

2019 KCA Media Issue & Trend

01 · 3

트렌드리포트

딥페이크 기술의 빛과 그림자

1. 들어가며
2. 딥페이크 기술의 가능성
3. 딥페이크 기술의 폐해
4. 마치며

딥페이크 기술의 빛과 그림자

요약문

딥페이크(Deepfake) 기술이 영상 산업에 적극 활용되고 있다. 보다 쉽고 간편하게 특수효과를 만들어 내거나 AR 영상을 제작하는 등 산업 전반의 가능성을 키우고 있다. 하지만 한편으로는 가짜 뉴스 제작이나 성범죄에 악용되며 혼란을 야기하고 있기도 하다. 딥페이크의 명과 암을 살펴본다.

1. 들어가며

딥페이크(Deepfake)란, 인공지능 기술인 딥러닝(deep learning)과 ‘가짜’를 의미하는 단어인 페이크(fake)의 합성어로, 인공지능 기술을 이용하여 진위 여부를 구별하기 어려운 가짜 이미지나 영상물을 뜻한다. 딥페이크라는 단어가 등장한 시기는 2017년으로, 미국 온라인 커뮤니티 Reddit의 한 회원이 기존 영상에 유명인의 얼굴을 입혀 가짜 콘텐츠를 게재한 데서 유래되었다. 이후 온라인 커뮤니티와 SNS를 중심으로 급속히 확산된 딥페이크 콘텐츠는 최근 DeepFaceLab, Faceswap 등 오픈 소스 형태의 영상 합성 제작 프로그램이 배포되면서 더욱 성행하고 있다.



딥페이크의 기반은 2014년에 등장한 머신러닝 기술인 ‘적대관계생성신경망(Generative Adversarial Networks, 이하 GAN)’이다. GAN은 AI 모델을 생성모델¹과 분류모델²로 구분하며, 각 모델의 학습을 반복하는 과정을 거친다. 이러한 과정에서 생성모델과 분류모델은 서로를 적대적 경쟁자로 인식하며 상호 발전하게 된다. 생성모델은 실제와 유사한 데이터를 생성할 수 있게 되고, 분류모델은 데이터의 진위여부를 구별할 수 없게 되는 것이다. 이 과정을 반복하면서 원본 영상과의 구별이 거의 불가능할 정도로 정교한 합성 영상이 만들어진다.

그림 1 원본 영상(왼쪽)과 딥페이크 영상(오른쪽)

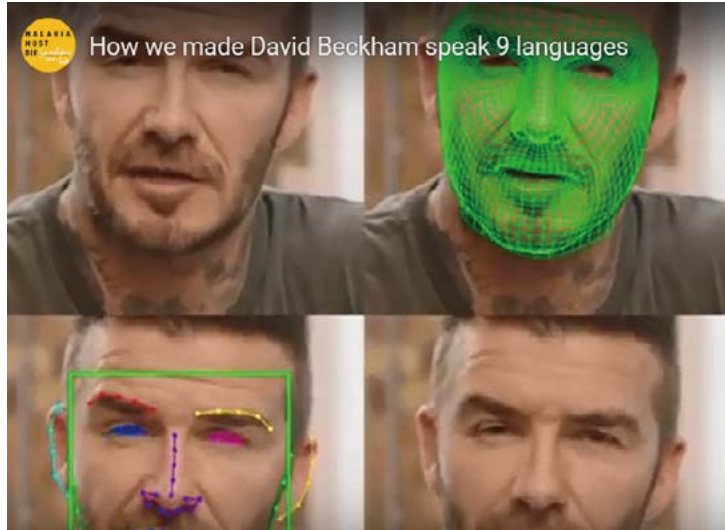


출처: TheNextWeb

딥페이크 콘텐츠 분석업체인 Deeptrace에 따르면, 딥페이크 관련 온라인 커뮤니티는 2019년 기준으로 20개, 이용자는 약 9만 5,000명에 달한다. 딥페이크는 실제 산업 현장에서도 널리 활용되고 있다. 영국 축구선수 출신인 David Beckham은 말라리아 퇴치 캠페인 홍보 영상에서 중국어, 힌디어, 아랍어 등 9개 언어를 구사하며 내용을 전달한다. 해당 영상 기술을 지원한 기업인 Synthesia는 딥페이크 기반의 더빙 기술을 활용하여 Beckham의 언어를 구현했다.

