

2月15日，美国人工智能研究公司OpenAI发布人工智能文生视频大模型Sora，该模型可深度模拟真实的物理世界，并生成包含多个角色及特定运动的复杂场景。目前，根据用户的文本提示，Sora可以创建长达60秒的逼真视频，这是近日互联网上最火爆的科技新闻。两年来，从ChatGPT3.5、GPT-4 Turbo到今日的Sora，可以说OpenAI公司正以惊人的迭代速度创造着人工智能的奇迹。

近几日，笔者仔细翻看网络上流传的Sora模型制作的短视频，高质量的画面与周围环境的匹配度无比“丝滑”，即便是笔者也拍不出这个效果，坦率地说，如果让我去鉴别其真假，相信大多数人和笔者一样没有能力区分。

需要说明的是，从工业化时代到互联网时代的无数次技术变革所带来的真假难辨现象，与今天的Sora短视频在本质上是不同的，不同的是随着技术的迭代与升级，鉴别真假的难度越来越高，这就引发了一个具有普遍性的、令人焦虑的问题——Sora之后我们如何拯救真相？

Sora之后，我们如何拯救真相

■李侠 李双

当真与假的界限崩塌之后

毫无疑问，以Sora模型为代表的人工智能是新一轮科技革命的前奏，随着应用场景的渐次绽出、新的盈利模式的涌现以及技术迭代的加速，人工智能带来的科技革命势不可挡。

其广阔的商业前景暂且不表，笔者更加关心的是其对社会观念与群体认知造成的直接冲击。因为Sora模型所构建的视频或将直接模糊真与假的界限。再加上日益临近的现实与虚拟界限的双向渗透，这一切或将导致整个社会的认知架构崩塌，随之而来的一系列急剧的连锁反应可能使人猝不及防。可以说，技术又一次让人类被动地站在转变的关口。为此，我们不妨先设想一下当真与假的界限崩塌之后会出现什么情况。

假设诈骗分子掌握了这项技术，利用专门收集的个人信息制作短视频，视频中的人、地点、时间全部是真实的，就连声音也可以通过技术处理，按照脚本设计。比如谎称受害人遭遇车祸急需打款救治，此时受害者家属如何鉴别真伪？相比之下，以往的电话诈骗十分低级。再比如政敌/仇家利用虚假的不良短视频陷害竞争者，公众何能辨之真假？在这种场景下，“眼见为实”的传统鉴别标准已经失灵。这类场景还可以列出很多，所有这些潜在危害的实质在于真假界限崩塌所带来的真相隐遁现象。

如果缺乏相应的强有力的监督机制，未来每个人都有可能成为新技术的受害者，这就是技术发展的吊诡之处：很多人没有享受到技术带来的红利，却无端承受了技术滥用带来的不便，比如各种数字技术对于老年人的不友好表现。而真假界限崩塌引发的更为严重的问题是，它会摧毁维系社会有序运行的一系列概念与认知框架，并以多米诺骨牌的方式将整个社会推向失范与无序状态。

众所周知，所有的社会秩序框架与规则的底层所依托的基本概念大抵有三：真、善、美。其中“真”是最底层的基础概念，它决定了“善”与“美”的标准的变化方向。换言之，如果真相不存在就会出现事实的建构主义，会导致认识论上的相对主义，而这种相对主义延伸到道德领域就会出现道德相对主义，即否认存在普遍的道德事实。在这种相对主义的主宰下，原本属于主观领域的审美就会被扭曲塑造。

具体而言，“真”在社会运行中起着确

立信任体系的作用。一个社会的真相越是得到彰显，其信任体系的半径就越大，整个社会的交易成本也就越低，反之亦然。此即Sora等人工智能产品模糊真假边界后，人们会产生普遍性焦虑的深层原因所在，即当矛盾律不再可靠之时，我们该以何种标准判断世间对错。

真相的背后尚且有真理作支撑，其标准是恒定的，如2+2=4在什么时候都不会变，但依托真理形成的“善”与“美”的标准是可变的，这就使得一旦真相隐遁，以Sora为代表的短视频对于后者的冲击将更为剧烈，甚至可以完全改变特定时代人们对善与美的标准。

