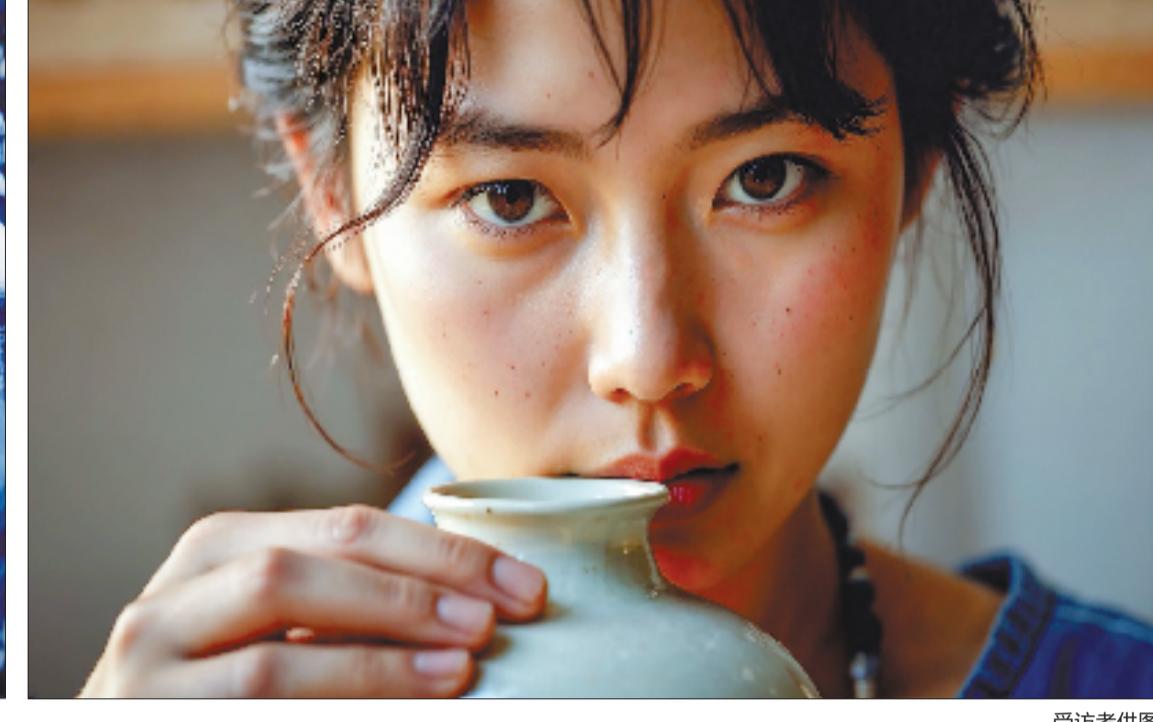


AI视频时代,谁在闻风而动

当内容产业开始进入AI模型化阶段,大家拼的不仅是流量,也是数据、算法和核心创意。我们面临的挑战不仅是某种技术的变革,也是思维和意识的转型。



左右图均为AI创作者罗翀用“可灵”等工具,制作的AI非遗文化主题片《瓷韵》的截图。



受访者供图

■ 潮声 | 执笔 谢丹颖

人工智能(AI)推动视频生成技术又迈出新的一步。

不久前一条发布在社交媒体平台的AI视频中,角色集体开口戏谑:“我们不过是0和1的排列组合?醒醒吧,伙计。”

让它们“开口说话”的是美国谷歌公司在今年5月发布的视频生成模型Veo 3。其最大的特点是在视频中融合音频,直接生成话语流畅、口型自然的人物,且自带符合场景特征的音效。而此前,AI视频一直是默片,需要后期配音,再借助工具让角色嘴唇动作看起合理。

2022年以来,以ChatGPT聊天机器人程序为代表的生成式人工智能引发社会关注。行业像被按下了快进键,几乎每个月都有相关热点出现。相比之下,视频生成技术在最初一段时间里不温不火。不过,历经近3年的发展,AI视频已逐渐从最初类似PPT、动画的形态,进化至能够直接产出合理视频。

基座模型能力的迭代,带来了画面真实感、视频情绪度和流畅度的提升,影响面不断扩展,渗透性持续增强。

如今,学界、业界已在眺望AI视频迎来“技术奇点”的时刻。我们该如何理解视频生成模型的核心突破意义?它又会给人们生产生活带来什么影响?

