文生视频类人工智能的风险与三维规制：以Sora为视角

邓建鹏[1，2]　赵治松[1]

（1. 中央财经大学 法学院；2. 金融科技法治研究中心，北京 100081）

摘　要：文生视频类人工智能Sora一经发布即引发万众瞩目，其具有的强理解能力、高度仿真性及多模态融合能力为社会带来视觉、听觉震撼的同时，引发诸多法律风险。与此前的生成式人工智能大模型相比，Sora的潜在法律风险在人格权保护、网络犯罪及社会信任等方面更为突出。面对前沿科技给个人权益、刑事犯罪及社会稳定等领域带来的挑战，要及时采取相应的多维规制对策。一是加强对人格权的民法保护，明确个人信息使用的授权，强化数据采集和视频内容监管；二是优化刑法适用与归责，完善刑事法律的解释、适用及责任制度；三是通过规范监管，提升社会信任，推动人工智能由规制对象转向规制工具，助推人工智能系统的安全性和可靠性。

关键词：Sora；法律风险；三维规制；人工智能；文生视频类人工智能

中图分类号：TP18；D9　　**文献标识码：**A　　**文章编号：**1005-9245（2024）06-0092-09

一、Sora的技术突破

美国人工智能公司OpenAI发布ChatGPT-4后，在不到半年的时间内再度发布新的人工智能模型Sora。Sora作为新型文本转视频模型，可依据文本指令创作时长接近一分钟的逼真且内容丰富的视频，能够生成具有多个角色、特定运动以及不同主题和背景的复杂场景[①]。目前的内测和实践表明，Sora不仅能够了解使用者要求的具体内容，而且掌握了不同物体在现实世界中存在和运行的规律[②]。尽管现阶段Sora对复杂场景和空间细节的理解、模拟尚不完全准确[③]，但其出色的文本理解能力和视频生成能力推动了自然语言处理技术的前沿发展。

区别于此前发布的Runway、Pika及Make-A-Video等文生视频类人工智能模型，在视频处理方面，Sora大模型突破了循环网络、生成对抗网络、自回归变压器等技术带来的狭义的视觉数据类别和视频时限的局限[④]。Sora的学习能力和多模态处理能力使其能够全面理解文本提示等信息，超强的整合功能使其能够智能理解和分析视频中各要素的外

收稿日期：2024-03-20

基金项目：本文系中央高校基本科研业务费项目、中央财经大学重大研究支持计划“数字经济与数字治理研究”的阶段性成果。

作者简介：邓建鹏，中央财经大学法学院教授、博士生导师，金融科技法治研究中心主任；赵治松，中央财经大学法学院硕士研究生。

①③ Creating video from text，https://openai.com/sora.

② Video generation models as world simulators，https://openai.com/research/video-generation-models-as-world-simulators.

④ Aidan Clark，Jeff Donahue，Karen Simonyan.Adversarial video generation on complex datasets，https://doi.org/10.48550/arXiv.1907.06571；Wilson Yan，et al. Videogpt：Video Generation using Vq-vae and Transformers，https://doi.org/10.48550/arXiv.2104.10157；Jonathan Ho，et al. Imagen Video：High Definition Video Generation with Diffusion Models，https://doi.org/10.48550/arXiv.2210.02303.

在表现与内在含义，彰显其在语言模型、计算机视觉、图像生成等领域的强大功能，为视觉媒体领域注入创新力量。这一技术变革不仅在提高内容生产效率和创造力等方面具有显著影响，而且为各类创作提供了新渠道，为多领域的应用提供了广泛潜力，包括教育、电影制作、营销，等等[①]。Sora 的功能和社会价值远超短视频等行业，学术界有观点认为，Sora 解决了世界上最复杂的事——真实世界的视觉问题。事实上，其真正的意义在于依靠技术能够生成如同自然界一般复杂、有细节的视频或图像，换言之，如果能创造视频，就能创造世界[②]。因此，Sora 的技术变革为人工智能应用领域带来广泛机遇和创新空间。

相较此前的文生文、文生图等模态的大模型，以 Sora 为代表的文生视频人工智能技术具有突出特点。首先，文生视频代表人工智能模型的理解能力进一步增强，表现为模型对输入文本的理解程度不断加深，模型能够捕捉上下文关系和语境，使生成的视频更加贴近原始文本的含义。这种关键技术的突破全面提高了文生视频内容的连贯性和准确性。其次，视频效果具有高度仿真性。通过 Sora 生成的视频能够在视觉和语境上模拟真实场景，不仅可以通过文字生成新的视频，而且可以利用原有的视频、图像进行视频扩展、调整框架和角度、图像动态化等多种操作[③]。只要输入现实的图像或视频素材，模型就能生成与之相符的视频内容，为使用者提供逼真的视觉体验，表现出内容生成式人工智能模型愈发强大的图像理解和创造能力。最后，Sora 模型体现了多模态融合，将文本、音频、视觉等多模态信息进行全面整合，即将文字描述转化为同时包含视觉、听觉甚至情感表达的复杂多媒体内容，实现跨模态的信息映射和转换。因此，Sora 将在视频生成、社交媒体和娱乐等领域拥有良好的商业和应用前景。

