

上海舞蹈迎来万紫千红的春天
2024新人新作之我见

► 10版·文艺百家

三条屏里的世界
《草木人间》的三重割裂

► 11版·影视

上巳节的起承转合
流风余韵还能不断激起共鸣和呼应

► 12版·传统节日里的中国

“世界模拟器”的文化偏见与AIGC时代的文化竞争

赵宣

OpenAI 开发的文(图)生视频模型 Sora 演示素材发布后,再次在全球范围引发了对生成式人工智能的迭代进化及内容生成能力的关注。它可以根据简单的文本指令生成长达 60 秒的高质量视频。这些视频不仅背景细致,还包含复杂的多角度镜头和富有情感的角色,并根据相应的物理原理在时空上向过去或未来推演。

其在演示视频中展现出出色的运动连贯性、时空统一性和形象逼真性,对传统意义上的视听内容生产行业形成了极大挑战,一时间,如“未来电影不再需要人类创意”“新闻业将消失”等标题成为自媒体吸引流量的密码,也从侧面反映出了 Sora 的技术突破确实为内容生产提供了几乎无限的可能性。

多模态视频生成 与新的世界模拟器

2023 年 ChatGPT 的出世吸引了全球对新一代生成式 AI 的注意力,确立了大语言模型作为新一代内容生成技术的重要基础地位。大语言模型特别强调模拟人类语言理解和生成方面的先进能力,为知识创造和信息传递开辟了新途径。通过其对话能力,大语言模型不仅改善了人机交互的自然度和效率,而且推动了以用户查询为中心的信息生成和交流新方式。

在此后一段时间,围绕基于大语言模型的多模态模型的探索尤为突出,如文生图、文生音乐等应用,这些模型通过整合视觉、文本等多种信息形式,丰富了 AI 的理解和表达能力,为构建更加细致和动态的世界模拟器奠定了基础,标志着今天的研发重点正向着通过人工智能生成复杂、多维世界的全面理解和再现迈进。

其中,文生视频或图生视频的应用一直作为主要的研发热点暗流涌动。2023 年底,谷歌的多模态模型 Gemini 1.0 正式上线。在其技术报告的描述中, Gemini 基于 Transformer 架构,同时在图像、音频、视频和文本数据中进行训练,目的是使模型既具有跨模态的强大通用能力,又有尖端的理解和推理能力;此后推出的 WALT 利用扩散模型和 Transformer 架构突破性解决了 AI 对视频对象的时间逻辑理解难题。2024 年, OpenAI 经历了戏剧性的“宫斗”戏码后,突然爆炸性地推出了在视频长度和稳定性上都形成突破的 Sora。

Sora 对三维空间、因果关系和物理逻辑等人类理解要素的模拟,使对生成式大模型的讨论被推进到了“世界模拟器”的层面,对现实世界物理时空的模拟生成,也再次激活了一度受到冷落的元宇宙场域概念。但就像由数字图像技术带来的“后真相”问题一样,作为世界模拟器的人工智能技术势必引发生成内容真实性的一系列伦理、道德和法律问题,并成为讨论的核心。而另一方面,在 Sora 展现出内容创造力的同时,也暴露出了更深层次的困境:当前广泛流行的 Sora 生成视频片段中,尽管开发者已经尝试性地融入了非西方文化元素,但依赖于西方中心化的数据源,其生成内容的审美重点却仍然深受欧美影视文化与视觉文化影响。

如果我们认为多模态生成视频的应用将作为世界模拟器发挥其潜力,那么这样的文化偏向就需要引起我们的重视。考虑到上一轮世界模拟器媒介——电影,特别是数字时代的电影——以的表现,我们就不能轻信“技术是中立的”这样的论调,而应正视今天 AIGC 实践中正在渐渐呈现出的这些问题。

