

最近，美国人工智能研究公司 OpenAI 的文本生成视频大模型 Sora 备受关注。扔进去一段文字，很快生成“大片”质量的视频，Sora 带给人们的感受无疑是炸裂级的。不过，有专家提出警告，Sora 等新的人工智能工具很可能会带来假视频的泛滥，而人们也会越来越难以分辨这些视频的真假。如果不法分子利用 Sora 造假、造谣，普通人又该如何识别呢？

Sora 的发布意味着
“假消息”“假新闻”不断迭代

有人利用Sora 炮制“假新闻” 该如何识别？

“假消息”“假新闻”不断迭代

Sora 的发布，意味着“假消息”“假新闻”正不断迭代。所以，对那些习惯于从短视频平台获取信息的网友来说，真的不能再随意相信那些视频了。

目前，很多人关注 Sora 在电影等娱乐内容生产上的应用，可实际上它也会改变新闻的生产方式和流程。眼下，用相关技术炮制的“假新闻”层出不穷，已经改变了很多惯有理念：图文时代，人们认为“有图有真相”，结果发现图可以 P；视频时代，很多人觉得“有视频有真相”，如今又出现了 Sora，用人工智能直接生成视频，连原始素材都不再需要。

中央民族大学新闻与传播学院教授郭全中认为，Sora 的出现导致视频生产门槛

大大降低，新闻伦理会受到很大冲击。

新技术、新应用虽然能帮助很多行业摆脱简单或重复的劳动，或者让创新灵感变得更加容易实现，但也会产生一定的负面影响。比如，别有用心之人利用深度造假技术，能更轻易地盗用他人身份、编造相关信息，实施商业诋毁、敲诈勒索、网络攻击和犯罪等违法行为。再比如，不法分子借助深度伪造技术，散布虚假视频，激化社会矛盾，煽动暴力行动，威胁国家安全和公共安全等。

可见，Sora 是一把双刃剑。管理部门、业界应当注重针对相关技术的监管，而对普通人来说，也要有所警觉：“有图有真相”“有视频有真相”都已经成为历史。

“打标签”有帮助吗？

如果肉眼无法分辨视频真伪，那有什么办法来区分呢？就目前业界的探索看，“打标签”似乎最简单、最直接。

多名社交平台从业者表示，不论国内外，现有的法律法规都没有禁止人工智能生成的内容在社交平台传播。事实上，从技术的角度看，平台也未必有能力判断相关内容是“人工智能生成的”还是“真实拍摄的”。所以，眼下各平台的普遍做法是要求发布者对这类生成内容打标签，而且不仅仅涉及大模型生成内容，还包括那些摆拍、编剧、剪辑的作品，都要打上标签。

作为大模型的开发者，谷歌和 OpenAI 也在研究“打标签”——通过后台设置，为所有通过网络其大模型生成的内容打上水印，告知用户。

然而，“打标签”或“打水印”无法从根本上解决深度造假的问题。一方面，按照目前社交平台的治理方式，有相当一部分“打标签”行为得依靠发布者。从近期澄清的一些谣言看，其中不乏断章取义、移花接木、摆拍编造的作品，但拍摄者和发布者并没有打上相关标签。可见，就治理成果而言，“打标签”并没有将造假行为一网

打尽。虽然有的视频经过平台自治或网友举报被标记、删除，但仍有大量假视频、假消息。

另一方面，即使大模型开发企业通过技术设置，使得大模型生成的文字、图片、视频等被强制“打标签”、告诉公众它们来自人工智能而非现实，但从现实传播看，相关标签、水印可能在分享中被人故意抹除，截屏、录屏、二次剪辑加工等，都可以轻松去除标签和水印，从而使得公众越发难以识别。