在古希腊哲学家亚里士多德看来，美德是一种适中，好的技艺可使人置于良好的状态，过度和不足都会使人走上恶的道路。比如爱可以使人变得柔和，可以更好地面对生活；但过度的爱会走向占有和嫉妒，不足的爱则会走向冷漠，这些坏的性质取决于人们对中庸的认知，进而影响人的行为。

人们判断自己的行为是否失格尚且不易，机器所生成的结果更不受人控制。短视频无数次虚假建构某种极端的行为标准，久

而久之人们便会认为那是善的，而这种善的标准可能与我们现行的社会秩序是矛盾的，比如中国古代的二十四孝被当作善的典范，而在今天看来这些事例几乎都是不道德的。

道德真理/道德真相是可以被塑造的。诚如英国思想家赫克托·麦克唐纳所言，强大的沟通者可以鼓励整个群体接受新的道德真理。当下，我们正处在善恶标准的变革期，如果真相以激进的方式消失了，善恶的标准将面临大洗牌的局面。

同理，美的标准也是会变的。假如Sora产出的短视频持续推出一种奇怪的美丑标准，人们也会在润物细无声的背景下改变对美的认知。如人们对于整容、钻石、奢侈品等的非理性迷恋多是商家多年广告营销的结果，这样隐蔽的推广形式以消费主义的姿态出现，更容易确立一种畸形的审美标准。

审美本应是多元化的个体体验，但如今的大模型凭借其强大的自主学习能力，基于对个人数据的充分掌握，可以实现围绕个体偏好的精准定向“投喂”，从而以信息茧房的模式塑造一个时代的美丑标准。这些标准没有人知道对错，甚至无人意识到它们的存在，这就使得它们带来的危害更难被消除。想想历史上环肥燕瘦的审美标准的变迁，而裹小脚的陋习又戕害了多少女性的一生。

从这个意义上说，拯救真相就是拯救我们的社会秩序及认知。

分布式真相拯救机制

哲学家波普尔在鉴别理论的真假时曾提出过一个概念——逼真度，即两个理论相比较时逼真度高的理论就是好的理论。在拯救真相的征途上，我们可以有限地借用逼真度这个概念。逼真度之所以有限，是因为未来即便技术专家去鉴别由Sora升级版制作的这些视频，也不能保证完全可以鉴别出来，只能说他们比普通人的鉴别更可靠一些，也就是逼真度更高一些而已。

因此，从根本上解决真相的隐遁现象，还是要从制度上来考量。例如可以设立第三方机构，由它们推出针对Sora短视频的辅助鉴别模型，从而形成双方力量的均衡，这样从理论上就可以最大限度拯救真相。

但是引入辅助鉴别模型又会带来两个潜在问题。

其一，技术的军备竞赛。短视频生产方会不断加快技术迭代的速度，从而超越鉴别模型现有的技术水平、引发鉴别模型的升级，如此循环往复。

Sora来了，真相远了

如果视频生产方的迭代速度超过鉴别模型的迭代速度，在这个间隔期仍然会存在真相隐遁现象。从过往的历史经验看，资本总会跑向生产方以超越监管方，利用时间差赚取超额利润，这就是技术自身发展无法克服的逐利逻辑。对于大力出奇迹的Sora模型更是如此，不用说，资本骨子里就青睐于大数据、算法与算力这些具有盈利能力的资源。

其二，监督的第三方会带来无穷倒退的可能性。换言之，如何保证作为第三方的鉴别模型拥有者不会谋取私利？为了防止这种情况的发生，我们又会被迫引入第四方来监督第三方，如果没有一种暂停机制，这个循环会导致无穷倒退，现实生活中不乏这样的事例。

鉴于以上种种，笔者建议，为了彻底拯救被技术吞噬的真相，可以采用分布式真相拯救机制，每一方都负起拯救真相的责任，从而全方位捍卫真相，即短视频生产方推出产品时必须标注真假（可以通过立法强制性要求）；作为监督方的第三方必须履行自身职责，为公众守护真相；整个社会作为消费者必须承担起监督与批评的责任（这是全方位、全天候的监督，体现集体智慧），由此，一个分布式的真相拯救链条得以确立。