-
- 1) 실제와 유사한 데이터를 생성하려는 모델이다
 - 2) 데이터의 진위여부를 판별하는 모델이다

그림 2 다국어 더빙에 적용되는 딥페이크 기술



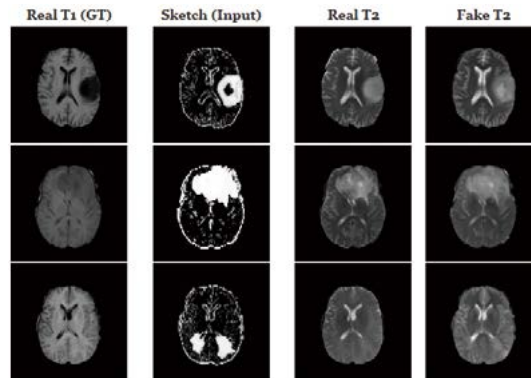
출처: Malaria Must Die

최근엔 의료계에서도 딥페이크 기술을 바탕으로 연구를 진행하고 있다. 2019년 7월에 독일 뤼벡대(University of Lübeck) 의료정보학연구소 연구진은 딥러닝 알고리즘의 이미지 패턴매칭 기능³⁾을 활용하여 질병 진단용 인공지능을 개발했다. 뤼벡대 연구진은 적대관계생성신경망(GAN)을 이용하여 원본 영상과 진위여부를 구별할 수 없는 정확도의 딥페이크 의료영상을 만들어냈다. 딥페이크 영상은 인공지능이 질병을 학습하고 정확히 진단할 수 있도록 딥러닝하는데 사용된다. 환자의 사생활 침해에 대한 우려와, 의료용 3D 이미지 합성에 필요한 비용 때문에 의료영상을 분석할 데이터가 충분치 않았는데, 딥페이크 기술로 이를 해결한 것이다. 딥페이크 기술은 영상뿐 아니라, 오디오, 사진, 문서 등에서 다방면으로 활용되고 있다.

3) · 의료용 이미지(X선, MRI, CT)를 분석하여 신체의 이상 징후를 감지하는 데 활용된다.



그림 3 딥페이크 기술이 적용된 의료영상



출처 : Ground AI

하지만 한편으로 여론을 조작하려 가짜 뉴스를 제작하거나 유명인의 얼굴을 무단 도용해 성인물을 만드는 등 그 위험성 또한 심화되고 있다. 유명 할리우드 배우들이 위조된 성인물 영상으로 피해를 보았으며, 국내에서도 온라인 커뮤니티와 SNS를 중심으로 유명인뿐 아니라 일반인까지 딥페이크의 악용 피해를 보고 있다. 최근에는 딥페이크 기술을 이용한 음성, 문서 위조 사건까지 발생하며 불안감이 고조되고 있다.

이처럼 정교한 딥페이크 기술은 미디어 업계에서 특수효과로 활용되어 콘텐츠 품질 개선에 기여할 수 있지만, 한편으로는 성인물과 가짜 뉴스 제작 등 각종 범죄의 도구로 쓰이고 있다. 이에 본고는 ‘양날의 검’으로 불리는 딥페이크 기술에 주목하여, 딥페이크 기술의 긍정적 방향과 악용 사례를 동시에 고찰한다.

