音频深度伪造

音频深度伪造(也称为语音克隆或深度伪造音频) 是人工智能的产物[1]，用于创建令人信服的语音句子，听起来像特定的人在说他们没有说过的话。[2][3][4]这项技术最初是为改善人类生活而开发的各种应用。例如，它可以用于制作有声读物，[5]还可以帮助因喉咙疾病或其他医疗问题而失声的人恢复声音。[6][7]在商业上，它为许多机会打开了大门。这项技术还可以创建更加个性化的数字助理和听起来自然的文本转语音以及语音翻译服务。

欺诈事件

从 2020 年代初开始，音频深度伪造 (audio deepfakes) 就变得非常普遍，人们可以通过简单的移动设备或个人电脑进行此类伪造。[8]这些工具还被用于通过音频传播虚假信息。[3]这导致全球公众对使用音频深度伪造的副作用产生了网络安全担忧，包括其可能在基于音频的社交媒体平台上传播虚假信息和虚假信息。[9]人们可以将其用作一种合乎逻辑的语音欺骗技术，[10]并利用它们来操纵舆论，以进行宣传、诽谤或恐怖主义。每天都有大量的语音记录通过互联网传输，检测欺骗行为极具挑战性。[11]音频深度伪造攻击者的目标是个人和组织，包括政客和政府。[12]

2019 年，诈骗者利用人工智能冒充一家德国能源公司首席执行官的声音，指示其英国子公司的首席执行官转账22 万欧元。[13] 2020 年初，同样的技术冒充了一名公司董事，实施了一项精心策划的骗局，说服一名分行经理转账 3500 万美元。[14]

根据 2023 年McAfee 的一项全球调查，十分之一的人报告称自己曾遭受过 AI 语音克隆骗局的攻击；其中 77% 的人报告称自己因骗局而损失了金钱。[15][16]音频深度伪造还可能对金融机构目前使用的语音识别系统构成威胁。[17][18] 2023 年 3 月，美国联邦贸易委员会向消费者发出警告，禁止使用人工智能伪造陷入困境的家庭成员的声音来乞求钱财。[19]

2023 年 10 月，在英国工党于利物浦举行的会议开幕式上，一段深度伪造的工党领袖基尔·斯塔默（Keir Starmer）的音频被发布，其中错误地描绘了他辱骂工作人员并批评利物浦的场景。[20]同月，一段深度伪造的斯洛伐克政治家米哈尔·西梅奇卡（Michal Šimečka）的音频被错误地声称捕捉到了他讨论操纵即将到来的选举的方法。[21]

在2024 年新罕布什尔州民主党总统初选竞选期间，超过 20,000 名选民接到了人工智能冒充总统乔·拜登的自动电话，敦促他们不要投票。[22][23]新罕布什尔州总检察长表示，此举违反了州选举法，Life Corporation 和 Lingo Telecom 涉嫌参与其中。[24] 2024 年 2 月，美国联邦通信委员会禁止使用人工智能在自动电话中伪造声音。[25][26]同月，政治顾问史蒂夫·克雷默承认他以 500 美元的价格委托了这些电话。 他说，他想呼吁人们关注制定规则来管理人工智能在政治竞选中的使用。[27] 5 月，联邦通信委员会表示，克雷默通过冒充当地政治人物的电话号码违反了联邦法律，并提出 600 万美元的罚款。新罕布什尔州的四个县以压制选民和冒充候选人等重罪起诉了克莱默。[28]

类别

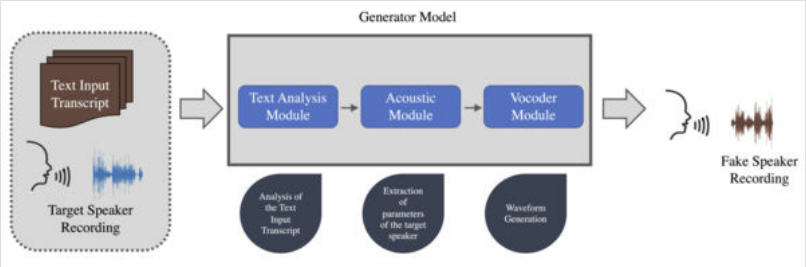
音频深度伪造可以分为三类：

基于重放

基于重放的深度伪造是恶意作品，旨在重现对话者声音的录音。[29]

有两种类型：远场检测和剪贴检测。在远场检测中，受害者的麦克风录音会作为测试片段在免提电话上播放。[30]另一方面，剪贴涉及伪造来自文本相关系统的请求句子。[11]基于文本的说话人验证可用于防御基于重放的攻击。[29][31]目前检测端到端重放攻击的技术是使用深度卷积神经网络。[32]

合成基



基于合成的方法图

基于语音合成的类别是指利用软件或硬件系统程序人工生成人类语音。语音合成包括文本转语音，旨在实时将文本转换为可接受的自然语音，[33]使用文本的语言描述规则，使语音听起来符合文本输入。

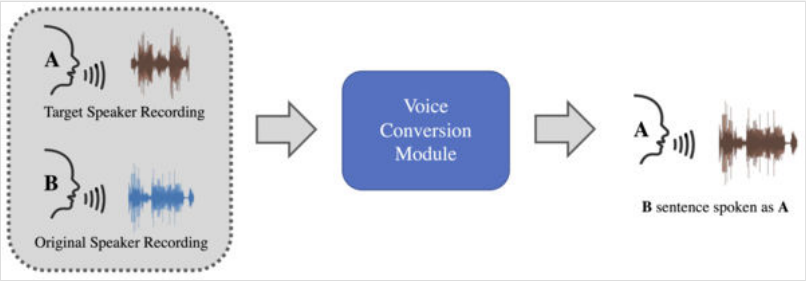
这种类型的经典系统由三个模块组成：文本分析模型、声学模型和声码器。生成通常必须遵循两个基本步骤。首先需要收集干净且结构良好的原始音频以及原始语音音频句子的转录文本。其次，必须使用这些数据训练文本转语音模型以构建合成音频生成模型。