借用英国咨询专家麦克唐纳的名言，这是一个后真相的时代。但是，只要我们未雨绸缪，每一次技术变革所带来的冲击终将被人类所化解，所不同的只是成本与代价的大小而已。

（作者单位：上海交通大学科学史与科学文化研究院）

除了“用魔法打败魔法”，辨识AI视频还有哪些办法

■本报记者 赵广立

“Sora之后，眼见不一定为实了。”

上面这句感慨，恐怕许多人都感同身受。只需输入一段文字描述，Sora就能生成一段长达一分钟、画面逼真、稳定连贯的高清视频。由于画面过于真实，人们仅凭肉眼很难辨别它们竟出自AI之手。

不仅能生成视频，AI还能“魔改”视频。最近，小鹏汽车研究团队提出了一个名为“任何物体在任何场景”的新型通用视频模拟框架，它能无缝地将任何物体插入到现有的动态视频中。同样，肉眼难以分辨。

被小鹏汽车团队“魔改”的视频，就可以使用这种方法鉴别。董晶说，他们团队做了初步分析后发现，在“塞入”目标物体后，视频/图像不同帧的颜色、纹理会略有变化，“可以此为线索，在收集相关数据后进行训练和检测测试”。

不过，董晶谈到，随着Sora等工具在AI生成视频细节与多元化处理方面的能力增强，生成视频中的显式伪造痕迹会越来越少，仅依赖传统的视频分析与伪造检测方法甄别视频内容的真假，无疑会变得更加困难。

“目前针对性技术的进展还比较初步，仍需加强各类检测技术的开发和优化。”董晶告诉《中国科学报》，当前，技术方面还是沿用常规检测技术思路，因此要在构建新型伪造视频数据集的基础上，提高模型的识别能力。

同时，还需要及时更新已有视频检测模型对新型生成视频算法的兼容性。此外，还可借助数字水印、数字签名、视频检索等技术手段，加强对生成视频数据生命周期的追踪与管理。

“总体而言，目前视频内容鉴别仍处于相对被动的状态，需要针对不断迭代升级的视频合成新算法去博弈验证。”董晶表示，尽管越来越难，但AI视频在生成过程中仍不可避免地会产生一些特定的模式或痕迹，相关检测技术也会持续利用这些肉眼不易察觉的线索来反制、分析和鉴别。

她和团队从多个角度提出了新的检测算法。这些算法或基于重建误差，或基于多模态对比学习，或基于伪造特征纯化，均是对“新的特定鉴别线索挖掘”的不断尝试。

推动建立具有国际共识的标准与规范

为避免引发混乱，“从源头约束”等非技术方案频频被提起。例如，有人提出，可以约定如OpenAI等相关AIGC技术主体，在生成视频之初就埋下AI生成的印记。

董晶对《中国科学报》表示，埋设印记的方法是目前可推荐的应对策略之一，但仍需克服技术上的挑战和限制，如标记的可靠性、隐蔽性、普适性等，并综合考虑隐私和安全等因素。

相对于对视频的被动检测，水印或标记属于主动防御。董晶告诉记者，其团队目前也围绕视觉生成式水印开展了一些研究工作——他们希望在目前生成式模型中加入“鲁棒水印嵌入模块”，的是让生

董晶说，这种方法较为通用，一旦检测模型参数确定，部署简单，批

真假难辨之下，越来越多的人开始担心，AI视频可能引发更多混乱。比如视频证据不再可信：“未来也许你不得不坐在法庭被告席上，观看一段连你自己都不知道的‘作案视频’。”

中国科学院自动化研究所研究员董晶研究的就是图像篡改、深度伪造等人工智能内容安全与对抗技术，她和研究团队的许多成果已应用于多媒体智能鉴别。面对AI越来越强大的功能，技术上有哪些应对的方法和手段？普通公众面对视频内容时如何“多留几个心眼儿”，以防上当受骗？为此，《中国科学报》采访了董晶。

“用魔法打败魔法。”董晶说，目前技术上主要有两类智能检测方法去甄别一段视频是否为AI生成。

一种是基于数据学习的方法。这通常需要提前收集伪造视频和真实视频（最好是配对数据）作为训练数据集，训练出强大的深度网络。只要模型能够“记住”视频帧中的异常或痕迹，例如图像噪点、帧间不连续的运动轨迹等，就能辨别真伪。