二、新模态的法律风险

生成式人工智能模型的底层逻辑是通过互联网规模的数据培训获得文字解读和视频理解、生成能力，因此，文生视频的新模态仍存在原有大模型带来的知识产权、数据来源合法性、生成不良或违法信息、算法偏见等威胁和风险[④]。相较之前存在的法律风险，由于以 Sora 为代表的人工智能具有理解能力、视频真实性显著提升以及多模态融合与同步等特征，其对人格权、网络犯罪、社会信任等方面将带来更加严重的威胁和挑战。

（一）侵犯人格权的风险

生成式人工智能大模型依托海量训练语料和高质量的微调语料进行学习与理解[⑤]。对于文生文、文生图模态而言，训练语料的获取或网络爬虫数据更多指向著作权等知识产权、数据安全以及有价值的数据信息等[⑥]；对于文生视频模态，关注焦点应向公民的肖像权、名誉权、荣誉权等人格利益或死者的人格利益转移。

1. 侵犯肖像权等人格权

网络中显示的公民肖像包括公民本人上传的图像、视频，也包括因展示公共环境或新闻报道等合理使用情况下使用的图像、视频。文生视频技术的进步使个体的肖像可以被轻松提取和合成，进而引发侵权问题。肖像权系个人就自己的肖像是否制作、公开、使用的权利，体现了个人的尊严和价值[⑦]，公民对其自身肖像享有决定权，可以选择在何时、何地、以何种方式展示自己的肖像。公民将个人肖像上传至个人媒体或网络，并不意味着他人

① Y.X.Liu，et al.Sora：A Review on Background，Technology，Limitations，and Opportunities of Large Vision Models，https://doi.org/10.48550/arXiv.2402.17177.

②《专访王坚：Sora意义非凡，只谈对短视频行业影响是羞辱它》，https://static.nfapp.southcn.com/content/202403/10/c8676224.html。

③ Video generation models as world simulators，https://openai.com/research/video-generation-models-as-world-simulators.

④ 刘艳红：《生成式人工智能的三大安全风险及法律规制 —— 以ChatGPT为例》，《东方法学》，2023 年第 4 期；邓建鹏、朱怿成：《ChatGPT模型的法律风险及应对之策》，《新疆师范大学学报（哲学社会科学版）》，2023年第5期。

⑤ 钱力、刘熠、张智雄等：《ChatGPT的技术基础分析》，《数据分析与知识发现》，2023 年第 3 期。

⑥ 丁晓东：《论人工智能促进型的数据制度》，《中国法律评论》，2023 年第 6 期；焦和平：《人工智能创作中数据获取与利用的著作权风险及化解路径》，《当代法学》，2022 年第 4 期；刘玲霞：《数字主权安全的理论内涵、现实挑战与应对路径》，《陕西师范大学学报（哲学社会科学版）》，2024 年第 1 期。

⑦ 王泽鉴：《人格权法：法释义学、比较法、案例研究》，北京：北京大学出版社，2013 年版，第 139 页。

可以未经许可制作或使用其肖像。但文生视频新技术的出现打破了这一权利界限，使网络上的大量个人肖像可能在未经个人同意的情况下被用于各种数据分析、特征提取、视频制作，等等。特别是公务人员或明星等社会公众人物，其因信息公开、剧照等原因，具有较高的辨识度和数量较多的公开肖像，例如，ModelScope 模型曾生成美国电影明星威尔·史密斯吃意大利面的虚拟视频[①]。如果未经授权，模型通过语料库训练或网络爬取这类数据，可能侵犯他人肖像权。当大模型的理解能力和识别能力进一步提升，文生视频技术就会将这一情况转化为对公民的潜在威胁。

大模型可以根据使用者输入的图片生成相关的模拟视频，例如，阿里巴巴发布的大模型 EMO 可以通过输入单个角色的图像和声乐音频，生成具有表现力的面部表情和各种头部姿势的视频[②]。此类模型不仅仅对辨识度较高的公众人物产生威胁，而且对普通公民具有潜在的风险隐患，如果第三人恶意使用他人肖像生成相关视频，极有可能对个人的肖像权等人格利益造成侵权。此时，大模型背后的研发机构可能负有相关的道德和法律义务，如果大模型对所有输入的图片不经审查核实便按指令生成相关视频，模型将成为违法者的“帮凶”。

2. 侵害名誉权和荣誉权

文生视频技术的误导性使用可能对个体人格名誉、荣誉构成严重威胁。首先，此类大模型技术可以将真实的人物置于虚构环境中，或将个体的脸部肖像嵌入他人身体，即“换脸”。随着以 Sora 为代表的大模型技术的涌现，深度合成和“换脸”技术将更加难以被人们识别。未来，普通使用者只需输入图片或文字指令，就可以生成真实感极强的视频，视频内容可能包括虚构事件、活动或社交场景，形成对个体真实性格和行为的误导性内容。这种误导性内容会通过社交媒体、视频平台或其他在线渠道传播，迅速为广大受众知晓。大众在未经辨别的情况下轻信虚假信息，进而形成对特定个体的错误看法，导致大众对其产生误解，错误地评价其品德、行为或社会地位，损害了个体人格和名誉。制作误导性视频内容成为有意攻击或中伤他人的手段：通过将他人的肖像置于虚构的、可能被侮辱、羞辱的情境中，或现实世界中某一特定个体在视频中的内容表达并非其本意，引发大众对个体的恶意攻击、曲解、侮辱或诽谤，使被害人百口莫辩。