具体来说，以目标说话人的声音转录的文本作为生成模型的输入。文本分析模块处理输入的文本并将其转换为语言特征。然后，声学模块根据文本分析模块生成的语言特征从音频数据中提取目标说话人的参数。[8]最后，声码器根据声学特征的参数学习创建声音波形。最终生成音频文件，其中包括波形格式的合成模拟音频，以许多说话人的声音创建语音音频，甚至包括那些不在训练中的说话人。

这方面的第一个突破是由WaveNet引入的[34]，这是一种用于生成原始音频波形的神经网络，能够模拟许多不同说话者的特征。多年来，该网络已被其他系统[35][36][37][38][39][40]所取代，这些系统可以合成每个人都能听到的高度逼真的人工声音。[41]

文本转语音在很大程度上取决于用于实现该系统的语音语料库的质量，而创建整个语音语料库的成本很高。另一个缺点是语音合成系统无法识别句号或特殊字符。此外，歧义问题一直存在，因为以相同方式书写的两个单词可能具有不同的含义。

模仿型



基于模仿的方法图

基于模仿的音频深度伪造是一种将一个说话人（原始人）的原始语音转换为另一个说话人（目标人）说话的方式。[42]基于模仿的算法将语音信号作为输入，通过改变其风格、语调或韵律来改变它，试图在不改变语言信息的情况下模仿目标语音。[43]这种技术也称为语音转换。

这种方法经常与之前的基于合成的方法相混淆，因为在生成过程方面，这两种方法之间没有明确的区分。事实上，这两种方法都会修改语音音频信号的声谱和风格特征，但基于模仿的方法通常保持输入和输出文本不变。这是通过改变这句话的朗读方式来匹配目标说话者的特征来实现的。[44]

模仿声音的方法有很多种，比如使用声音相似的人类模仿原说话者。近年来，最流行的方法是使用一种称为生成对抗网络 (GAN)的神经网络，因为它既灵活又能产生高质量的结果。[29][42]

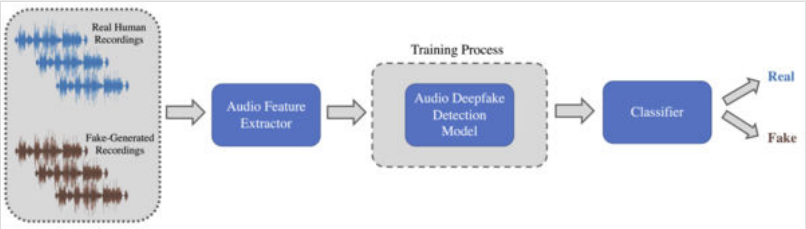
然后，使用模仿生成方法将原始音频信号转换为目标音频中的语音，从而生成新的语音，如假语音所示。

检测方法

音频深度伪造检测任务确定给定的语音音频是真实的还是假的。

最近，这已成为法医研究界的热门话题，以跟上伪造技术的快速发展。

一般来说，深度伪造检测方法可以根据其执行检测任务所利用的方面分为两类。第一种侧重于低级方面，寻找由生成器在样本级别引入的伪像。第二种则侧重于代表更复杂方面的高级特征，如语音录音的语义内容。



通用音频深度伪造检测框架

人们开发了许多机器学习和深度学习模型，采用不同的策略来检测假音频。大多数情况下，这些算法遵循三个步骤：

1. 每段语音录音都必须经过预处理，转换成适当的音频特征；

2. 计算出的特征被输入到检测模型中，该模型执行必要的操作，例如训练过程，这对于区分真实和虚假的语音音频至关重要；
3. *输出被输入到最终模块中，以产生假*类或*真*类的预测概率。按照 ASVspoof[45]*挑战命名法，假音频用术语“Spoof”*表示，而真音频则被称为*“Bonafide”*。

多年来，许多研究人员已经证明，无论使用什么特征，机器学习方法都比深度学习方法更准确。[8] 然而，由于过度训练和手动特征提取，机器学习方法的可扩展性尚未得到证实，尤其是在许多音频文件的情况下。相反，当使用深度学习算法时，需要对音频文件进行特定的转换，以确保算法能够处理它们。

有几种不同检测方法的开源实现，[46][47][48]，通常许多研究小组将它们发布在GitHub等公共托管服务上。

开放挑战和未来研究方向

音频深度伪造是一个非常新的研究领域。因此，它有很多发展和改进的可能性，以及采用这项技术可能给我们的日常生活带来的潜在威胁。下面列出了最重要的威胁。

Deepfake 生成

关于生成，最重要的方面是受害者的可信度，即音频深度伪造的感知质量。

有几个指标决定了音频深度伪造生成的准确度，其中最广泛使用的指标是MOS（平均意见得分），即用户评分的算术平均值。通常，要评分的测试涉及对不同语音生成算法所生成的句子的感知评估。该指标表明，由单个说话者训练的算法生成的音频具有更高的 MOS。[44][34][49][50][39]

采样率在检测和生成音频深度伪造方面也起着至关重要的作用。目前，可用数据集的采样率约为 16 kHz，这大大降低了语音质量。提高采样率可以提高生成质量。[37]

Deepfake 检测

重点关注检测部分，影响近期模型的一个主要弱点是所采用的语言。

大多数研究集中于检测英语中的音频深度伪造，而没有太多关注中文和西班牙语等使用最广泛的语言[51]以及印地语和阿拉伯语。

还必须考虑更多与不同口音相关的因素，这些因素代表与特定个人、地点或国家密切相关的发音方式。在音频的其他领域，例如说话人识别，口音被发现对性能有显著影响，[52]因此，即使在这种检测任务中，这一特征也可能影响模型的性能。

此外，对音频数据进行过度预处理会导致非常高且通常难以承受的计算成本。因此，许多研究人员建议采用自监督学习方法[53]，处理未标记数据以有效地完成检测任务并提高模型的可扩展性，同时降低计算成本。

