

深度合成技术应用与风险应对

李婧文¹, 李雅文²

(1. 北京交通大学电子信息工程学院, 北京 100044; 2. 中国信息通信研究院政策与经济研究所, 北京 100191)

摘要: 深度合成技术被广泛应用于智能服务、音视频产品制作、媒体传播和信息服务, 已成为人工智能领域的热点应用之一。剖析了深度合成技术原理和实现方式, 分析了深度合成技术在机器人写稿、语音合成、智能客服、声音复制、人脸合成、姿态操纵、虚拟人物和虚拟场景等典型应用场景中的应用情况, 并总结了深度合成技术应用带来的生成和传播虚假信息、侵害他人合法权益、被其他违法犯罪活动利用、影响国家安全等风险。研究了欧盟、美国等采取措施开展深度合成技术应用治理的情况和我国加强深度合成应用管理的情况, 并从深度合成技术应用的智能性、拟真性、泛在性特点出发, 分析了防控深度合成技术应用风险的举措, 提出了构建全方位监管规则体系, 强化技术监督, 提高全社会风险防范意识等具体应对措施。

关键词: 深度合成; 深度伪造; 应用场景; 风险; 应对

中图分类号: TP391

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023032

Application and risk response of deep synthesis technology

LI Jingwen¹, LI Yawen²

1. School of Electronic Information Engineering, Beijing Jiaotong University, Beijing 100044, China

2. Institute of Policy and Economics, China Academy of Information and Communications Technology, Beijing 100191, China

Abstract: Deep synthesis technology is a widely used application in the field of artificial intelligence, particularly in intelligent services, audio and video production, media communication and information services. A basic research was conducted on the principle and implementation of deep synthesis technology. And the application of deep synthesis technology were analyzed, such as robot writing, voice synthesis, intelligent customer service, voice replication, face synthesis, posture manipulation, virtual characters and virtual scenes. Based on the analysis of the characteristics and development trend of the application of deep synthesis, the risks brought by the application of deep synthesis technology were summarized, such as the generation and dissemination of false information, infringement of the legitimate rights and interests of others, utilization by other illegal and criminal

收稿日期: 2023-01-16; 修回日期: 2023-04-17

通信作者: 李婧文, bjtuljw@sina.com

基金项目: 工业和信息化部电信市场开放与监管问题研究项目 (txs2022003)

Foundation Item: Research on Telecommunications Market Opening and Regulation of the Ministry of Industry and Information Technology (txs2022003)

引用格式: 李婧文, 李雅文. 深度合成技术应用与风险应对[J]. 网络与信息安全学报, 2023, 9(2): 184-190.

Citation Format: LI J W, LI Y W. Application and risk response of deep synthesis technology[J]. Chinese Journal of Network and Information Security, 2023, 9(2): 184-190.

activities, and impact on national security. The measures taken by the EU, United States and China to govern the application of deep synthesis technology were studied and the countermeasures were recommended such as building a comprehensive regulatory rule system, strengthening technical supervision, and improving risk prevention awareness across society. Furthermore, the need for intelligence, authenticity, and universality in the application of deep synthesis technology was highlighted. Based on the analysis of the characteristics and development trends of deep synthesis technology, a comprehensive insight into the potential risks and countermeasures associated with its application governance were provided.

Keywords: deep synthesis, deepfake, scenario, risk, response

0 引言

随着人工智能技术的迅猛发展和计算机算法的深入应用,深度合成技术在信息服务和信息传播中得到了广泛应用。深度合成技术提升了服务质量,提高了传输效率,增强了传播效果,不仅方便了人们生活,也提高了生产效率。同时,作为一种新型的技术,深度合成技术带来了一定的应用风险^[1],需要在积极利用的同时加强治理,防范技术应用给社会和个人带来的危害^[2]。为了规范深度合成技术应用,各国积极采取相应管理措施。2022年11月,国家互联网信息办公室、工业和信息化部、公安部联合发布《互联网信息服务深度合成管理规定》,对深度合成技术应用作出规定,将对这一热点技术应用的监管提升到新的高度。