数字图形技术与 电影的偏见

现代电影的全球技术、产业与美学在 1970 年代确定下来。与新好莱坞电影的产业和市场体系确立下来几乎同时,北美为中心的计算机图形学研究也发生了重要的产业转移。随着越战后美国政府对军工领域的计算机图形学研究经费投入大幅削减,这个领域迎来了新的投资人:1972 年,汇聚了该领域顶尖科学家的 PDP 小组在犹他大学成



立,标志着计算机图形学的研究重心从麻省理工学院逐渐转向了西海岸,好莱坞开始持续在这一领域投入资金,著名的 3D 艺术品“犹他茶壶”就在这一时期诞生;1974 年, Triple-I 公司建立了图形产品组,并进一步将计算机图形学的应用从学术研究推向了商业领域,特别是电影制作。 Triple-I 利用大型工业计算机 PDP-10 为《西部世界》和《星球大战》等电影制作了数字图形,也创作了如《亚当斯一家》和《电子世界争霸战》这样的数字动画短片。这一时期,既是计算机图形学在技术和应用的进步时期,也是计算机科学的研发重点向民用和娱乐产业倾斜的时期。

于是,改由好莱坞牵头的计算机图形学研发和应用,站在冷战时期欧美计算机科学研究的优势基础上,成为全球引领性的科创单位。这一技术优势经由 1970—1980 年代形成的好莱坞全球市场,经历了 1990 年代的“好莱坞数字之夏”。以三维动画为代表的数字动画,成为了新世纪以后视听领域再现世界的主流手段。

从技术审美的角度来看,一方面,好莱坞在《侏罗纪公园》和《泰坦尼克号》等电影中突破了数字形象与真实影像之间时空统一的表现难题,并进而而在以皮克斯为代表的三维动画实践中完成了全三维动画图形呈现(《玩具总动员》系列)和越发精确的物理逻辑呈现(《怪兽公司》系列);好莱坞也在新千年头十年开始推动电影胶片的数字化转型,以更符合数字影像技术便利与经济效益的 DCP 储存制式以及三维数字化的放映单位,确立了全球影视行业新的制播技术标准;并在新世纪第二个十年以后,确立了以无限逼近真实效果的三维数字动画为主流的全球视听影像的审美范式。在第一部全三维动画长片《玩具总动员》中对犹他茶壶的彩蛋式致敬,确认了这一“犹他大学—好莱坞—皮克斯”的研发、应用、传播体系。

正是因其在 1970 年代技术研发上的先发优势,好莱坞迅速占领了上一代世界模拟器的话语权和审美决定权。如果观察一下这一时期全球主流影视

文化出海应包括 数据出海,语料库是未来文化软实力

如果仔细分析 Sora 生成内容的美学要素,就能够轻易发现,它的审美范式依旧是这一“犹他大学—好莱坞—皮克斯”系统的延续。不仅 Sora 所代表的、目前吸引了全球目光的多模态视频生成模型延续了电影工业的视觉文化偏向,事实上,从 ChatGPT 这样的大语言模型的应用实践中,也有学者发现了目前主流大语言模型中普遍存在的数据选择偏见,即由选择构成训练语料库的文本语言偏见问题。简单说,即以英语为主要训练语料库的大语言模型中,其使用“母语”的任务完成度更高,并优先基于英语世界的文化逻辑或意识形态逻辑进行“思考”与回应。

今天的多模态模型,无论是 Dell-E、Midjourney 等文生图模型,还是 Sora 等视频生成模型,其产生高质量视觉内容的语料库前提,依旧以“犹他大学—好莱坞—皮克斯”的美学系统为基础。若我们在今天已经建立起了多模态生成模型将成为未来世界模拟器的一般认识,那么对这一问题的回应就变得迫切