使用真实音频数据训练和测试模型仍是一个尚未开发的领域。事实上，使用具有真实世界背景噪音的音频可以提高假音频检测模型的稳健性。

此外，大部分努力都集中在检测基于合成的音频深度伪造，很少有研究分析基于模仿的音频，因为它们在生成过程中存在内在困难。[11]

防御深度伪造

多年来，用于防御音频深度伪造可能带来的恶意行为（如身份盗窃和操纵国家州长的演讲）的技术不断增加。

为了防止深度伪造，一些人建议使用区块链和其他分布式账本技术 (DLT) 来识别数据来源并追踪信息。[8][54][55][56]

为了对抗深度伪造，人们还提出了从数字内容中提取和比较与感知情绪相对应的情感线索的方法。[57][58][59]

另一个关键方面涉及如何缓解这一问题。有人建议，最好只为需要它们的人保留一些专有检测工具，例如记者的事实核查工具。[29]这样，那些创建生成模型的人（也许是为了邪恶的目的）就不会确切知道哪些特征有助于检测深度伪造，[29]从而阻止潜在的攻击者。

为了改进检测效果，研究人员正在尝试推广这一过程，[60]寻找能够提高性能的预处理技术，并测试用于训练的不同损失函数。[10][61]

研究项目

全球各地有许多研究小组致力于识别媒体操纵行为，例如音频深度伪造、图像和视频深度伪造。这些项目通常由公共或私人资金支持，并与大学和研究机构保持密切联系。

为此，美国国防高级研究计划局 (DARPA)运行了语义取证 (SemaFor) 项目。[62][63]利用同样来自 DARPA 的媒体取证 (MediFor)[64][65]项目的一些研究成果，这些语义检测算法将必须确定媒体对象是否已被生成或操纵，以自动分析媒体来源并揭示伪造各种内容背后的意图。[66][62]

另一项研究项目是“人工智能时代媒体可信度维护”（PREMIER）[67]计划，由意大利教育、大学和研究部 (MIUR)资助，并由五所意大利大学运营。PREMIER 将采用新颖的混合方法来获得更易于解释和更安全的取证检测器。[68]

DEEP-VOICE[69]是一个公开可用的数据集，旨在用于研究目的，以开发系统来检测语音是否由神经网络通过基于检索的语音转换 (RVC) 的过程生成。初步研究表明，人类语音中的特征与人工智能算法生成的特征之间存在许多统计学上显著的差异。

公众挑战

在过去的几年中，已经组织了无数的挑战来进一步推动音频深度伪造研究领域的发展。

最著名的世界挑战赛是 ASVspoof[45]，即自动说话人验证欺骗和对策挑战赛。该挑战赛每两年举办一次，由社区主导，旨在促进对欺骗的思考和对策的开发。[70]

另一项近期的挑战是 ADD[71] ——音频深度伪造检测，它考虑更现实生活中的假象。[72]

语音转换挑战赛[73]也是每两年举办一次的挑战赛，其目的是使用相同的语音数据比较不同的语音转换系统和方法。

参见

- 15.人工智能
- Deepfake
- 深度学习
- 数字克隆
- 数字信号处理
- 人工智能
- 语音分析
- 语音识别
- 语音合成

参考

1. 史密斯，汉娜；曼斯特德，凯瑟琳 (2020 年 4 月 1 日)。*武器化的深度伪造：国家安全与民主*。第 28 卷。澳大利亚战略政策研究所。第 11-13 页。ISSN 2209-9689 (https://search.worldcat.org/issn/2209-9689)。
2. Lyu, Siwei (2020)。 “ Deepfake 检测：当前挑战和后续步骤” (https://ieeexplore.ieee.org/document/9105991)。2020*年 IEEE 国际多媒体会议和博览会研讨会 (ICMEW)*。第 1-6 页。arXiv ：2003.09234。doi (https://arxiv.org/abs/2003.09234)：10.1109 (https://doi.org/10.1109%2Ficmew46912.2020.9105991)/ icmew46912.2020.9105991。ISBN (https://ieeexplore.ieee.org/document/9105991) (https://arxiv.org/abs/2003.09234) (https://doi.org/10.1109%2Ficmew46912.2020.9105991) 978-1-7281-1485-9。S2CID 214605906。 检索日期2022-06-29 (https://api.semanticscholar.org/CorpusID:214605906)。
3. Diakopoulos, Nicholas；Johnson, Deborah (2020 年 6 月)。 “预测和解决选举背景下深度伪造的伦理影响” (http://journals.sagepub.com/doi/10.1177/1461444820925811)。新*媒体与社会*。23**7** (2020-06-05 出版) ：2072–2098。doi (https://api.semanticscholar.org/CorpusID:226196422)：10.1177/1461444820925811。ISSN (https://doi.org/10.1177%2F1461444820925811)1461-4448。S2CID226196422 (https://search.worldcat.org/issn/1461-4448)。 (http://journals.sagepub.com/doi/10.1177/1461444820925811) (https://doi.org/10.1177%2F1461444820925811) (https://search.worldcat.org/issn/1461-4448) (https://api.semanticscholar.org/CorpusID:226196422)
4. 玛吉·墨菲 (2024 年 2 月 20 日) 。 “Deepfake Audio Boom 利用了价值 10 亿美元的初创公司的人工智能” (https://www.bloomberg.com/news/articles/2024-02-21/biden-deepfake-and-other-audio-fakes-were-made-with-elevenlabs-ai)。布隆伯格。 (https://www.bloomberg.com/news/articles/2024-02-21/biden-deepfake-and-other-audio-fakes-were-made-with-elevenlabs-ai)

