深度合成十大趋势报告

(2022)

清华大学人工智能研究院
北京瑞莱智慧科技有限公司
清华大学新传院智媒研究中心
国家工业信息安全发展研究中心

学术顾问（按姓氏首字母）

沈昌祥　中国工程院院士，中央网信办专家咨询委员会顾问，国家保密战略专家咨询委员会主任委员
张　钹　中国科学院院士，清华大学人工智能研究院创始院长

指导专家（按姓氏首字母）

陈昌凤　清华大学新闻与传播学院教授、常务副院长
陈树丽　中国科学院声学研究所助理研究员
戴　娇　中国科学院信息工程研究所高级工程师
段伟文　中国社科院哲学所科技哲学研究室主任、研究员，科学技术和社会研究中心主任
赫　然　中国科学院自动化研究所研究员
何小龙　国家工业信息安全发展研究中心副主任
胡清华　天津大学智能与计算学部副主任，天津大学人工智能学院院长
李建民　清华大学计算机系副研究员
梁　正　清华大学公共管理学院教授，清华大学人工智能国际治理研究院副院长
许　可　对外经济贸易大学法学院副教授，对外经济贸易大学数字经济与法律创新研究中心主任
曾　毅　中国科学院自动化研究所研究员，人工智能伦理与治理研究中心主任
张凌寒　北京科技大学文法学院副教授
张　震　国家互联网应急中心高级工程师
赵志耘　中国科学技术信息研究所党委书记、所长，科技部新一代人工智能发展研究中心主任
朱　军　清华大学计算机系教授，清华大学人工智能研究院基础理论研究中心主任，瑞莱智慧RealAI首席科学家

研究机构

清华大学人工智能研究院

清华大学人工智能研究院于2018年6月28日正式成立，张钹院士担任研究院的院长，聘请图灵奖得主姚期智院士作为学术委员会主任。人工智能研究院依托清华大学优势学科，以未来人工智能的原创性基础理论为发力点，力求在探究智能本质的基础上，产生人工智能基础理论和关键技术上的颠覆性创新成果，着力打造具有清华特色影响力的基础性、源头性的新高地，积极推进大跨度学科交叉融合，积极推进大范围技术与产业、学校与企业的融合。

北京瑞莱智慧科技有限公司

瑞莱智慧RealAI是全球领先的第三代人工智能技术基础设施和解决方案提供商，致力于以安全、可靠、可解释、可扩展的第三代人工智能，为高价值场景智能化升级提供一站式赋能平台。RealAI孵化自清华大学人工智能研究院，由张钹院士、朱军教授共同担纲首席科学家。团队坚持源头创新和底层研究，在国际测评和竞赛中多次斩获冠军、发表顶会期刊论文百余篇，并参与多项国家及行业标准制定。目前，RealAI已在政务、金融、能源、制造、互联网等领域落地，为合作伙伴提供了人脸识别系统安全检测与增强、隐私保护计算等全套产品和解决方案。

清华大学新传院智媒研究中心

清华大学新闻与传播学院智媒研究中心于2020年1月成立，旨在加强智能时代的媒体变革研究，大力扩展跨学科的研究格局，以适应新技术变革时代的新闻传播学科发展。智媒研究中心承担了国家社科基金重大项目、教育部重点实验室建设项目，以及多项国家级、省部级和相关产业的研究课题。新闻与传播学院常务副院长、国家万人领军计划人才陈昌凤任中心主任，数据传播国际项目主任、青年拔尖人才蒋俏蕾任副主任，全国多所大学新闻与传播学、计算机科学、哲学、社会学、法学等学科专家担任学术顾问和研究员。

国家工业信息安全发展研究中心

国家工业信息安全发展研究中心（工业和信息化部电子第一研究所）简称国家工信安全中心，是工业和信息化部直属事业单位，是我国工业领域国家级信息安全研究与推进机构。中心始终坚持以“支撑政府、服务行业”为宗旨，经过60余年发展与积淀，构建了以工业信息安全、产业数字化、软件和知识产权、智库支撑四大板块为核心的业务体系，涵盖工业信息安全、两化融合、人工智能、大数据、工业互联网等业务领域，服务对象包括工业和信息化部、中央网信办、科技部、发改委等政府机构，以及相关科研院所、企事业单位和高等院校等各类主体，致力于发展成为集“智库咨询、基础研发、技术服务、生态培育”为一体的工信领域国家高端智库与安全保障权威机构。

目录 CONTENTS

引言

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。2017年，一位名叫“Deepfakes”的用户在美国Reddit网站上分享了篡改人脸的色情视频，将深度合成技术带到了大众面前。

近年来，深度合成技术在影视、传媒、教育等领域呈现出大量正向应用，伴随着不断涌现的使用需求，各类商业化产品层出不穷，深度合成内容数量和关注度呈现飞速增长，受众接受度显著提升。但恶意使用该技术生成的音视频亦显现出了巨大的破坏力，给个人、企业造成了声誉损害和财产损失，乃至给社会、国家安全造成威胁。

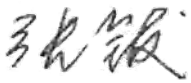
技术是价值中立的，但围绕技术的人类行为并非中性，良性和恶性的影响也并非对称。科技进步，不仅要挖掘出机器的智能，更应展现出人类“驾驭”技术的智慧。因此，《深度合成十大趋势报告（2022）》将从技术研究、领域应用、发展趋势等方面，全面深入地介绍、研判深度合成技术与应用、风险与挑战、趋势与展望等，并就其发展与治理给出建议。

院士寄语

在人类的数千年文明史中，科学技术的革新已经多次带来跨时代的社会变革。从远古时代的火到现代社会的电，每一次技术变革在造福人类的同时，也不可避免地带来了安全风险。

当今方兴未艾的人工智能，又一次为人类带来了新的安全风险。与其他技术一样，人工智能也会被人为的误用和滥用。与其他技术不同，由于人工智能的安全性问题主要来源于算法本身的不安全，它更容易被误用和滥用。以深度合成技术为例，不法分子可以利用它给社会带来巨大的威胁。未来人工智能产业的发展，一方面必须扩大应用的场景，另一方面也必须保证数据和算法的安全性、鲁棒性。科技发展和应用如何才能在最大程度上做到趋利避害，是人类需要长期探索和研究的课题。

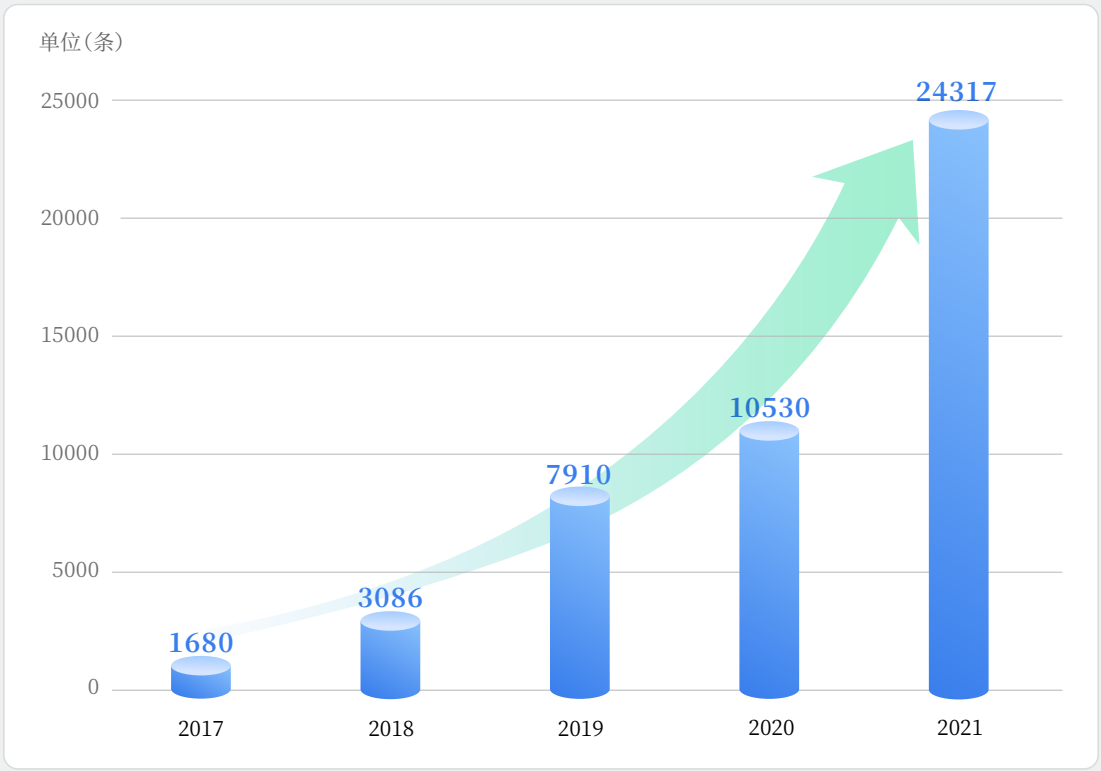
人工智能的治理也是一项长期任务，要两手抓、两不误，既要走创新发展的道路，坚持发展更加安全、可靠、可控的人工智能，还要抓人工智能的治理，从法律法规、伦理规范、行业共识等不同层面去发力。希望此份研究报告，能够为人工智能的健康发展提供参考和指引。

