

2020. 2. 25. IS-090

빅데이터로 본 딥페이크(Deepfake) : 가짜와의 전쟁

Big data Analysis for Deepfake : The War Against Fakes

이승환 책임연구원
(seunghwan.lee@spri.kr)[†]

- 이 보고서는 「과기정통부 정보통신진흥기금」을 지원받아 제작한 것으로 과기정통부의 공식의견과 다를 수 있습니다.
- 이 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문사항 또는 수정·보완할 필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.
 - 소프트웨어정책연구소 이승환(seunghwan.lee@spri.kr) 책임연구원

《 요 약 문 》

인공지능을 활용하여 제작한 콘텐츠, 딥페이크가 확산 중이다. 딥페이크는 딥러닝(Deep Learning)을 이용해 원본 이미지나 동영상 위에 다른 영상을 중첩하거나 결합하여 원본과는 다른 콘텐츠를 생성하는 기술이다. 딥페이크는 산업적 활용가치가 높아 영화, 음반 등 다양한 분야에서 활용이 가능하나, 포르노, 가짜뉴스 등 부작용도 존재한다. 본고에서는 딥페이크 영상이 확산되기 시작한 최근 2년간의 소셜 빅 데이터를 분석하고 시사점을 도출하였다.

분석결과, 딥페이크에 대한 언급 양이 2017년 말을 기준으로 증가하였고, 2018년 말부터는 급증하고 있다. 언급 양 증가 시점마다 사회·문화적으로 이슈를 발생시키는 딥페이크 콘텐츠가 등장하였다. 딥페이크에 대한 언급 양은 남성 비중이 여성의 약 2배에 달하는 수준이며, 예술, 경영관련 직군의 언급 비중이 높고, 분야별로는 영화·음악, 정치와 연관되어 주로 언급이 이루어지고 있었다. 채널측면에서는 주로 트위터를 통해 언급된 내용에 답변을 달거나, 공유하는 형식으로 전파되고 있다. 감성분석 결과, 현재 딥페이크에 대한 부정적인 인식이 강하고, 시간의 경과에 따라 강도가 증가 추세이다. 세부적으로는 슬픔과 두려움에 대한 감정이 강하며, 즐거움에 대한 감성도 내재되어 있었다. 딥페이크 발생 초기에는 즐거움에 대한 감정이 높았으나, 시간의 경과에 따라 슬픔과 두려움의 감정이 증가하는 것으로 분석되었다.

딥페이크의 확산으로 가짜 콘텐츠와의 전쟁이 본격화되고, 적용대상, 분야 등도 다변화 될 전망이다. 다양한 SNS와 딥페이크 기술이 결합되면서 확산 폭은 확대되고, 대상도 유명인사에서 일반인 및 사물로 다변화 될 전망이다. 적용 분야도 현재는 성인물과 정치 이슈 중심이나, 향후 영화, 음반, 게임, 스포츠 등 다양한 분야로 확대되고, 규제이슈도 더욱 부각될 것으로 보인다. 딥페이크 기술은 명과 암이 존재하나, 현재 부정적인 인식이 더 강해 이를 고려한 기업전략 및 정책 수립이 요구되는 시점이다. 기업은 딥페이크를 활용한 다양한 사업모델 발굴에 주력하고 위험을 최소화해야 하며, 정부는 딥페이크 산업 활성화를 고려한 정책수립, 모니터링 강화, 선거관련 가짜 뉴스에 대비할 필요가 있다.

《 Executive Summary 》

Deepfake, a content created using artificial intelligence, is spreading. Deep fake is a technology that uses deep learning to create processed content that is different from the original by overlaying or combining other images on top of the original image or video. Deepfake has a high industrial value and can be used in various fields such as movies and recordings, but there are also side effects such as pornography and fake news. In this paper, we analyzed big data of the recent two years when deep fake images began to spread and draw implications.

As a result of analysis, the volume of deep fakes (Mention Volume) increased as of the end of 2017, and has increased rapidly since the end of 2018. Every time the volume of mentions increased, deep fake content emerged that raised issues socially and culturally. Men's mention of Deepfake is about twice that of women, and the share of arts and management positions is high, and it is mainly related to film, music, and politics. On the channel side, it is mainly spread in the form of replying or sharing the contents mentioned through Twitter. In terms of sentiment analysis, there is a strong negative perception of deep fakes, and the intensity is increasing over time. In detail, feelings of sadness and fear are strong, and emotions of pleasure are inherent. In the early stages of deep fake development, feelings of joy were high, but feelings of sadness and fear increased with time.