5. Chadha, Anupama；Kumar, Vaibhav；Kashyap, Sonu；Gupta, Mayank (2021)，Singh, Pradeep Kumar；Wierzchoń, Sławomir T.；Tanwar, Sudeep；Ganzha, Maria (eds.)，《Deepfake：概述》 (https://link.springer.com/10.1007/978-981-16-0733-2_39)，*第二届计算、通信和网络安全国际会议论文集*，网络与系统讲义，第 203 卷，新加坡：Springer Singapore，第 557-566 页，doi：10.1007/978-981-16-0733-2_39 (https://doi.org/10.1007%2F978-981-16-0733-2_39)，ISBN (https://link.springer.com/10.1007/978-981-16-0733-2_39) (https://doi.org/10.1007%2F978-981-16-0733-2_39) 978-981-16-0732-5，S2CID 236666289 (https://api.semanticscholar.org/CorpusID:236666289)，检索日期 2022-06-29
6. 基尔默 (https://search.worldcat.org/issn/0190-8286)重拾声音。但批评者担心该技术可能被滥用” (https://www.washingtonpost.com/technology/2021/08/18/val-kilmer-ai-voice-cloning/)。*《华盛顿邮报》*。ISSN 0190-8286 。检索日期：2022-06-29。 (https://www.washingtonpost.com/technology/2021/08/18/val-kilmer-ai-voice-cloning/) (https://search.worldcat.org/issn/0190-8286)
7. Etienne, Vanessa (2021 年 8 月 19 日)。 “瓦尔·基尔默在与喉癌抗争后利用*人工智能*技术恢复声音：听听结果” (https://people.com/movies/val-kilmer-gets-his-voice-back-after-throat-cancer-battle-using-ai-technology-hear-the-results/)。PEOPLE.com 。检索于 2022-07-01。 (https://people.com/movies/val-kilmer-gets-his-voice-back-after-throat-cancer-battle-using-ai-technology-hear-the-results/)
8. Almutairi, Zaynab； (https://search.worldcat.org/issn/1999-4893)**2022-05-04**)。 “现代音频深度伪造检测方法回顾：挑战和未来方向” (https://doi.org/10.3390%2Fa15050155)。*算法。15*(5): 155。doi： 10.3390/a15050155。ISSN (https://doi.org/10.3390%2Fa15050155) 1999-4893。 (https://doi.org/10.3390%2Fa15050155) (https://doi.org/10.3390%2Fa15050155) (https://search.worldcat.org/issn/1999-4893)
9. Caramancion, Kevin Matthe (2022 年 6 月)。 “对播客中传播的音频格式中的错误/虚假信息的探索：以 Spotify 为例” (https://ieeexplore.ieee.org/document/9795760)。*2022年 IEEE 国际物联网、电子和机电一体化会议 (IEMTRONICS)* 。第1-6页。doi ：10.1109/IEMTRONICS55184.2022.9795760。ISBN (https://doi.org/10.1109%2FIEMTRONICS55184.2022.9795760) (https://ieeexplore.ieee.org/document/9795760) (https://doi.org/10.1109%2FIEMTRONICS55184.2022.9795760) 978-1-6654-8684-2.S2CID 249903722 。 (https://api.semanticscholar.org/CorpusID:249903722)
10. Chen, Tianxiang；Kumar, Avrosh；Nagarsheth, Parav；Sivaraman, Ganesh；Khoury, Elie (2020-11-01)。 “Deepfake检测的泛化” (https://www.isca-speech.org/archive/odyssey_2020/chen20_odyssey.html)。*说话人和语言识别研讨会 (Odyssey 2020)*。ISCA：132–137。doi：10.21437/ (https://doi.org/10.21437%2FOdyssey.2020-19)Odyssey.2020-19。S2CID219492826 (https://api.semanticscholar.org/CorpusID:219492826)。 (https://www.isca-speech.org/archive/odyssey_2020/chen20_odyssey.html) (https://doi.org/10.21437%2FOdyssey.2020-19) (https://api.semanticscholar.org/CorpusID:219492826)
11. Ballesteros，多拉·M.；罗德里格斯-奥尔特加、约翰娜；伦扎，迭戈；贡萨洛阿尔塞 (2021-12-01)。 “Deep4SNet：用于虚假语音分类的深度学习” (https://www.sciencedirect.com/science/article/pii/S0957417421008770)。*具有应用程序的专家系统*。**184**：115465。doi：10.1016/ (https://doi.org/10.1016%2Fj.eswa.2021.115465)j.eswa.2021.115465。ISSN0957-4174 (https://search.worldcat.org/issn/0957-4174)。S2CID237659479 (https://api.semanticscholar.org/CorpusID:237659479)。 (https://www.sciencedirect.com/science/article/pii/S0957417421008770) (https://doi.org/10.1016%2Fj.eswa.2021.115465) (https://search.worldcat.org/issn/0957-4174) (https://api.semanticscholar.org/CorpusID:237659479)