▲ 因其在 1970 年代技术研发上的先发优势,好莱坞迅速占领了上一代世界模拟器的话语权和审美决定权。图为《捉妖记》海报。

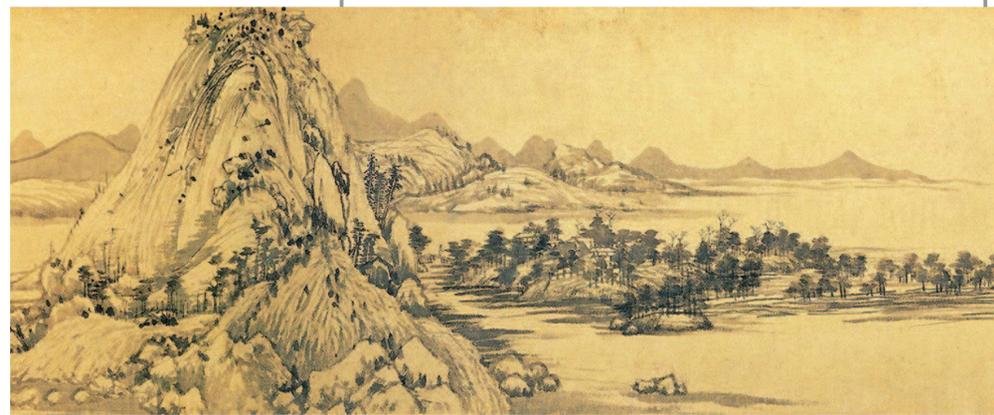
▲ 在三维动画领域,好莱坞依靠其经济和技术优势引领,决定着文化与审美表达的偏向。如改编自《山海经》故事的中国电影《捉妖记》中,妖兽的形象却是典型的“怪物史莱克”式的。图为《捉妖记》海报。

起来。

事实上,数据要素化与语料库建设正是这一轮以人工智能技术为核心的全球技术、文化竞争中的“软实力”。正如郑永年指出的,当前中国互联网公司、硬件发展都不输于美国,但中国发展人工智能的短板,主要体现在数据质量方面。这首先意味着我们在当下讨论“文化出海”时,也要考虑“数据出海”,跨文化交流必须包括数据交流。来自中国的要素化数据应更主动参与全球生成式人工智能的语料库建设,参与内容生成,确保 AI 训练数据的多样性和平衡性。

另一方面,建设高质量的中国文化数据库——包含广泛的语言、文化、历史和艺术等方面的数据——将成为这一轮人工智能文化交往与世界模拟建设中的核心竞争力。这既是对中华优秀传统文化进行创造性转化和创新性发展的最前沿任务,也是在“人工智能+”时代从源头上保证技术自主可控、维护文化领域意识形态安全的迫切需要。

(作者为上海师范大学影视传媒学院教授)



▲ 元代黄公望《富春山居图》局部

文学和绘画,皆是隐藏的艺术,用一句话隐藏不便说出的千万句,用一棵树隐藏千万棵树。

隐士与富春山居图

刘亮程

作家陆春祥将自己的书院建在隐士严子陵故乡,这个坐落在富春江畔叫桐庐的地方,因为文学而聚集了许多作家和读书人。那晚我在书院讲堂给几十位写作者讲课时,窗外鸣叫了一天的麻雀、乌鸦和喜鹊,好似都安静下来。鸟也有嘴叫累的时候。这些听过历代文人隐士高谈阔论的原住鸟儿们,或许早已习惯了人的聒噪。这地方自从有了隐士严子陵,所有一切都隐藏不住了,人的声音更是没有消停过。

年轻时我读严子陵的故事,想象不出他隐在怎样的地方。其时我不为人知的生活中,靠写诗想让人知道。就像树梢上一整天扯嗓子鸣叫的鸟,想让更多的鸟听到。我在那个偏僻地方也扯嗓子叫过。后来我靠写作让自己的声音传到了外边。如我写的,“人把一件件事情干完,干好,人就渐渐出来了。”

