Sora生成的视频人物面部特写。Sora视频截图

突然发布“王炸”模型Sora：
OpenAI首个文生视频模型将颠覆现实？

“隆重介绍Sora，我们的文本转视频模型。”当地时间2月15日，OpenAI突然发布首款文生视频模型——Sora，震惊程度可以让熬夜族们彻底清醒到睡不着觉，直呼“王炸来了”。

据OpenAI在社交平台发文称，Sora可以创建长达60秒的视频，其中包含高度详细的场景、复杂的摄像机运动以及充满活力的情感等多个角色。

为了展示这个王炸级技术，OpenAI还配上了一个带有提示词的视频：“美丽、白雪皑皑的东京城很繁华。镜头穿过熙熙攘攘的城市街道，跟随几个人享受美丽的雪天并在附近的摊位购物。美丽的樱花花瓣随着雪花在风中飞舞。”视频中，60秒一镜到底的画面中，不仅做到了画面主角表现流畅——一对情侣携手悠闲地漫步在东京街道，甚至摊贩两边的背景人物，都流畅真实得难以置信。从大中景无缝切换到脸部特写。

“60秒超长长度”“单视频多角度镜头”“这怎么办啊！”“太卷了吧！”……当“世界模型”越来越真实，人类社会虚拟和现实的界限，还能区分得清吗？

Sora生成的视频人物漫步东京街头。Sora视频截图

Sora诞生让现实不存在了？

“这是我们的视频生成模型Sora，今天，我们开始为有限数量的创作者提供访问权限。”OpenAI创始人兼CEO山姆·阿尔特曼在社交媒体说。

目前OpenAI在官网上已经更新了Sora生成的很多视频，这些视频不仅准确呈现出指令细节，还能理解物体在物理世界中的存在，并生成具有丰富情感的角色。该模型还可以根据提示、静止图像甚至填补现有视频中的缺失帧来生成视频。

例如，一个大语言模型的提示词描述是：在东京街头，一位时髦的女士穿梭在充满温暖霓虹灯光和动感城市标志的街道上。

在Sora生成的视频里，女士身着黑色皮衣、红色裙子在霓虹街头行走，不仅主体连贯稳定，还有多镜头，包括从大街景慢慢切入到对女士脸部表情的特写，以及潮湿的街道地面反射霓虹灯的光影效果。

更令诸多网友热议的视频，是一只会“踩奶（注：猫的一种恋母的表现）”的猫：一只猫试图叫醒熟睡的主人，要求吃早餐，主人试图忽略这只猫，但猫尝试了新招，最终主人从枕头下拿出藏起来的零食，让猫自己再多待一会儿。在这个AI生成视频里，猫甚至都学会了“踩奶”，对主人鼻头的触碰甚至都是轻轻的，接近物理世界里猫的真实反应。

但值得注意的是，在这个视频中也有一些小瑕疵：猫主人翻身时胳膊肘跟被子融为一体。

根据OpenAI解释的工作原理，Sora是一个扩散模型，它生成的视频一开始看起来像静态噪音图形，通过多个步骤逐渐去除噪声后，视频也从最初的随机像素转化为清晰的图像场景。

与GPT模型类似，Sora使用了Transformer架构（基于自注意力机制的神经网络模型），因此可以实现极强的扩展性。

OpenAI将视频和图像表示为称作“补丁”的较小数据单位集合，每个“补丁”都类似于GPT中的一个“标记”，通过统一的数据表达方式，能实现在更广泛的视觉数据上训练和扩散变化，包括不同的时间、分辨率和纵横比。

Sora是基于过去对DALL·E和GPT的研究基础构建，利用DALL·E 3的重述提示词技术，为视觉模型训练数据生成高描述性的标注，因此模型能更好地遵循文本指令，实现用户想要的视频场景。

除了能够仅根据文字说明生成视频外，该模型还能根据现有的静态图像生成视频，并准确、细致地对图像内容进行动画处理。该模型还能提取现有视频，并对其进行扩展或填充缺失的帧。

技术破壁之后会“深度造假”吗？

随着人工智能成为世界各地科技界的焦点，新工具Sora进一步引发了人们对深度造假的担忧：根据简单的文本提示生成高度逼真的60秒视频，这不是大大提高了人工智能视频和已被用来欺骗民众“深度造假”内容的质量吗？