12. , Supasorn；Seitz, Steven M.；Kemelmacher-Shlizerman, Ira (2017-07-20)。"合成奥巴马：从音频中学习口型同步" (https://doi.org/10.1145/3072959.3073640)。ACM图形学报。36 (4): (https://search.worldcat.org/issn/0730-0301)95 (https://api.semanticscholar.org/CorpusID:207586187) : 1–95:13。doi ：10.1145/ (https://doi.org/10.1145%2F3072959.3073640) 3072959.3073640。ISSN 0730-0301。S2CID 207586187。 (https://doi.org/10.1145/3072959.3073640) (https://doi.org/10.1145%2F3072959.3073640) (https://search.worldcat.org/issn/0730-0301) (https://api.semanticscholar.org/CorpusID:207586187)
13. Stupp, Catherine。"诈骗者利用人工智能模仿首席执行官的声音，引发不寻常的网络犯罪案件" (https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402)。*《华尔街日报》*。检索日期：2024-05-26。 (https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402)
14. Brewster, Thomas。"警方发现，诈骗者在 3500 万美元银行抢劫案中克隆了公司董事的声音" (https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/)。*《福布斯》*。检索日期：2022-06-29。 (https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/)
15. "生成式人工智能让语音骗局更容易被相信" (https://www.axios.com/2023/06/13/generative-ai-voice-scams-easier-identity-fraud)。Axios。2023年 6 月 13 日。*2023*年6 月 16 日检索。 (https://www.axios.com/2023/06/13/generative-ai-voice-scams-easier-identity-fraud)
16. Bunn, Amy (2023 年 5 月 15 日)。"人工智能冒名顶替者——网络犯罪分子利用人工智能语音克隆进行新型诈骗" (https://www.mcafee.com/blogs/privacy-identity-protection/artificial-imposters-cybercriminals-turn-to-ai-voice-cloning-for-a-new-breed-of-scam/)。*迈克菲博客。2023*年6 月 16 日检索。 (https://www.mcafee.com/blogs/privacy-identity-protection/artificial-imposters-cybercriminals-turn-to-ai-voice-cloning-for-a-new-breed-of-scam/)
17. 考克斯，约瑟夫 (2023 年 2 月 23 日)。"我如何利用人工智能生成的声音闯入银行账户" (https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-ai-generated-voice)。Vice。2023年*6*月 16 日检索。 (https://www.vice.com/en/article/dy7axa/how-i-broke-into-a-bank-account-with-an-ai-generated-voice)
18. Evershed，Nick；Taylor，Josh (2023 年 3 月 16 日)。"人工智能可以欺骗 Centrelink 和澳大利亚税务局用于验证身份的语音识别" (https://www.theguardian.com/technology/2023/mar/16/voice-system-used-to-verify-identity-by-centrelink-can-be-fooled-by-ai)。*《卫报》*。2023 年6 月 16 日检索。 (https://www.theguardian.com/technology/2023/mar/16/voice-system-used-to-verify-identity-by-centrelink-can-be-fooled-by-ai)
19. "诈骗者利用人工智能增强家庭应急计划" (https://consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes)。*消费者建议*。2023-03-17 。检索日期：2024-05-26。 (https://consumer.ftc.gov/consumer-alerts/2023/03/scammers-use-ai-enhance-their-family-emergency-schemes)
20. "工党会议第一天发布了基尔·斯塔默爵士的 Deepfake 音频" (https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181)。 (https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181)

21. Meaker, Morgan。 “斯洛伐克选举深度伪造表明人工智能对民主构成威胁” (https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/)。*连线*。 (https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/)
22. “假冒拜登人工智能自动电话背后的政治顾问在新罕布什尔州面临指控” (https://www.cnn.com/2024/05/23/politics/new-hampshire-ai-robocall-biden-charges/index.html)。 (https://www.cnn.com/2024/05/23/politics/new-hampshire-ai-robocall-biden-charges/index.html)
23. “政治顾问被指控雇用魔术师向选民发送拜登深度伪造电话” (https://lawandcrime.com/high-profile/political-consultant-hired-transient-magician-to-spam-voters-with-deepfake-calls-using-joe-bidens-voice-urging-them-not-to-cast-ballot-lawsuit/)。*法律与犯罪*。2024-03-15 。检索于2024-05-23。 (https://lawandcrime.com/high-profile/political-consultant-hired-transient-magician-to-spam-voters-with-deepfake-calls-using-joe-bidens-voice-urging-them-not-to-cast-ballot-lawsuit/)
24. 大卫·赖特；布莱恩·冯；布莱恩·冯 (2024 年 2 月 6 日)。 “新罕布什尔州总检察长宣布，拜登的假自动拨号电话与德克萨斯州的公司有关” (https://www.cnn.com/2024/02/06/tech/nh-ag-robocall-update/index.html)。CNN。 (https://www.cnn.com/2024/02/06/tech/nh-ag-robocall-update/index.html)
25. Brian Fung (2024 年 2 月 8 日)。 “FCC 投票禁止使用人工智能生成的声音的诈骗自动呼叫” (https://www.cnn.com/2024/02/08/tech/fcc-scam-robocalls-ai-generated-voices/index.html)。CNN。 (https://www.cnn.com/2024/02/08/tech/fcc-scam-robocalls-ai-generated-voices/index.html)
26. *FCC*规定自动拨号中人工智能生成的声音为非法 | 联邦通信委员会” (https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal)。www.fcc.gov。2024-02-08 。检索于2024-05-26。 (https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal)
27. Kramer, Marcia (2024-02-26)。 “史蒂夫·克莱默解释为何他使用人工智能在新罕布什尔州冒充拜登总统 - CBS 纽约” (https://www.cbsnews.com/newyork/news/steve-kramer-explains-why-he-used-ai-to-impersonate-president-biden-in-new-hampshire/)。*www.cbsnews.com*。检索于2024-05-23。 (https://www.cbsnews.com/newyork/news/steve-kramer-explains-why-he-used-ai-to-impersonate-president-biden-in-new-hampshire/)
28. “一名政治顾问因拜登深度伪造自动电话而面临指控和罚款” (https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative)。 (https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative)
29. Khanjani, Zahra；Watson, Gabrielle；Janeja, Vandana P. (2021-11-28)。 “伪造程度有多深？重点关注音频 Deepfake：一项调查” 。arXiv: 2111.14203 (https://arxiv.org/abs/2111.14203) [cs.SD (https://arxiv.org/archive/cs.SD)]。 (https://arxiv.org/abs/2111.14203) (https://arxiv.org/archive/cs.SD)
30. ；Qiu, Lili (2019-09-09)。 “打击针对语音助手的重放攻击” (https://doi.org/10.1145/3351258)。ACM交互式、移动、可 (https://api.semanticscholar.org/CorpusID:202159551)*穿戴和普适技术论文集*。3 (3): 100:1–100:26。doi : 10.1145/ (https://doi.org/10.1145%2F3351258) 3351258。S2CID 202159551。 (https://doi.org/10.1145/3351258) (https://doi.org/10.1145%2F3351258) (https://api.semanticscholar.org/CorpusID:202159551)

