

N 新京报 澎湃新闻

文生视频刷屏 影响有多大?



OpenAI官网上Sora生成的首个视频全景画面截图



“现实不存在了，彻底端掉行业饭碗”？OpenAI 首个视频生成模型 Sora 发布，一夜之间刷屏朋友圈。据 OpenAI 官网，Sora 模型能根据提示词，生成长达 1 分钟的高清视频。因其“逼真”和“富有想象力”被广泛赞誉，也颠覆了传统视频生成领域平均只有 4 秒的视频生成长度。

这比横空出世的 ChatGPT (OpenAI 研发的一款聊天机器人程序) 更令人震惊。虽然 Sora 模型尚未面向公众开放，但是已经有人感叹：视频生产颠覆性革命来了……那么，Sora 能做什么，牛在什么地方，会带来什么挑战？Sora 生成的视频合理吗？事情真的这么简单吗？记者进行了相关调查。

焦点1

Sora 能做什么？

“一位时尚女性走在充满温暖霓虹灯和动画城市标牌的东京街道上。她穿着黑色皮夹克、红色长裙和黑色靴子，拎着黑色钱包。她戴着太阳镜，涂着红色口红。她走路自信又随意。街道潮湿且反光，在彩色灯光的照射下形成镜面效果。许多行人走来走去。”

OpenAI 在官网陆续公布了十几段长短不一的视频，均根据文字提示生成。这是第一段视频的文字提示，Sora 根据提示生成的一分钟视频在网上广为流传，可以视为 Sora 的代表作，其展示的信息量极大，从主要人物到背景人物，从远处的街景到地面，看起来十分逼真，镜头切换也非常流畅自然。

从这张全景画面截图，可以看到和主人公同时走动的路人，湿漉漉的地面，眼花缭乱的霓虹灯和招牌。再看后面几个特写镜头：女主脸上斑点、脖纹清晰可见，显得非常真实，还有墨镜里反射的街景，这些细节都没放过。

据 OpenAI 发布的信息，Sora 主要功能包括：根据文字提示生成视频；静态图片生成视频；对视频进行填充和扩充等。显然，上述功能均并非 Sora 首创，很多其他工具可以不同程度实现。不过，其生成效果并不能令人满意，并且很容易出现错误。而此次，Sora 则将生成视频的质量提升了一大步。

总之，Sora 的 AI 视频效果，几乎和人工拍摄不相上下。也因此，虽不是从 0 到 1 的原始创新，却已足够令人震惊。

马斯克在 X 上两次转发网友评论时写到“AI 增强的人类将创造出最好的作品”“人类认赌服输”。而 360 集团创始人、董事长周鸿祎日前也在微博发文表示，这意味着 AGI 实现将从 10 年缩短到 1 年。

焦点2

Sora 牛在什么地方？

这些看上去是可以假乱真的视频，Sora 是怎么做到的？

根据 OpenAI 官方介绍，用户可以通过使用自然语言提示词 (Prompt)，生成长达 1 分钟的视频。Sora 模型对语言有深刻的理解，因而能够准确地理解提示词，并生成令人信服的人物，而且可以在一个生成的视频中创建多个镜头，在这些镜头中准确地保持角色和视觉风格的一致性。

在 360 创始人周鸿祎看来，Sora 只是小试牛刀，它展现的不仅仅是一个视频制作能力，而是大模型对真实世界有了理解和模拟之后，会带来新的成果和突破。他认为，AI 不一定那么快颠覆所有行业，但它能激发更多人的创造力。

中央民族大学新闻与传播学院教授郭全中在接受记者采访时说，相比 ChatGPT，Sora 有了质的飞跃：一是 Sora 具备推理能力，人类长期以来孜孜以求的 AGI (通用人工智能) 成为可能；二是在当今视频时代，Sora 作为图生视频的利器将助力传媒业降本增效，极大地扩大能力范围和边界，给传媒业带来的影响远远超过 ChatGPT。

中国人民大学吴玉章讲席教授、国家发展与战略研究院研究员刘永谋在新京报评论撰文认为，Sora 的 AI 视频效果，几乎和人工拍摄不相上下。他预测，未来，各种不同用途的 Sora 短视频，或许会很快充斥各种视频社交平台，其对现实的影响与改变，不可不谓之巨大。

从业界到学界，几乎是异口同声认为，Sora 将会深度改变视频相关领域，从影视、传媒到社交平台，内容生产将会发生巨变。

焦点3

Sora 是“通用世界模型”吗？

为何 Sora 能够在视频生成的质量上明显领先同业产品？OpenAI 在官方网站上表示，Sora 是能够理解和模拟现实世界的模型基础，相信这一能力将是实现通用人工智能的重要里程碑。不过，Sora 还存在很多不完善之处，仍然处于世界模型研究应用的初期阶段。