The implications are as follows. The proliferation of deep fakes will lead to a full-fledged war against fake content, as well as diversification of applications and fields. As various SNS and deep fake technologies are combined, the spread of spread will expand, and the target will also be diversified from celebrities to the general public and things. The field of application is now centered on adult and political issues, but it will be expanded to various fields such as movies, recordings, games, and sports in the future. Although deep fake technology exists in light and dark, it is time to establish corporate strategy and policy considering the negative perception. Companies need to focus on discovering various business models using DeepFake and minimize risks, and the government needs to prepare for policy mix, strengthen monitoring, and prepare for fake news related to election.

《 목 차 》

1. 딥페이크(Deepfake)의 부상	1
2. 빅데이터로 본 딥페이크(Deepfake)	5
2.1 주요 키워드 및 양(Volume) 분석	5
2.2 감성분석(Sentiment Analysis)	11
3. 시사점	14
참고문헌	17

《 Contents 》

1. Emergence of Deepfake	1
2. Big data Analysis for Deepfake	5
2.1 Keyword and Volume Analysis	5
2.2 Sentiment Analysis	11
3. Implication	14
Reference	17

1. 딥페이크(Deepfake)의 부상

- 인공지능을 활용하여 제작한 콘텐츠, 딥페이크(Deepfake)가 확산 중
 - 딥페이크는 딥러닝(Deep Learning)을 이용해 원본 이미지나 동영상 위에 다른 영상을 중첩하거나 결합하여 원본과는 다른 가공 콘텐츠를 생성하는 기술¹⁾
 - 딥페이크는 딥러닝(deep learning)과 가짜(fake)의 혼성어
 - 딥페이크 용어는 美 커뮤니티 레딧(Reddit)의 이용자 「deepfakes」가 2017년 12월 유명인의 얼굴을 성인물에 합성한 동영상을 처음 유포시킨 데서 유래
 - 2018년 12월 7,964개였던 딥페이크 콘텐츠 수가 2019년 9월 1만 4,678개로 증가하였으며, 20개의 딥페이크 제작 커뮤니티가 존재하고 이용자 수는 95,791명에 달함(2019년 9월 기준)²⁾
- 딥페이크를 구성하는 핵심기술은 생성적 적대 신경망(GAN: Generative Adversarial Network)
 - GAN은 생성모델과 식별모델이 공존하며, 각 모델이 학습을 반복하고, 이러한 과정에서 생성모델과 식별모델은 서로를 적대적 경쟁자로 인식하며 상호 발전
 - 생성모델이 데이터를 기반으로 훈련한 후 가짜 동영상을 만들고 식별모델은 가짜를 감지하며, 다른 모델이 감지할 수 없을 때까지 가짜를 만드는 방식
 - GAN을 NIPS³⁾에서 발표한⁴⁾ 구글 브레인의 Ian Goodfellow는 생성자를 위조지폐범에, 식별자를 경찰에 비유⁵⁾

1) Ruben Tolosana et al, Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. JOURNAL OF LATEX CLASS FILES, VOL. 13, NO. 9, MARCH 2016; 최순욱, 오세욱, 이소은(2019). 딥페이크 이미지 조작: 심층적 자동화에 따른 사실의 위기와 폰크툼의 생성. 미디어, 젠더&문화, 34(3), 339-380