1 深度合成技术概要

1.1 深度合成技术的含义

深度合成技术是指利用以深度学习、虚拟现实为代表的生成合成类算法制作文本、图像、音频、视频、虚拟场景等信息的技术^[3],最初源于一个名为“Deepfakes”的红迪网用户2017年年底在社交网站上发布了一段利用名人面孔合成的视频,该虚假视频在网络空间广泛传播^[4]。

深度合成技术的使用者通过使用深度学习、虚拟现实等技术手段完成“换脸”“换声”“虚拟播报”等,以实现更好的信息服务和传播效果。深度合成技术的重要实现基础是卷积神经网络(CNN, convolutional neural network),该技术具有极强的表征学习能力,能够通过仿照生物知觉皮层和视觉皮层,针对像素和音频进行深度学习,起初主要应用于医学影像。此后,随着深度学习、

虚拟现实,特别是生成对抗网络(GAN, generative adversarial network)^[5]等技术的不断发展,深度合成技术快速发展,并被广泛应用到文本、图像、音频、视频、虚拟场景的生成制作^[6]。

1.2 深度合成的技术实现

利用深度合成技术,不仅可以模拟和仿制已有的文本、图像、音频、视频,还可以制作和生成原本并不存在的文本、图像、音频、视频,不仅可以“以假乱真”,还能够“无中生有”^[7]。

在技术实现方式方面,深度合成技术依赖于深度学习模型,识别和训练既有数据提升数据精准识别和预测能力,并在此基础上生成高度仿真的合成数据,实现深度伪造和深度合成的目的。具体的技术应用包括自动编码器、生成对抗网络、变分自编码器、扩散模型等。

1) 通过自动编码器对数据进行处理,去除虚假数据、噪声数据及低质量数据,随后根据数据特征进行分类,对原始数据进行标准化处理。自动编码器用于学习对于特定网络的无标记数据的有效编码。自动编码器由编码器和解码器两部分组成,编码器需要从原始高维空间中提取有效信息,并将数据从原始高维空间映射为低维空间(隐层空间)的编码表示,而解码器确保提取出的隐层编码中所蕴含的信息量足以被恢复回原始高维空间。

2) 利用生成对抗网络进行数据合成。生成对抗网络起源于博弈论,包括生成器和鉴别器两部分,生成器负责学习并拟合真实数据的分布特征,鉴别器负责判断输入数据是真实数据还是合成数据,二者在一个极小极大的博弈游戏中相互对抗不断变强。当生成器能够由随机噪声生成足以“骗”过鉴别器的图像时,可以认为生成器较好地拟合了真实数据的分布;在后续使用中,通过采

样随机噪声并输入该生成器以生成大量高度仿真的合成数据^[8]。除了直接从随机噪声生成内容外,还可以将条件(如分类标签)作为输入加入生成器和鉴别器,使得生成结果符合条件输入的属性,让生成内容得以控制。

3) 利用变分自编码器提升数据合成质量。由于自动编码器仅关注隐层编码的重建能力,其隐层空间分布往往缺乏一定规律,在连续的隐层空间随机采样得到的一组编码通常会产生无意义和不可解释的生成结果。为了构建一个有规律的隐层空间,变分自编码器不再将输入映射成隐层空间中的一个固定编码,而是转换为对隐层空间的概率分布估计,一般假设先验分布是一个标准高斯分布。同样地,训练一个概率解码器,实现从隐层空间分布到真实数据分布的建模。当给定一

个输入,通过后验分布估计出关于分布的参数(多元高斯模型的均值和协方差),并在此分布上采样,最后通过概率解码器输出合成数据。

4) 一个标准扩散模型有两个主要的过程:正向扩散和反向扩散。在正向扩散过程中,图像被逐渐引入的噪声污染,直到图像成为完全随机噪声。在反向扩散过程中,利用一系列马尔可夫链在每个时间步逐步去除预测噪声,从而从高斯噪声中恢复数据。扩散模型最近在图像生成任务中表现出了显著的性能,并在图像合成等任务上取代了生成对抗网络。扩散模型能够产生更多样化和更细粒度的图像,且不会受到模式崩溃的影响,这是由于扩散模型保留了数据语义结构。

