



视频生成模型 Sora 横空出世,让影视、广告业乃至教育界人士顿生危机感——

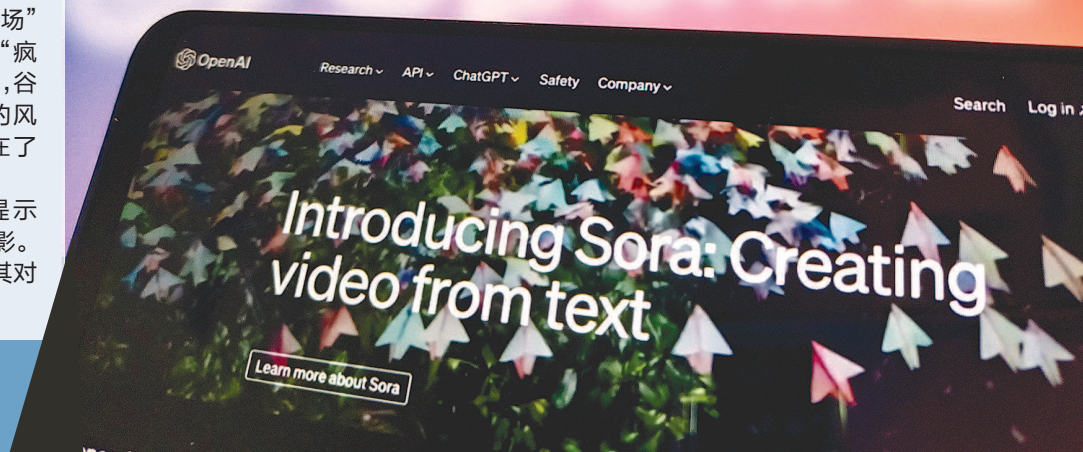
“阿拉丁神灯”?“潘多拉魔盒”?



“现实不存在了?”用“你方唱罢我登场”来形容春节假期最后两天大洋彼岸的“疯狂”,再贴切不过了。北京时间2月16日,谷歌的 Gemini Pro 1.5 还没出几个小时的风头,天一亮,全世界的聚光灯就集中在了 OpenAI 的 Sora 身上。

Sora, 新型视频生成模型,它可以根据简短的文字提示(prompt),将其转化为长达一分钟的高清视频,镜头感堪比电影。

未来,各种 Sora 视频或将现身并“占领”视频社交平台,其对现实的影响与改变,可谓巨大。



▲ OpenAI 近日发布视频生成模型 Sora

▲ 谷歌日前推出名为 Gemini 的新人工智能模型,声称其在一系列智力测试中的表现优于 OpenAI 的 GPT-4 模型和“专家级”人类

图 IC



AI 视频要“变天”了

Sora “逆天”在哪儿? 它不仅能够根据文字指令创造出既逼真又充满想象力的场景,而且能生成长达1分钟的超长视频,还是一镜到底那种。

从发布的视频来看,60秒的视频中,视频主体与背景保持了高度的流畅性与稳定性,在一个视频里实现多角度镜头,分镜头切换符合逻辑且流畅。此外,Sora 对于光影反射、运动方式、镜头移动等细节处理非常到位。

“这座城市洁净无瑕,到处可见先进的未来式有轨电车、绚丽的喷泉、巨型的全息投影以及四处巡逻的机器人。”……OpenAI 首席执行官山姆·奥特曼也亲自“下场”,在社交媒体上根据网友给出的 prompt 输出视频。

Sora 的视频效果,几乎和人工拍摄不相上下,由此带来的强大冲击让无数人感叹:“现实,不存在了。”OpenAI 是怎么做到的? 根据官网介绍,“通过一次性为模型提供多帧的预测,我们解决了一个具有挑战性的问题。”数小时后,Sora 的技术报告发布,尽管依旧“很 OpenAI”——不包括模型和实现细节。技术报告中,“里程碑”是关键词。Sora 还能将静态图像转化成视频,以该图像为背景,在视频中添加一些动态的元素,使图像变得更加生动。

据悉,OpenAI 目前没有向公众发布 Sora 的计划,但将首次开始与第三方安全测试人员共享该模型。

对相关行业具颠覆性

纽约大学助理教授谢赛宁直言:“Sora 将改写整个视频生成领域。”谢赛宁分析,Sora 应该是建立在 DiT 这个扩散 Transformer 之上;在视频压缩网络上可能采用

VAE 架构,区别就是经过原始视频数据训练。他推算 Sora 可能有大约 30 亿个参数,并表示:“Sora 或许并不需要人们想象中的那么多 GPU 来训练,如果真是如此,Sora 的后期迭代将会非常快。”

英伟达人工智能研究院首席研究科学家 Jim Fan 认为,Sora 是一个数据驱动的物理引擎,称这是视频生成领域的 GPT-3 时刻,“回到2020年,GPT-3 不是一个很完美的模型,但是它有力证明了上下文学习的重要性,所以不要纠结于 GPT-3 的缺陷,多想想后面的 GPT-4”。

360 创始人周鸿祎也发布微博,提到自己对 Sora 的看法。在他看来,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。周鸿祎写道:“这次 OpenAI 利用它的大语言模型优势,让 Sora 实现了对现实世界的理解和对世界的模拟两层能力,这样产生的视频才是真实的,才能跳出 2D 的范围模拟真实的物理世界。”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,“今天 Sora 可能给广告业、电影预告片、短视频行业带来巨大的颠覆,但它不一定那么快击败 TikTok,更可能成为 TikTok 的创作工具”。