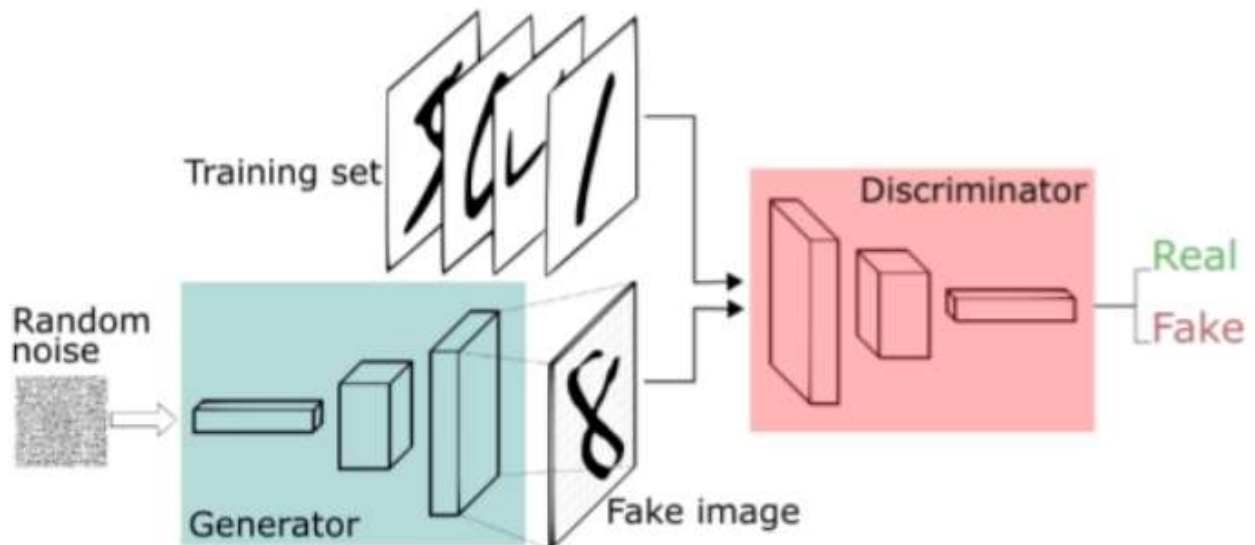
2) Deepttrace(2019) “The state of Deepfakes, Landscape, Threats and Impact”

3) Neural Information Processing Systems

4) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, 『Generative Adversarial Networks』

5) 생성자는 감별자를 속이지 못한 데이터를, 감별자는 생성자에게 속은 데이터를 입력받아 학습하고, 이 과정이 반복되면서 위조지폐가 정교해지듯 점점 더 실제에 가까운 거짓 데이터를 만들 수 있게 되는 것

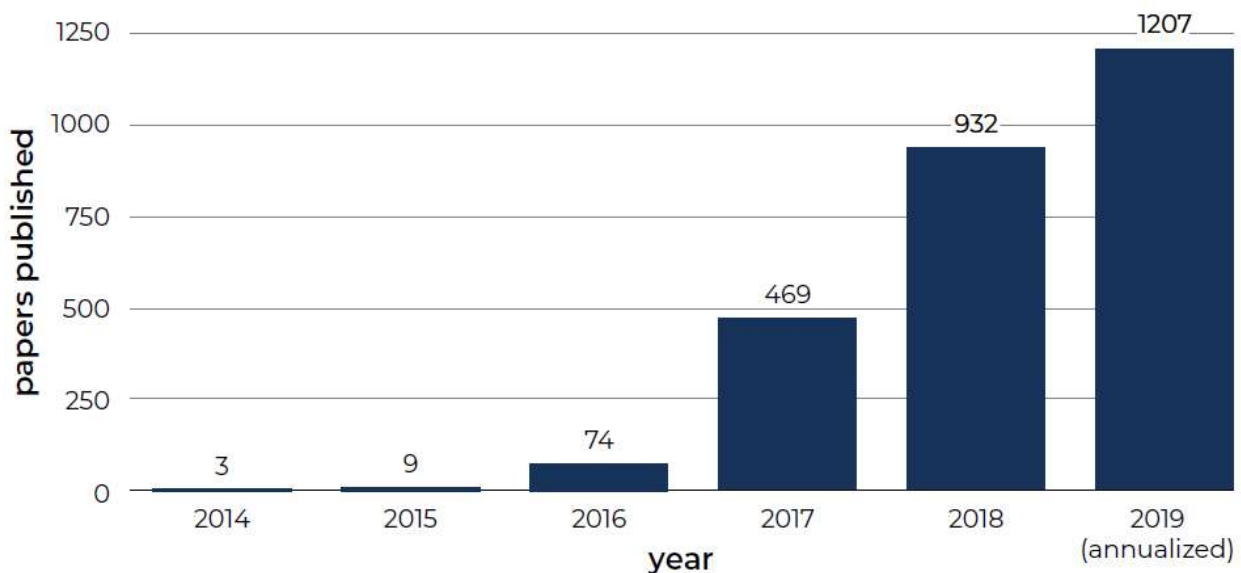
[그림] 생성적 적대 신경망(GAN) Architecture



자료: Palacode Narayana Iyer Anantharaman(29 Oct 2018), Generative Adversarial Networks

