动画制作、渲染……通过高级计算机程序实现。最终，数字人能还原真人80%以上的外貌，在外观、动作以及声音方面，能做到像现实世界里一样自然。

直播电商头部机构谦寻控股技术人员去年曾说，在现今的数字技术帮助下，通过采集真人形象及语言、肢体数据，再对采集的数据进行训练开启直播，一个真

人形象的AI数字人主播的诞生，最快只需三天。

如果是非真人定制数字人主播，不仅制作成本低，制作效率还能进一步提升。数字人制作头部企业硅基智能近日表示，在硅基智能一开始做数字人时，行业成本还是几万元甚至数十万元一个，但现在已经下降至几百甚至几十元一个。而且，数字人在执行各种选择器、任务器和反馈器任务时，如在直播间人数变少时生成降价促销的直播话术等，具有一定优势。

随着数字人成本平民化，不少企业都愿意抱着试一试的心态来定制一个数字人。2023年，中国AI数字人行业产品需求量TOP5的行业分别是电子商务及卫生、社会保障和社会福利业、教育、金融和交通运输业。

针对电商行业，艾媒咨询CEO兼首席分析师张毅表示，“目前，AI数字人在电商前端扮演主播的角色，与真人主播形成互补，拉长直播时长，从而提供更大的流量；在后端，AI数字人可以提高系统办公的效率，例如快速上架商品，

智能数据统计等。”

近两年，数字人产业得到了迅速发展。艾媒咨询数据显示，2022年中国虚拟人核心市场规模为120.8亿元，同比增长94.2%，预计2025年将达480.6亿元。



后台
揭秘

“采销东哥”是如何炼成的？

“分身”也有了鲜活的生命力。

记者从京东了解到，“采销东哥”背后依托的，是京东言犀大模型对形象与声音的不断学习与训练优化。

京东云言犀团队原本打算使用一段演讲音频作学习素材，但经过测试发现，当时演讲的语气用于直播带货过于正式。因此，京东云言犀团队又录制了一段20分钟闲谈，让东哥畅聊自己的旅行经历，提取出声学特征后，通过言犀语音大模型合成出人工语音。目前，言犀团队已经可以做到最短使用6秒素材，就能复现一个真人的声音。

说好普通话并不难，但要让“采销东哥”AI数字人学习本人的发音习惯，比如要不要加鼻音，哪里要连读，就需要大模型给出判断。言犀语音大模型在训练时，被“喂”入5万小时鲜活的语音数据，这让言犀数字人可以智能匹配不同直播风格，比如用沉稳的音色营造专业的氛围，又或者用极具感染力的声音吸引用户下单，还赋予言大姿态肢体

表现。

言犀也对语音风格进行了精细控制，捕捉并复制了说话时的情感，使得合成的语音不仅在音色上与真人相似，而且在情感表达上也保持一致，包括口音和节奏的细微差别、语速的快慢和语调的抑扬顿挫。结合语义理解，系统进行口唇、肢体驱动，让数字人说话时的动作与语义匹配，更加自然，比如在讲话时他偶尔会搓动手指，强调某件事时会配合更大幅度的手部动作，还有时不时的点头等。

实验表明，绝大部分用户在120秒内难以察觉这是数字人，怀疑是真人录制或者AI换脸。

在回应部分网友认为“采销东哥”当晚表情过于严肃时，言犀团队表示，为使数字人从形象到声音，更贴合真实直播场景，团队打造“声音供应链”，基于多颗粒度韵律的增强语音合成技术，准确把握用户情绪，提高合成语音的韵律丰富度和可控性，并将根据反馈持续迭代。