2. 딥페이크 기술의 가능성

할리우드를 비롯한 영상 제작 업계에서는 딥페이크 기술로 특수효과를 만들어 내고 있다. 특히 과거를 재현하거나 더 이상 실존하지 인물을 그리고자 할 때 유용하게 활용되고 있다. 2016년에 개봉한 영화 <Rogue One>⁴에는 1977년 작 <Star Wars: A New Hope>에 출연한 배우가

4) <Star Wars> 시리즈의 스핀오프 영화이다.

당시의 모습 그대로 등장했다. 〈Rogue One〉 제작진은 배우와 외형적으로 유사한 대역 배우를 섭외한 후, 모션 캡처 기법과 딥페이크 기술을 활용하여 대역 배우의 얼굴에 과거 배우의 얼굴을 합성하는 방식으로 영화를 촬영했다. Netflix도 딥페이크 기술을 활용해 주연 배우의 현재와 과거 모습을 동시에 재현해 낸 〈The Irishman〉의 출시를 예고했다.

그림 4 1977년 작 〈Star Wars〉의 배우(왼쪽)와 2016년 작 〈Rogue One〉에서 구현된 배우의 모습(오른쪽)



출처: ZIMBIO

그림 5 주연배우의 현재와 과거를 동시에 구현한 〈The Irishman〉



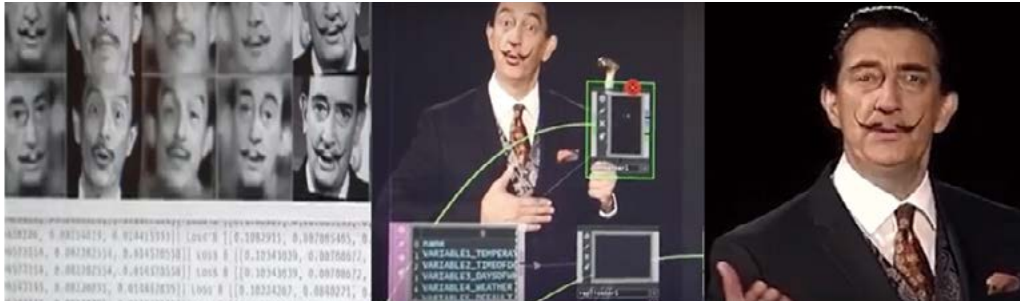
출처: IMDb

딥페이크 기술은 AR·VR 콘텐츠 제작에도 활용되고 있다. 미국 플로리다에 있는 화가 Dali 박물관은 미국의 광고사인 Goodby, Silverstein & Partners(GS&P)와 협업해 몇십 년 전에 작고한 예술가를 되살려냈다. GS&P는 Dali의 영상을 수집한 후 머신러닝 기법을 활용해 표정 알고리즘을 분석했다. 이후, Dali와 신체적 특징이 유사한 대역 배우를 섭외해 배우의 얼굴에 Dali의 얼굴·표정을 합성하고 Dali와 비슷한 성우의 목소리를 입혔다. 전시회 관람객들은 실물 크기의 키오스크 속에 등장하는 Dali와 촬영하고, 대화하는 등 AR Dali의 안내를 받으며 작품을 감상할 수 있다.