□ 생성적 적대 신경망(GAN) 연구의 발전이 딥페이크에 확산에 기여

- 생성적 적대 신경망에 대한 연구는 2014년 4건에서 2019년 9월 1,207건으로 급속히 증가

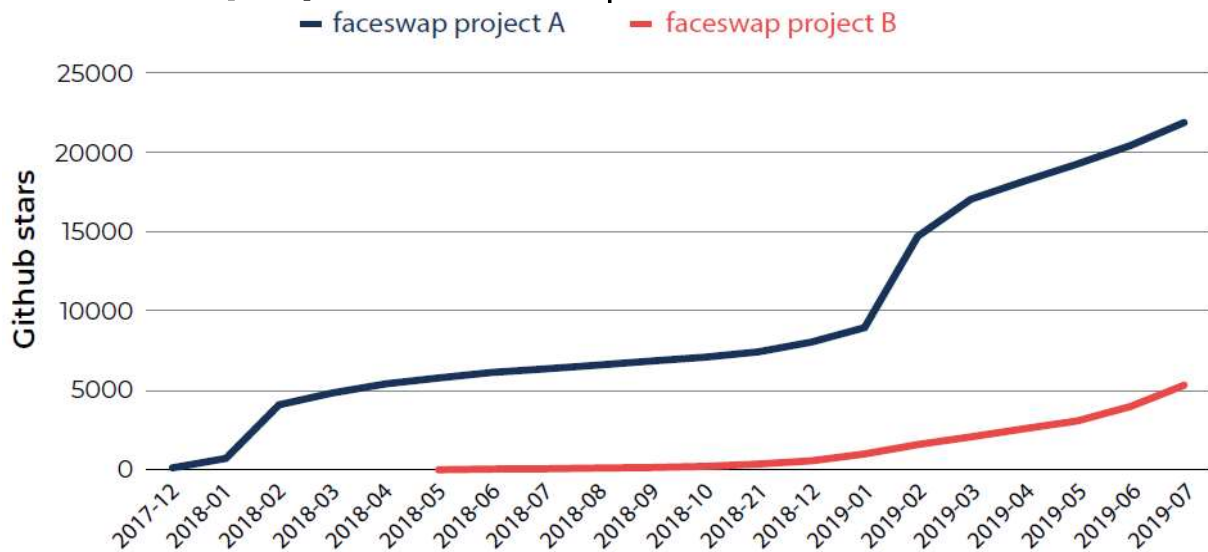
[그림] GAN관련 arXiv⁶⁾에 게재 연구 수

자료: Deepttrace(2019) "The state of Deepfakes, Landscape, Threats and Impact"

6) arXiv is an online archive of scientific and mathematical research papers maintained by Cornell University

- 딥페이크 콘텐츠는 GAN을 활용한 DeepFaceLab, Faceswap 등 오픈 소스 영상 합성 프로그램이 배포되면서 더욱 성행
- GitHub를 통해 딥페이크 기술로 영상의 얼굴을 바꾸는(Face-swap) 오픈 소스 코드 공유도 확대
- * GitHub Star는 '북마크' 혹은 '좋아요'와 비슷한 역할을 하며, 보통 Star개수를 통해 프로젝트의 신뢰도와 확산 가능성을 가늠

[그림] GitHub의 faceswap 프로젝트 star증가 사례



자료: Deeptrace(2019) "The state of Deepfakes, Landscape, Threats and Impact"

- 딥페이크(Deepfake)는 산업적 활용가치가 높아 영화, 음반 등 다양한 분야에서 활용 중
- 딥페이크는 특수효과를 구현하고 제작비용을 낮추는데 기여⁷⁾
 - 2019년 4월 개봉한 월 스미스 주연의 영화 「제미니맨」 속에서 51세 월 스미스가 맡은 「헨리」는 23세 월 스미스인 「주니어」에게 쫓기며, 주니어는 뉴질랜드 기업 Weta Digital이 만들어낸 딥페이크 영상
- 美 기업 스페이스 150은 딥페이크를 통해 뮤직비디오와 신곡을 출시⁸⁾
 - 美 힙합가수 Travis Scott의 가사와 멜로디를 인공지능으로 학습하여 신곡 「Jack Park Canny Dope Man」 과 뮤직 비디오를 딥페이크로 제작

7) FINANCIAL TIMES(2019.10.10.) "Deepfakes: Hollywood's quest to create the perfect digital human"

8) Fast Company(2020.2.13.) "Meet Travis Bott, the Travis Scott twin whose music and lyrics were created with AI"

- 英 기업 Synthesia는 2019년 4월 축구스타 베컴이 말라리아 퇴치 홍보 캠페인 영상을 딥페이크를 활용하여 9개 언어로 제작⁹⁾
- 새로운 비디오를 촬영하지 않고, 기존 자료를 편집하여 개인화된 영상 제작이 가능하며 제작비용도 기존의 1/10 수준으로 절감 가능¹⁰⁾