3. 侵犯死者的人格利益

近年来，随着人工智能技术飞速迭代，人们可以利用人工智能寄托自身情感。Sora 和 EMO 等具有高度仿真的大模型能够以数字化方式“复活”逝去的亲人，但此类行为也有可能引发巨大的风险隐患。人们通过模拟数字人“复活”逝去亲友寄托哀思，于情、于理、于法都不能予以谴责，但难以把握此类行为的范围边界。例如，有人利用人工智能“复活”逝世明星李某，并表示只是表达爱，不具有其他目的[③]。这类未经死者近亲属同意便“复活”死者的行为，或许会损害死者的人格利益，也可能对死者近亲属造成精神上的伤害。例如，已逝演员乔某某的父亲和某刑事案件被害人江某的母亲均表示，这类行为对其造成严重伤害，令人难以接受[④]。这引发有关死者的人格利益保护与侵犯死者近亲属权利等法律问题。我国传统法律文化高度重视对死者利益的保护，例如，《大清律例》“发塚”律条规定：“若塚先穿陷及未殡埋，而盗尸柩者，杖九十、徒二年半。开棺椁见尸者，亦绞。”[⑤]当代法律将人格利益延伸至死者[⑥]，《中华人民共和国民法典》（以下简称《民法典》）第九百九十四条明确对死者人格利益的保护。在现实生活中，数字化“复活”死者可能出于寄托哀思的目的，但也可能被用于其他非法目的，人工智能模型难以确定使用者与死者的身份关系，难以判断使用者的主观意图，即使使用者主观上具有善意，但也可能对死者近亲属造成精神上的伤害。

（二）助长网络犯罪

人工智能模型除自身技术特点和缺陷带来的风

① 《更乱了！已经有真人视频冒充Sora了，威尔·史密斯吃意大利面玩梗》，https://weibo.com/ttarticle/p/show?id=2309405003525233967143#_loginLayer_1710486630676。

② EMO：Emote Portrait Alive—Generating Expressive Portrait Videos with Audio2Video Diffusion Model under Weak Conditions，https://humanaigc.github.io/emote-portrait-alive/.

③ 《AI复活李玟博主回应侵权质疑》，https://k.sina.com.cn/article_2104693617_m7d7313710330199wh.html。

④ 《乔任梁父亲回应“儿子被AI复活”：尽快下架！》，https://hqtime.huanqiu.com/article/4H0zRg5xq7l。

⑤ 《大清律例》，田涛等点校，北京：法律出版社，1998年版，第408-410页。

⑥ 王泽鉴：《人格权法：法释义学、比较法、案例研究》，北京：北京大学出版社，2013年版，第250页。

险外，还存在人为原因造成的滥用情况。文生视频等模型在刑事风险领域除文生文、文生图带来的侵犯著作权罪、假冒专利罪、侵犯商业秘密罪、侵犯公民个人信息罪、侵犯知识产权罪、传授犯罪方法罪外[①]，以 Sora 为代表的视频高度仿真技术还可能在某些领域改变犯罪的形态或成为重要犯罪工具，助长各类网络犯罪。

1. 改变部分犯罪的形态

Sora 在生成高度仿真视频内容方面取得显著进展，使人们难以区分生成视频与真实拍摄的视频，为网络犯罪提供了新渠道，部分传统犯罪行为可能出现新形态，值得高度重视。例如，《中华人民共和国刑法》第二百九十一条之一第二款规定的“编造、故意传播虚假信息罪”等，传统的虚假信息指可能造成恐慌的谣言或虚假新闻，但 Sora 等文生视频大模型可以直接把虚假文字信息转化为看似“真实”的视频。通过伪造高度仿真的视频传播违法和有害信息，包括宣传恐怖主义、极端主义内容等，影响社会稳定，甚至威胁国家安全。此外，“深度伪造”技术的起源与色情信息业密切相关，部分深度伪造者关注某些阴暗面，以便制作与名人相关的色情视频[②]，若无严格规制，文生视频等大模型的广泛应用将使色情信息或其他违法内容更易传播。这些模型能够生成逼真的虚构场景，使用者可能采取各种手段规避传统的过滤和审查机制，制作淫秽、色情内容，构成侮辱罪与传播淫秽物品牟利罪等多种犯罪，使犯罪形态更为复杂。

2. 成为网络犯罪的工具

高度仿真性的文生视频模型可能成为恶意制作与传播诈骗、色情、恐怖主义等内容的网络犯罪工具。生成式人工智能具有极强的创造性、仿真性和互动性，不法分子可以利用其输出内容实施极具迷惑性的犯罪行为[③]。Sora 在制作社交媒体视频方面特别是在短视频内容生成上具有优势，能够模拟真实人物的外貌、语言以及特定场景、背景，其生成的虚构视频可能被用于欺骗、虚假宣传甚至具有煽动性，导致个人信息泄露、财务损失，对个体和社会造成危害。虽然研发者对各类生成式人工智能设置了禁令规则，但现实中屡屡发生使用者规避禁令或无视禁令的事件，使其生成违法信息。随着生成式人工智能技术的不断成熟，开源模型的技术越来越易于获取，文生视频类大模型易沦为网络犯罪工具。