董晶说，这种方法较为通用，一旦检测模型参数确定，部署简单，批

居里夫人的科学家形象深入人心，不论对于科学共同体还是普通公众来说，玛丽·居里都可谓20世纪的标杆人物。爱因斯坦在《悼念玛丽·居里》一文最后说，“居里夫人的毅力与热忱，哪怕只有一小

除了“用魔法打败魔法”，辨识AI视频还有哪些办法

■本报记者 赵广立

量检测效果良好。但是，该方法相当依赖训练数据的体量和完备性，对于未知或未训练的数据检测通常会失效。

另一种是基于特定线索的方法。它首先需要定义出视频中一些不合常理或逻辑的视觉“线索”，如光照不一致、人脸视频中应有的活体生理信号、说话人的口型和发音时序不匹配等细节，然后设计相应的算法去提取并定位这些线索，进而取证。这种方法可解释性更好，对视频段的定向检测性能佳，但对数据本身的多样性兼容较差。

被小鹏汽车团队“魔改”的视频，就可以使用这种方法鉴别。董晶说，他们团队做了初步分析后发现，在“塞入”目标物体后，视频/图像不同帧的颜色、纹理会略有变化，“可以此为线索，在收集相关数据后进行训练和检测测试”。

不过，董晶谈到，随着Sora等工具在AI生成视频细节与多元化处理方面的能力增强，生成视频中的显式伪造痕迹会越来越少，仅依赖传统的视频分析与伪造检测方法甄别视频内容的真假，无疑会变得更加困难。

“目前针对性技术的进展还比较初步，仍需加强各类检测技术的开发和优化。”董晶告诉《中国科学报》，当前，技术方面还是沿用常规检测技术思路，因此要在构建新型伪造视频数据集的基础上，提高模型的识别能力。

同时，还需要及时更新已有视频检测模型对新型生成视频算法的兼容性。此外，还可借助数字水印、数字签名、视频检索等技术手段，加强对生成视频数据生命周期的追踪与管理。

“总体而言，目前视频内容鉴别仍处于相对被动的状态，需要针对不断迭代升级的视频合成新算法去博弈验证。”董晶表示，尽管越来越难，但AI视频在生成过程中仍不可避免地会产生一些特定的模式或痕迹，相关检测技术也会持续利用这些肉眼不易察觉的线索来反制、分析和鉴别。

她和团队从多个角度提出了新的检测算法。这些算法或基于重建误差，或基于多模态对比学习，或基于伪造特征纯化，均是对“新的特定鉴别线索挖掘”的不断尝试。

推动建立具有国际共识的标准与规范

为避免引发混乱，“从源头约束”等非技术方案频频被提起。例如，有人提出，可以约定如OpenAI等相关AIGC技术主体，在生成视频之初就埋下AI生成的印记。

董晶对《中国科学报》表示，埋设印记的方法是目前可推荐的应对策略之一，但仍需克服技术上的挑战和限制，如标记的可靠性、隐蔽性、普适性等，并综合考虑隐私和安全等因素。

相对于对视频的被动检测，水印或标记属于主动防御。董晶告诉记者，其团队目前也围绕视觉生成式水印开展了一些研究工作——他们希望在目前生成式模型中加入“鲁棒水印嵌入模块”，的是让生

董晶说，这种方法较为通用，一旦检测模型参数确定，部署简单，批

部分继续存在于欧洲的知识分子中间，欧洲也将面临一个更加光明的未来”。

然而，作者通过梳理居里夫人本人的演讲著述及有关传记后，发现居里夫人的科学家形象在科学传播和科学界有明显区别，尤其是科学史领域传记研究潮流的兴起，居里夫人的形象有了新的变化。作者认为，这种变化正是回应了一种时代需求。

作者首先分析了科学传播领域的英雄版本形象。科学传播领域最深入人心的居里夫人形象，是由居里夫人本人设计和管理的。居里夫人在培育自身科学家形象时，非常强调纯粹性。比如居里夫人在《皮埃尔·居里传》中将“为爱好而进行的科学研究”视为“人类最高尚的情操”。

不过，这一形象经久再塑，成为了单向度的科学精神楷模，而这种单向度形象忽略了居里夫人作为一个在国家、民族、性别、学科不对等时代的弱势群体、受压迫者，从事科学工作时所面临的复杂挑战。