OpenAI 在官网表示，目前 Sora 模型也有弱点：它可能难以准确模拟复杂场景的物理特性，并且可能无法理解因果关系的具体实例。例如，一个人可能会咬一口饼干，但之后，饼干可能没有咬痕。同时，该模型还可能混淆提示的空间细节，例如，左右混淆，并且可能难以精确描述随时间推移发生的事件，例如遵循特定的相机轨迹。

而什么是世界模型？记者了解到，实际上 runway 公司在去年 12 月就提出过要开发通用世界模型 (General World Model)，用其旗下的 Gen-2 模型来模拟整个世界，“我们相信，人工智能的下一个重大进步将来自理解视觉世界及其动态的系统，这就是为什么我们要围绕通用世界模型开始一项新的长期研究工作”。

从效果上看，目前 OpenAI 已经通过 Sora 部分做到了这一点，因为只有理解物理世界的运行法则，文生视频模型才能创造出更加逼真的视频。英伟达高级科学家 Jim Fan 就对此表示，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，还是虚构的。该模拟器通过去噪和梯度学习方式，学习了复杂的渲染、直观的物理、长期推理和语义理解”。而理解现实世界的物理法则，也正是通往通用人工智能这一“终极目标”的必经之路。

焦点4

Sora 带来哪些挑战？

至于 Sora 的出现，可能导致大量的影视等相关行业从业者失业，更是不可忽视的重大社会风险。其实，关于 AI 大规模应用可能引发的失业问题，已经有汗牛充栋的深入研究，也提出了很多应对方案，包括延缓 AI 技术的应用速度、帮助再就业、征收 AI 税等。现在，Sora 一夜刷屏提醒我们，要将这些措施稳步细化实施，而不能再仅仅停留在“刷屏”上了。

去年，针对 ChatGPT，就有学者指出，人工智能快速发展，令我国基础教育优势荡然无存。只掌握知识，而缺乏运用知识解决问题的能力，没有创新精神与创新能力，将难以应对人工智能快速发展的挑战。

舆论普遍预测，人工智能会带来未来职业格局的颠覆性变化。为此，仅调整学校办学，新增专业、课程的思维是行不通的。因为，这也跟不上社会变化的速度。

能适应社会变化的，唯有转变重视知识教育的传统模式，教育应从让学生学会知识转变到学会学习，学会创新、创造，这就必须扭转知识化、应试化的功利教育倾向。

也就是说，如果不推进破除基础教育存在的功利化、短视化问题，AI+教育等技术变革，也难以带来人才培养模式的创新，更遑论提高人才自主培养质量，造就拔尖创新人才。

而对于传媒从业者来说，则有两挑战值得思考：一就是对视频内容生产的颠覆，或导致大量人员失业；二就是对新闻传播伦理的挑战，模糊了真实与虚拟的边界。

刘永谋认为，当前的短视频时代，Sora 既然可与人工媲美，制造短视频的专门团队就可能面临下岗危险。要知道，短视频的摄影、导演、剪辑等任务，Sora “一个人”费点电就全包了。

郭全中也认为，Sora 的出现，会导致视频生产门槛大大降低，新闻伦理肯定会受到很大冲击。

焦点5

Sora 横空出世 哪些公司有布局？

浙商证券认为，短期内，Sora 及同类产品可大幅提升图像和短视频的制作效率，改变创意生产及营销 workflow，提升短视频产品生产力，对于业态更加复杂的长视频和游戏，受限于模型目前还无法很好理解因果关系和其他技术难点，现阶段或以提供美术灵感支持为主。

据不完全统计，有十多家 A 股上市公司之前已经布局多模态大模型或对 AI 视频的相关研发。

海康威视 (002415) 在去年 4 月 20 日回复投资者提问时披露，公司从几年前就开始研发视觉大模型，现在已经到了多模态大模型的研发阶段，包括视觉、语音、文本等多模态信号的融合训练及处理。

大华股份 (002236) 于 2023 年 10 月发布“星汉大模型”，该模型融合点云、语音、图像等输入，构建了多模态融合的行业视觉大模型。

万兴科技 (300624) 此前发布音视频多媒体大模型万兴“天幕”。据了解，万兴“天幕”是一个以音视频生成式 AI 技术为基础的多媒体创作垂类大模型，由视频大模型、音频大模型、图片大模型、语言大模型组成，聚焦数字创意垂类创作场景。

博汇科技 (688004)、易点天下 (301171)、数码视讯 (300079)、汉王科技 (002362)、当虹科技 (688039)、东方国信 (300166)、神思电子 (300479)、因赛集团 (300781)、拓尔思 (300229)、国脉文化 (600640)、佳都科技 (600728) 等上市公司也布局相关视频研发领域。