如此看来，“打标签”只是防范深度造假的初步手段或基础手段，效果未必理想。

“樱花雪景”是假的

那么，有什么方法能更加精准地识别深度造假呢？一个案例是让谷歌发布的人工智能大模型 Gemini-ni1.5 观看并分析 Sora 生成的视频。在观看完一段主题为“樱花雪景”的视频后，Gemini 判断这段视频是假的，理由如下：樱花一般在春天开放，那时不下雪；雪下得太均匀了；虽然是下雪天，但视频中的人物穿得很单薄。Gemini 给出的理由未必都站得住脚，不过提供了

另一种思路：面对人工智能造假，技术打假变得越发重要。

Gemini 对 Sora 生成内容的判断，提供了另一条思路——用人工智能识别人工智能。不少业界人士将此比喻为“用魔法打败魔法”，认为通过技术手段，有机会从根本上识别深度造假并降低相关风险。

很多第三方企业与机构开启了“鉴别深度伪造”技术的研发之路。合合信息图像算法研发总监郭丰俊

介绍，国内外已有不少企业专注于鉴别人工智能篡改。他举例，比较典型的是在金融领域，已有不少金融机构利用国内技术企业自主研发的人工智能甄别伪造人脸图像技术，来判断相关人脸图像是真实的还是合成的，或者被 AI 换脸。这一技术主要应用在银行及金融行业的客户图像、视频真伪检测上，能避免身份盗窃和欺诈，保护财产安全。（综合上海网络辟谣等）

清华大学 土木工程专业停招？ 清华：纯属误解！

近日，多个自媒体账号称“清华大学停止土木工程专业招生”，引发公众关注。

清华大学2月26日回应称：这纯属误解。清华大学自2017年起实施大类招生培养与管理，部分招生专业名称发生变化，原来的土木、水利以及工程管理等单一专业已合并使用“土木水利海洋工程”的大类专业名称，不再单独招生。这是教育改革中的正常专业更新，并非相关报道所称的“清华大学停止土木等专业的辅修学位招生”。

据清华大学官网发布的《2023—2024学年度专业设置、当年新增专业、停招专业名单》，文末明确提到：“我校自2017年起实施大类招生培养与管理，部分招生专业名称发生变化，详见清华大学本科招生网”。

查询清华大学本科招生网，在“土木类”招生专业下，明确说明了清华土木“现已成为拥有土木工程、水利科学与工程、工程管理、交通工程、海洋科学与工程等五个方向的大土木类专业”。网站称，清华大学在全国高校率先获准设立了宽口径本科专业“土木、水利与海洋工程”，实行大类招生、大类培养和大类出口。本科按大类统一招生，进入大类后的前两年实施大类通识教育，后两年需完成一个主修专业方向和至少一个辅修专业方向的课程学习。学生在第四学期进行大类内主修专业方向确认和辅修专业方向选定。主修专业方向确认方法：填报五个志愿的优先次序，各专业方向根据学生填报的志愿进行择优录取。（来源：光明网）

“长沙县黄花镇 三个孩子被偷”？ 安徽警方：查处！

近日，安徽阜阳警方曝光近期三起典型案例，对网络造谣者进行处罚。警方依法查处王某侠散布“山东到安徽等地将发生地震”谣言案。2023年12月12日，网民王某侠在某视频平台发布视频称，“年内中国可能发生七级以上地震，山东到安徽等地将发生地震”，造成恶劣影响。目前，太和警方依法对王某侠予以行政拘留处罚。

警方依法查处王某静、王某春摆拍并发布“卖票人员被打”谣言案。2023年12月28日，网络主播王某静在某视频平台发布摆拍的“八里河梦里雪乡卖票人员被打，雪地里被暴打十分钟，现还在雪堆躺着”谣言视频，造成恶劣影响。目前，颍上警方依法对二人予以行政拘留处罚。

警方依法查处李某英散布“长沙县黄花镇被偷三个孩子”谣言案。目前，界首警方依法对李某英予以行政处罚处罚。

（综合大皖新闻等）