그림 6 딥페이크 기술을 활용하여 AR 영상으로 구현된 Dali의 모습



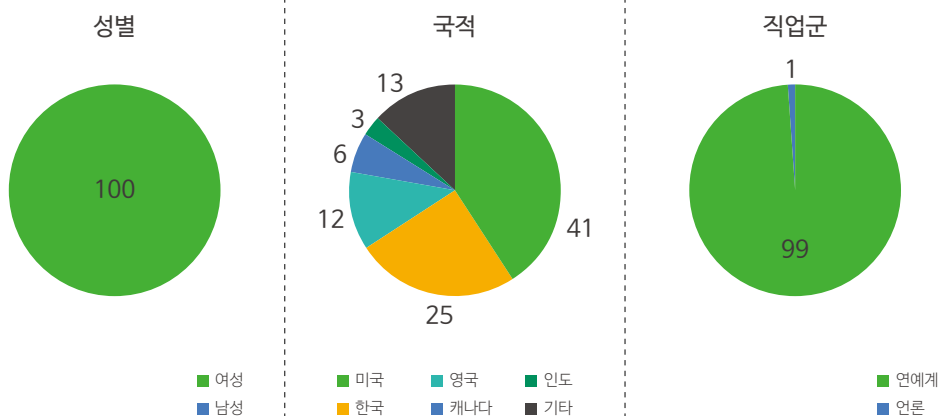
출처: Dali Lives

3. 딥페이크 기술의 폐해

한편, 유명인의 신체를 무단 도용해 만든 딥페이크 포르노도 확산되고 있다. Deeptrace에 따르면, 온라인에 퍼져 있는⁵ 딥페이크 영상은 2018년 12월 기준 7,964개에서 2019년에는 14,678건으로 전년도보다 84% 상승했다. 딥페이크 영상 중에서도 성인물이 전체의 96%를 점유하며 가장 큰 비중을 차지하고 있다. 딥페이크 성인용 콘텐츠를 중점적으로 게시하는 사이트 중에서 규모가 가장 큰 4개 사이트의 총 영상 조회 수는 1억 3,436만 4,438회에 달한다.

그림 7 딥페이크 성인용 영상물 대상의 특성

단위: %

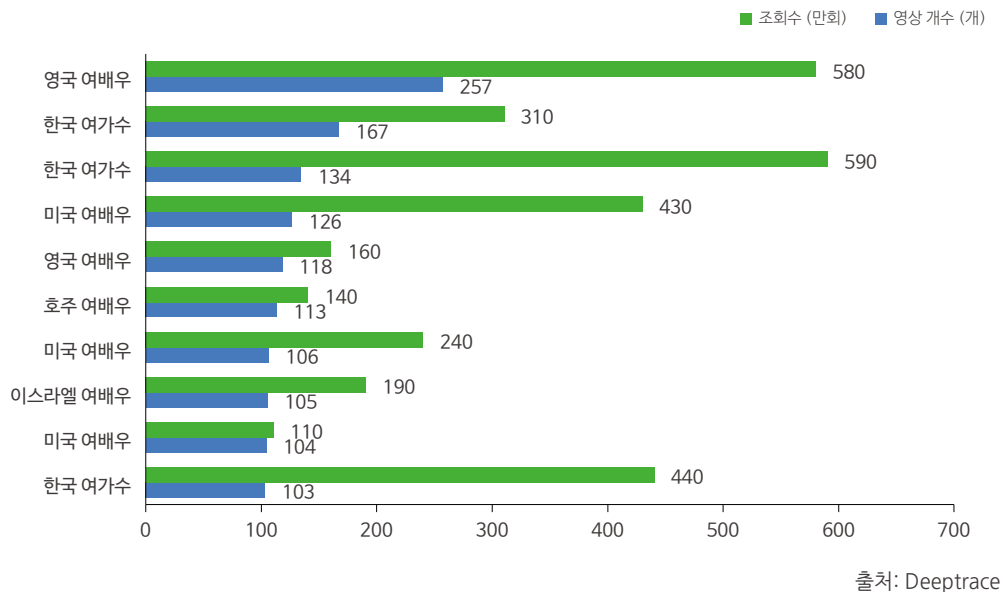


출처: Deeptrace

5) YouTube, Vimeo, LiveLeak, Dailymotion 등 동영상 전문 사이트와 Google Search를 통해서 찾은 결과물이다.