忽如一夜春风来

2023年初,AI生成视频与生成图像,几乎在同一时期进入公众视野——英国人工智能公司Stability AI的“稳定扩散模型”(Stable Diffusion)带火了“文生图”,美国人工智能公司Runway的视频生成模型“Gen-1”则是激起了“人人都能制作电影大片”的无限遐想。

彼时,“稳定扩散模型”用时数月,让AI生成的图像几近照片级真实。Runway创始人不由兴奋:“我们已经看到了图像生成模型的爆发,相信2023年将会是视频之年。”

然而,图像生成模型的成功并没能引发AI生成视频技术的迅速成熟。

起初,AI视频主要分为两条技术路径:或是与“文生图”的技术一脉相承,着重“还原呈现”,打上“扩散模型”(Diffusion Model)的烙印;或是沿用ChatGPT的技术脉络,采用“自回归模型”(Autoregressive Model)方法,讲究“逻辑推理”,靠大参数、大数据,从零开始构建模型体系。

“走纯粹的‘自回归’路径,至今尚未出现很好的产品。”浙江大学艺术与考古学院副教授沈华清说。同时,他认为使用更普遍的“扩散模型”缺陷也明显。

沈华清自称“无门无派、兴趣广泛的教书先生”,从“好奇尝鲜”变成“深度沉迷”。沈华清类比帧动画原理,向记者解释“扩散模型”的技术难点:“按最低的24fps(帧/秒)算,5秒的视频需要120张图。每张图间都要保持相互的人物一致、动作连贯,连光影的斑驳细节都不能穿帮——这相当于让120位画家同时画同一幅画,每一幅每一笔都要严丝合缝。”

的确,静态图像生成只需解决“是什么”的问题。视频却要在此基础上,在回答“如何变化”命题的同时,保证主体的统一以及符合常识的运动。2023

年底,与美国人工智能初创公司Midjourney同名的“文生图”工具已经能生成以假乱真的图像。同期,美国AI初创公司Pika Labs发布的AI视频产品Pika 1.0还局限在风格特定的3秒片段上。

沈华清回忆起那段使用经历,即便先用“文生图模型”生成了不错的分镜图,再通过“图生视频模型”拼接成片,但在最终呈现的几秒视频里,人物总是畸形、画面常有畸变,“这哪是人在动,分明是算法在抽搐,看得人脊背发凉”。

生成视频技术始终“小步快走”,没有根本性的突破。就在大家快失去耐心时,时间来到2024年2月:美国开放人工智能研究中心OpenAI继ChatGPT后,发布“文生视频模型”Sora。

这一全新模型用ChatGPT背后的神经网络架构Transformer替换传统“扩散模型”中的卷积U-Net网络,迭代出一条新路径DiT(Diffusion Transformer)。如此,Sora可以精准根据文本指令,生成最长为1分钟的高清视频,画面逼真、丝滑连贯。

业内人士称:“AI视频的‘GPT时刻’来了。”

忽如一夜春风来。眼下,腾讯“混元”、华为“诺亚”……各家厂商在大模型领域谋局落子,可谓“神仙打架”。其中,又以拥有海量视频数据的字节跳动、快手增势最为迅猛,其分别在2024年3月、6月推出“即梦”和“可灵”,迅速跻身AI视频产品的第一梯队。

但即便是Sora问世一年后的今天,一键“文生视频”仍难有良品。“毕竟,语言是高度压缩的信息。”沈华清说,一千个读者眼中尚且有一千个哈姆雷特,将抽象文字直接转为具体的时空连续体,这对算力和工程化的要求实在太高,“不是谁都能做的,也不是在短时间内就能做好的。”

是助手,更是共创者

令人振奋的AI技术革新终归要落地产业,才能产生实际价值。

一位技术人员告诉记者,不同于此前大模型简单直接的“文本输入—文本输出”交互逻辑,视频生成技术因视觉模态的复杂性,用起来要棘手得多。而“能用”与“好用”之间,又横亘着训练数据、算力、成本控制等多重障碍。

眼下,单个的AI视频生成工具还处在“宣传视频都很好,但实际一点都不好用”的阶段。

“套用多种工具很有必要。”中国传媒大学导演系科班出身的罗翀,拍过豆瓣8.3分传记式宣传片、拿过中国纪录片学院奖。今年年初,他从杭州某大厂离开,转型自由AI导演。在制作多个商片的过程中,他迅速摸到了生成视频的一条路子。