（三）降低社会信任

文生视频的仿真性和虚拟性除可能侵害个人权益、助长网络犯罪外，因其易用便捷性和技术迅速普及受到自媒体以及商业机构的青睐。人工智能的出现在一定程度上影响了人们对客观世界的了解[④]，虽然以 Sora 为代表的文生视频类大模型可以将人工智能对世界的理解与认知投射到现实，但这种投射受使用者意念的控制，因此，高度仿真性的背后存在社会信任降低的风险。

1. 社交媒体的滥用

随着移动短视频受众数量的急剧上升，国内的抖音、微信公众号以及国外社交媒体 Meta、YouTube 等平台正在成为虚假视频的重灾区。Sora 不仅使视频生成方式、创新内容及制作成本大幅降低，而且加速了视频内容的生产节奏，尤其是在全民短视频时代，个体影视创作者能够快速将创意转化为视觉内容[⑤]。首先，自媒体、博主等在追逐高点击率和高关注度的过程中，可能倾向使用 Sora 制作更“吸睛”的虚构视频内容。这将导致用户难以辨别虚构与真实，使社交媒体上信息的真实性和可信度受到挑战，社会信任也因此受到影响。其次，自媒体对此类模型的滥用可能导致虚假视频信息在社交媒体平台蔓延。自媒体可能故意误导公众，传播虚假观点和虚构事件，影响社会舆论，不仅损害社交媒体平台的信息生态，而且会引发公众对真实信息的质疑，降低社会信任度，使人们对社交媒体上传播的信息产生更强质疑和不信任感。

2. 不正当商业利用

文生视频类大模型生成的视频内容高度逼真，

① 侯跃伟：《生成式人工智能的刑事风险与前瞻治理》，《河北法学》，2024 年第 2 期。

② 蔡士林：《“深度伪造”的技术逻辑与法律变革》，《政法论丛》，2020 年第 3 期。

③ 盛浩：《生成式人工智能的犯罪风险及刑法规制》，《西南政法大学学报》，2023 年第 4 期。

④ 任晓明、林艺霏：《人工智能视野下的知识体系修正理论》，《陕西师范大学学报（哲学社会科学版）》，2022 年第 2 期。

⑤ 令小雄、王鼎民、唐铭悦：《ChatGPT到Sora：Sora文生视频大模型对影视创作的机遇、风险及矫治》，《新疆师范大学学报（哲学社会科学版）》，https://doi.org/10.14100/j.cnki.65-1039/g4.20240318.001。

使虚构广告和产品演示更具欺骗性。首先，部分商家或广告商可以利用 Sora 或类似技术生成虚构的广告内容，通过模拟产品的性能、效果吸引消费者，这一行为可能扰乱正常的信息传播秩序。例如，现阶段的各类短视频平台利用短视频吸引流量，继而直播带货。这类运营模式下产生的“信息茧房”会损害消费者的合法权益，并对商业环境产生负面影响。其次，不正当的商业利用会加剧混淆真假商业信息。企业借助文生视频技术制作误导性内容，歪曲竞争对手形象，扭曲市场竞争规则，降低投资者、消费者对企业的信任，冲击市场秩序，损害企业声誉。这不仅破坏商业竞争规则，导致市场不正当竞争，损害商业信任，而且会对营商环境造成重大不良影响。

3. 造成群体性歧视

生成式人工智能运用基于人类反馈的强化学习技术，具有填补算力缺失的现实意义，但也会使开发者将自身价值偏好输出给机器[①]，而机器并不能对开发者附加的价值倾向进行客观判断和筛选。有学者将人工智能的偏见总结为系统偏见、统计和计算偏见以及人类偏见三类[②]。在文生图类大模型测试过程中，出现过男女性别偏见问题[③]，而文生视频内容的高度仿真性在一定程度上扩大了系统偏见带来的影响。通过将反映真实世界群体的肖像置于虚构环境中，创造出虚假的场景或事件，可能使观众误解该群体的文化、行为和价值观，加深对该群体的刻板印象。例如，通过制作虚构的视频将某个群体描绘成暴力分子、懒惰者或其他负面形象，可能导致观众对该群体产生偏见。文生视频技术的滥用，还可能以具象而生动的方式强化社会既有的分化和对立。通过刻意编辑和篡改群体形象，制作具有歧视性的虚构内容，可能引发社会不同群体间的紧张关系，滋生明显的社会隔离、加剧群体间的对立、降低社会凝聚力，背离社会伦理道德准则。

三、法律风险的三维规制

Sora 等文生视频类大模型可能威胁个人权益、刑事安全以及社会信任和稳定。面对不同维度的风险隐患，应结合技术机理，分别从个人权益保护、刑法适用归责、规范监管等维度出发，加强有效应对，将前沿科技的法律风险和潜在威胁置于可控范围。

