由 AI 生成的短片图片，基本上成功完成了王超下达的所有指令。

4 分 27 秒（亦即 6408 帧）的视频，最后由 Elevenlabs 为英文旁白配音，加上开源版权的配乐，完成整个视频的制作。”

在王超给 GPT 的指令中，他希望 AI 能将最后一句旁白翻译成莎士比亚式的诗意古英语；在给 Midjourney 的指令中，他希望 AI 可以在片头做出“末世废土风”，带一点手绘的感觉，而中间的叙事段落则要以普通人的视角切入，影像风格偏真实，“用 33 毫米电影镜头，采用 Imax 摄像机拍摄”；而到了图生视频阶段，又希望 PixVerse 为猫咪加上伸懒腰、眯眼睛的动态——结果 AI 都成功办到了。

不过，这并不代表过程中就不需要人工参与；相反，与人们想象中的“一键生成”相比，人力花费的时间长达 8 天。

“每个环节 AI 都会暴露一些问题。”王超解释说，“最大的问题是 AI 每一次的回应都带有随机性，我们戏称为‘开盲盒’：即使你每次都输入同样的提示词，它出来的结果依然会是不同的。”

比如他想生成一张黑猫的图片，第一次出来是绿眼睛，第二次出来是蓝眼睛；第一次瘦些，第二次胖些；女孩子身上穿的“白底碎花裙”，每次也都有细微差别。为了生成最终能用的 88 张分镜图，王超一共试了 600 多次才成功。“我们行话叫‘roll 图’，就是图片生成出来，人工要手动挑选，挑选出来的图，很多细节如果不符合现实世界的物理规则，也要手动用笔刷修改：小猫的爪子怎么动，叶片如何随风摇摆，都要去设定参数。很多时间就花在这里。所以用 AI 做视频，虽然硬件成本是降低了，但体力成本是一点没减少。”

前 Sora 时代的 AI 并不高效，那么 Sora 能绕开这些原始问题吗？在王超看来，部分可以：“我们从样片中可以看到，Sora 在光线、动力、风感等很多自然界的规律上，能够和现实世界匹配。它生成视频的长度和精度也远超当下技术，而且从单机位变成了多机位，且跳过了图片阶段，这些都是质的飞跃。”

但 Sora 的底层逻辑和 GPT 相似，因此也存在概率和随机性，它不是故意做得每次都不同，而是没法做到每次都相同。“Sora 目前能做到的是 60 秒内的场景一致性和情节连贯性，再长就难说了。如果将来要应用在影视剧，麻烦就大了：你不能今天故事发生在这个场景，明天发生在那个吧？主角的脸，第一集是一个，第二集是另一个，那当然也不行啊。”

萧飞也认为，AI 目前在品质上并不能取代传统，但它让很多囿于时间和经费的点子成为了可能，这也许会爆炸式地提升视频内容的数量和质量：“我们可以把它看作是手机高清摄影取代了传统专业摄影，让不具有专业器材和专业培训的内容创作者有了实现梦想的可能。”

最近传出某影视从业者说要打造全 AI 剧，但这种鼓吹“一键生成”的，大部分是骗子。以 Sora 目前的时长，影响短视频行业或许还有可能，但拍电影电视剧，即使微短剧也够呛。

把它当作工具，而不是做工具人

OpenAI 也并不避谈 Sora 的缺点，官网上承认：“交互是目前系统最大的短板之一，AI 还不能完全把握时间的因果关系和物理世界的法则，例如人咬了一口饼干后，饼干的形状会发生怎样的变化。”

看过 Sora 样片的观众应该都注意到了其中的“穿帮”之处：打翻水杯的时候，水从杯壁而不是杯口流出；橘猫向主人伸出第三只手；女子左右脚互换；樱花树无根系地浮在半空……本来这些穿帮并不算什么，也许是随机生成中的一次小失误——但考虑到官方样片一定是精挑细选之后的产物，记者眼前就出现了《致命魔术》里那一堆帽子——在“大变活人”震撼世界的同时，背后可能有海量的失败堆骨成山。这无疑给 Sora 的可靠性打了一个问号。