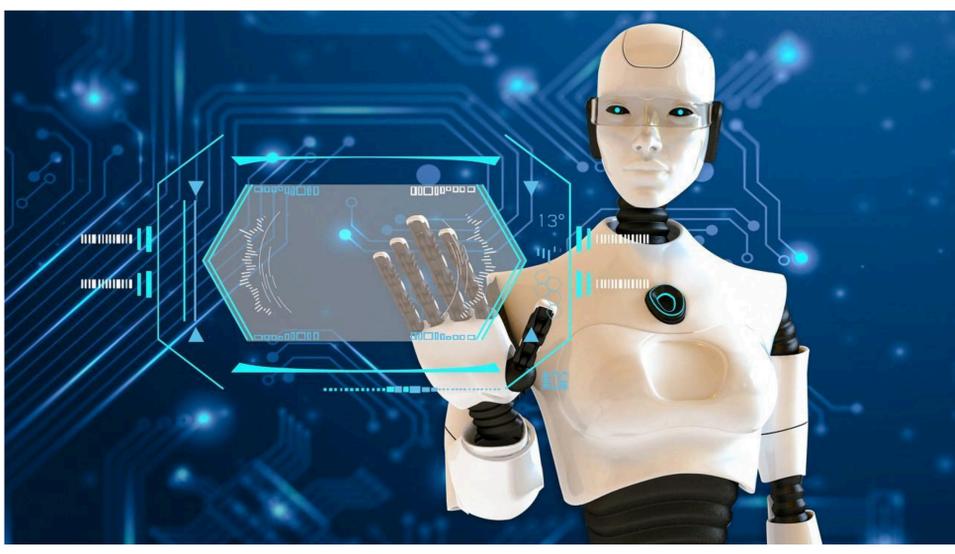


## OpenAI 推出 Sora: 重新定义现实的突破性视频生成模型

作者: 微信公众号:【架构师老卢】

2-19 7:43

~ 11



**概述:**就在几天前, 当大伊万瞥了一眼手机快速查看社交媒体时, 他被一条新闻惊呆了, 差点让他难以置信地吐出饮料! OpenAI 是去年发布的开创性 ChatGPT 背后的著名美国 AI 初创公司, 刚刚公布了他们最新的奇迹: Vincent 视频模型 Sora。他们不仅介绍了 Sora, 而且还发布了一些令人瞠目结舌的 AI 生成视频供全世界观看。根据 OpenAI 的说法, 这些视频拥有三个显著的特点: 它们长达 60 秒, 在单个视频中提供多角度拍摄, 并包含世界模型。从本质上讲, OpenAI 在视频生成效果方面实现了质的飞跃, 使 Sora 有别于其前辈。那么, 为 ChatGPT 做出贡献的 OpenAI 这次是如何

就在几天前, 当大伊万瞥了一眼手机快速查看社交媒体时, 他被一条新闻惊呆了, 差点让他难以置信地吐出饮料! OpenAI 是去年发布的开创性 ChatGPT 背后的著名美国 AI 初创公司, 刚刚公布了他们最新的奇迹: Vincent 视频模型 Sora。

他们不仅介绍了 Sora, 而且还发布了一些令人瞠目结舌的 AI 生成视频供全世界观看。根据 OpenAI 的说法, 这些视频拥有三个显著的特点: 它们长达 60 秒, 在单个视频中提供多角度拍摄, 并包含世界模型。

从本质上讲, OpenAI 在视频生成效果方面实现了质的飞跃, 使 Sora 有别于其前辈。



那么, 为 ChatGPT 做出贡献的 OpenAI 这次是如何实现 AI 视频生成, 并且几乎实现实物的呢? 这种 AI 视频生成的技术优势是什么, 它可能对我们的生活甚至人类社会产生什么影响?

### Sora 的技术

让我们先解决第一个问题。目前, OpenAI 对其大型 AI 连续视频模型的内部工作原理和算法仍然相当保密。

他们只是在社交媒体上暗示, OpenAI 在“为模型提供多帧预测”方面取得了突破。

根据大伊万的个人解释, 这似乎意味着空已经获得了通过自学深入和情感理解和构建人类发出的通用命令语言的能力。



当然, 这是一个修订版本, 其中包含所需的单词和表情符号, 以及为清楚起见而重新格式化:

“通用语言”: 在 OpenAI 的演示模式下, Sora 生成视频的方法非常简单。它所需要的只是英语 (或其他语言) 来描述所需的场景, 完全绕过了传统 3D 视频生成的复杂性。

“深刻”: Sora 理解人类语言并基于这种理解构建视频的能力确实很深远。视频生成所需的细节和深度远远超过文本描述。如果 ChatGPT 可以通过文字实现某些细节, 那么这些描述就必须在视频中充分体现出来, 展示出指数级的复杂性和深度。