——中国科学院院士、清华大学人工智能研究院创始院长 

趋势一：深度合成内容制作与传播数量高速增长

深度合成的图像、视频、音频、文本等内容具有极强的娱乐性与传播性，伴随着技术的发展成熟，越来越多的创作者在互联网中发布和分享深度合成内容，包括面部替换、表情操纵、语音合成等不同的技术分支。目前，在社交媒体、音视频网站等平台上，通过搜索相关关键词，可以发现大量的深度合成内容在互联网中传播。根据不完全统计，创作者在互联网平台中发布的深度合成内容的数量呈高速增长，以视频为例，2021年新发布的深度合成视频的数量，较2017年已增长10倍以上。

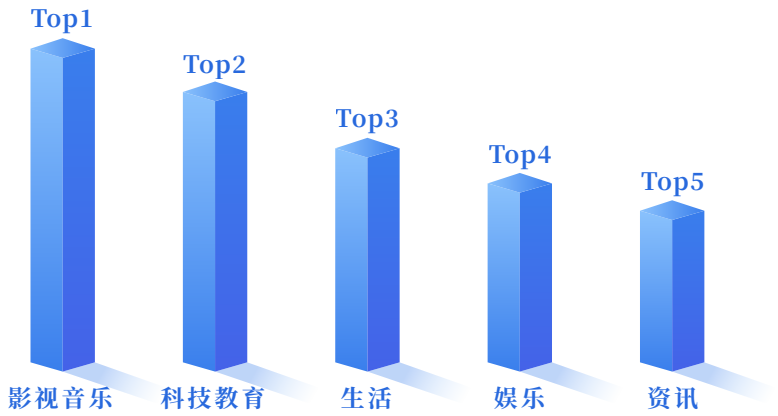
互联网中深度合成视频的发布数量变化趋势



数据说明：在10家国内外平台中（爱奇艺、腾讯视频、优酷、哔哩哔哩、抖音、快手、微博、YouTube、Twitter、TikTok），以“Deepfakes”等10个中英文关键词进行检索，并通过URL去重后，统计出数据结果

互联网中传播的深度合成视频涉及了丰富的内容领域。其中，深度合成视频数量最多的类型是影视音乐，包含电影、电视剧、音乐等方面的内容。排名第二的视频类型为科技教育类，该类视频关注对深度合成技术的讲解和讨论，分享最新的研究成果等。除此之外，排名第三到第五的视频类型分别为生活、娱乐和资讯类。

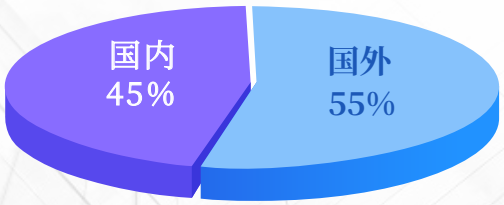
不同类型内容深度合成视频数量排序



数据说明：在10家国内外平台中（爱奇艺、腾讯视频、优酷、哔哩哔哩、抖音、快手、微博、YouTube、Twitter、TikTok），以“Deepfakes”等10个中英文关键词进行检索，并通过URL去重后，将具有视频分类标签的数据进行统计得到的分析结果

无论在国内或国外，深度合成视频的数量均呈现相似的变化规律。分别选取国内和国外各2个平台（哔哩哔哩、抖音、YouTube、TikTok），基于4个平台进行统计分析。结果发现，当前国内外的深度合成的视频数量，在通过关键词可检索到的范围内，数量基本持平。

国内外深度合成视频数量对比



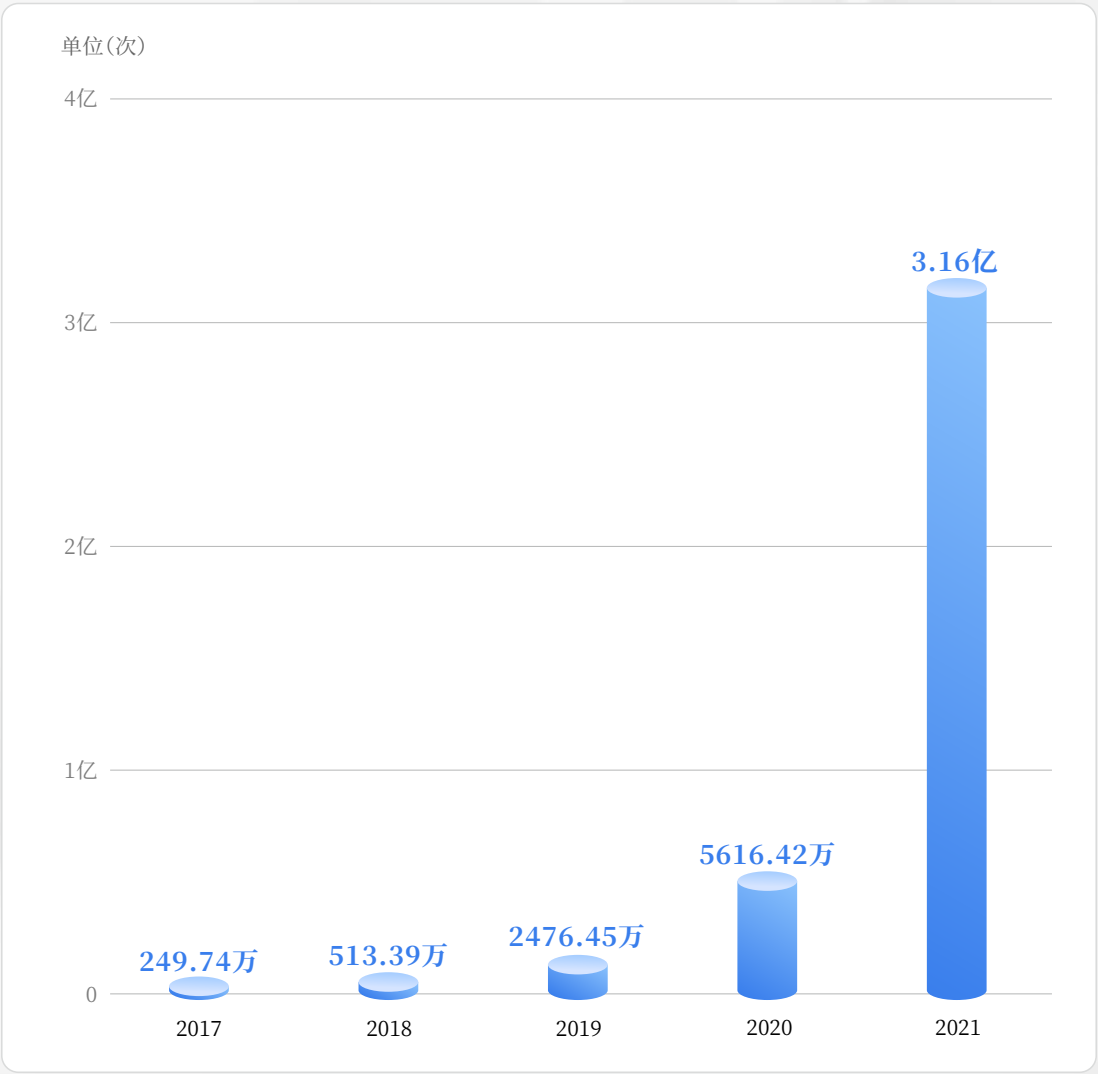
数据说明：在4家国内外平台中（哔哩哔哩、抖音、YouTube、TikTok），以“Deepfakes”等10个中英文关键词进行检索，并通过URL去重后，统计得到国内外数量对比的分析结果