接着，作者阐述了科学史界近十年对居里夫人形象的重新挖掘。新挖掘

成的视频本身携带可见或不可见的数字水印。

他们最近还尝试了在真实图像或视频中加入“对抗噪声”，这样一来，生成模型就不能在这些源数据上进行AI合成。

除技术手段之外，董晶还提到一些非技术层面的措施。

“人们需要完善AI数据治理与AI工具使用的监管法规，同时开展科普教育，加强行业规范和公众的相关防范意识等。”董晶说，对于境外AI生成服务主体，如OpenAI，“我们呼吁推动建立具有国际共识的AI数据技术标准与规范，形成共同应对生成视频的合理标记和协同监管方案”。

董晶认为，通过规范Sora这类新型视频生成工具的使用，如对其训练所依赖的源数据集做好管理和收集、规范可能产生敏感或虚假内容的生成视频的输出和安全性测试、规范治理与管控措施，便可降低AI生成视频的滥用风险，“甄别难度不会一直增加”。

增强对虚假视频的“免疫力”

尽管认同“不应将辨别视频是否为AI生成的工作交给公众”，但董晶坚持认为，普通人还是可以在面对视频内容时“多留几个心眼儿”，以防上当受骗。

为此，董晶支了几招。首先，观察视频细节的逻辑真实性，例如视频中人物动作、背景环境等是否与客观世界相符，人物的生理特征（如牙齿、手指、皮肤纹理、虹膜颜色等）是否符合常理。

她表示，现阶段Sora算法等能否简单便捷地大量生成高质量图像视频还是未知，从已公布的视频片段来看，其在运动上的瑕疵，仔细观察还是能够辨别的。

其次，观察视频的质量和清晰度是否均衡。一般而言，AI生成的视频可能会在画面质量、清晰度等方面存在一些瑕疵，例如图像模糊、画面抖动等。

最后，检查视频的内容逻辑是否合理，比如内容和情节是否合理和连贯。如有疑问，可进一步查看视频来源、发布平台、评论、格式和制作时间等信息是否可信或一致。还可以借助一些专门用于检测AI生成视频的技术工具和软件交叉验证。

董晶表示，在视频聊天等交互场景下，可以主动要求对方转为视频、靠近或远离镜头等加以甄别，因为现有伪造技术对较大运动变化的预测和生成效果相对较差。

除此之外，董晶提醒，在当下复杂的媒介和舆论环境之下，普通公众应积极学习相应知识、适当了解AI生成的机制和破绽，以备不时之需。

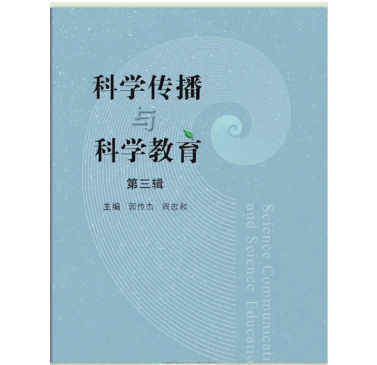
“其作用就像定期接种最新流感疫苗一样，能增强对虚假视频的‘免疫力’，变得不再盲从。”董晶对记者说，“虽然我个人认为民众无须承担辨别AI生成内容的工作，但出于公益，提高全民网络素养和安全防范意识，最大限度地避免虚假信息传播、经济诈骗、舆情误导并促进社会信任，是我们每个人应尽的义务。”

速读

论文题目：科学家精神融入剧本杀：面向Z世代的科学游戏与思想教育融合路径

作者：张文艺 张增一

出处：《科学传播与科学教育（第三辑）》



剧本杀是一种情景角色扮演游戏，也是Z世代喜闻乐见的社交形式。科学家精神是否可以结合剧

本杀，让Z世代的人共情、共鸣、共振？

Z世代泛指1995年至2009年间出生、经历世纪之交的一代青少年，他们在互联网文化的浸润下长大，有着不同于其他年代群体的价值观念、生活态度和行为方式。

当前市场已经出现了一些科普类“剧本杀”作品，如国内首款流行病学调查主题剧本杀《手账》，内容主要是防疫和健康知识科普；又如一款名为《拯救宇航员》的剧本杀产品。

剧本杀市场中，科普类、科幻类都属于小众类型。作者从剧本叙事、游戏机制与情感传播三个方面，描述了剧本杀融入科普教育的可能性。

题材与叙事是剧本杀的重要组成部分，玩家作为游戏主体，在不同题材的叙事中不断建构主人翁意识。作者通过一款应用软件，调研了网络市场口碑较好的剧本杀题材。结果显示，以历史题材、人物题材和