Deepttrace가 딥페이크 성인용 콘텐츠 전용 사이트 5개(규모가 가장 큰 사이트 5개이다)의 영상을 분석한 결과, 영상에 등장하는 특정 인물의 성별이 모두 여성이었다. 인물의 국적을 살펴보니 41%가 미국인, 25%가 한국인, 12%가 영국인, 6%가 캐나다인, 3%가 인도인이었으며, 인물의 99%가 연예계 종사자였다. 문제는 해당 영상의 대부분이 대상자의 허락을 받지 않고 불법 도용된 이미지로 파악된다는 것이다. 여성, 그리고 유명인에 대한 피해가 가장 큰 실정이다. 영국의 한 배우를 합성한 성인용 딥페이크 영상물이 257개가 유포되고, 조회수가 580만 회를 넘긴 사례도 있다. 불법 제작된 딥페이크 성인용 콘텐츠 제작 방지 방안을 마련하는 일이 시급하다고 볼 수 있다.

그림 8 각국 연예인이 등장하는 딥페이크 성인용 영상의 개수와 조회수



가짜 뉴스도 심각한 딥페이크 악용 사례이다. 유명인사의 얼굴을 합성한 딥페이크 뉴스가 난무함에 따라 정치적·사회적 불안감이 가중되고 있다. 2019년 9월, 이탈리아에서 Matteo Renzi 전 총리가 다른 정치인들을 모욕하는 딥페이크 영상이 한 프로그램에서 방영된 후, SNS를 통해 급속히 확산되었다. 해당 영상은 배우의 몸에 총리의 얼굴을 합성한 조작 영상이었지만, 영상의 내용을 실체처럼 인식한 일부 SNS 이용자가 총리를 향해 거센 비판을 제기하는 해프닝이 일어나기도 했다.



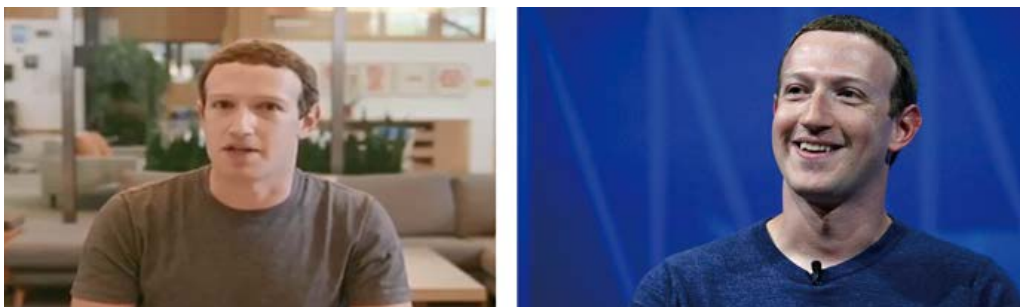
그림 9 Matteo Renzi 이탈리아 전 총리의 모습이 담긴 조작영상(왼쪽)과 실제 총리의 모습(오른쪽)



출처: Observers, Bloomberg

그보다 3개월 전인 2019년 6월에는 Facebook CEO인 Mark Zuckerberg가 등장하는 14초 분량의 영상이 Instagram에서 화제를 모았다. 영상에서 Mark Zuckerberg는 Facebook 이용자 데이터를 통제할 힘이 본인에게 있다는 과시적 내용의 발언으로 시청자의 뭇매를 맞았다. 이용자 데이터 유출사건으로 신뢰도 하락이라는 위기를 맞닥뜨린 Facebook의 CEO에게는 당치않은 발언이라는 비판이 줄을 이었다. 하지만 해당 영상은 실체가 아닌 조작된 영상이었으며, 영국의 한 디지털 예술가가 시 프로젝트의 일환으로 제작·게재한 영상이었던 것으로 드러났다.