罗翀介绍,不同视频生成模型的优缺点各异。比如,快手的“可灵”强在对多镜头、物理规律的理解;美国人工智能公司Runway的模型性价比更高,生成速度更快。

但他也告诉记者,基于AI生成产

品的不稳定性,即便经过多种AI视频制作工具的多次打磨、筛选,还是需要借助PS等编辑软件再进行微调,才能得到更满意的结果,以生成“完全可以和传统商片媲美手腕的成品”。

“虽然缺少故事线,但未来,意识流的赛博视频或将成为一个门类。”罗翀说,自己仿佛遇见了共创者,“我负责想象、尝试,AI负责调整、展现,降本增效的同时,极大地拓宽创作的自由度。”

院线影视讲究起承转合,质量要求更高。但在浙江,利用AI制作的视频仿佛距离“院线水准”不算太远。一家老牌影视企业,便提供了可供分析的落地样本。

第27届上海国际电影节启幕前夕,浙江博采传媒有限公司一条3分钟创意短片《两代悟空对战》,在B站传播量近百万。网友纷纷赞叹所用技术之精妙——无论是六小龄童饰演的86版美猴王,还是网游《黑神话:悟空》的天命人,“大圣风采依旧”。

记者也去凑了个热闹。在湖州市德清县博采AI虚拟影视基地,导演的监视器中,两代悟空对战正酣,远处宫殿群光影流动。但现场,只有两位动捕老师,拿着特殊棍棒,在一块“空地”“打”得激

烈,无论是角色样貌、服饰,还是建筑、云雾,都是LED显示屏“附上”的画面。工作人员用鼠标一点,切换只在分秒间,演员置身其中,画面真假难辨。

“单靠AI,肯定跑不出这样的片子。”博采传媒研发中心总监王伟东告诉记者,《两代悟空对战》融合了影视行业所用的4D扫描、Holo身体扫描、LED拍摄等技术,“我们称之为‘虚拟制片’。”据他介绍,团队专门开发了一套虚拟制片管理软件Kmoke,融合各类AI工具,“效率直接提高了3倍、成本起码节省了三分之一。”

博采传媒总裁办项目统筹牛聪说,在电影创作中,相比导演和制片,AI其实是做好了一份助理的工作——通过AI实时预演,让创意的好坏“尽在眼前”;传统影视的各个环节也从“线性等待”转为“动态协同”,无论是调整剧本还是场景,在AI的“协同”下都能更高效完成。

“今年,我们引入AI大模型,继续迭代一整套AI创作系统‘墨客’,实现连贯性的剧本创作,并产出精准可控的视频。”牛聪坦言,针对现在AI视频像素细节不够的问题,“通过改进的AI增分技术,分辨率能从720p跃升为5K,

直接达到电影放映级画面质量。”

拉平起始点,重新竞争

在一份技术报告中,美国开放人工智能研究中心OpenAI对AI视频的定义是“世界模拟器”。这个远景表明,AI视频有潜力成为一种通用人工智能,理解世界、改变世界。

这种颠覆性在技术细节中确有显露。有从业者根据Sora产品表现出的不错的“3D一致性”,推测它或许能通过参数的再叠加,冲破智能阈值,触摸到世界完整理解和创造的边界。

“大力出奇迹”真能无往不利?学界对此的质疑声也不少。

北京通用人工智能研究院院长朱松纯曾明确:过去,“大数据+大算力+大模型”的思维定式,过度简化了通用人工智能的复杂性。美国互联网公司Meta人工智能研究负责人杨乐昆更是直言:“生成式模型是建立世界模型的死胡同,通过生成像素来模拟世界的动作,既浪费又注定失败。”

实践也证明,AI视频实现如此“暴力美学”的代价极高:运行一个动辄超百亿参数规模的视频生成模型,尖端显卡要“跑”数十秒甚至数分钟,才能制作一个一分钟、分辨率高达1080p的视频,算力成本高昂得惊人。

不可忽视的还有生成式AI的固有症状——“幻觉”。“0.8<0.09”的数学对答、四条腿的蚂蚁图、在跑步机上倒着跑的人——这些都是AI制作可能导致的可笑错误。它没有自主意识,对现实世界“知之甚少”,擅长计算却拙于纠错。北京大学新闻与传播学院教授胡泳坦言,这类低级错误或许会在技术的迭代中减少,但永远无法彻底修复,失败风险始终存在。