自2017年“Deepfakes”概念出现以来，互联网中深度合成视频的数量逐年高速增长，且制作的内容更加丰富多元。这些最先在美国Reddit网站上引发关注的深度合成视频，经过4年多的时间，在国内也形成了几乎同等规模的发布和传播数量。

趋势二：深度合成内容关注度指数级增长

在社交媒体、音视频网站上发布与传播的深度合成内容，获得了越来越多的关注，以视频为例，其播放、点赞、转发等互动数据逐年攀升。一些以深度合成内容为主要产出的账号，逐渐积累起了“粉丝”人群。以视频的“点赞/喜欢”数据为例进行统计，自2017年以来，该项数据呈现出指数级的显著增长，2021年新发布的深度合成视频的点赞数已超过3亿。

互联网中深度合成视频的点赞/喜欢数量变化趋势



数据说明：在10家国内外平台中（爱奇艺、腾讯视频、优酷、哔哩哔哩、抖音、快手、微博、YouTube、Twitter、TikTok），以“Deepfakes”等10个中英文关键词进行检索，并通过URL去重后，统计出视频获得的“点赞”或“喜欢”的数量，进行求和后统计出数据结果

选取超过4000条深度合成视频的互动数据，分析“播放量、点赞/喜欢数、收藏数、转发数、评论数”这5项互动数据之间的关系后发现，平均每1000次播放，可产生约8.92次的点赞/喜欢，同时会产生约3.51次的转发，将深度合成的内容进行新一轮的传播。

深度合成视频的互动数据关系

数据说明：在检索到的深度合成视频中，选取超过4000条分析其互动数据，统计5种互动数据分别的总和，计算5种数据的相对关系

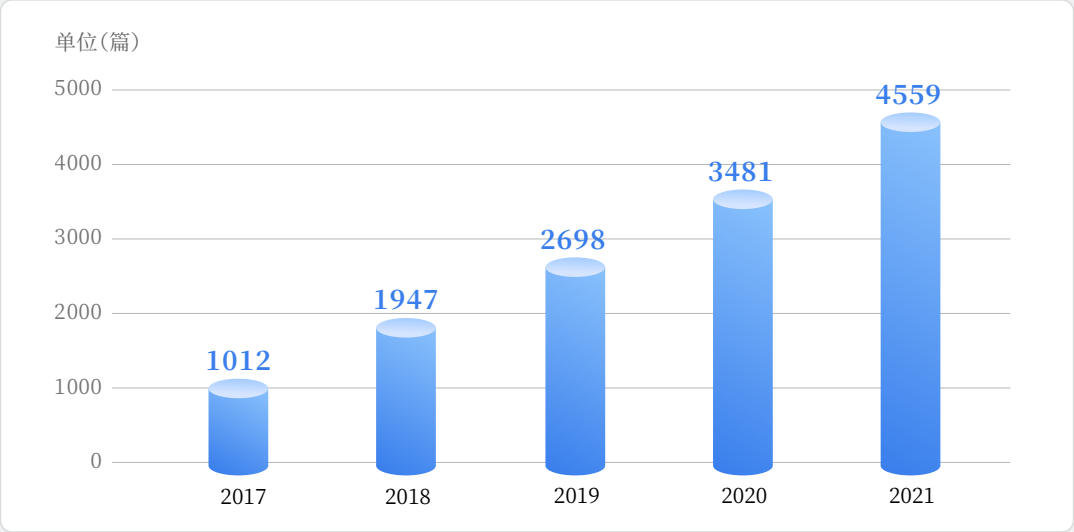
随着内容创作者发布越来越多的深度合成视频，加之“英国女王发表圣诞贺词”、“阿汤哥表演硬币魔术”等一系列深度合成视频的火爆“出圈”，都让深度合成内容引发了平台用户大量的讨论和关注。通过趋势一中所展示的科技教育类深度合成内容处于分类排序Top2的位置也可以看出，公众在对深度合成内容本身的关注之余，还有探索和分享其技术原理的意愿。这些因素均使得深度合成视频的互动数据不断攀升，并在2021年形成爆发式增长。

趋势三：深度合成领域研究论文数量高速增长

技术研究成果是深度合成内容广泛传播的底层推动力。其中，蒙特利尔大学在2014年提出了生成对抗网络（GAN），将数据生成的逼真度推进到一个新的高度，大大降低了深度合成的门槛。生成对抗网络由生成器和判别器两部分组成，生成器通过机器自动生成欺骗判别器的数据，而判别器判断数据对象是真实的或是机器生成的，目的是找出生成器做的假数据。在固定判别器训练生成器和固定生成器训练判别器的不断循环下，生成器所生成的数据真实度将得到大幅提升，使得深度合成的结果难以被人眼辨别。

GAN的出现在深度学习领域掀起了一场革命。以GAN为基础的深度合成技术在近些年中日新月异地发展，已经能够生成出高质量的图像、视频、音频以及文本。相关统计数据显示，2017年以来深度合成领域的论文数量正持续增长。

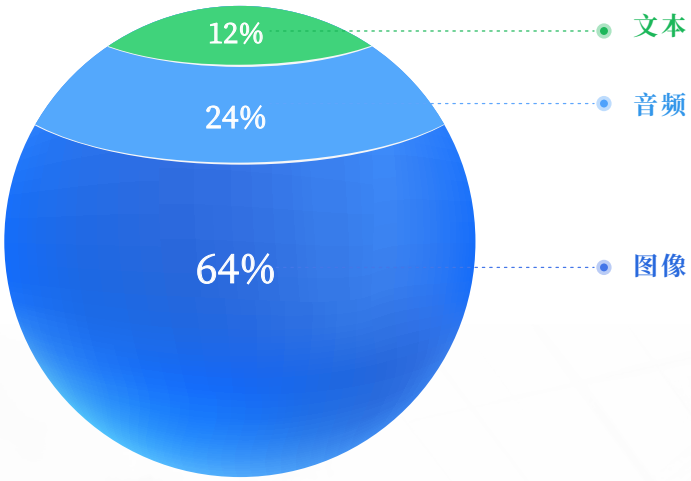
深度合成领域的论文发表数量变化趋势



数据说明：在arXiv中，论文摘要范围内以“GAN”等16个关键词进行检索，统计图像、视频、音频、文本合成领域的论文发表数量

这些深度合成领域的论文，包含对图像、视频、音频、文本等不同模态的合成方面的技术研究。按照图像（包含视频）、音频、文本方向进行划分可以发现，针对图像类生成模型的研究占据最高的比例，达到64%。图像合成的部分，包含了对人物的嘴部、表情、面部整体的合成与替换，以及对人体姿态的全身合成，除了人物之外，也有针对物品、环境与场景的合成研究。此外，语音类生成模型研究成果数量，高于文本类的研究。