深度合成技术的实现过程如图1所示。

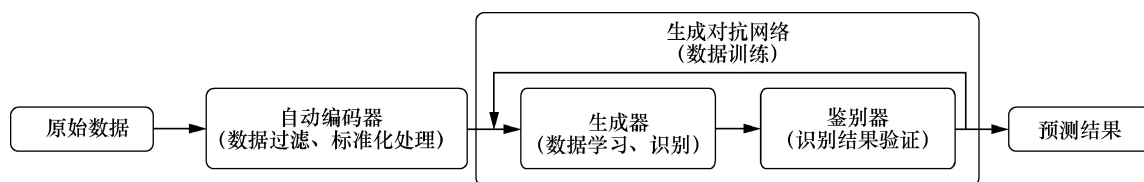


图1 深度合成技术的实现过程

Figure 1 Implementation process of deep synthesis technology

2 深度合成技术应用和应用场景

2.1 深度合成技术应用

当前深度合成技术应用主要包括以下几类。

一是篇章生成、文本风格转换、问答对话等对文本内容进行生成或者编辑。这类应用属于文本信息编辑,利用文本信息编辑可以快速生成文本信息和文本作品。

二是文本转语音、语音转换、语音属性编辑等对语音内容进行生成或者编辑的应用。这类应用能够快速实现文本信息和语音信息的相互转换,实现不同语音格式和标准之间的自动转换。这类应用被广泛用于语音、文字等编辑和后期制作^[9]。

三是音乐生成、场景声编辑等对非语音内容进行生成或者编辑的应用。利用这类应用能够编辑和生成新的音频,如生成音乐作品、场景声音。

四是人脸生成、人脸替换、人物属性编辑、人脸操控、姿态操控等对图像、视频内容中人脸等生物特征进行生成或者编辑的应用。在这类应

用中,可以借助软件程序实现生成人脸、更换人脸等“造脸”“换脸”,也可以通过重塑表情、口型和语速等实现操控人脸动作的“以假乱真”甚至“无中生有”^[10]。

五是图像增强、图像修复等对图像、视频内容中非生物特征进行编辑的应用。这类应用属于图像修复技术,能够还原、修复、再造图像。

六是三维重建等对虚拟场景、虚拟人物进行生成或者编辑的应用。三维重建是通过单视图或者多视图从二维图像中恢复物体三维结构,在计算机中对客观世界进行虚拟现实表达,这一应用可以用于虚拟空间中编辑或者生成虚拟物体、虚拟人物、虚拟场景。

由于不同应用的作用不同,在深度合成技术的具体应用场景中,使用者可能单独使用某一种技术,也可能同时使用多种合成技术,达到深度合成效果。

2.2 深度合成应用场景

从应用实践看,深度合成技术已被广泛应用于生产生活中,具体如下。

(1) 机器人写稿

使用者利用深度合成技术,同时借助数据挖掘、机器学习和自然语言处理等技术,能够自动生成和编辑新闻消息和文字作品。基于大数据分析平台,智能机器人能够在短时间内挑选出新闻点,抓取和分析相关资料,并按指令要求形成一定格式要求的新闻稿件。例如,今日头条新闻写作机器人 Xiaomingbot 在 2016 年 6 月上线后,不仅在里约奥运会期间,撰写了大量的赛事报道稿件,而且撰写了大量涵盖科技、财经、房地产等十几个类别的新闻稿件,迄今已创作两万多篇文章;2017 年,微软虚拟机器人“微软小冰”出版了《阳光失了玻璃窗》诗集;腾讯新闻机器人 Dreamwriter 能够从海量新闻资讯中全面抓取和提取分析相关数据,并在第一时间生成新闻资讯,形成新闻作品,满足了新闻信息即时传播的需要。

(2) 语音合成

深度合成技术已被广泛应用于有声阅读、资讯播报、订单播报、智能硬件、音乐和场景音制作等。首先,利用深度学习、计算机识别、分析处理等技术将给定的文本信息转化成语音形式,再通过计算机识别技术等对语音信息进行转换,并对语音的语速、音调、音量等要素进行编辑处理,合成有声读物和音乐、场景声音等音频产品。例如,百度在线语音合成软件能够通过转换软件,将文字转换为语音文件,并应用于阅读听书、资讯播报、电子图书、背景音/场景音制作等。