（一）加强对人格权的民法保护

1. 个人信息使用的明确授权

在训练语料选择方面，生成式人工智能的语料通常源于正版语料库与网络爬取等，包括网页、博客和社交媒体，等等[④]。从公共利益或社会福利视角看，收集更多数据、进一步加强人工智能的预测能力有助于大幅提高科技增长效率。但为模型训练而采集、处理个人信息不属于《民法典》第九百九十九条规定的合理使用范围，若不对其加以限制，将人格利益让步于科技与经济，不仅意味着现代宪法体制的退让，而且可能损害个人自由，丧失人格尊严[⑤]。

在确保 Sora 等文生视频类大模型对人格权的保护方面，首先要关注训练语料中使用的个人肖像权等人格权益是否有明确授权。无论大模型训练使用的正版语料库还是网络爬取数据，当商业化运用涉及个人肖像权等个人权益数据时，应得到明确授权与同意，同时授权应采取清晰明确的方式，并允许被使用者撤销授权。文生视频类大模型涉及个人肖像权，EMO 等模型可容纳各种语言的口语音频，虽然关于声音究竟是具体人格权、一般人格权还是特殊的法定的人格利益仍存在争论，但声音作为法定的特殊人格利益，应采取法定人格利益的保护方式保护声音权益[⑥]。随着科技

① 喻国明、苏健威：《生成式人工智能浪潮下的传播革命与媒介生态——从ChatGPT到全面智能化时代的未来》，《新疆师范大学学报（哲学社会科学版）》，2023年第5期。

② Reva Schwartz，Apostol Vassilev，Kristen Greene，et al. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence，NIST Special Publication1270，2022：6-9.

③ Nestor Maslej，Loredana Fattorini，Erik Brynjolfsson，et al. The AI Index 2023 Annual Report，Institute for Human-Centered AI，Stanford University，2023：8-29.

④ Pablo Villalobos，et al. Will we run out of data? An analysis of the limits of scaling datasets in Machine Learning，https://doi.org/10.48550/arXiv.2211.04325.

⑤ 季卫东：《数据、隐私以及人工智能时代的宪法创新》，《南大法学》，2020年第1期。

⑥ 王利明：《论声音权益的法律保护模式》，《财经法学》，2024年第1期。

的发展，或许有更多个人法益被人工智能采集利用，要明确隐私权和个人信息权益虽存在诸多相同之处，但二者是《民法典》确认的不同的人格权益，具有诸多差异[①]，不能因公民将个人信息发布在网络上，就视为默许任意采集使用。法律和各类企业规范都应结合人工智能特征与具体场景，在个人信息的存储、使用、加工、传输、提供、公开等环节强化人工智能企业的主体责任和社会义务[②]。

2. 强化数据采集和视频内容的监管

语料库和网络数据的采集监管旨在建立合理、多元、无歧视的语料库，以降低模型产出群体性歧视的风险。首先，应确保训练语料来源广泛、多元化，涵盖不同文化、社会背景的语言表达，有助于防止模型偏向性，进而提高其适应性和文化敏感度，降低对人格权的潜在侵害，避免生成带有侮辱性、诽谤性的视频内容。其次，应实施有效的数据筛选机制，排除包含有害、歧视性或不道德内容的视频。可以通过自动化反馈模型和人工标注审核相结合的筛选方式[③]，确保训练语料和学习过程的纯净性，减少对不当行为的学习。最后，在模型运营过程中，应持续加强对视频的对抗和监测。例如，OpenAI 的安全报告表明，研发团队将通过对抗性地测试 Sora 模型，构建工具帮助检测误导性内容，并设置文本分类器检查并拒绝违反使用政策的文本输入提示，包括极端暴力、性内容、仇恨图像、名人肖像或使用他人 IP 的文本输入[④]。这类企业内部规制内容带有成为行业自律章程的价值，当条件成熟时，可将其上升为针对此类人工智能的法律规则的参考原则。因此，将数据采集和视频内容监管作为研发和维护的重要内容，能够在一定程度上保证模型的持续健康输出。

3. 严格技术研发和使用资格准入与监管

与普通大模型处理文字、文档、图片功能的应用不同，文生视频类大模型的诸多应用通常直接指向公民个人的肖像权等人格权益或涉及死者的人格利益。为保障个体的人格权，可以引入运行资格的准入或备案，以加强对可能侵害公民人格权的商业机构运营行为的监管。在技术开发层面，《互联网信息服务深度合成管理规定》第十五条、第十七条规定，生成或编辑人脸、人声等生物识别信息的，应依法自行或委托专业机构开展安全评估，同时，应在生成或编辑的信息内容的合理位置、区域进行显著标识。文生视频带来的风险可能会对这一标准提出更高要求，应通过更加严格的技术资格准入，强化企业在技术研发时自查技术风险的力度。