不同模态的生成式模型论文占比



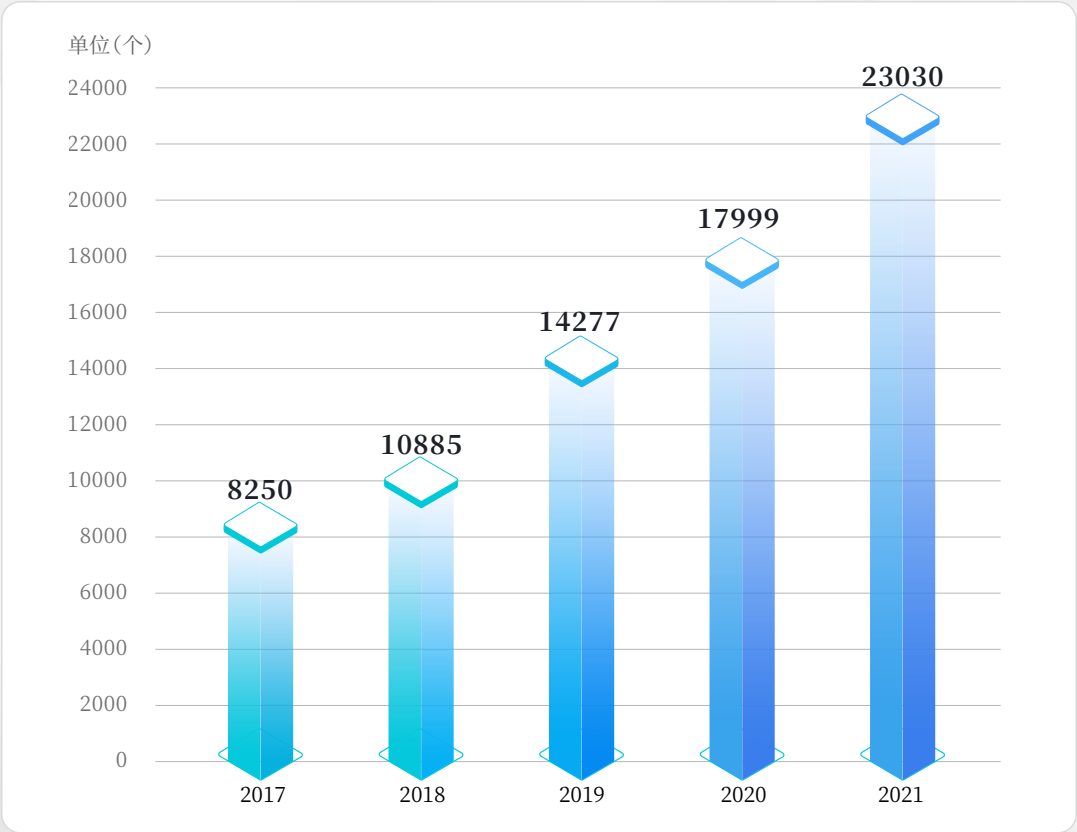
数据说明：在arXiv中，将检索出的深度合成领域论文，按照图像（包含视频）、音频、文本分类统计数量

越来越多的深度合成方向研究成果，推动图像、视频、音频、文本的内容合成朝向效果更加逼真、制作更加高效的方向不断发展。2021年“元宇宙”概念的爆发，为深度合成技术提供了更加宽广的应用场景，同时也更加促进了深度合成技术的加速发展。未来，2D图像合成向高质量3D合成的跨越以及多模态合成内容的协同制作等将成为技术突破的热点方向。

趋势四：深度合成领域开源项目数与讨论度持续攀升

2017年，名为“Deepfakes”的用户利用深度合成技术制作的成人视频在Reddit社区中疯狂传播，迫于公众舆论压力，Reddit网站将该用户封号。该用户随即在全球最大的代码开源平台GitHub上公开了实现该技术的源代码，瞬时引发了技术爱好者的广泛关注与讨论。此后，开发者们不断创造和丰富更多的深度合成相关技术的项目与代码，统计数据显示，2017年以来的深度合成领域的项目数量显著增长。

深度合成领域的开源项目发布数量变化趋势

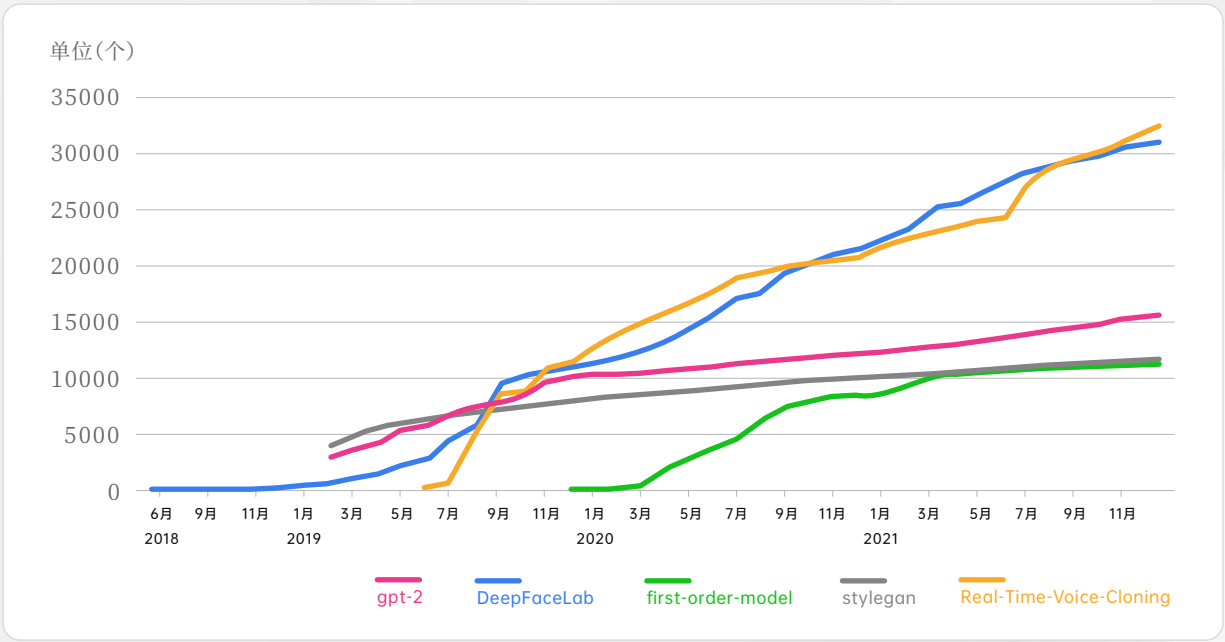


数据说明：在GitHub中，以“GAN”等16个关键词进行检索，统计图像、视频、音频、文本合成领域的开源项目发布数量

在这些开源项目中，也涌现出了一些很有代表性的方法，这些方法受到越来越多的关注，也被应用在了不同的场景之中。如faceswap项目，可实现识别和交换视频中的人脸的功能，自2018年初开源，曾一度登上GitHub排行榜第二位，目前已获得超过4万的关注量。基于这些开源方法，更多的技术爱好者持续讨论并且做出补充和贡献，共同促进深度合成方法在合成质量、制作效率等方面不断升级和迭代。

以图像方向、音频方向、文本方向中具有代表性的5个开源项目为例，其Star数量在2021年均已突破1万（GitHub中Star数超过1万的项目占总项目数的比例低于百万分之九）。其中，DeepFaceLab、first-order-model、stylegan、Real-Time-Voice-Cloning、gpt-2项目可分别实现面部替换、动作或表情操纵、人脸等图像生成、声音复刻、文本生成的功能。