“情感”: 人工智能生成的图像和视频历来难以传达人类情感。例如, 目前人工智能生成的肖像往往缺乏情感深度, 只有程式化的微笑, 导致一些评论家将它们贴上“死气沉沉”的标签。然而, Sora 模型呈现的角色表现出令人难以置信的自然、合乎逻辑和微妙的情感, 可以无缝地适应周围的环境。

借用《普罗米修斯》中的一句台词, 他们就像“这些视频中的演员……有灵魂”, 比真实的人更像真实的人。



当然, 在这些功能中, Big Ivan 认为“构建”是最关键的方面, 正如 OpenAI 的“世界模型”概念所强调的那样。

我们都认识到, 现实世界是按照特定的物理定律运作的: 物体服从重力, 风吹乱头发, 易碎物品在掉落时可以预见地破碎。

传统的 3D 建模和 AI 模型在这一领域举步维艰, 在尝试忠实地复制现实世界的物理时经常会遇到问题。构建我们的现实带来了重大挑战, 在这个领域实现准确性仍然是一项艰巨的任务。



在这方面, 空表现出了几乎与现实相提并论的非凡能力。

例如, 它发布了人工智能视频, 展示了一只在地里嬉戏的金毛小狗, 一个玩蜡烛的异想天开的生物, 以及人们在日本樱花盛开的季节悠闲漫步。这些视频严格遵循物理定律, 展示了与因果关系完全一致的无缝过渡和逻辑序列。

此外, 根据 OpenAI 的说法, Sora 的“构建”能力可以实现类似于现实世界的无限细节。与受帧数限制的传统 3D 建模不同, Sora 可以无休止地复制细节。再加上其近乎真实的感知和听觉能力, 这种构建世界的能力确实令人敬畏。

### 索拉的影响

因此, 像 ChatGPT 一样, Sora 无疑会对我们的日常生活产生重大影响。前者对我们日常生活的影响远远超出了仅仅产生文本输出或进行随意对话。一些公司已经开始利用 ChatGPT 来制定计划, 从而节省了大量时间并提高了运营效率。



与 Sora 相比, ChatGPT 对社会的影响可以忽略不计。Sora 的出现将导致广泛采用, 成本将随着时间的推移迅速下降。短视频创作者将感受到最初的影响, 因为 Sora 无需真人即可实现几乎免费的视频创作。

Sora 的功能将得到改善, 可能会影响中长视频创作者。如果它的功能扩展到一个多小时, 甚至可以使用 Sora 制作电影和电视剧。



许多人可能会怀疑 Sora 的能力, 认为它只会创造虚拟世界和角色。然而, 这与事实相去甚远。Sora 在视频制作中无缝地连接了现实和虚拟。即使是 AI 视频平台, 虽然仍处于起步阶段, 但也可以提取真人的数据来生成 AI 渲染的图像。Sora 作为数据黑匣子, 毫不费力地处理了这项任务。

例如, 您可以将偶像的图像输入到 Sora 中进行自学迭代。您提供的数据越多, 与您的偶像相似的虚拟角色就越接近。然后, 您可以描述动作, 见证您的偶像在屏幕上栩栩如生。

例如, 当前的技术可以准确地捕获身体数据以实现精致的特征, 当这些特征输入到 Sora 中并进行自我迭代时, 可以让您的偶像无缝地表演。



从这个角度来看, Sora 将比 ChatGPT 更快、更深刻地催化我们生活的变革, 重塑人类社会。此外, 最重大的转变尚未展开。

想象一下, 在图形表示中见证您的计划快速发展, 直观地设想城市和道路建设、机场航站楼和其他大型基础设施项目。对于军事战略家来说, 借助人工智能将作战计划可视化不是无价的吗?

有了 Sora, 这变得毫不费力。只需输入数据, Sora 就会直观地表现出来。您提供的数据越多, 表示就越复杂和详细。



当然, 这种进步在很大程度上依赖于计算能力和存储容量。凭借强大的计算能力, 我们最终可能会目睹一个非凡的奇观: 整个人类社会以 Sora 的形式在我们面前的超级计算机上以图形方式显示。

这就引出了一个问题: 现实世界与虚拟世界的区别是什么? 这些领域中哪些是真实的, 哪些是模拟的? 此外, 我们的现实可能只是另一个文明精心制作的模拟吗?

在 Sora 的领域中, 我们的认知是自主的, 还是我们只是一系列从属于 Sora 的算法和电信号? 这些问题既令人敬畏又令人恐惧, 引发了深深的思考。

但是, 现在, 让我们退后一步。预计到 2024 年, 将出现众多类似于 Sora 的 AI 视频生成工具, 邀请大家进行观察和推测。