“一键生成”能“生”万物？

Sora诞生半年后，它的“挑战者”纷至沓来，连“等不及”又“追不上”的英伟达也亲自下场。

迄今为止，Sora依旧只释放小样、未开放使用，而快手可灵、智谱清影、Vidu已率先打开体验大门，走向大众。

尽管“一键生成”的初体验称不上完美，却搅动了内容产业的一池春水。身边不少微短剧、广告、动画已经开始用上AI这个“效率搭子”。人工智能生成技术，从前不久的文生图，到如今的文生视频、图生视频、视频生成视频，“AIGC宇宙”不断扩张。

AI，是华夏神话里的“神笔马良”吗？它能让多少想象力、创造力活起来、动起来？

▲2024年7月26日，中国科技企业智谱AI面向全球用户发布其自主研发的人工智能生成视频模型清影（Ying）。图为用户登录界面。

▲将一张图片发给国产视频大模型Vidu，一段动画视频随即自动生成。图为视频截图。

◀7月6日，2024世界人工智能大会上的工作人员（左）向参观者介绍AIGC古籍修复技术。

新华社记者方喆 摄

“文生视频”，如何“生”

“文生视频是一颗重磅炸弹。”半年来，从大厂到独角兽的Sora复现潮无一不在说明产业界对“生成”的重视。

视频生成，简言之是通过生成式人工智能技术，将文本、图片等多模态输入，转化为视频信号。

当前，视频生成的技术路线主要有两种。一种是扩散模型，其中又分为两类，一类是基于卷积神经网络的扩散模型，如Meta的EmuVideo、腾讯等推出的VideoCrafter；一类是基于Transformer架构的扩散模型，如OpenAI的Sora、快手的可灵AI、生数科技的Vidu等。另一种是自回归路线，如谷歌的VideoPoet、Phenaki等。

目前，基于Transformer架构的扩散模型是视频生成模型的主流选择，也称“DiT”（Di为Diffusion缩写，T为Transformer缩写）。

文本“扩散”为视频？“扩散在此指一种建模方式。”北京大学信息工程学院助理教授、博士生导师袁粒举了一个生动例子——

米开朗琪罗在凿刻著名的大卫雕像时，说过这样一句话：雕塑本来就在石头里，我只是把不要的部分去掉。“这句话很形象地形容了‘扩散’这一建模过程。原始的纯噪声视频好比未经雕琢的石块。如何敲打这个大石块，敲除多余的部分，直到把它敲成轮廓清晰的‘大卫’，这样的方式就是‘扩散’。”袁粒说。

袁粒进一步解释：“Transformer就是一个神经网络，遵从‘规模规则’，执行敲石块的过程。它能处理输入的时空信息，通过理解其内部复杂关系来理解现实世界，使模型具备推理能力，既能捕捉视频帧之间的细微联系，也能确保视觉上的连贯、时间上的流畅。”

“效率搭子”，有多快

一只憨态可掬的北极熊被闹钟叫醒，背起行囊，乘坐直升机、转乘高铁、换乘出租车、登上轮船，跨越山河湖海、历尽艰难险阻，终于到达南极，与企鹅相会……

这部时长1分半、名为《一路向南》的动画短片，由视频生成模型Vidu完成。原本1个月的工作量，有了AI这个“效率搭子”的加入，仅用1周时间就制作出精良作品——效率是过去的4倍。

这让北京电影节AIGC短片单元最佳影片得主、Ainimate Lab AI负责人陈刘芳心生感慨：视频生成技术，让高水平动画不再是大厂才敢玩的“烧钱游戏”。

AI动画《一路向南》的创作团队仅由3人构成：一名导演、一名故事版艺术家、一名AIGC技术应用专家。而以传统流程制作的话，需要20人。算下来，仅制作成本就降低90%以上。

正如快手视觉生成和互动中心负责人万鹏飞所言，视频生成的本质是从目标分布中采样计算得到像素。这种方式能以更低的成本，达到更高的内容自由度。

进入Vidu的视频生成页面，笔者也体验了一把“一键生成”的自由。上传一张照片设置为“起始帧”或作为“参考的人物角色”，在对话框里输入想要生成的场景的文字描绘，点击“生成”，一条灵动精彩的短视频就自动生成了。从进入页面到下载完毕，不足1分钟。

“‘人人成为设计师’‘人人成为导演’的时代将会到来，就像当年‘人人拥有麦克风’一般。”智谱AI首席执行官张鹏说。

“世界模拟器”，有戏吗

视频生成，颠覆的仅仅是内容产业吗？这显然不是OpenAI的初衷。“生成视频”只是一道“开胃菜”。

Sora诞生之前，OpenAI并未将其定位为AIGC的实现工具，而是复刻物理世界的“容器”——世界模拟器。这一容器里，运行着真实世界的物理规律、环境行为、交互逻辑，恰似《黑客帝国》描绘的虚拟世界，冲击着我们的想象与感官。

然而，物理世界是三维的，目前的Sora等模型还只是基于二维运作，并非真实物理引擎，也就谈不到深层次的物理世界模拟。

“多年来，我一直表示，‘看到’世界即为‘理解’世界。但是现在我愿意将这个概念推进一步，‘看到’不仅仅是为了‘理解’，而是为了‘做到’。”斯坦福大学讲席教授李飞飞公开表示，空间智能的底线是将“看到”和“做到”联系在一起，有一天，AI将会做到这一点。

当“看到”还不等于“做到”时，人工智能的创造就不能停。最近，又有新的技术路线出现了。不同路线之间你追我赶，共同向前，推进这个由向量与模型构造的智能世界。

未来的“世界观”，依旧是一道尚未揭晓的谜题。正如美国物理学家费曼所说：“我不能创造一个我不理解的世界。”但这并不意味着，理解了一个世界，就一定能够创造出一个世界。

此刻，依旧是颠覆到来的前夜。这就是为什么当我们向技术探索者抛出关于未来的问题时，会得到截然不同的答案。也许“不确定”，正是这个时代的幸事。

据半月谈

8月28日，参会人员在2024中国国际大数据产业博览会上体验AIGC智影奇境。 新华社记者杨文斌 摄