代表性深度合成开源项目的Star数量趋势图



数据说明：在GitHub中，以5种不同类型的具有代表性的深度合成项目为例，统计出5个开源项目的Star数量变化

纵观计算机行业的发展，开源项目已经成为推动产业进步的强大力量，开源项目聚焦的方向也是产业发展的风向标。深度合成在开源社区中持续走高的热度，将继续推动该技术的发展与在产业中的落地。

趋势五：深度合成需求场景趋于多元且成熟

随着深度合成技术的发展，其正向应用场景也在不断丰富，如影视制作、广告营销、电子商务、社交娱乐等。深度合成技术在一些场景中升级了传统的内容制作方法，显著地提升了制作效率和质量；深度合成技术也为一些全新的场景提供了技术可能，支撑实现更多具有想象力的虚实交互的空间。用户多持开放态度并予以较高评价和接受度。

案例1：手语AI合成主播“小聪”解说北京冬奥会谷爱凌夺冠

2022年2月8日，“小聪”用流畅的手语解说了中国女子雪上项目夺得首金后激动人心的瞬间。“小聪”是全球首个手语AI合成主播，帮助听障人士常态化、高质量地接收资讯信息，克服理解障碍，更好享受数字化生活。根据世界卫生组织发布的数据，全球听力障碍人群高达4.66亿，在我国则有2700万人。在聋人可懂度测评中，“小聪”的手语播报可懂度达到85%以上。

案例2：虚拟偶像从“小众亚文化”走进大众视野

2021年洛天依登上央视春晚舞台，虚拟偶像进入主流视野。根据哔哩哔哩CEO陈睿在12周年主题演讲中公布的信息，2020年6月-2021年5月，共有32412位虚拟主播在哔哩哔哩开播，同比增长40%。而根据数据平台vtbs.moe抓取的数据测算，过去一年，哔哩哔哩针对虚拟偶像的订阅打赏比上年同期增长了350%。

案例3：绘制像素画，辅助动画和游戏设计

像素绘画是视频游戏中最受欢迎的美学之一。在《Trajes Fatais：Suits of Fate》游戏中，每个角色平均要绘制五百个精灵，采用传统方法设计每个精灵，绘制时间大约需要一小时。该游戏通过使用深度合成技术，自动完成草图、线条图、底纹、区域上色和索引的流程，将每个精灵的绘制时间减少了15分钟。

案例4：修复老照片，让历史“活”过来

2021年9月2日，第八批在韩中国人民志愿军烈士遗骸回国。次日，《人民日报》微信公众号发布了名为《他们正年轻！AI修复给志愿军老照片上色》的短视频，用深度合成技术将老照片中的志愿军黑白面孔逼真还原，让大家一睹当年“最可爱的人”的青春风采。视频发布后，仅《人民日报》公众号平台播放量就已达到130万人次，点赞数达4万人次，转载量达1000余次。

案例5：失独母亲重新听见“女儿”

2019年9月，一位母亲向阿里巴巴人工智能实验室求助，希望能够将她患癌离世的14岁女儿做成可以交互的软件。3个月后，他们帮助这位妈妈合成了一段女儿长达20秒的语音存在天猫精灵里。这位妈妈认为，这对于失独家庭是一种新的疗愈手段，如同旅行、心理咨询。在知乎的“失独妈妈把女儿做成 AI，这样留下挚爱的方式你能接受吗？”问题下的166个回答中，持接受态度的网友约达70%以上。

案例6：AI换脸“拯救”被劣迹艺人殃及的影视作品

2021年上映的电视剧《突围》，其中一名演员因行为失当被曝光，剧组后期不得不删掉一些重要的戏份。而一些删除会影响剧情的片段，剧组则运用了AI换脸方式解决。

特别是在影视制作领域，近些年来，深度合成技术已成为受艺人劣迹行为拖累作品的救场工具。电视剧《长安十二时辰》、《光荣时代》、《了不起的儿科医生》、《突围》等多部作品均使用了该技术。而观众对影视作品AI换脸后的效果评价，已从最初的“惨不忍睹”变成了“瑕不掩瑜”。

从上述案例可以看出，随着技术成熟度的不断提高，尤其是其伴随场景的“多点开花”，深度合成快速进入大众视野，亮眼应用频现，彰显了技术背后的温情和人文关怀，被接受度也在持续提升。

趋势六：深度合成商业化产品类型逐渐丰富

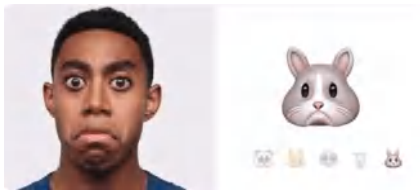
越来越多的企业机构开始利用深度合成技术提供面向公众的产品和服务。这些产品和服务在增强趣味性、提高传播性等方面均提供了新型解决方案，也获得越来越多用户的认可。目前，依托深度合成技术的产品和服务已涵盖图像、视频、音频、文本等多个维度、多个领域。

一、图像和视频方向

图像和视频方向的产品和服务在深度合成应用初期最为普遍，但是由于产品质量良莠不齐且容易侵犯用户隐私，当监管规范到位后其数量逐渐减少。如2021年被广泛讨论的FacePlay等应用，只需用户上传一张人脸照片，就可快速生成不同服装场景的图片和视频。

案例1：

Animoji是自苹果iPhone X开始推出的一项新功能——定制表情，使用者可以通过iPhone前置摄像头来捕捉人的脸部表情，即时转为手机可用的表情包。苹果工程师为此开发了一整套算法，通过一系列面部数据和表情数据训练算法，让它们建立多个可以去描述的静态3D模型，同时借助人工智能技术自主学习来实现对面部表情的识别和建模。



案例2：

FacePlay是一款换脸视频制作软件，在FacePlay里有数十种不同风格的特效短视频模板，包括古风、民族风、现代风等，用户只需要上传一张照片，即可用自己的形象进行制作。使用FacePlay，可快速生成特效视频。2021年8月，该软件更新版本后连续5天登上App Store免费榜Top2，掀起模仿潮流，一周下载量达200w+。

二、音频方向

语音方向的深度合成技术发展较快、普及迅速，合成音的音质和自然度越来越高，并衍生出声音复刻、歌声合成、方言合成、情感合成等更多功能。语音合成已经成为人机交互的重要一环，且被广泛应用于智能硬件、智能客服、语音导航、有声读物、机器人、语音助手、自动新闻播报等场景。业界多家企业机构已发布基于深度神经网络技术合成音频的商用开放平台。

案例：

2019年央视大型文化节目《经典咏流传》在第二季中，推出了AI小工具“读诗成曲”。用户仅需要朗读一段诗词，就可以听到用自己声音唱诵的经典诗词唱段。该个性化合成技术，让更多大众参与到传颂经典的环节中来。据统计，该工具互动总量已超过821万人次，有90万线下经典传唱人参与互动。

音乐方向的深度合成包括词曲写作、伴奏生成、歌声合成等不同环节。音乐合成涵盖古典音乐、爵士、流行等各种风格，同时因模型带有各种乐器和作曲家的元信息，使其可以分清不同乐器和作曲家风格，生成结果更加可控。互联网行业已出现如Amper ScoreTM、AIVA、Jukedeck MAKE、ecrett music、Melodrive、Orb Composer等音乐创作工具。在歌声合成方向，深度合成技术正同市场产生更加紧密的联系，帮助虚拟歌手唱出更加自然、动人的音乐。

案例：

在2020年7月9日世界人工智能大会（WAIC）云端峰会开幕式上，由微软亚洲互联网工程院开发的虚拟歌手小冰，携手小米小爱、百度小度、B站泠鸢共同演唱大会主题曲《智联家园》。

三、文本方向

以文本为形式的深度合成在新闻报道、诗文创作、聊天问答等方面都得到越来越多的应用，并显现出巨大的创作效率和未来潜力。比如2015年新华社就启用了“快笔小新”，一个通过深度合成技术生成类人创作稿件的程序。目前，类似的文本深度合成技术已在社会上有了更广泛的应用和普及，公众均可通过相关应用程序获得快速创作文本的体验。