周鸿祎也直言:“一旦人工智能能接上摄像头,把所有的电影都看一遍,把 YouTube 和 TikTok 上的视频都看一遍,对世界的理解将远远超过文字学习,一幅图胜过千言万语,这就离 AGI 真的不远了,不是 10 年、20 年的问题,可能一两年很快就可以实现。”

人才培养模式需创新

全国政协委员、上海科技馆馆长倪闽景在《Sora 只是开始,教

育真正的挑战来了》一文中指出,Sora 的出现,首先会让影视行业、广告行业的专业人士瑟瑟发抖,但是需要瑟瑟发抖的显然是教育行业。

“无论是 ChatGPT 还是 Sora,都仅仅是开始,以强大算力和算法支持的数字大脑,将在学习能力、学习范围和学习深度上远超人类,人类只有依靠人工智能一起协同学习,才是教育的出路。”倪闽景认为,人类学习将进入超级学习阶段,教育的逻辑将改变:首先,重新定义知识,从知识到知与识;其次,重新定义学习,从工具到本能需要;最后,重新定义创造,从创造东西到创造体验。

倪闽景写道:“Sora 在英文中有小鸟在空中咏唱的意思,这个浅淡的人工智能咏唱,如果能够让所有教育者听懂,那么我们是否会面对各种无谓的内卷释然并放下呢?”

21 世纪教育研究院院长熊丙奇认为,比感慨 AI 淘汰传统职业更重要的是,如何直面这一挑战,应对这一新的时代发展趋势。其中,最为关键的是,必须改革教育、创新人才培养模式。

他一针见血地指出,近年来,我国高校新增不少人工智能、大数据管理与应用专业;各地亦在大力推进教学数字化变革。不过,这些教育变革,存在两大值得关注的倾向:知识化倾向、应试化倾向。

熊丙奇表示:“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。”

他感慨,过去 20 多年来,我国教育界一直在求解钱学森世纪之问——“为什么我们的学校总是培养不出杰出人才?”在 Sora 刷

上海科研团队联合研发 AI 音频生成平台安菲翁

本报讯(记者 郢阳)日前,上海人工智能实验室与香港中文大学(深圳)联合团队推出 AI 音频生成平台安菲翁(Amphion)。

安菲翁现已开源并提供免费商用,其不仅具备语音及歌声合成转换、音效及音乐生成等多种能力,更可实现转换过程可视化,有效地降低了应用门槛,助力更广泛的开发者研发 AI 音频。

当前,文本驱动的生成模型在图像和视频领域均已取得显著成果。然而,由于音频领域知识体系相对独立,长期以来“文生音”的研究仍面临重重壁垒。针对这一现状,安菲翁创新性集成了经典模型架构,并提供歌声转换过程可视化能力。

记者了解到,在安菲翁的“文生语音”模块主要采用了深

度学习技术,将文本转换成自然流畅的高拟真度语音;而“文生音频”集成了当下主流的文本驱动音频生成模型架构,即基于 VAE Encoder、Decoder 和 Latent Diffusion 的文本驱动的音频生成算法。与传统的音频开源工具不同,安菲翁提供了生成过程可视化及音频可视化功能。联合团队旨在通过可视化,使初级开发者更好地理解模型的原理和细节。

据悉,安菲翁为古希腊神话中的传奇音乐家,传说其弹奏的优美琴声可让顽石感灵。上海人工智能实验室团队借此为 AI 音频生成平台命名,希望通过 AI 技术的创新为音频领域注入全新的研究思路,开源开放,“声”生不息。

屏,AI 又取得突破后,如何求解钱学森世纪之问也更显紧迫。

“对此,要形成基本共识,那就是,只靠技术手段难以推进人才培养模式改革,必须改革教育管理与评价体系。”他表示。

警惕“潘多拉魔盒”

在 Sora 横空出世前,OpenAI 首席执行官山姆·奥特曼就曾表示,人工智能有望在推理能力和可靠性、多模态、可定制化和个性化三个领域大幅提升,并认为,至少在未来 5 到 10 年内,AI 大模型技术将处于一个非常陡峭的成长曲线上。

值得关注的是,就在 Sora 发布的同日,谷歌也发布其大模型矩阵的最新力作——Gemini 1.5,其将稳定处理上下文的上限扩大至 100 万 Tokens(词元),这代表它能一次处理大量的信息——包括 1 小时的视频、11 小时的音频、超过 3 万行代码或超过 70 万字的代码库。

新事物,显然无法做到“完美”。OpenAI 明确指出,Sora 可能难以准确模拟复杂场景的物理原

理,并且可能无法理解因果关系。这一模型还可能混淆提示的空间细节,并且可能难以精确描述随着时间推移发生的事件——而这些缺陷或将使 Sora 生成一些不合逻辑的东西。

据了解,OpenAI 内部仍在开展模型伦理侧的对抗性测试,比如错误信息、仇恨内容、偏见内容,至于色情暴力等,会在文本输入时就被拒绝。

近两日,也有一些人工智能业内人士向记者表示,从技术角度来看,Sora 是非常重大的飞跃,但“硬币有两面”,它存在潜在的滥用可能,“在 AI 视频构建的世界中,‘眼见为实’行不通了,人们眼睛看到的,或许无法作为判断的依据”。

是“阿拉丁神灯”还是“潘多拉魔盒”,显然现在下结论为时尚早。该业内人士同时表示,对于新技术的到来,需要社会足够的包容性;同时,在大规模推向社会前,务必“系好安全带”,做好应对风险的准备。

本报记者 郢阳