技术障碍外,高质量训练数据又从何而来?一如ChatGPT问世引发的深度造假、版权侵权、隐藏偏见等法律伦理争议,AI视频同样绕不开这些熟悉而严重的“老问题”。

但不可否认,AI视频正加速被市场接纳,其价值与影响力持续攀升。《2025中国广告主营销趋势调查报告》显示:今年4月,超50%的广告主已将生成式AI纳入创意生产,AI营销内容占比超过10%。

同期,美国电影艺术与科学学院宣布,2026年第98届奥斯卡金像奖将正式允许AI参与创作的电影参评。这标志着AI正式进入主流评价体系。但评审标准中,“将综合考虑人类创作者在创意创作过程中所发挥的核心作用,来决定最终的获奖者”的微妙表述,也暗示着人类艺术本质的不可替代性。

AI视频正处落地的“中场哨”阶段。热潮过后,玩家们纷纷沉下心来,打磨产品的基础能力、拓宽模型的适应边界、重构与用户的长期关系。

可以确定的是,AI正成为创作的基底。正如快手联合创始人程一笑将“可灵”定位为“更多行业创作的新基础设施”,AI将拉平所有人的起始点。

“我们不要放大,也不要低估AI的影响。”5年前,沈华清将AI带入课堂,鼓励学生借助工具,在学习与创作中尝试拓展、延伸、融合。他认为,在这个技术愈发平权的时代,竞争已转向快速捕捉创意并落地的能力,考验的是使用者的积累——“拥有审美、分析、判断能力,成为跨领域、跨学科的融合通才,是AI时代的新要求”。

● 链接

那些用AI“拍”的影视剧

迅猛发展的AI技术,正在颠覆许多行业的格局,当然也包括影视业。从《2001太空漫游》中的“哈尔”到《流浪地球》的“莫斯”,我们发现,AI已然不是影视作品中的某一个角色形象,甚至一种辅助生产技术。它,正在成为荧幕画面背后,那只游走在创作、拍摄、剪辑等产业全流程的“手”——

《乌鸦》

2022年,由爱尔兰艺术家主导、使用OpenAI工具生成的短片《乌鸦》讲述了一个舞者变乌鸦的故事,获得戛纳电影节短片最佳短片奖。尽管这一奖项不算特别权威,但该短片凭此奖,获得角逐英国电影学院奖的资格。这给一些艺术创作者也带来了“可能被AI替代”的焦虑。

《冰霜》

2023年,美国“路标”公司推出了世界上第一支影像完全由AI生成的故事短片,长度不到13分钟。故事讲述了一个全球极端气候变迁背景下的灾难片剧情。为了制作生成这部影片,该团队运用了多项主流的AI生成技术,包括大语言模型协助制作脚本、利用图像模型生成画面、人脸图像识别技术让人物“活起来”,以及AI语音合成配音等。

《中国神话》

2024年3月,国内首部AI全流程微短剧《中国神话》上线。该剧由央视频、中央广播电视台总台人工智能工作室联合清华大学新闻与传播学院元宇宙文化实验室合作推出,其美术、分镜、视频、配音、配乐全部由AI完成。全剧共6集,分别为《补天》《逐日》《奔月》《填海》《治水》《尝百草》,由一个个经典神话故事起笔,借助AI技术拓展人们对神话的常规想象,再通过经典意象和当下人类社会的深度链接,展现民族精神的时代回响。

《三星堆:未来启示录》

2024年7月,由北京市广电局指导,博纳影业出品制作的AI全流程制作科幻短剧集《三星堆:未来启示录》上线,并同步亮相北京大运河博物馆举行的“探秘古蜀文明——三星堆与金沙”展览。该短剧第一季共12集,每集时长3至4分钟。故事放在了近未来,围绕地球古文明遗迹的异变展开,讲述了三股力量在数字重建的古蜀国中探寻解救文明危机的密钥,并逐步揭开古蜀国神秘面纱的传奇故事。从场景、人物到道具,剧集应用了十几种最先进的AIGC视频技术。

(本报记者 严粒粒 整理)

不同类型视频内容制作成本



数据来源:专业AI研究机构量子位智库《2024年AI视频生成研究报告》 潘泓璇 制图



博采AI虚拟影视基地影棚内,虚实结合拍摄。 受访者供图