数字图像处理

Fake Media 调研报告

信息与电子工程学院　信息工程

2023 年 5 月 28 日

1　Fake Media 的概念

随着互联网技术的不断发展，人们获取信息的手段从过去的报纸、广播、电视，变为了现在的图片、视频和各种社交媒体。相比与传统媒体，新媒体数字化的特点不仅使其更方便传播，同时内容也更加多元丰富。不再像过去大部分人只能做被动的接收者，现如今每个人都可以是内容的生产者。

但这也间接导致了很多问题。例如随着数字图像处理的技术不断发展，我们可以对图像进行编辑。这包括了对摄影作品的修图美化，也包括一些不怀好意的人修改图像造谣、诈骗等。近年来人工智能技术不断发展，对于图像的编辑手段更上一个台阶。前不久特朗普被关进监狱的新闻大火，伴随着特朗普被捕的照片，但事后作者才声明照片实际上是 AI 合成的结果。

由这些例子可以看出，Fake Media 在技术层面上可以理解为利用数字手段对一些信息的数字载体进行编辑的手段，在传播的角度上看，这些编辑的结果往往是为了达成欺骗的效果。

2　矛与盾

2.1　制作 Fake Media 的典型技术、发展现状

以数字图像为例，制作 Fake Media 的典型技术有传统数字图像处理技术，具体包括了图像增强、图像分割、特征提取等算法，这些算法的基础有直方图均衡化、图像滤波、图像形态学分割等技术。现如今传统的数字图像处理技术发展已经十分成熟，其核心是对于数字图像像素点进行合适的变换处理，以达到某种效果。

在传统的数字图像处理手段之后便是随着人工智能技术不断火热的卷积神经网络。卷积神经网络的快速发展是 2012 年以后，以 LeNet 和 AlexNet 网络结构的提出为开端，新的网络模型和学习方法不断出现发展，使得图像处理达到了新的高度，也诞生了人工智能领域中十分重要的计算机视觉。在 2015 年 ResNet 出现，其核心思想在于最后一层卷积的结果加上输入图像，这使得网络学习的内容从最开始是一整张图像，到现在学习的内容是目标图像和原始图像的差，也就是“残差学习”，该网络在图像分类和目标检测任务上表现比过去的网络更好。

2014 年伊恩 • 古德费洛等人提出了生成对抗网络 GAN，该网络由一个生成网络和一个判别网络组成，可以使其输出结果尽可能模仿训练集的真实样本。GAN 的出现使得制作 Fake media 的能力更加强大，以此方法可以生成以假乱真的图片，例如 AI 作画、AI 作图都可以该网络为基础进行实现。

随着卷积神经网络的发展，也诞生了旁支循环神经网络 RNN 的出现。循环神经网络可以对于图像序列、视频进行处理。在 RNN 之后，2015 年又在此基础上提出了 LSTM(Long Short Term Memory) 模型，该模型解决了 RNN 随着训练时间的加长以及网络层数增多导致的梯度爆炸和梯度消失等问题。LSTM 在文本生成、机器翻译、语言识别、视频标记等应用场景具有很大优势，而自然语言处理也成为了人工智能研究的重要领域。

而在 2017 年，*Attention Is All You Need* 发布，提出了 Transformer 模型，统一了图像处理和自然语言处理模型，几乎所有任务都可以基于该模型进行实现，现在有关 Transformer 的研究也还在不断进行中。

结合以上技术，我们可以完成现在绝大多数 Fake Meida 的制作，比如：AI 翻唱，将原曲歌手的声线替换为其他人的声线；AI 作图，输入一段文字生成相应的图像；AI 换脸，将视频或图像中的人脸替换为其他人的人脸等等。

2.2 良性应用实例以及对人类发展的意义

Fake Media 的发展伴随着的是数字图像处理、自然语言处理等技术的发展，这些技术的使用已经很大程度上融入了我们的日常生活。

例如将真实人脸映射到动漫人物上的 Live2D 技术，催生了 Vtuber、Vup 这一新兴的直播模式。将直播和二次元结合开辟了新的直播赛道，在商业上取得了极大的成功。还有近些年来开始逐渐使用的 AI 配音，我们可以利用 AI 技术对某个人的声音特征进行提取，并用该声线进行配音任务，目前已经有游戏厂商将这一技术用于实践。在有声书领域，AI 配音已经可以被广泛见到，并且有调查说明观众中几乎很少有人能分辨出哪些是真人配音，哪些是 AI 配音。

从以上例子可以看出，Fake Meida 的发展以及使用有广阔的前景，并且具有开辟新赛道的能力，这对于经济和市场是一块巨大的油田，能够带动许多行业的发展。此外某种程