(3) 智能客服

智能客服机器人借助机器学习技术,通过语音识别、语义理解技术,判断用户提问意图,然后通过文语转化和知识图谱,找到相关问题的解决方案,再借助深度训练等给予用户相应解答。相关技术已应用于金融、保险、汽车、电子商务、政府政务等领域。例如,华为公司开发的 QAbot 能够在多种客服服务问答场景中自如应答,及时拦截处理高频易混淆的问题,大大地提升了问答效率和准确率,大幅度降低了客服运维的人力成本。

(4) 声音复制

声音模拟和声音复制是语音合成技术的个性化应用,操作者可通过少量的录音进行模型训练,

得到与被模拟者在音色和发音风格上相似的声音样本,然后通过语音转化和合成技术快速复刻被模拟者的声音^[11]。通过声音复刻,可以应用至播报、配音等场景中。语音合成技术可以用于制作与样本声音完全不同的声音,达到“换声”或者“变声”的效果和目的,还可以制作操作者任何需要的声音。

(5) 人脸合成

利用图像深度合成技术,可以对图像、视频中的生物特征进行生成和编辑,合成人脸图像或者替换人脸图像,实现“造脸”“换脸”的目的^[12],再利用图像增强、场景编辑等技术对图像、视频中的非生物信息进行生成或者编辑,以实现以假乱真的效果。

(6) 姿态操纵

操作者利用深度合成技术,通过改变目标对象的眉毛、眼睛、嘴部、头部位置等信息,操纵目标对象的面部特征,实现对目标对象的人脸操纵;同时,可以通过改变目标对象的身体部位位置,操纵目标对象的身体姿态,实现对目标对象的姿态操纵^[13]。

(7) 虚拟人物和虚拟场景

虚拟主播是虚拟人物的典型应用,虚拟主播综合运用多种深度合成技术,通过多模态合成技术生成虚拟形象,同时采用人脸合成、实时面部表情控制、姿态与语音驱动等技术,实现虚拟主播的互动播报。利用深度合成技术生成的虚拟人物被大量应用于新闻播报、场馆讲解、产品介绍、迎宾接待、在线教育等场景。同时,可以利用三维重建等深度合成技术,生成和编辑虚拟化的场景,用于产品宣传、影视制作^[14]。例如,2018 年 11 月,新华社首次推出虚拟主播,实现新闻信息播报,此后又进一步推出虚拟主播与其他虚拟人物对话,实现了 AI 合成主播面对面采访虚拟动漫人。

深度合成技术作为人工智能领域的新应用、新实践,成为多个行业场景中的热点应用,未来随着深度学习和虚拟现实(VR)、增强现实(AR)等新型技术和新型应用场景的发展,将催生出更多应用和业态,成为信息技术发展应用的重要方向之一。

3 深度合成技术应用特点与发展趋势

作为人工智能领域的新应用,深度合成技术已广泛应用于智能服务、音视频产品制作、媒体传播和信息服务等业务场景和业务活动。

3.1 深度合成技术应用特点

从深度合成技术在不同场景中的具体应用情况看,深度合成技术应用最显著的特征是它的智能性、拟真性和泛在性。

(1) 智能性

深度合成技术基于已有算法模型自动生成产品,用户只需提供相应数据和信息,就可以通过算法软件自动生成所需要的产品,智能化、自动化程度高,合成过程中需要人工干预的环节少,一些应用甚至不需要人工干预就能完成,这为技术的深入推广和广泛应用提供了便利。

(2) 拟真性

深度合成综合利用深度学习等多种技术生成产品,真实程度高,拟真性强,辨别难度大。美国纽约大学和密歇根州立大学通过 AI 系统 DeepMasterPrints 复制了 23% 的指纹数据,通过了扫描仪测验,系统错误率仅为 1/1000^[15]。深度合成技术的这一特点,使得它能在不同的场景中实现以假乱真的效果。