[그림] 말라리아 퇴치 캠페인의 딥페이크 영상 제작 과정



자료: Techcrunch.com(April 25, 2019). "The startup behind that deep-fake David Beckham video just raised \$3M"

□ 하지만, 포르노, 가짜뉴스 등 딥페이크(Deepfake)로 인한 부작용도 존재

- 2019년 9월 기준, 14,678건의 딥페이크 영상 중 성인물이 전체의 96%를 차지
 - 딥페이크 성인용 콘텐츠를 중점적으로 게시하는 사이트 중에서 규모가 가장 큰 4개 사이트의 총 영상 조회 수는 1억 3,436만 4,438회에 달함
- 딥페이크를 활용한 가짜 뉴스로 정치·사회적 불안감이 가중
 - 2019년 9월, 이탈리아 Matteo Renzi 전 총리가 다른 정치인들을 모욕하는 딥페이크 영상이 게시되었고, 내용을 실체처럼 인식한 이용자가 총리를 비판하는 해프닝이 발생¹¹⁾

□ 본고에서는 딥페이크 영상이 확산되기 시작한 최근 2년간의 빅 데이터를 분석하고 시사점을 도출¹²⁾

9) www.prolificlondon.co.uk, Malaria Must Die campaign uses video AI to make David Beckham speak nine languages(9 April 2019)

10) Techcrunch.com(April 25, 2019). "The startup behind that deep-fake David Beckham video just raised \$3M"

11) Bloomberg(2019.09.17.) Italy's Renzi Seeks Kingmaker Role in Bid to Start New Party.

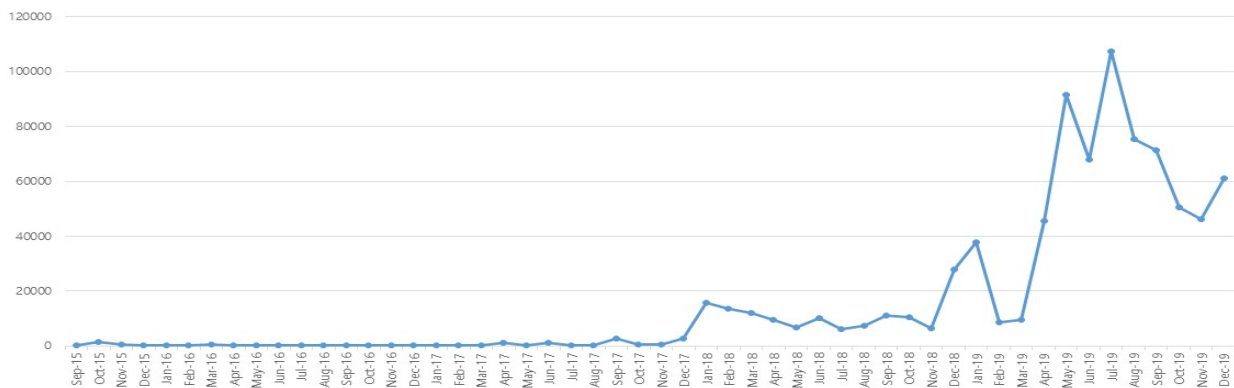
12) 트위터 등 소셜 데이터를 활용, 2017년 말~2020년 초까지 분석, 지역은 글로벌, Text는 영어 기준

2. 빅데이터로 본 딥페이크(Deepfake)

2.1. 주요 키워드 및 양(Volume) 분석

- 딥페이크(Deepfake)에 대한 언급 양(Mention Volume)이 최근 급증
○ 2017년 말을 기준으로 언급 양이 증가하였고, 2018년 말부터 급증

[그림] 딥페이크 Mention Volume 변동(2015.10~2020.1)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 언급 양이 증가한 시점마다 사회·문화적으로 이슈를 발생시키는 딥페이크(Deepfake) 콘텐츠가 등장
 - 2017년 말 미국의 Reddit 사이트에서 「딥페이크」라는 가명을 가진 회원이 다수의 가짜 포르노 비디오를 올렸고 그 중, 유명 배우들의 콘텐츠가 주목받음
 - 키워드 : Deepfake, Reddit, Hollywood, Celebrity, Porn, FakApp, AI