度上 Fake Media 的使用也可以为我们的日常生活带来便捷，并且个人创作的门槛降低，我们可以利用这些技术以图像、视频、声音等形式完成我们心中的内容创作。

2.3 恶意使用实例以及对人类社会产生的危害

利用 GAN 模型，能轻松地将一类图片转换成另一类图片。利用 AI 技术，已经能够伪造极度仿真的人物声音、神态及动作，生成难以识别的虚假视频、图像。不法分子可以利用 AI 技术伪造出相关人员的样貌，进而冒用其身份骗过身份认证系统，窃取钱财或机密数据

不久前，内蒙古包头市公安局电信网络犯罪侦查局发布一起使用智能 AI 技术进行电信诈骗的案件，福州市某科技公司法人代表郭先生 10 分钟内被骗 430 万元。4 月 20 日中午，郭先生的好友突然通过微信视频联系他，自己的朋友在外地竞标，需要 430 万保证金，且需要公对公账户过账，想要借郭先生公司的账户走账。基于对好友的信任，加上已经视频聊天核实了身份，郭先生没有核实钱款是否到账，就分两笔把 430 万转到了好友朋友的银行卡上。郭先生拨打好友电话，才知道被骗。骗子通过智能 AI 换脸和拟声技术，佯装好友对他实施了诈骗。

显然，随着 Fake Media 的发展，仅凭人脸、语音、指纹识别，已经不能严格确认一个人的身份。除了可能带来的财产危害，使用 Fake Media 创建有针对性的图片和视频，可能会扩大与侵犯隐私和操纵社交相关的威胁。同时也存在发布虚假新闻，对国家政治产生影响的可能。

2.4 防范恶意使用的技术手段、技术规范与法律建设情况

目前，Fake Media 背后的人工智能风险已然成为社会和政策关注的焦点。对于 Fake Media 的运用所引发的风险，需要从技术手段、伦理准则、治理模式、决策机制等多层次构建体系化的风险治理机制。

但动态最优政策的又具有滞后性的困境。政策制定是一个时间过程，从政策提议到一个完整的政策出台，包括政策咨询、政策论证等多项议程。任何新技术的发展政策制定都是基于某个时间点而对未来所作的决策，基本规则是依据现有信息对未来后果的认知来进行决策。此种决策在逻辑上存在一定的风险，未来是变动不居的，此时合理的政策未必在未来就合理，技术的快速发展极有可能会对现有决策基础造成冲击。如果决策时间过于冗长，政策制定机构间未有效沟通，就会出现时滞现象，从而在影响政策执行力的同时，造成执行结果偏离决策，增加公众利益损害的风险。而有关技术的快速发展，也也导致了当下相关技术规范和法规难以跟上技术进步的节奏。

目前提出的协同治理机制的实现路径是帮助政策不断更新的有效措施。首先是建立以政府为主导，由科学共同体、企业、公众、社会组织等多元主体组成的专家咨询委员会，尤

其是提高公众和非政府组织的参与力度，借力于协商使来自不同行业、阶层的专家关注相关技术的风险本质，以社会理性能接受的程度作为人工智能政策决策的基础，消解新技术对社会的各方面冲击。其次是建立面向专家咨询委员会的听证、咨询以及协商机制和有效的信息沟通机制，为各主体参与协商提供保障，如制定咨询委员会职责条例、定期信息通报制度等。最后是将网络、人工智能等新技术嵌入民主协商。在实践性、信任感和民主供给层面深刻改变协商民主的运行机制的同时，优化参与协商民主途径，降低协商的交易成本，提高协商民主的效率。如通过网络、数据挖掘、人工智能等技术手段收集、处理和分析公众意见，实践“网络协商”和“微协商”，不仅可弥补场外公众“缺场”，且可利用算法对数据求得最优结果，保障协商的有效性和可靠性。

3 我对 Fake Media 的认识

首先，伴随 Fake Media 有关技术的不断发展，社会的分工体系将逐渐改变。人们所从事的大量重复、单调的基础活动将被替换，例如基本的绘图作品可以利用 AI 生成，再用人力进行细节修改。这对于资本方是成本的降低，但对于劳动者是岗位的减少。

其次，Fake Media 也将带来隐私困境和私域数据产权侵犯。隐私原意是指不愿他人知晓的私人信息，且与他人和社会利益无关，即“不受干涉或免于侵害的独处的权利”。而进入以数据和算法为核心特征的智能社会，数据成为信息的一种表达方式，而智能社会的内在本性是不停“追逐”数据，即“从一切事物中提取尽可能多的数据，特别是关于个人特殊信息的数据”。面对一切可数据化、可计算的世界，包括人在内的一切变成数字、符号，人的声音、面容等特征都可以被“模仿”、可以被“创造”、可以被“编辑”。

总之，Fake Media 背后的技术没有好坏，但其如何应用是衡量其意义的关键。目前有关的防范手段以及法律建设相较于技术还有着严重的滞后，希望在未来我们能看到约束后的技术使用，为人类社会带来更多的价值和积极的影响。