在技术应用层面，各类互联网平台需承担积极的主体责任，确立平台规范，填补监管漏洞，同时，保障不确定性法律概念、倡导性条款的有效适用[⑤]。对于网络平台，允许平台内经营者提供利用人工智能技术数字化“复活”他人的服务，作为平台经营者应承担相关技术来源、运营和监管责任，如果相关平台内经营者存在侵权行为，在自身监管范围内要承担连带责任。人工智能具有强大的通用性，其应用范围和具体方式可能远超研发者本意，需要特定使用资格的准入和应用层面的规范，预防潜在风险。例如，欧盟《人工智能法》规定进口商、分销商、部署者等不同主体在人工智能应用、部署、商业使用等方面的原则性义务[⑥]。同时，用户应自我培育数智素养，能够鉴别有价值的信息，避免恶意信息和虚假信息的变相诱导[⑦]。还应强化平台的审查责任、明确使用者的道德和法律义务界限以及培养公民的个人权利意识，三种措施协同推进，互促互进[⑧]。

（二）优化刑法适用与归责

1. 完善刑法的解释与适用

近十余年来，为应对科技风险，《中华人民共和国刑法》增设了诸如破坏计算机信息系统罪、帮助信息网络犯罪活动罪等罪名。随着人工智能时代的到来，有学者主张增设滥用人工智能罪、人工智

① 程啸：《个人信息向数据互联发展中的法律问题研究》，《政治与法律》，2020 年第 8 期。

② 王锡锌：《个人信息国家保护义务及展开》，《中国法学》，2021 年第 1 期。

③ 夏润泽、李丕绩：《ChatGPT 大模型技术发展与应用》，《数据采集与处理》，2023 年第 5 期。

④ Creating video from text，https://openai.com/sora.

⑤ 刘权：《论互联网平台的主体责任》，《华东政法大学学报》，2022 年第 5 期。

⑥ Artificial Intelligence Act，第二十三条、第二十四条、第二十六条，https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html。

⑦ 孙那、鲍一鸣：《生成式人工智能的科技安全风险与防范》，《陕西师范大学学报（哲学社会科学版）》，2024 年第 1 期。

⑧ 王禄生：《论“深度伪造”智能技术的一体化规制》，《东方法学》，2019 年第 6 期。

能事故罪，以应对人工智能的安全与风险问题[①]。但漫长的立法周期与科技的极速迭代存在矛盾。面对科技快速发展带来的犯罪形态变化，应通过不断优化法律解释与适用，应对科技迅速发展带来的犯罪形态变化[②]，适应不断发展的人工智能技术带来的刑事犯罪问题，确保法律能够有效覆盖 Sora 等文生视频类大模型带来的风险，防范可能存在的法律漏洞。同时，面对人工智能应用的普及与网络信息的快速传播，应着力促进国际法律的监管与合作；为减少网络犯罪和模型滥用问题，应建立跨界信息共享机制，提高对违规行为的识别和处理效率，共同应对人工智能带来的刑事风险。

2. 完善刑事责任制度

人工智能不完全等同于技术工具，其在设计之初就具有鲜明的社会与政治性格[③]。如果 Sora 等文生视频类大模型以技术中立为由免责，无疑会成为犯罪分子青睐的犯罪工具，出现将其广泛应用于诈骗、制作色情、虚假信息和恐怖信息等犯罪行为。面对大模型等犯罪工具智能化时代的到来，应健全人工智能作为犯罪工具的刑事责任制度，区分行为人故意利用人工智能实施犯罪行为、人工智能研发者和使用者的过失责任等情况[④]。核心技术开发团队也应受到关注，因其塑造了人工智能的运行逻辑、价值偏好和应用功能[⑤]。将模型的研发者、推广者和使用者纳入刑法规制的责任体系范围，能够有效规制大模型避免成为网络犯罪的工具。此外，应对文生视频和真人模拟等功能建立严格的审查机制、限制应用场景，确保 Sora 等文生视频类大模型的应用符合法律与道德伦理。通过政府监管审查、定期技术审查以及研发机构的公开报告，有效监测模型在实际应用中的潜在风险，避免模型的非正常使用。

（三）规范监管以提升社会信任

1. 加强研发机构和平台的合规建设

在短视频风靡的时代，制作精美、细致的虚假视频内容可能会以惊人的速度广泛传播。对此，社交媒体平台应建立强有力的模型应用审查机制，包括网络应用服务平台自身的监管职责，确保其使用的大模型如 Sora 等符合法律和行业标准。随着应用的普及与技术的开源，人工智能生成视频逐渐普及，不能仅依靠政府审查视频内容，强化平台对视频内容的审查义务是数字时代的必然要求。Sora 等文生视频类大模型呈现新技术、新模态和新模式，但法律规则时常滞后甚至缺位，因此，平台应积极承担主体责任，弥补数字时代法律治理的缺陷[⑥]与监管资源的不足。例如，抖音平台规定人工智能技术生成的数字化形象必须备案，在举报反馈选项中增加“AI 生成内容问题”[⑦]。Sora 等文生视频类大模型研发机构可通过添加水印等方式对人工智能生成视频内容设置明显标记或“人工智能生成”的风险提示。例如，OpenAI 的使用政策要求模拟自然人的产品必须获得本人的明确同意或明确表示为人工智能生成[⑧]，防止用户对上传的虚拟视频产生误解。除社交平台对上传、发布的视频具有审查义务外，广告媒体、商业使用等平台也应建立健全合规审查，确保建立良好的公众信任和公平竞争的商业环境。无论社交媒体平台还是人工智能视频应用商家，均需确保其运营合规合法。