(3) 泛在性

深度合成技术可以加工处理文本、图像、音频、视频等信息,可广泛应用于多种场景和活动,应用丰富、领域宽、范围广。为普及技术应用,深度合成可采用无监督学习模式,其数据集无须标定,自动进行特征归并,节约了人力与物力成本,极大降低了应用门槛。

3.2 深度合成技术的发展趋势

深度合成技术一经推出就得到高度的社会认可度,具有广泛的应用前景。从未来发展趋势看,深度合成技术和应用将处于高速增长阶段,在技术成熟度和复杂性、应用深度与广度、对社会产生的影响力等方面将进一步加大。

2022 年 2 月,《深度合成十大趋势(2022)报告》提到,深度合成技术在内容制作和传播数量方面将呈高速增长态势,未来需求场景将多元化并逐渐趋于成熟,商业化产品类型将日益丰富,

但随着技术和应用的演进和深入,深度合成识别的难度将进一步增大,深度合成内容产生的负面影响将持续加剧并对个人和社会产生实质性危害。

4 深度合成技术应用的风险与应对

深度合成技术应用方便了人们的生产生活,提高了生产的效率,同时带来的风险也引起了人们的关注,一些国家纷纷采取措施,防范深度合成技术应用带来的风险。

4.1 深度合成技术应用存在的风险

深度合成技术尽管发展迅猛,但其带来的风险才刚刚开始。受深度合成技术自身具有的复合性强、技术迭代快、透明性差等因素的影响,在互联网平台的聚合放大效应下,深度合成技术的应用带来了生成和传播虚假信息、侵犯他人合法权益、被违法犯罪活动利用甚至直接危害国家安全等风险^[16]。

(1) 生成和传播虚假信息

机器人写稿、虚拟主播等能够提高信息传播速度、提升传播效率和质量,但如果使用不当将可能产生虚假信息并快速和大面积传播虚假信息^[17]。虚假信息的广泛传播,不仅可能误导公众,降低数据信息的可信度^[18],还可能破坏传播秩序,导致社会混乱,甚至扰乱社会公共秩序^[19]。

(2) 侵犯他人合法权益

“换声”“换脸”“姿态操控”等应用可能在当事人没有授权的情况下非法使用他人肖像,甚至可能故意丑化或者恶搞当事人形象,侵犯他人的肖像权、名誉权、隐私权等合法权益。另外,这些应用也可能被用于非法收集和使用他人信息、非法攫取大众数据,侵犯他人个人信息和隐私权^[20]。技术的不当使用,使其成为侵犯他人合法权益的工具。

(3) 被其他违法犯罪活动利用

“声音复制”“换脸”等可能被不法分子用于人脸识别、身份识别等场景,冒充身份后为造假、诈骗等违法犯罪活动提供便利^[21],成为违法犯罪活动的工具^[22]。

(4) 对国家安全产生影响

深度合成技术产生的虚假舆情信息会干扰公

众对政府决策的判断,降低政府的权威性^[23],降低政府部门对公共事件的处理能力,对维护国家安全产生严重影响。2020年,澳大利亚和美国分别发布的《深度伪造武器化——国家安全与民主》和《深度伪造:基本威胁评估》报告提出,深度伪造技术降低了国家之间信息战的成本,恶意行为主体会利用深度伪造技术编造虚假信息发动舆论战,影响国家安全。

4.2 深度合成技术应用治理的措施

深度合成技术带来的风险引起了各国的广泛关注,各国积极完善管理制度,采取管理措施,规范深度合成技术应用活动,防范其给社会和个人带来的风险^[24]。

欧美等国家与地区率先意识到深度合成技术应用带来的风险,通过法律法规制定等方式加强对深度伪造等技术和应用的管理,防范深度合成技术应用给个人和社会带来的风险。美国国会提出《2018年恶意伪造禁令法案》等,尝试从法律层面上规范深度伪造等深度合成技术应用^[25]。欧盟委员会制定发布《应对线上虚假信息:欧洲方案》《反虚假信息行为准则》等文件^[26],将利用深度合成技术制作不实信息纳入虚假信息打击范围,防范虚假信息的生成和传播^[27]。我国有关部门先后制定《网络音视频信息服务管理规定》《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》,对包括生成合成类算法在内的算法技术应用活动进行规范。同时,国家互联网信息办公室等部门开展“清朗·2022年算法综合治理”等专项行动,打击违法算法应用行为,推进深度合成等算法应用的规范化和透明化。