31. Villalba, Jesus；Lleida, Eduardo（2011）。“防止对说话人验证系统的重放攻击”。2011年 *Carnahan 安全技术会议* (https://ieeexplore.ieee.org/document/6095943)。第 1-8 页。doi：10.1109/ (https://doi.org/10.1109%2FCCST.2011.6095943)CCST.2011.6095943。ISBN (https://ieeexplore.ieee.org/document/6095943) (https://doi.org/10.1109%2FCCST.2011.6095943) 978-1-4577-0903-6。S2CID 17048213。 检索日期2022-06-29 (https://api.semanticscholar.org/CorpusID:17048213)。
32. Tom (https://api.semanticscholar.org/CorpusID:52187155) , Francis；Jain, Mohit；Dey, Prasenjit (2018-09-02)。 “使用带注意力机制的深度卷积网络进行端到端音频重放攻击检测” (https://www.isca-speech.org/archive/interspeech_2018/tom18_interspeech.html)。Interspeech *2018*。*ISCA*：681–685。doi ：10.21437 /Interspeech.2018-2279。S2CID (https://doi.org/10.21437%2FInterspeech.2018-2279)52187155。 (https://www.isca-speech.org/archive/interspeech_2018/tom18_interspeech.html) (https://doi.org/10.21437%2FInterspeech.2018-2279) (https://api.semanticscholar.org/CorpusID:52187155)
33. Tan, Xu; Qin, Tao; Soong, Frank; Liu, Tie-Yan (2021-07-23). "A Survey on Neural Speech Synthesis". arXiv:2106.15561 (https://arxiv.org/abs/2106.15561) [eess.AS (https://arxiv.org/archive/eess.AS)].
34. Oord, Aaron van den; Dieleman, Sander; Zen, Heiga; Simonyan, Karen; Vinyals, Oriol; Graves, Alex; Kalchbrenner, Nal; Senior, Andrew; Kavukcuoglu, Koray (2016-09-19). "WaveNet: A Generative Model for Raw Audio". arXiv:1609.03499 (https://arxiv.org/abs/1609.03499) [cs.SD (https://arxiv.org/archive/cs.SD)].
35. Kuchaiev, Oleksii; Li, Jason; Nguyen, Huyen; Hrinchuk, Oleksii; Leary, Ryan; Ginsburg, Boris; Kriman, Samuel; Beliaev, Stanislav; Lavrukhin, Vitaly; Cook, Jack; Castonguay, Patrice (2019-09-13). "NeMo: a toolkit for building AI applications using Neural Modules". arXiv:1909.09577 (https://arxiv.org/abs/1909.09577) [cs.LG (https://arxiv.org/archive/cs.LG)].
36. Wang, Yuxuan; Skerry-Ryan, R. J.; Stanton, Daisy; Wu, Yonghui; Weiss, Ron J.; Jaitly, Navdeep; Yang, Zongheng; Xiao, Ying; Chen, Zhifeng; Bengio, Samy; Le, Quoc (2017-04-06). "Tacotron: Towards End-to-End Speech Synthesis". arXiv:1703.10135 (https://arxiv.org/abs/1703.10135) [cs.CL (https://arxiv.org/archive/cs.CL)].
37. Prenger, Ryan; Valle, Rafael; Catanzaro, Bryan (2018-10-30). "WaveGlow: A Flow-based Generative Network for Speech Synthesis". arXiv:1811.00002 (https://arxiv.org/abs/1811.00002) [cs.SD (https://arxiv.org/archive/cs.SD)].
38. Vasquez, Sean; Lewis, Mike (2019-06-04). "MelNet: A Generative Model for Audio in the Frequency Domain". arXiv:1906.01083 (https://arxiv.org/abs/1906.01083) [eess.AS (https://arxiv.org/archive/eess.AS)].
39. Ping, Wei; Peng, Kainan; Gibiansky, Andrew; Arik, Sercan O.; Kannan, Ajay; Narang, Sharan; Raiman, Jonathan; Miller, John (2018-02-22). "Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning". arXiv:1710.07654 (https://arxiv.org/abs/1710.07654) [cs.SD (https://arxiv.org/archive/cs.SD)].
40. Ren, Yi; Ruan, Yangjun; Tan, Xu; Qin, Tao; Zhao, Sheng; Zhao, Zhou; Liu, Tie-Yan (2019-11-20). "FastSpeech: Fast, Robust and Controllable Text to Speech". arXiv:1905.09263 (https://arxiv.org/abs/1905.09263) [cs.CL (https://arxiv.org/archive/cs.CL)].
41. Ning, Yishuang; He, Sheng; Wu, Zhiyong; Xing, Chunxiao; Zhang, Liang-Jie (January 2019). "A Review of Deep Learning Based Speech Synthesis" (https://doi.org/10.3390%2Fapp9194050). *Applied Sciences*. **9** (19): 4050. doi:10.3390/app9194050 (https://doi.org/10.3390%2Fapp9194050). ISSN 2076-3417 (https://search.worldcat.org/issn/2076-3417).