案例：

2017年底，清华大学自然语言处理与社会人文计算实验室发布了基于深度学习技术的诗歌写作系统“九歌”。该系统能根据用户输入的关键词句，自动创作出集句诗、藏头诗、律诗和不同风格的绝句，系统上线至今的访问量已超过1000万人次。

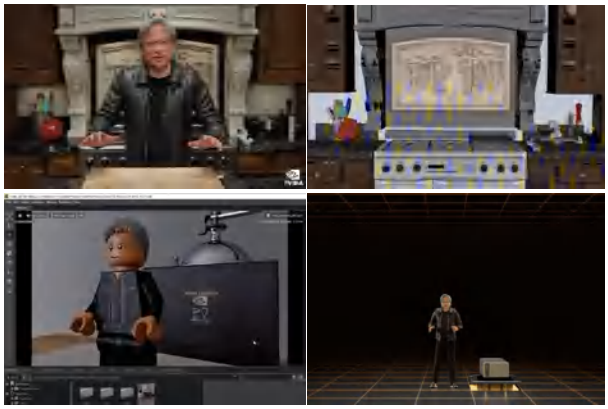
综上，以深度学习技术为基础的图像、视频、音频、文本等多种合成类型产品和服务已经走向大众，并在诸多实践领域显现出应用潜力，给公众带来前所未有的数字体验，正在重塑人类未来生产、生活方式。

趋势七:深度合成重新定义虚拟化数字生存空间

深度合成技术极大地丰富了虚拟数字空间的信息内容,为更加多样化的传播行为提供了可能性。对多个视频网站中高热度视频内容进行统计分析可以发现,深度合成技术已经渗透到了新闻、娱乐、学术等多个信息传播领域。伴随着自动数据生成、全身合成、3D建模等技术的逐渐成型,信息挖掘、信息制作、信息传播的时间与资金成本被大大压缩,进而推动更多数字应用场景的拓展与落地。

案例1:

2021年8月11日,英伟达在计算机图形学顶级会议Siggraph 2021上发布纪录片,还原了4月GTC线上峰会中“数字黄仁勋”制作的全过程。“数字黄仁勋”基于英伟达虚拟协作平台Omniverse制作。11月的开发者会议上,英伟达进一步推出了Omniverse Avatar,将建模与AI技术相结合,帮助“元宇宙”创作者轻松建立能够理解自然语义、甚至能实时转换语言的虚拟人物形象。

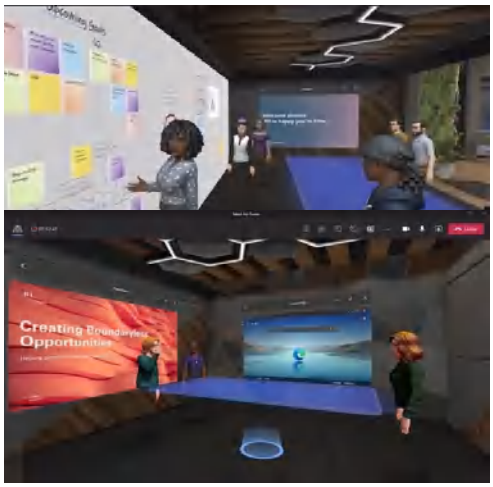


案例2:

2021年10月28日,脸书宣布更名为“Meta”,此次改名宣告了扎克伯格进军元宇宙的决心,紧接着Meta在2021年12月09日宣布旗下元宇宙平台Horizon Worlds正式面向美国与加拿大地区18岁以上的成年人开放。该平台允许用户构建自己的虚拟世界,并提供了一系列的模板和工具。在Horizon Worlds中,Oculus虚拟现实头戴设备用户可以创建一个没有双腿的化身形象,在虚拟世界中四处游荡。在那里,他们可以与其他用户的化身进行互动,甚至可以一起玩游戏。

案例3：

2022年1月6日，微软在Microsoft Ignite China 2022（微软在线技术大会中国站）上，围绕元宇宙发布了两项重要功能——Mesh for Microsoft Teams和Dynamics 365 Connected Spaces，务实性地提出了元宇宙办公环境之下的诸多细节，使在现实世界会议室中的人，跟线上办公的人有一个相对平等包容的体验。



案例4：

2021年12月28日，百度Create大会2021元宇宙论坛在希壤虚拟空间多人互动平台的第一座城市CreatorCity举行。借助深度合成技术，首个国产元宇宙产品“百度希壤”打造出了跨越虚拟与现实、永久续存的多人互动空间。“百度希壤”包含智能Avatar、智能语音助手等互动功能，提供方位语音、大场景混音等“声临其境”的体验，同时也打造了朋克区、生态湖等沉浸式空间。

从更广泛的传播社会学意义上看，一个新的人类生存场景将以深度合成技术为基石展开。在虚拟数字空间中，原本因媒介技术而产生的人类社交沟通障碍得以被进一步消除，智能化、视觉化、场景化、虚拟化成为新的交往常态，使人们进一步沉浸在由媒介构建的数字虚拟环境当中。

基于深度合成技术打造的未来人类数字虚拟空间——元宇宙，完成了对现实空间和时间的多重复制和延伸，跳脱了传统物理空间的局限，提供了一个虚拟人、自然人和机器人融生的逼近现实且超越现实的新世界。

趋势八：深度合成内容负面风险持续加剧且产生实质危害

随着深度合成技术的开放开源、深度合成产品和服务的增多，深度合成内容制作的技术门槛越来越低，实现了技术的“平民化”，普通人仅需少量的图像、视频、音频、文本等样本数据，利用简便易用的合成工具，就能模糊真实和虚拟（虚假）的边界，解构“眼见为实”的认识论权威，进而冲击社会信任、媒体信任、政治信任。事实上，通过深度合成技术制造虚假视频、虚假音频进行诬陷、诽谤、诈骗、勒索等违法行为和事例已屡见不鲜，深度伪造内容数量不断增多、危害性不断增强。

由于深度合成技术需要大量数据训练以实现更高逼真度的特性，明星、高管、学者、政要等公众人物的图像、视频、音频资料极易通过网络获取，他们也因此成为了被恶意攻击的首选。

一、社会及国家安全风险

案例：美国政要被深度伪造，威胁政治秩序和选举

2018年4月，有技术团队制作了涉及美国前总统奥巴马的换脸视频，视频中“奥巴马”称美国现总统特朗普为“彻头彻尾的白痴”，该视频在网上已获得480余万次的浏览量；2018年5月，有人利用深度伪造技术制作了特朗普的视频，批评比利时环保政策等。美国众议院情报委员会于2019年6月13日组织听证会，重点分析了深度伪造技术影响下的国家和选举安全风险，认为其已威胁到美国的政治秩序和选举。

可以预见，极端组织一旦利用深度合成技术，将形成新威胁、新挑战。例如，捏造国家政要言论，伪造公务人员虚假视频，继而煽动他人实施破坏社会公德、扰乱社会秩序的行为；也可成为训练恐怖分子的“洗脑”教材，教唆进行恐怖活动，危害国家安全、破坏民族团结，严重威胁社会安宁。

二、公司名誉、财产损失

案例：阿联酋某银行被伪造客户语音诈骗3500万美元

2021年10月，《福布斯》杂志根据发现的一份法庭文件报道：欺诈者用深度伪造语音冒充客户，电话联系阿拉伯联合酋长国某银行经理，声称公司即将进行收购，需要银行批准一个3500万美元的转账。银行经理确认转账金额和收款账户后进行了转账。随后，这笔被窃取的资金被转移到全球各地的银行账户。


《法庭文件》：
UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA
APPLICATION OF THE UNITED STATES FOR AN ORDER
FOR A COMMISSIONER'S APPOINTMENT PURSUANT TO 18 U.S.C. § 3512

