

Sora 不再是一个创意玩具，而是一个“数据驱动的物理引擎”，一个可学习的模拟器或“世界模型”，可以对真实或虚拟世界进行模拟。

Sora 是不是横空出世？

过去一年多，以 ChatGPT 为代表的大模型现身江湖，引发科技界一片沸腾。2024 年中国农历新年还没过完，AI 界又发生了三件大事：谷歌推出 Gemini 新版本，支持超过 100 万个 token 的输入窗口；英伟达推出 ChatWith RTX，誓把每个人电脑都变成一个私有大模型；当然，最震撼的还是 Sora 华丽登场。

近年来，AI 生成文字、生成图片的技术飞速发展。相比之下，AI 生成视频领域虽有 Runway、Pika、Meta、谷歌等多个明星公司入局，但因技术难度更大，发展仍处于早期。

2023 年 12 月，谷歌团队发布视频生成模型 VideoPoet，一次能生成 10 秒超长、连贯大动作的视频，超过其他智能生成 3 ~ 4 秒的模型，已足以让业内振奋。谁能想到，2 个月不到，Sora 已经可以一次生成 60 秒钟的连贯视频，并且包含高度细致的背景、复杂的多角度镜头，以及富有情感的多个角色。

OpenAI 首席执行官奥特曼在 X 平台上发布了一系列视频，精美的场景、逼真的人物细节让用户惊叹不已。而这些视频全都是通过 OpenAI 2 月 15 日发布的最新视频生成模型 Sora 制作的，人们将其描述为“绝无仅有”和“游戏规则改变者”。

1 分钟的视频并不长，但对于 AI 文字生成视频却是一个巨大的飞跃。在此之前，文字生成视频的时长只有短短几秒，一大原因就在于，AI 不知道接下来要发生什么，因而不知道该生成怎样的内容。

Sora 发布的黑衣红裙女视频很好地保持了人物的连贯性，让观众也明显感受到其几乎可以“以假乱真”。当然，如果仔细观察，可以发现红裙女视频中人物的脚步在某几个帧会出现不自然的扭曲，视频中的背景广告牌虽然酷似日文，但由于目前 AI 还无法直接在视频中“认识”文字，其只能生成似是而非的“日文”。在另一则中国龙年舞龙的视频中，出现的也并非真正的汉字“龙”，只是像汉字的图形而已。这些都是 AI 生成

视频的特点之一。不过瑕不掩瑜，Sora 已经用事实吊打了同类型的文生视频应用。它不仅更加符合逻辑，还在一定程度上“展现”了对现实世界的理解能力。

英伟达人工智能研究院首席研究科学家 Jim Fan 在社交平台感叹，Sora 不再是一个创意玩具，而是一个“数据驱动的物理引擎”，一个可学习的模拟器或“世界模型”，可以对真实或虚拟世界进行模拟。

那么，Sora 具体厉害在哪？

如果说，之前的 AI “文生视频”工具是“模拟现实”，Sora 则是“构建现实”。两者的区别在于，前者是对现实的模仿，难以捕捉现实世界的物理规则、动态变化。后者，则是在虚拟世界里，构建另外一种现实。其学习的不仅是像素与画面，还有现实世界的“物理规律”。

Jim Fan 对咖啡杯里的海盗船这则视频进行了拆解分析。首先，在这个视频中模拟器生成了两只带不同装饰的精美海盗船，这需要 Sora 在其潜在空间中解决文本到 3D 的隐含问题；第二，这两艘船需要在航行和避开对方路径时始终保持动画效果；第三，从流体动力学上来看，咖啡的流动以及船只周围形成的泡沫都非常自然。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；第四，在逼真度上，整体的效果看起来几乎就像使用光线跟踪渲染一样；第五，模拟器

咖啡杯里的海盗船视频，展现了人类所期望的正确物理规则。