42. Rodríguez-Ortega, Yohanna; Ballesteros, Dora María; Renza, Diego (2020). "A Machine Learning Model to Detect Fake Voice" (https://link.springer.com/chapter/10.1007/978-3-030-61702-8_1). In Florez, Hector; Misra, Sanjay (eds.). *Applied Informatics*. Communications in Computer and Information Science. Vol. 1277. Cham: Springer International Publishing. pp. 3–13. doi:10.1007/978-3-030-61702-8_1 (https://doi.org/10.1007%2F978-3-030-61702-8_1). ISBN 978-3-030-61702-8. S2CID 226283369 (https://api.semanticscholar.org/CorpusID:226283369).
43. Zhang, Mingyang; Wang, Xin; Fang, Fuming; Li, Haizhou; Yamagishi, Junichi (2019-04-07). "Joint training framework for text-to-speech and voice conversion using multi-source Tacotron and WaveNet". arXiv:1903.12389 (https://arxiv.org/abs/1903.12389) [eess.AS (https://arxiv.org/archive/eess.AS)].
44. Sercan, Ö Arık; Jitong, Chen; Kainan, Peng; Wei, Ping; Yanqi, Zhou (2018). "Neural Voice Cloning with a Few Samples" (https://papers.nips.cc/paper/2018/hash/4559912e7a94a9c32b09d894f2bc3c82-Abstract.html). *Advances in Neural Information Processing Systems (NeurIPS 2018)*. **31** (published 12 October 2018): 10040–10050. arXiv:1802.06006 (https://arxiv.org/abs/1802.06006).
45. "| ASVspoof" (https://www.asvspoof.org/). *www.asvspoof.org*. Retrieved 2022-07-01.
46. *resemble-ai/Resemblyzer* (https://github.com/resemble-ai/Resemblyzer), Resemble AI, 2022-06-30, retrieved 2022-07-01
47. mendaxfz (2022-06-28), *Synthetic-Voice-Detection* (https://github.com/mendaxfz/Synthetic-Voice-Detection), retrieved 2022-07-01
48. HUA, Guang (2022-06-29), *End-to-End Synthetic Speech Detection* (https://github.com/ghuawhu/end-to-end-synthetic-speech-detection), retrieved 2022-07-01
49. Kong, Jungil; Kim, Jaehyeon; Bae, Jaekyoung (2020-10-23). "HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis". arXiv:2010.05646 (https://arxiv.org/abs/2010.05646) [cs.SD (https://arxiv.org/archive/cs.SD)].
50. Kumar, Kundan; Kumar, Rithesh; de Boissiere, Thibault; Gestin, Lucas; Teoh, Wei Zhen; Sotelo, Jose; de Brebisson, Alexandre; Bengio, Yoshua; Courville, Aaron (2019-12-08). "MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis". arXiv:1910.06711 (https://arxiv.org/abs/1910.06711) [eess.AS (https://arxiv.org/archive/eess.AS)].
51. Babbel.com; GmbH, Lesson Nine. "The 10 Most Spoken Languages In The World" (https://www.babbel.com/en/magazine/the-10-most-spoken-languages-in-the-world). *Babbel Magazine*. Retrieved 2022-06-30.
52. Najafian, Maryam; Russell, Martin (September 2020). "Automatic accent identification as an analytical tool for accent robust automatic speech recognition" (https://linkinghub.elsevier.com/retrieve/pii/S0167639317300043). *Speech Communication*. **122**: 44–55. doi:10.1016/j.specom.2020.05.003 (https://doi.org/10.1016%2Fj.specom.2020.05.003). S2CID 225778214 (https://api.semanticscholar.org/CorpusID:225778214).
53. Liu, Xiao; Zhang, Fanjin; Hou, Zhenyu; Mian, Li; Wang, Zhaoyu; Zhang, Jing; Tang, Jie (2021). "Self-supervised Learning: Generative or Contrastive" (https://ieeexplore.ieee.org/document/9462394). *IEEE Transactions on Knowledge and Data Engineering*. **35** (1): 857–876. arXiv:2006.08218 (https://arxiv.org/abs/2006.08218). doi:10.1109/TKDE.2021.3090866 (https://doi.org/10.1109%2FTKDE.2021.3090866). ISSN 1558-2191 (https://search.worldcat.org/issn/1558-2191). S2CID 219687051 (https://api.semanticscholar.org/CorpusID:219687051).

54. Rashid, Md Mamunur; Lee, Suk-Hwan; Kwon, Ki-Ryong (2021). "Blockchain Technology for Combating Deepfake and Protect Video/Image Integrity" (http://koreascience.or.kr/article/JAKO202125761199587.page). *Journal of Korea Multimedia Society*. **24** (8): 1044–1058. doi:10.9717/kmms.2021.24.8.1044 (https://doi.org/10.9717%2Fkmms.2021.24.8.1044). ISSN 1229-7771 (https://search.worldcat.org/issn/1229-7771).
55. Fraga-Lamas, Paula; Fernández-Caramés, Tiago M. (2019-10-20). "Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality". *IT Professional*. **22** (2): 53–59. arXiv:1904.05386 (https://arxiv.org/abs/1904.05386). doi:10.1109/MITP.2020.2977589 (https://doi.org/10.1109%2FMITP.2020.2977589).
56. Ki Chan, Christopher Chun; Kumar, Vimal; Delaney, Steven; Gochoo, Munkhjargal (September 2020). "Combating Deepfakes: Multi-LSTM and Blockchain as Proof of Authenticity for Digital Media" (https://ieeexplore.ieee.org/document/9311067). *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*. pp. 55–62. doi:10.1109/AI4G50087.2020.9311067 (https://doi.org/10.1109%2FAI4G50087.2020.9311067). ISBN 978-1-7281-7031-2. S2CID 231618774 (https://api.semanticscholar.org/CorpusID:231618774).
57. Mittal, Trisha; Bhattacharya, Uttaran; Chandra, Rohan; Bera, Aniket; Manocha, Dinesh (2020-10-12), "Emotions Don't Lie: An Audio-Visual Deepfake Detection Method using Affective Cues" (https://doi.org/10.1145/3394171.3413570), *Proceedings of the 28th ACM International Conference on Multimedia*, New York, NY, USA: Association for Computing Machinery, pp. 2823–2832, doi:10.1145/3394171.3413570 (https://doi.org/10.1145%2F3394171.3413570), ISBN 978-1-4503-7988-5, S2CID 220935571 (https://api.semanticscholar.org/CorpusID:220935571), retrieved 2022-06-29
58. Conti, Emanuele; Salvi, Davide; Borrelli, Clara; Hosler, Brian; Bestagini, Paolo; Antonacci, Fabio; Sarti, Augusto; Stamm, Matthew C.; Tubaro, Stefano (2022-05-23). "Deepfake Speech Detection Through Emotion Recognition: A Semantic Approach" (https://ieeexplore.ieee.org/document/9747186). *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore, Singapore: IEEE. pp. 8962–8966. doi:10.1109/ICASSP43922.2022.9747186 (https://doi.org/10.1109%2FICASSP43922.2022.9747186). hdl:11311/1220518 (https://hdl.handle.net/11311%2F1220518). ISBN 978-1-6654-0540-9. S2CID 249436701 (https://api.semanticscholar.org/CorpusID:249436701).
59. Hosler, Brian; Salvi, Davide; Murray, Anthony; Antonacci, Fabio; Bestagini, Paolo; Tubaro, Stefano; Stamm, Matthew C. (June 2021). "Do Deepfakes Feel Emotions? A Semantic Approach to Detecting Deepfakes Via Emotional Inconsistencies" (https://ieeexplore.ieee.org/document/9523100). *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Nashville, TN, USA: IEEE. pp. 1013–1022. doi:10.1109/CVPRW53098.2021.00112 (https://doi.org/10.1109%2FCVPRW53098.2021.00112). hdl:11311/1183572 (https://hdl.handle.net/11311%2F1183572). ISBN 978-1-6654-4899-4. S2CID 235679849 (https://api.semanticscholar.org/CorpusID:235679849).
60. Müller, Nicolas M.; Czempin, Pavel; Dieckmann, Franziska; Froghyar, Adam; Böttinger, Konstantin (2022-04-21). "Does Audio Deepfake Detection Generalize?". arXiv:2203.16263 (https://arxiv.org/abs/2203.16263) [cs.SD (https://arxiv.org/archive/cs.SD)].