除诈骗外，企业高管如被行业竞争对手篡改发言内容、散播虚假消息，将给其本人和企业形象带来严重冲击，进而影响公司经济利益，甚至致使股价下跌。知名学者如被合成视频篡改发言内容，亦可能导致市场恐慌，引发金融市场动荡。

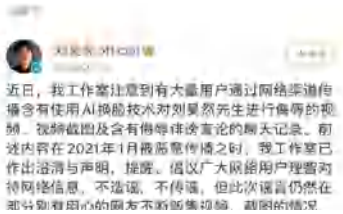
三、个人名誉、财产、精神损害

案例1：常州男子好奇“裸聊”诈骗套路被骗11万

2021年5月，常州王某报警称自己被裸聊诈骗11万余元。当时，一女网友请他下载某软件并提出视频邀请，他出于好奇接受了邀请，只露出了自己的脸。挂电话后不久，王某就收到了一段有自己头像的不雅视频。王某表示，骗子用AI换脸技术把他的脸合成到他人的视频上，并以此威胁其转账。

案例2：艺人刘昊然被伪造不雅视频并恶意传播

2021年1月23日，有网民发现刘昊然被深度合成技术伪造出不雅视频，之后视频被恶意传播。之后，刘昊然工作室公开作出澄清和声明。8月，视频再次在社交软件上大肆传播。8月28日晚，刘昊然工作室发文称，工作室已于当日向警方报案。


近日，我工作室注意到有大量用户通过网络渠道传播含有使用AI换脸技术对刘昊然先生进行售卖的视频、视频截图及含有侮辱诽谤言论的聊天记录。前述内容在2021年1月被恶意传播之时，我工作室已作出澄清与声明，提醒、倡议广大网络用户理智对待网络信息，不造谣、不传谣，但此次谣言仍然在

案例3：犯罪团伙伪造他人人脸动态视频，提供给黑灰产

2021年10月14日，央视CCTV-1《焦点访谈》栏目报道：安徽合肥警方查获一起非法利用深度合成技术，伪造他人人脸动态视频，为黑灰产业链提供注册虚拟手机卡等技术支撑的案件。目前，8名犯罪嫌疑人因侵犯公民个人信息罪已被移送起诉。

人类的传播活动，也因深度合成技术逐渐步入“深度后真相”时代。首先，“深度伪造”深刻影响了新闻对真相的记录，对虚假内容高难度的甄别影响了事实核查的有效性。其次，在社会重大突发事件或政治事件节点上，深度合成技术如若被恶意使用，将会借助社交媒体，使虚假信息短时间内在互联网上产生病毒式的扩散和蔓延。第三，在日常事件的信息发布与追踪中，深度伪造的信息还会造成舆论意见的不断翻转，冲击主流媒体对舆论的引导效能，激化社会不同群体的矛盾，削弱主流意见对社会的弥合作用。

亟需警惕的是，深度合成技术的恶意伪造内容通常迎合大众猎奇心理，具有极强的意识塑造能力。而受年龄、教育水平等因素影响，不同群体对于虚假信息的敏感程度也不相同。未成年人的意识形态正处于形成的关键时期，其分辨能力较弱，模仿能力较强，最易受到不良影响。

趋势九：深度合成鉴别需求逐渐增大且难度提升

随着研究的深入，深度合成技术得到快速发展。从早期基于卷积神经网络到越来越成熟的Transformer，通过在对抗生成网络中不断自我优化、升级换代，深度合成内容质量得到大幅提升。比如早期的换脸视频，由于技术不成熟、不完善，尚存在“微表情不自然”、“面部边缘有锯齿”等明显换脸痕迹，可作为参照辨别真假。但近年来在社交媒体上广泛传播的换脸视频，动作的逼真度、自然度，以及视频整体的清晰度、流畅度都得到大幅提升，足以达到以假乱真水平，传统基于生物特征的鉴别方式越来越难以发挥作用。

案例1：

2020年，英国女王圣诞节致辞的视频被故意合成荒诞的内容，但是视频的自然度已经大大改善。

案例2：

2021年，演员“阿汤哥”变魔术、打高尔夫的合成视频，已经十分逼真，很难找到瑕疵。

为了应对深度合成内容越来越逼真且多元的问题，采用技术方案进行自动化鉴别的需求也应运而生，如基于人工智能技术实现自动化检测。常用的方法包括基于伪造内容数据集完成对模型检测器的训练、基于帧间不一致性实现对伪造内容的判别等。这些方法在开源数据集中均能达到很高的准确率。但伴随新型伪造方法的层出不穷、网络传播环境的日趋复杂，加上基于深度神经网络的检测算法存在结构性缺陷等，反深伪检测技术也面临“强对抗性”，需要持续更新与迭代优化。类似于“猫鼠游戏”，深度合成和检测在不断学习攻防过程中会自我进化，规避上一代的对抗技术。为了能在对抗攻防中掌握主动权，未来反深伪检测技术的发展需融合多模态内容的取证分析、基于数字水印的溯源技术等多方面能力，实现伪造内容的精准识别，打造可信内容体系。

为了促进深度伪造检测相关技术的发展，国内外已发布专项研究计划以及开展相关的学术竞赛。研究计划方面，美国国防部高级研究计划局（DARPA）先后发布了两项具有代表性的研究计划。媒体取证项目Media Forensics致力于探索并识别深度伪造中存在的视听不一致的技术，包括像素不一致（数字完整性）、物理定律不一致（物理完整性），以及与其他信息源的不一致（语义完整性）等问题。语义取证项目Semantic Forensics旨在开发伪造媒体资产的自动检测、归因和表征方面技术。学术竞赛方面，2019年底，亚马逊、Meta（原Facebook）、微软等机构共同发起了DeepFake Detection Challenge深度伪造检测挑战赛，主办方投入100万美元奖金池，吸引了全球超过2000支队伍参赛。在国内，由中央网信办和公安部等部门共同指导的2020年“第二届中国人工智能·多媒体信息识别技术大赛”和2021年“第三届中国人工智能大赛”中，均设置了音频与视频方向的深度伪造检测相关赛道。世界顶级黑客赛事GeekPwn（极棒），也分别在2020年和2021年中发起了虚假人脸识别的挑战赛。

目前，学术界和产业界均已对深度合成鉴别检测技术的研发进行了大量投入，包括Meta（原Facebook）、谷歌、微软、马普所等机构均推出视频认证的方法或产品。在国内，清华大学、中科大等高校均在深伪检测方面取得显著成果。瑞莱智慧RealAI、腾讯优图实验室等企业机构也已构建人脸合成检测平台并发布针对性的检测产品，支持对多种换脸方法进行检测。例如，瑞莱智慧的深度伪造内容检测平台DeepReal拥有工业级的检测性能和应对实网环境对抗变化的检测能力。

趋势十：深度合成治理监管机制逐步建立

最大限度地保护公民、企业权益，维护社会正常秩序，使深度合成等新兴技术造福于人类，是现代政府的重要职责。近几年来，针对深度合成技术恶意使用所带来的挑战，世界各国纷纷出台管理法律法规，探索对深度合成的依法管理。

国际方面：

国际方面，以美国和欧盟为例详述其针对深度合成技术的治理监管机制。此外，德国、新加坡、英国、韩国、俄罗斯等国家，均有适用于深度合成技术相关犯罪案件审理的法律法规。

美国：从联邦和州层面进行专门立法

2018年12月

美参议院提出《2018年恶意伪造禁令法案》（Malicious Deep Fake Prohibition Act of 2018），对制作深度伪造内容引发犯罪和侵权行为的个人，以及明知内容为深度伪造还继续分发的社交媒体平台，进行罚款和长达两年的监禁。如果伪造内容煽动暴力、扰乱政府或选举，并造成严重后果，监禁将长达10年。