그림 10 Mark Zuckerberg가 등장한 딥페이크 영상(왼쪽)과 실제 그의 모습(오른쪽)



출처: Cnet, Forbes

이처럼 딥페이크 기술 확산으로 진실과 거짓의 구분이 어려워지고 있다. 2019년 6월에는 말레이시아 경제부 장관의 동성애 장면이 담긴 영상이 알려지면서 말레이시아 정계가 둘로 나뉘어 갑론을박을 펼쳤다. 말레이시아는 동성애를 법적으로 금기시하기 때문에 장관을 체포해야 한다는 의견과, 해당 영상은 딥페이크 기술로 조작되었다는 주장이 첨예하게 다투었다. 하지만 국제 전문가의 검토 결과, 해당 영상에서는 딥페이크 기술의 흔적을 발견할 수 없었다.

딥페이크 기술이 스캔들의 진위여부에도 영향을 미치는 등, 사회적 파급력이 높아지는 가운데 딥페이크 기술에 대한 경고의 필요성이 높아지고 있다. 2018년 4월에 온라인 매체인 BuzzFeed와 Jordan Peele 감독은 딥페이크 기술의 위험성을 경고하기 위한 영상을 공동 제작했다. 해당 영상에는 Obama 미국 전 대통령이 등장하여 Trump 현 대통령을 향해 거침없는 비판을 행했다. Obama 대통령이 실제로 출연한 것처럼 보이는 해당 영상은 사실 교묘히 조작된 딥페이크 영상이었다. Obama 대통령의 모습을 영상 속에 합성해서 만들어 낸 결과물인 것이다. 제작진은 시청자들이 딥페이크 기술의 사회적, 정치적 파급력에 대해 인식할 만한 계기를 부여하고자 경고성 영상을 제작한 것이다.

그림 11 Obama 대통령의 조작 영상과(왼쪽) 영상을 제작한 Jordan Peele(오른쪽)



출처: Vox

Facebook과 Amazon, Google 등 글로벌 플랫폼도 딥페이크 기술의 악용을 저지하고자 나섰다. Facebook은 2019년 9월에 Microsoft, MIT, 옥스퍼드대학, 코넬공과대학 등 전 세계 우수 대학의 인공지능 전문가들과 손을 잡고 딥페이크 감지 기술 경연 대회인 Deepfake Detection Challenge(DFDC)를 개최하기로 했다. 해당 경연은 딥페이크 동영상 데이터셋을 이용하여 영상 해독 알고리즘을 개발하는 것이다. Facebook은 연구비·상금 수여 등 명목으로 해당 대회에 1,000만 달러를 투자하겠다고 발표했다. 2019년 10월에는 Amazon도 DFDC에 합류했다. Amazon은 영상 해독 알고리즘의 개발을 위해 향후 2년간 개발자들에게 100만 달러에 달하는 Amazon Web Services⁶ 이용권을 제공한다고 발표했다. Google은 3,000개의 딥페이크

6) Amazon이 운영하는 클라우드 컴퓨팅 솔루션 제공 플랫폼이다.



영상을 촬영한 후, 딥페이크 감지 기술 연구에 제공했으며, twitter는 조작된 콘텐츠를 공유하지 못하게 하거나, 조작된 콘텐츠가 허위임을 나타내는 링크를 삽입하는 등 딥페이크 악용 사례를 방지하는 정책을 펼치고 있다.

딥페이크 기술로 제작한 성인용 콘텐츠에 대한 법적 단속도 강화되는 추세이다. 유명인뿐만 아니라 일반인을 대상으로 한 딥페이크 콘텐츠의 불법 제작·유통을 방지하기 위함이다. 미국 Virginia 주(州)에서는 2019년 7월 1일부터 딥페이크 성인용 콘텐츠를 보복성 음란물(리벤지 포르노)의 영역에 포함한다는 내용의 법안이 발효되었다. 딥페이크 기술이 적용된 보복성 음란물을 제작·유통하면 최대 12개월의 징역과 2,500달러의 벌금이 부과된다. Virginia 주 정부는 개인적 원한 때문에 딥페이크 기술의 폐해를 입는 일반인이 발생하지 않도록 엄정히 처벌하겠다는 입장을 고수하고 있다.