在合规领域，一方面，监管机构应重点督促网络平台与 Sora 等文生视频类大模型研发机构进行合规培训，确保平台上传的视频内容遵守法律法规，减少违法违规行为的发生；另一方面，监管机构可以促进相关商业机构在现行《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》《互联网信息服务算法推荐管理规定》等规范体系下的管理导向型合规，并结合目前 Sora、EMO 等文生视频类大模型的法律风险，健全合规风险防控的风险导向型合规[⑨]，促进企业合规建设。

2. 从规制对象到规制工具的转变

除事后惩治外，应通过事前监管预防，缓解人

① 刘宪权：《人工智能的安全与风险问题》，《比较法研究》，2018 年第 4 期。
② 刘艳红：《Web3.0 时代网络犯罪的代际特征及刑法应对》，《环球法律评论》，2020 年第 5 期。
③ 余盛峰：《临界：人工智能时代的全球立法变迁》，北京：清华大学出版社，2023 年版，第 2 页。
④ 刘宪权：《人工智能时代的刑事风险与刑法应对》，《法商研究》，2018 年第 1 期。
⑤ 邓建鹏：《区块链的法学视野：问题与路径》，《学术论坛》，2023 年第 3 期。
⑥ 刘权：《论互联网平台的主体责任》，《华东政法大学学报》，2022 年第 5 期。
⑦《抖音社区自律公约》，https://www.douyin.com/rule/policy?activeId=self_decipline#heading-3。
⑧ Building with the OpenAI API platform，3（b，d），https://openai.com/policies/usage-policies.
⑨ 陈瑞华：《企业合规管理体系建设的两种模式》，《法学评论》，2024 年第 1 期。

工智能生成虚假视频冲击社会信任的问题。人工智能研发机构与社交网络平台开设面向公众的投诉举报途径，公众发现违法或不良内容后，向研发机构或平台投诉举报，研发机构或平台应删除违法或不良视频内容。各类社交平台都应积极引入用户参与虚假信息反馈机制，在用户层面打通模型反馈机制的渠道，方便用户报告虚假信息。此外，研发机构和平台可引入用户激励机制，鼓励用户参与虚假信息的反馈。

在系统应用层面，模型的研发机构、社交媒体和商业应用平台应在发展和引入技术的同时，发展智能化的虚假信息检测系统，在代码中适当嵌入规范要求，提高监测效率和准确性。大模型系统自身具有较强的学习能力，通过使用自然语言处理和机器学习算法等反馈模型，能够快速识别和过滤虚假信息，因此，也可以利用数据信息的匿名化采集和实时监控，侦测危险或虚假信息，防范虚假信息对社会造成的不良影响。有学者认为，法律规则通常以事后惩罚、赔偿或恢复性措施支持其规定的要求，技术管理的重点是在事前，旨在预测和防止不当行为，通过将规范嵌入架构，可显著增加规范的事前应用，并相应减少对规范事后应用的依赖[①]。

据此，研发机构和网络平台应依据法律规范的基本要求，建立严格的审核和过滤系统，通过自动化的反馈模型和人工审核，检查生成的视频内容，识别和处理可能违反规定的信息，确保发布的视频内容合法合规。结合人工智能的主要应用场景，设立独立的审查机构或评价机构，由专业人员和具有代表性的使用者组成，提供专业判断，负责审查模型生成的视频内容是否符合相关准则，从而使模型输出可以更加主动地纠正潜在偏向，提高生成视频内容的合规性。通过上述方法，有助于使基于规则的秩序转向基于技术管理的秩序，将对不法行为的纠正和惩罚转向对不法行为的预防和排除，将依赖规则和标准转向采用技术解决方案[②]。这一转变过程将使监管者认识到，法律可以将人工智能作为规制对象，也可以将之作为有效的规制工具。

3. 推进人工智能的透明和公正应用

在 Sora 等文生视频类大模型的技术层面，有效规制是确保模型设计与实现符合伦理道德、公平和透明原则的关键。技术设计应融入伦理规则，只有坚持科技向善，采取复合性措施矫正算法歧视，才能在数据观念改变的前提下重构伦理体系[③]，给予公民人格权充分的尊重与保障。通过法律、行政法规、规章以及行业标准等明确规定模型训练过程中应遵循的法规与道德伦理准则，并将相应准则贯穿数据获取、训练、应用和监管全过程，从源头规范文生视频类大模型的行为，保证其公正性。

透明度和可解释性是增强社会信任的关键，以 Sora 为代表的新兴生成式人工智能同样存在算法披露和解释不足等问题，容易显现“算法黑箱”弊端，长此以往，在算法偏见的冲击下将出现算法信任危机[④]。有学者指出，有的研发机构并不关心人工智能的安全问题，也无意作出任何限制人工智能权力范围的规定。因此，要确保人工智能与人类保持紧密的配合与协作[⑤]。出于对技术和商业秘密的保护，不应强制各类人工智能模型开源和进行完全透明的解释，但模型开源可以成为增强人们信任和有效监管的途径。例如，2024 年 3 月 17 日，x.AI 公司开源了人工智能模型 Grok[⑥]，为行业树立了良好示范。