4.3 应对深度合成技术的建议

深度合成技术给社会发展带来的风险需要全社会共同努力去面对,尤其需要充分发挥管理部门、网络平台、网络用户等各方的作用。

(1) 构建全方位监管规则体系

以对深度合成技术应用全过程管理为目标,在已有国家立法的基础上,涵盖深度合成技术的数据收集、数据处理、算法运行等环节的全过程各方的使用行为,制定一系列可操作性强的相关标准和规范,完善具有操作性的管理要求和合规

指引,形成包括法律法规、标准规范、政策指引等全方位监管规则体系,明确各方的权责,实现对深度合成技术的引导促进和监督管理,促进深度合成技术应用规范有序发展。

(2) 强化技术监管

深度合成技术应用带有鲜明的专业性、时效性、快速迭代性等特征,难以单纯依靠人工实现管理,应进一步强化技术监管,防范深度合成技术应用带来的风险,具体建议如下。一是建立开发者承诺备案机制。新应用上线前应先向主管部门备案,将备案内容在对应的网络平台上予以公布,并承担相应的法律责任;同时,新应用应提供举报功能,应将举报信息及时推送给对应的网络平台和主管部门指定的接收者。二是压实平台主体责任。网络平台应在新应用上线时对其备案的承诺与应用的一致性进行复查,并签订责任协议;网络平台应接收用户投诉并处理,及时将投诉与处理结果上报给主管部门。三是提升第三方评测机构的技术鉴别能力。对投诉和处理结果有争议时,将由主管部门认证的第三方测评机构进行技术鉴定,给出专业性的鉴定结论。四是提高主管部门对深度合成技术全社会应用的监测与处置能力。主管部门接收网络平台的投诉与处理结果,基于此对违规事件进行研判和及时处置,并对第三方监测机构进行认证和指导管理。

(3) 提高全社会风险防范意识

深度合成技术应用面广、隐蔽性强、潜在风险大,防范深度合成技术带来的风险需要提高广大互联网用户对深度合成技术应用的识别能力和风险防范能力。具体建议如下:一是典型案例识别方法的科普宣传,提高用户识别能力;二是广泛宣传投诉违规应用的公益义务。只有广大用户积极参与,才能对开发者、网络平台等形成有效及时的监督。

5 结束语

与其他技术应用一样,深度合成技术同样具有两面性,为社会经济发展和生产生活提供便利的同时也带来一定的风险。如何对深度合成技术应用带来的风险进行有效监管,应对技术发展带来的挑战,需要认真研究新技术、新应用的特点

和未来发展趋势,并随着技术的发展演进不断调整和完善政策措施。

参考文献:

- [1] ELIZABETH C. Reject the evidence of your eyes and ears: deepfakes and the law of virtual replicants[J]. Seton Hall Law Review, 2019(50): 177-186.
- [2] QAZIEU H, ZIA T, ALMORJAN A. Deep learning-based digital image forgery detection system[J]. Applied Sciences, 2022, 12(6): 2851.
- [3] JUDGE H B, DIXON J R. Deepfakes: more frightening than photoshop on steroids[J]. Judges' Journal, 2019, 58(3): 35.
- [4] 李明鲁. “深度伪造”的刑法治理路径[J]. 科技与法律, 2021(6): 40-47.
LI M L. The criminal law governance path of "deep forgery" [J]. Science Technology and Law, 2021 (6): 40-47.
- [5] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the Annual Conference on Neural Information Processing Systems 2014. 2014: 2672-2680.
- [6] GUEAD, DELP E J. Deepfake video detection using recurrent neural networks[C]//Proceedings of 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS). 2018.
- [7] JESSICA I. Defamatory political deepfakes and the first amendment[J]. Case Western Reserve Law Review, 2019(70): 427-438.
- [8] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.
- [9] PURINGTON A, TAFT J G, SANNON S, et al. Alexa is my new bff: social roles, user satisfaction, and personification of the amazon echo[C]//Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017: 2853-2859.
- [10] CHAWLA R. Deepfakes: how a pervert shook the world[J]. International Journal of Advance Research and Development, 2019, 4(6): 4-8.
- [11] ZHOU Y, LIM S N. Joint audio-visual deepfake detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 14800-14809.
- [12] ZOBAAED S, RABBY F, HOSSAIN I, et al. Deepfakes: detecting forged and synthetic media content using machine learning[M]//Artificial Intelligence in Cyber Security: Impact and Implications. Springer, Cham, 2021: 177-201.
- [13] 罗常伟, 於俊, 汪增福. 视频驱动人脸动画合成[J]. 自动化学报, 2014, 40(10): 2245-2252.
LUO C W, YU J, WANG Z F. Synthesizing performance driven facial animation[J]. Acta Automatica Sinica, 2014, 40(10): 2245-2252.
- [14] BOOTH J, ROUSSOS A, PONNIAH A, et al. Large scale 3D morphable models[J]. International Journal of Computer Vision, 2018, 126(2): 233-254.
- [15] 魏薇, 牛金行, 景慧昀. 人工智能安全[M]. 北京: 化学工业出版社. 2020.
WEI W, NIU J X, JING H Y. Artificial intelligence security[M]. Beijing: Chemical Industry Press, 2020.
- [16] CHESNEY R, CITRON D K. 21st century-style truth decay: deepfakes and the challenge for privacy, free expression, and national security[J]. Maryland Law Review, 2019, 78(04): 882-891.
- [17] CHESNEY R, CITRON D K. Deepfakes and the new disinformation war: the coming age of post-truth geopolitics[J]. Foreign Affairs, 2019, 98(01): 147-155.
- [18] MALIK A, KURIBAYASHI M, ABDULLAHI S M, et al. Deepfake detection for human face images and videos: a survey[J]. IEEE Access, 2022.
- [19] PESETSKI A. Deepfakes: a new content category for a digital age[J]. William & Mary Bill of Rights Journal, 2020, 29(2): 507-508.
- [20] CHESNEY R, CITRON D K. Deepfakes: a looming challenge for privacy, democracy, and national security[J]. California Law Review, 2019, 107(6): 1753-1819.
- [21] CHADHA A, KUMAR V, KASHYAP S, et al. Deepfake: an overview[C]//Proceedings of Second International Conference on Computing, Communications, and Cyber-Security. 2021: 557-566.
- [22] MARAS M H, ALEXANDROU A. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos[J]. International Journal of Evidence and Proof, 2018, 23(3): 255-262.
- [23] 刘国柱. 深度伪造与国家安全: 基于总体国家安全观的视角[J]. 国际安全研究, 2022, 40(3): 3-31.
LIU G Z. Deep forgery and national security: from the perspective of overall national security view[J]. Journal of International Security Studies, 2022, 40(3): 3-31.
- [24] DOUGLAS H. Deepfakes: false pornography is here and the law cannot protect you[J]. Duck Law and Technology Review, 2019, 17(1): 99-127.
- [25] JESSICA S, WOODROW H. The upside of deep fakes[J]. Maryland Law Review, 2019, 78(4): 960-966.
- [26] 史安斌, 胡宇. 向虚假信息宣战: 欧盟的探索[J]. 青年记者, 2018(34): 76.
SHI A B, HU Y. Declaring war on false information: exploration of the European Union[J]. Youth Journalist, 2018(34): 76.
- [27] DURACH F. Tackling disinformation: EU regulation of the digital space[J]. Romanian Journal of European Affairs, 2020, 20(1): 10-11.

[作者简介]



李婧文(2002—),女,北京人,主要研究方向计算机技术应用、电信技术与应用。



李雅文(1990—),女,山东潍坊人,中国信息通信研究院高级工程师,主要研究方向为数字治理、网络法治、数字经济规则。