2019年6月

美众议院提出《深度伪造责任法案》（Deepfakes Accountability Act），要求任何创建深度伪造视频媒体文件的人，必须用“不可删除的数字水印以及文本描述”来说明该媒体文件是篡改或生成的，否则将属于犯罪行为。2019年6月，众议院提出的《2020财年Damon Paul Nelson和Matthew Young Pollard情报授权法案》中提出，建议针对深度伪造技术鉴别开展相关的技术竞赛，以刺激鉴伪技术的研究和商业化。

两党议员分别在众议院、参议院同时提出《2019年深度伪造报告法案》（Deepfakes Report Act of 2019），明确了“数字内容伪造”定义，规定国土安全部定期发布深度伪造技术相关报告。该法案将国土安全部作为主管协调部门，规定报告的制作由国土安全部分管科学和技术事务的副部长负责，其分管的司局也参与深度伪造技术研究，如科学技术局、网络安全司、国土安全高级研究计划局等。

与联邦层相比，美国很多州的“深度伪造”规制明显走在前面。目前已有一些州通过了正式法律，对“深度伪造”进行规制，比较有代表性的包括加利福尼亚州、弗吉尼亚州和德克萨斯州。

欧盟：纳入既有法律框架进行规制

2018年4月

欧盟发表致欧洲议会、欧盟理事会、经济和社会委员会以及地区委员会的长篇公开信《应对线上虚假信息：欧洲方案》。《方案》集中阐释了欧盟委员会面对线上虚假信息挑战的基本观点，提出改进信息来源及其生产、传播、定向投放和获得赞助方式的透明度，改善信息多样性，提高信息可信度，制定包容性解决方案等原则，以实现全面防范视频、图像和文字等虚假信息，避免信息发布者违法操纵舆论等状况。

2018年5月

欧盟正式实施《通用数据保护条例（GDPR）》，把个人数据置于现行欧盟法律保护之下，范围包括可能被用于制作深度伪造内容的公民图片等数据，并适用于社交媒体、软件公司换脸软件可能引发的个人隐私泄露问题。未来，GDPR还将通过进一步确立“数据保护法案”进行补充。

2018年9月

欧盟发布历史上首份《反虚假信息行为准则(Code of Practice on Disinformation)》，旨在加强互联网企业对平台内容的自我审查，从源头打击网络虚假内容。

2020年1月

欧盟宣布于年初启动对《反虚假信息行为准则》实施效果的全面评估。本次评估主要针对签署了《准则》的互联网企业，以检验社交媒体和搜索平台等在打击网络谣言方面的行动力度，包括针对深度伪造等音视频文件的管控力度。

我国方面：积极探寻建设有效治理机制

2019年11月18日

国家互联网信息办公室、文化和旅游部、国家广播电视总局三部门联合颁布《网络音视频信息服务管理规定》，明确提出对基于深度学习、虚拟现实等技术具有媒体属性或社会动员功能的音视频信息服务开展安全评估，对非真实音视频信息进行标识，不得利用深度学习技术制作并传播虚假新闻信息，部署鉴别技术，尽快建立辟谣机制等措施。此规定于2020年1月1日正式执行。

2019年12月15日

国家互联网信息办公室颁布《网络信息内容生态治理规定》，规定网络信息内容服务使用者和网络信息内容生产者、网络信息内容服务平台不得利用深度学习、虚拟现实等新技术新应用从事法律、行政法规禁止的活动。此规定2020年3月1日起实施。

2020年5月28日

十三届全国人大三次会议表决通过了《中华人民共和国民法典》。其中人格权编明确规定：不论是否出于营利目的，均不得利用信息技术手段伪造他人肖像、声音。这部法律自2021年1月1日起施行。

2022年1月4日

国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局联合颁布《互联网信息服务算法推荐管理规定》，其中明确要求“不得生成合成虚假新闻信息”。此规定2022年3月1日起施行。

2022年1月28日

国家互联网信息办公室颁布《互联网信息服务深度合成管理规定(征求意见稿)》，旨在促进深度合成技术依法合理、有效利用，规范发展互联网信息服务深度合成活动，是具有系统性、针对性和可操作性的专门管理规定。

从以上各国针对深度合成技术采取的法律措施和治理手段看，各国均在不断建立健全伦理体系与法律法规，引导深度合成技术良性发展，同时针对利用深度合成技术进行违法犯罪的行为予以严厉限制。

发展及治理建议

综上，深度合成技术已展现了其强大的能量和可能性，加速应用已成为现实趋势。但在享受深度合成技术所带来便利的同时，也不得不直面技术的伴生风险。有鉴于此，特提出以下发展及治理建议：

一、加快推进法治建设，逐步完善配套法规政策

当前，世界各国普遍面临着法律规范的速度滞后于技术发展步伐的挑战。监管部门需要提前进行前瞻布局，在保护深度合成技术良性发展的基础上，制订针对不良深度合成应用的配套法规、管理条例等，形成科学严密、有机衔接的法规制度体系，增强法规的针对性和实效性。在立法与监管中，应进一步明确深度合成服务与应用相关方的法律责任，强化媒体平台等传播媒介的法律和社会责任，阻止危害社会及个人合法权益的深度伪造内容扩散。

二、倡导产研发展自律自治，健全常态化联动机制

行业层面，产研界自身在法律法规不完全成熟及体系化之前，应强化“伦理先行”意识，开展跨学科的前瞻性对策研究，以伦理道德和公众反馈作为牵引，共同促进善用，防范滥用、恶用。同时，相关行业组织应联合院校、企业、研究机构等技术主体牵头制定标准、公约、共识、指南、准则等行业制度规范，开展制度宣传、标准推广应用等活动。此外，平台间还应建立用户黑名单机制和危机应对联动机制，在发现深度伪造内容传播时快速响应、及时处置、协同应对，避免造成恶劣影响扩大化。

三、加强检测相关技术研究，全面提升监测和处置能力

技术层面，应引导人工智能学术界、产业界不断加强技术研究，利用技术创新、技术对抗等方式，持续提升和迭代深度合成检测能力，并扩展深度合成溯源、深度合成鉴定等方面的研究，防范伦理安全风险和合规风险。一方面，应加大对深度合成内容检测技术的研究支持，推动技术的推广应用；另一方面也可以通过举办技术竞赛、发布科研项目等方式引导企业加强源头技术攻关，促进科技创新。

四、加大宣传普及力度，加快形成全社会共同治理格局

社会层面，要持续加大深度合成风险普及力度，强化公民对深度合成等人工智能技术的认识。一方面推动公民作为负责任的深度合成技术使用者，主动标识合成内容；另一方面引导公众对互联网中传播信息进行多方验证，不轻信不实的音视频内容，提高全社会防范意识并鼓励公民监督。此外，政府部门还可向公众提供一些简单易操作且可靠的技术工具，使公众能够快速获知信息真伪，避免遭受损失。

五、支持鼓励正向应用，促进深度合成技术持续健康发展

在合规前提下鼓励深度合成技术创新、支持应用发展。近年来，我国相继在网络安全、数据安全、算法安全等领域出台法规标准，并在深度合成领域制定专门的管理规定。各相关方应当与时俱进落实好新的规范要求，并在此前提下不断追求技术突破，支持深度合成创新发展。同时，在帮助弱势或特殊群体、助力影视艺术行业等领域不断开拓深度合成技术应用场景，创立示范标杆，形成对人工智能行业整体的带动效应，不断提升核心竞争力。

研究团队（按姓氏首字母）

· 陈建益
· 黄阳坤
· 胡嵩智
· 纪林依
· 李　洋
· 刘　丹
· 刘永东
· 田　天
· 唐家渝
· 田成子
· 王　琦
· 武沛颖
· 萧子豪
· 杨彩虹
· 俞逆思
· 袁雨晴
· 闫　桥
· 朱　萌
· 张天奕
· 张雪靖
· 张　瑶