4. 마치며

딥페이크 콘텐츠가 확산되면서 각종 부작용을 양산하고 있지만, 딥페이크 기술이 암(暗)만 가지고 있는 기술은 분명 아니다. 딥페이크 기술이 영상 산업에 활용되면서 특수효과, 더빙, 과거 재현 등이 더 쉽고 간편해졌다. 어느 기술이 그렇듯, 딥페이크 기술도 쓰이는 방향과 목적에 따라 명과 암을 가진다. 사실을 조작하고, 무고한 사람을 공격하는 강력한 도구가 될 수 있음은 분명하다. 이를 방지하고 대책을 마련할 수 있는 논의가 필요함은 물론이거니와, 이를 유용하게 활용할 수 있는 방향 또한 놓쳐서는 안 된다. 딥페이크 기술이 우리 사회의 신뢰를 떨어트리는 끔찍한 허점이 되지 않도록 경각심을 가지되, 무궁무진한 발전 가능성도 함께 고려해야 한다.

REFERENCES

1. Bloomberg (2019.09.17.) Italy's Renzi Seeks Kingmaker Role in Bid to Start New Party.
<https://www.bloomberg.com/news/articles/2019-09-16/italy-s-renzi-to-leave-democrats-to-form-own-party-media-report>
2. Cnet (2019.06.11.) Deepfake video of Facebook CEO Mark Zuckerberg posted on Instagram.
<https://www.cnet.com/news/black-friday-2019-best-deals-today-250-ipad-175-switch-lite-best-macbook-air-deal-of-the-year/>
3. CNN Business (2019.10.07.) The number of deepfake videos online is spiking. Most are porn.
<https://edition.cnn.com/2019/10/07/tech/deepfake-videos-increase/index.html>
4. Dali Lives 홈페이지.
<https://thedali.org/exhibit/dali-lives/>
5. Deeptrace (2019.09) The State of Deepfakes: Landscape, Threats, and Impact.
https://regmedia.co.uk/2019/10/08/deepfake_report.pdf
6. Forbes (2019.03.09.) Who Got Rich This Week: Mark Zuckerberg's Net Worth Surges By Nearly \$3 Billion.
<https://www.forbes.com/sites/hayleycuccinello/2019/03/09/who-got-rich-this-week-mark-zuckerbergs-net-worth-surges-by-nearly-3b/#12c58fee6ea0>
7. Ground AI (2019.07.02.) Multi-scale GANs for Memory-efficient Generation of High Resolution Medical Images.
<https://www.groundai.com/project/multi-scale-gans-for-memory-efficient-generation-of-high-resolution-medical-images/1>
8. IMDb 홈페이지.
<https://www.imdb.com/title/tt1302006/mediaviewer/rm2103088385>
<https://www.imdb.com/title/tt1302006/mediaviewer/rm3633482497>
<https://www.imdb.com/title/tt1302006/mediaviewer/rm221025537>
9. Malaria Must Die 홈페이지.
<https://www.malariamustdie.com/>
10. Observers (2019.10.08.) Deepfake video of former Italian PM Matteo Renzi sparks debate in Italy.
<https://observers.france24.com/en/20191008-deepfake-video-former-italian-pm-matteo-renzi-sparks-debate-italy>
11. TheNextWeb (2019.09.06.) Facebook wants to combat deepfakes by making its own.
<https://thenextweb.com/facebook/2019/09/06/facebook-wants-to-combat-deepfakes-by-making-its-own/>
12. Vox (2018.04.18.) Jordan Peele's simulated Obama PSA is a double-edged warning against fake news.
<https://www.vox.com/2018/4/18/17252410/jordan-peepe-obama-deepfake-buzzfeed>
13. ZIMBIO 홈페이지.
<http://www.zimbio.com/'Rogue+One'+Easter+Eggs+You+Missed/articles/4fVTkALMvOJ/Grand+Moff+Tarkin>