严子陵是一个靠隐获得大名的人。他显然没做成隐士,因为跟刘秀是故友,一个被皇帝四处找寻的人,还有什么可隐的。只是他坚守住了不出山。刘秀也满足了想做“不招之臣”的心愿。后来慕隐士之名而来的那些名人,也都在隐士家乡留下了名字。当然还有更多的无名者,他们隐在自己的一生里。

我们驱车又乘船,游览富春江畔因严子陵而显名的钓台、祠堂、书院、寺庙时,我脑子里浮现的是《富春山居图》。严子陵和《富春山居图》的作者黄公望,隔了数百年,期间朝代更替,富春江也从汉朝流淌到元代。但他们同是心机深重的隐者,前者想隐逸自己,后者却用一幅画隐藏起一处江山。

在山居图这幅长卷中,那几个隐约的人:打鱼的渔夫、砍柴的樵夫、茅亭里戴着高高帽子的读书人,以及江船上面目模糊的过客,或许都是藏匿了真实身份的隐士。那两位渔樵,或许深藏起整顿山河的雄心,老老实实地垂钓砍柴。那读书人读的也许不是春花秋月的诗词,而是金戈铁马的兵书。这条江流过多少代人,江水会老,山木会腐朽新生。一条江在悠长时间里的模样被画家留下来。

早年我看这幅画,没想到过它跟哪个地方有关系,富春江只是一幅画里的江。即使后来我知道了有一条真的富春江,我也从未想过要来看看。我已经看过画中的江和山居了,一条在地上的江,怎么流淌也不会让人意外。

陪同的朋友说,《富春山居图》画的就是这一段山水。并在手机上搜索到山居图,让我对照着看。我说,古代画家不会像现代画家那样,对着山水写生作画。古人画的是心中之画。画家将万千山水藏于心中,待要画时,千里万里的山水争相涌来,一颗心灵成了世界中心,被造物安排过的山水,又被画家重新安排。这一次的再造,只因一颗心动。山野高低起伏皆是情绪,心中沧桑只委屈枯枝古树,而大江浩荡,再宽的波浪拍打在心之岸,多高远的云天,也在轻轻一抹的笔端。

在一幅名曰《富春山居图》的画中,真实的富春江早已被隐藏了,它成为一幅画的素材。一条被绘画当素材用过的山水,尽管不会因此减少什么,但肯定被攫取走了什么。那是

山水的魂。

我相信每一处山林都隐藏着一幅山水画,就像一个醒着的人心中藏着一场他夜晚会做的梦。山居图是富春江隐藏的一段山河大梦,千年前已经做成。

文学和绘画,皆是隐藏的艺术,用一句话隐藏不便说出的千万句,用一棵树隐藏千万棵树。就像一部小说的故事,是从一堆故事中活下来的一个故事。山居图中活下来的,是富春江岸无数年代里无数棵树的样子,无数座山还是山,无尽流淌的江水还是水。但是,一条被笔墨绘出的江已经完全不一样。那必是闭住眼睛看见的山水。画家需要牺牲掉多少棵树,才能让《富春山居图》的林木疏朗清透。画中那有数的几个人背后,是隐藏在世间的千万人。那是岁月和朝代更替里的人,跟草木一样,老了又年轻,枯了又荣。

富春江及沿岸山林,是被《富春山居图》隐藏的现实。再美的眼前自然,也必经过一个人的眼睛心灵,在人心灵中活下来的那些自然,成为了艺术。作为游客,我看见的是此刻的山水,而此时之外,是一条江和两岸林木漫长的生命枯荣期,画家所画,必是流了千年的江水,长了千百年的老树,和活过几世的人。就像作家会在一个句子里写尽一生一世。

富春江的波浪里满是古今流不尽的人声。看久了《富春山居图》,再看眼前的富春江,直觉得对岸山林,像是照着《富春山居图》的模样的,每一处都有点像,但都没完全长像。或许富春江要长成那幅画中的样子,还需千万年的光阴。因为艺术所表达的,是山尽头的山,水穷处的水。

(作者为新疆作协主席、茅盾文学奖得主)