对此，OpenAI表示，这款名为“Sora”的新工具最初只会供一小部分艺术家和电影制作人或试图找到将人工智能工具用于恶意目的方法的研究人员使用。

过去一年，人工智能生成的图像、音频和视频的质量迅速提高，OpenAI、Google、Meta和Stable Diffusion等公司竞相制造更强大的工具并寻找销售方式。与此同时，人工智能研究人员警告说，这些工具已经被用来欺骗民众。

实际上，其他公司也构建了自己的从文本到视频的人工智能生成器。谷歌正在测试一个名为Lumiere的模型，Meta有一个名为Emu的模型，人工智能初创公司Runway已经在开发产品来帮助电影制作人制作视频。但人工智能专家和分析师均表示，Sora视频的长度和质量超出了迄今为止所见的水平。

美国伊利诺伊大学厄巴纳-香槟分校信息科学教授特德·安德伍德表示：“我想象不出在接下来的两到三年内还能出现这种持续、连贯的视频生成水平。”虽然他推测OpenAI可能只是选择了模型最佳状态的视频进行展示，但他相信与其他由文本生成视频的工具相比，“Sora生成的视频质量似乎有所提升”。

如果你认为OpenAI的Sora只是像DALL·E一样，那可能就“略显肤浅”了。Sora是一个数据驱动的物理引擎，它是对许多世界的模拟，无论是真实的还是幻想的。模拟器通过一些去噪和梯度数学来学习复杂的渲染、“直观”物理、长期推理和语义基础。

“如果Sora使用虚幻引擎5对大量合成数据进行训练，我不会感到惊讶。”英伟达高级研究科学家兼人工智能代理负责人Jim Fan通过Sora生成的视频分析，提示词是两艘海盗船在一杯咖啡内航行时互相战斗的逼真特写视频。

“模拟器实例化了两种精美的3D资产：具有不同装饰的海盗船。Sora必须解决文本生成3D画面的问题；3D对象在航行中还要避开彼此路径并始终保持动画效果。”还有咖啡的流体动力学，甚至是船舶周围形成的泡沫。流体模拟是计算机图形学的一个完整子领域，传统上需要非常复杂的算法和方程，而视频的写实主义，几乎就像光线追踪渲染一样。

Jim Fan指出，视频中模拟器考虑到杯子与海洋相比尺寸较小，并应用移轴摄影来营造“微小”的氛围。“场景的语义在现实世界中并不存在，但引擎仍然实现了我们期望的正确物理规则，它将取代所有手工设计图形。”

“王炸”技术将带来行业落日？

技术的快速进步使得从电影制作到新闻行业等各个行业的人都在争先恐后地了解它可能会对自己工作产生怎样的影响。

在Sora诞生前，AI视频的工作流都是单镜头单生成，在一个视频中，多角度且连贯流畅的自由切换是无法想象的。“不管多么悲伤和恐惧，这就是所有工作的未来。”某电影后期制作人告诉华西都市报、封面新闻记者，技术的进步是不会止步的，不能更不会因为预感未来它可能取代我们的工作而停止。“对于行业来说也是好事，技术进步意味着能制作出更精良的影视作品，告别低质特效。”

AI视频生成器已在好莱坞引起了轰动。制作电影成本高昂、耗时，并且需要数十或数百人。一些技术专家推测，人工智能可以让一个人制作出与漫威大片具有相同复杂性视觉的电影。

“看看我们在图像生成技术发展的一年里取得了什么进展……一年后我们会在哪里？”电影导演兼视觉效果专家迈克尔·格雷西一直密切关注人工智能对行业的影响，他预测像Sora这样的人工智能工具将很快让电影制作者创建各种视频。“当技术剥夺了其他人的创造力、工作、想法和执行力，却没有给予他们应有的荣誉和经济报酬时，不是一件好事情。”他忧心忡忡地说。

Sora视频的质量，尤其是那些看起来像现实生活的视频，比大多数其他人工智能公司迄今为止能够制作的视频质量要高。普林斯顿大学计算机科学教授Arvind Narayanan表示，根据OpenAI发布的视频，Sora“似乎比任何其他视频生成工具都‘先进得多’。他表示，“这可能会导致‘深度伪造’视频，人们更难识别出人工智能生成的视频。如果你仔细观察一些视频，仍然可以发现许多不一致的地方。例如，他在社交平台的一篇帖子中指出，在东京街头的视频中，一名女子的左右腿交换了位置，背景中的人在有东西经过他们面前后消失了。”

但无论如何，OpenAI送上的“礼物”已经足够震撼了。

华西都市报-封面新闻记者 边雪