61. Zhang, You; Jiang, Fei; Duan, Zhiyao (2021). "One-Class Learning Towards Synthetic Voice Spoofing Detection" (https://ieeexplore.ieee.org/document/9417604). *IEEE Signal Processing Letters*. **28**: 937–941. arXiv:2010.13995 (https://arxiv.org/abs/2010.13995). Bibcode:2021ISPL...28..937Z (https://ui.adsabs.harvard.edu/abs/2021ISPL...28..937Z). doi:10.1109/LSP.2021.3076358 (https://doi.org/10.1109%2FLSP.2021.3076358). ISSN 1558-2361 (https://search.worldcat.org/issn/1558-2361). S2CID 235077416 (https://api.semanticscholar.org/CorpusID:235077416).
62. "SAM.gov" (https://sam.gov/opp/a8883be78ac1442e8a22924011fc13c4/view). *sam.gov*. Retrieved 2022-06-29.
63. "The SemaFor Program" (https://www.darpa.mil/program/semantic-forensics). *www.darpa.mil*. Retrieved 2022-07-01.
64. "The DARPA MediFor Program" (https://govtribe.com/file/government-file/darpabaa1558-darpa-baa-15-58-medifor-dot-pdf). *govtribe.com*. Retrieved 2022-06-29.
65. "The MediFor Program" (https://www.darpa.mil/program/media-forensics). *www.darpa.mil*. Retrieved 2022-07-01.
66. “DARPA 宣布选定语义取证项目研究团队” (https://www.darpa.mil/news-events/2021-03-02)。*www.darpa.mil*。检索日期：2022-07-01。 (https://www.darpa.mil/news-events/2021-03-02)
67. “PREMIER” (https://sites.google.com/unitn.it/premier/)。*sites.google.com*。检索日期：2022-07-01。 (https://sites.google.com/unitn.it/premier/)
68. “PREMIER - 项目” (https://sites.google.com/unitn.it/premier/project)。sites.google.com 。检索日期：2022-06-29。 (https://sites.google.com/unitn.it/premier/project)
69. Bird, Jordan J.；Lotfi, Ahmad (2023)。 “实时检测 AI 生成的语音以进行 DeepFake 语音转换”。arXiv ： 2308.12734 [(https://arxiv.org/abs/2308.12734) cs.SD] (https://arxiv.org/archive/cs.SD)。 (https://arxiv.org/abs/2308.12734) (https://arxiv.org/archive/cs.SD)
70. 山岸淳一；王鑫;马西米利亚诺·托迪斯科；萨希杜拉，医学博士；帕蒂诺，何塞；安德烈亚斯·诺奇；刘雪辰;李康益;托米·金努宁；埃文斯，尼古拉斯；赫克托·德尔加多 (2021-09-01)。 “ASVspoof 2021：加速欺骗和深度伪造语音检测的进展”。arXiv：2109.00537 (https://arxiv.org/abs/2109.00537) [eess.AS (https://arxiv.org/archive/eess.AS)]。 (https://arxiv.org/abs/2109.00537) (https://arxiv.org/archive/eess.AS)
71. “音频 Deepfake 检测：ICASSP 2022” (https://signalprocessingsociety.org/publications-resources/data-challenges/audio-deepfake-detection-icassp-2022)。IEEE*信号处理协会*。2021-12-17 。检索日期：2022-07-01。 (https://signalprocessingsociety.org/publications-resources/data-challenges/audio-deepfake-detection-icassp-2022)
72. 易姜岩;付瑞波;陶建华；聂帅；马浩鑫;王成龙;王涛；田正坤;白叶;范存航；梁山 (2022-02-26)。 “ADD 2022：首届音频深度合成检测挑战赛”。arXiv：2202.08433 (https://arxiv.org/abs/2202.08433) [cs.SD (https://arxiv.org/archive/cs.SD)]。 (https://arxiv.org/abs/2202.08433) (https://arxiv.org/archive/cs.SD)
73. “2020 年暴雪挑战赛和语音转换挑战赛联合研讨会 - SynSIG” (https://web.archive.org/web/20220702035041/https://www.synsig.org/index.php/Joint_Workshop_for_the_Blizzard_Challenge_and_Voice_Conversion_Challenge_2020)。*www.synsig.org*。原件 (https://www.synsig.org/index.php/Joint_Workshop_for_the_Blizzard_Challenge_and_Voice_Conversion_Challenge_2020)存档于 2022-07-02 。检索于2022-07-01。 (https://web.archive.org/web/20220702035041/https://www.synsig.org/index.php/Joint_Workshop_for_the_Blizzard_Challenge_and_Voice_Conversion_Challenge_2020) (https://www.synsig.org/index.php/Joint_Workshop_for_the_Blizzard_Challenge_and_Voice_Conversion_Challenge_2020)