情感题材为主的剧本杀受到市场68%的消费者的欢迎和喜爱。

作者认为，科学家事迹中的名人大家有故事情节、政治立场和教育意义，是科普剧本杀的绝佳题材。科学家精神可以与历史、人物、情感叙事相结合，突出时代感、责任感和家国情怀。此外，科学家的爱情、家庭故事更能体现人性，激发玩家共情。

在游戏机制方面，作者分别以老科学家和青年科学家为例展现了两个剧本杀中科学家精神与游戏机制的交融，让玩家在破解科学难题的过程中，学习专业知识、感受科研攻关的不易。

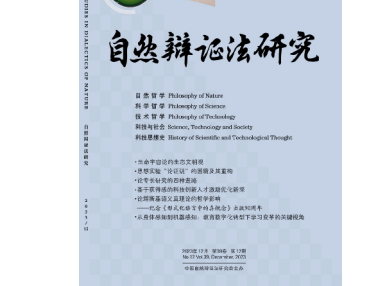
在情感传播方面，作者提出，此类剧本杀应以“共情”为实现路径，让科学家精神内化于Z世代心间。

最后，作者建议，从基于科学家正面形象的主体叙事、基于科学家人文情怀的情感渲染、基于“公众参与科学”的设计理念三个方面进行科普剧本杀设计。

论文题目：居里夫人的科学家形象辨

作者：刘元慧

出处：自然辩证法研究，2023，v.39



居里夫人的科学家形象深入人心，不论对于科学共同体还是普通公众来说，玛丽·居里都可谓20世纪的标杆人物。爱因斯坦在《悼念玛丽·居里》一文最后说，“居里夫人的毅力与热忱，哪怕只有一小

部分继续存在于欧洲的知识分子中间，欧洲也将面临一个更加光明的未来”。

然而，作者通过梳理居里夫人本人的演讲著述及有关传记后，发现居里夫人的科学家形象在科学传播和科学界有明显区别，尤其是科学史领域传记研究潮流的兴起，居里夫人的形象有了新的变化。作者认为，这种变化正是回应了一种时代需求。

作者首先分析了科学传播领域的英雄版本形象。科学传播领域最深入人心的居里夫人形象，是由居里夫人本人设计和管理的。居里夫人在培育自身科学家形象时，非常强调纯粹性。比如居里夫人在《皮埃尔·居里传》中将“为爱好而进行的科学研究”视为“人类最高尚的情操”。

不过，这一形象经久再塑，成为了单向度的科学精神楷模，而这种单向度形象忽略了居里夫人作为一个在国家、民族、性别、学科不对等时代的弱势群体、受压迫者，从事科学工作时所面临的复杂挑战。

接着，作者阐述了科学史界近十年对居里夫人形象的重新挖掘。新挖掘

的居里夫人形象更关注将她视作一个缔造机构的科学管理者，展示她成熟的科学组织能力和全面而前瞻的科学视野。

居里夫人作为科学管理者出场，与她的放射性实验室密不可分。用于实验室研究的镭盐在传统的实验室环境中无法满足制备要求，镭盐提取的半工业化处理模式实际上是一种工业活动，镭研究的推进必定带动镭工业的兴起。但镭被发现不足10年，镭市场的扩张就已经不受控制地膨胀，科学界很快就意识到必须有一套标准化的计量学，国际镭标准委员会把居里标准选为国际标。

今天的科学发展已经到了“大科学时代”，而居里夫人很早就有了更为精确的大科学时代的科学观、科学价值和科学实践体系。

作者最后总结，将居里夫人塑造为富有远见的科学规划者，正是为回应一种新的需要，打开居里夫人研究的复杂性，即打开了从科学个体触及科学整体背景的可能性。她的策略和经历、处理多重关系的经验等都是值得借鉴的。（尹一）