模型开源后，更多开发者和使用者将深度参与其中，可以发挥人工智能模型被众人监督的效果，模型的决策过程也可以在一定程度上被解释和理解，有助于公众特别是专业人士监督和纠正模型存在的潜在偏见。尽管监管规则不能强制要求所有类型的模型开源，但面对文生视频新模态的发展，要根据创新发展的情况及时作出调整。在不同的应用层面和使用方式中，实施相互转换禁止类、允许类以及提倡类策略[⑦]，法律和监管政策可以通过合规指导、合规强制、合规激励等综

①② [英]罗杰·布郎斯沃德：《法律 3.0：规则、规制和技术》，毛海栋译，北京：北京大学出版社，2023 年版，第 42、53 页。

③ 梅傲：《积极伦理观下算法歧视治理模式的革新》，《政治与法律》，2024 年第 2 期。

④ 安晋城：《算法透明层次论》，《法学研究》，2023 年第 2 期。

⑤ [美]沃尔特·艾萨克森：《埃隆·马斯克传》，孙思远、刘家琦译，北京：中信出版社，2023 年版，第 227 页。

⑥ Open Release of Grok-1，https://x.ai/blog/grok-os.

⑦ 王首杰：《创新规制的时间逻辑》，《华东政法大学学报》，2022 年第 3 期。

合措施，向模型开源适度倾斜[①]，在提升模型公开度和透明度的同时，鼓励大众参与模型监督和治理。

四、结　语

人工智能发展的速度令人震惊，如果说 ChatGPT 的发布揭开了通用人工智能时代的序幕，那么，Sora 的推出意味着人工智能的理解能力和生成能力得到质的飞跃，Sora 在令世界震撼的同时，也对现有秩序造成潜在冲击。Sora 的研发团队表示，“Sora 模型内测后迟迟未开放使用的原因是其能力过于强大，因此还有许多的安全工作需要完成”[②]。Sora 发布后不久，Figure 发布了可同人类进行完整对话，具有高级视觉和语言智能，能够快速、简单、灵活行动的人工智能机器人 Figure 01[③]。人工智能给人类带来太多惊讶和震撼，科技更新迭代的周期越来越短。我们要拥抱科技，及时调整规制的方向和重点，消除科技隐患，让科技更好地服务人类。

针对以 Sora 为代表的文生视频类人工智能，要在个人权益、刑事犯罪、社会稳定三个维度做好相应规制策略，涉及民法、刑法及规范监管等不同层面。在应对人格权可能带来的侵害时，应重点关注对肖像权、名誉权、荣誉权等个人权益的保护以及对死者人格权益的保护与尊重，明确相关权利的授权使用和撤销，规范公民肖像等数据信息采集以及数字化“复活”死者的合规监管；回应对刑事犯罪的影响，法律应主动适应科技时代的犯罪智能化和科技化，通过法律解释和刑事责任制度的完善，有效规制人工智能时代的犯罪新形态；面对社会信任遭受冲击，在健全研发机构和平台责任体系的同时，完善对虚假信息有效的反馈和处理机制，鼓励模型的透明化解释和开源，不断推进人工智能透明化和公正化。

Legal Risk and Three-dimensional Regulation of Artificial Intelligence: From the Perspective of Sora

DENG Jian-peng[1, 2]　ZHAO Zhi-song[1]

（1.Law School；2.Fintech Rule of Law Research Center，Central University of Finance and Economics，Beijing 100081）

Abstract: Sora，an artificial intelligence model which creates text-based videos，has attracted much attention since it was released. Its understanding ability，high level of simulation and multi-modal integration have brought visual and auditory shock to the society，and have also caused many legal risks. Compared with the previous AIGC models，Sora's potential legal risks are more prominent in the protection of personality rights，cybercrime and social trust. In the face of the risks and challenges of cutting-edge technology in the personal rights，crimes and social stability，it is necessary to take corresponding regulatory countermeasures. First，it is necessary to strengthen the civil law protection of personality rights，clarify the authorization of the use of personal information，and strengthen the supervision of data collection and video content. Second，it is urgent to optimize the application of criminal law. Third，it is necessary to enhance social trust by standardizing governance，improve the false information processing and feedback mechanism，transform artificial intelligence from as regulatory object to regulation tool，and promote the security and reliability of artificial intelligence system.

Key words: Sora；Legal Risk；Three-dimensional Regulation；Artificial Intelligence；Text-to-video AI Model

[责任编辑：李　蕾]
[责任校对：潘静静]

① 陈瑞华：《行政机关推进合规管理的三种方式》，《当代法学》，2024 年第 1 期。

②《Sora 团队首次专访：Sora 就是太强了，所以不让普通人用》，https://mp.weixin.qq.com/s/xRB6K-Q2vyal8n5f4paP6Tg。

③ Figure is the first-of-its-kind AI robotics company bringing a general purpose humanoid to life，https://www.figure.ai.