[그림] 딥페이크 Mention Word Cloud(2017년 12월~2018년 1월)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 2019년 5~6월 낸시 펠로시 의장, 저커버그 CEO 등 유명 인사의 영상이 게시
 - 미국 낸시 펠로시 하원의장(민주당) 미국 진보센터 행사에 참석해 술에 취한 듯한 어투로 말하는 조작된 동영상(13)이 소셜미디어를 통해 확산
 - 인스타그램에 마크 저커버그 페이스북 CEO의 인터뷰 발언을 조작한 딥페이크 영상이 게시(14)
 - 2018년 버즈피드가 딥페이크의 위험성을 알리기 위해 게시한 오바마, 트럼프 관련 영상 내용도 동시에 언급
 - 키워드 : Nancy Pelosi, Facebook CEO, Trump, Obama 등

[그림] 딥페이크 Mention Word Cloud(2019년 5~6월)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

[그림] 낸시 펠로시, 저커버그 딥페이크 영상



자료: www.cnn.com, "Doctored videos shared to make Pelosi sound drunk viewed millions of times on social media"; www.cnn.com, "deepfake video of Mark Zuckerberg presents a new challenge for Facebook"

13) 문제의 동영상이 재생 속도를 의도적으로 느리게 편집해 정상 어투와 표정을 만취한 모습으로 보이게 했다는 사실이 밝혀짐

14) 동영상에는 마크 저커버그 똑같이 생긴 인물이 등장해 "잠시 이런 상상을 해보라. 수십억 명의 도난당한 데이터와 그들의 모든 비밀, 그들의 생명, 그들의 미래를 완전히 통제하는 한 사람이 있다." 언급

- 2019년 8월에 37만 명의 구독자를 보유한 딥페이크 유튜브 채널 「Ctrl Shift Face」는 영화배우 Tom Cruise의 영상을 공개했고,¹⁵⁾ 약 730만 명이 조회
- 키워드 : Tom Cruise, YouTube, Ctrl shift Face, deep fake 등

[그림] 딥페이크 Mention Word Cloud(2019년 7~8월)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

[그림] Bill Hader Channels Tom Cruise Deepfake 영상



자료: Youtube Channel Shift Face, Bill Hader channels Tom Cruise [DeepFake]

- 2020년 1~2월에는 TikTok, Snapchat의 딥페이크 서비스 개발이 알려지며 딥페이크 정책 수립의 필요성이 논의
- 전 세계 15억 건의 다운로드가 이루어진 SNS 틱톡 운영사 중국 바이트댄스의 딥페이크 서비스 개발 소식이 보도¹⁶⁾

15) www.rollingstones.com: Watch Bill Hader Become Tom Cruise, Seth Rogen in Eerie Deepfake Video

16) www.techcrunch.com: ByteDance & TikTok have secretly built a deepfakes maker

- 미국 상원 의원은 TikTok이 국가안보를 위협하는 앱인지에 대해 조사의 필요성을 제기¹⁷⁾
- * 미국 내 틱톡 사용자는 1억 1,100만 명에 달하며, 바이트댄스가 중국 정보기관으로부터 사용자 정보를 공유하도록 강요받을 수 있다고 언급했으며, 최근 미 해군도 복무 장병들에게 틱톡 앱을 삭제하지 않으면 내부 네트워크에 접속할 수 없다고 공지
- 키워드 : TikTok, Byte Dance, Snapchat, Policy, Artificial intelligence 등

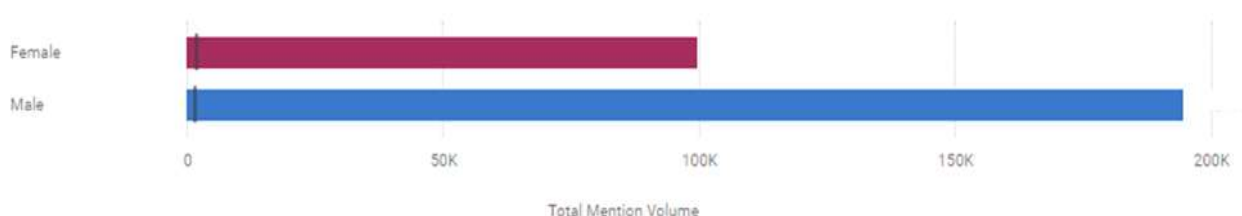
[그림] 딥페이크 Mention Word Cloud(2020년 1~2월)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

□ 딥페이크에 대한 언급 양은 남성의 비중이 여성의 약 2배에 달하는 수준

[그림] 성별에 따른 딥페이크 Mention Volume(2015.10~2020.1)



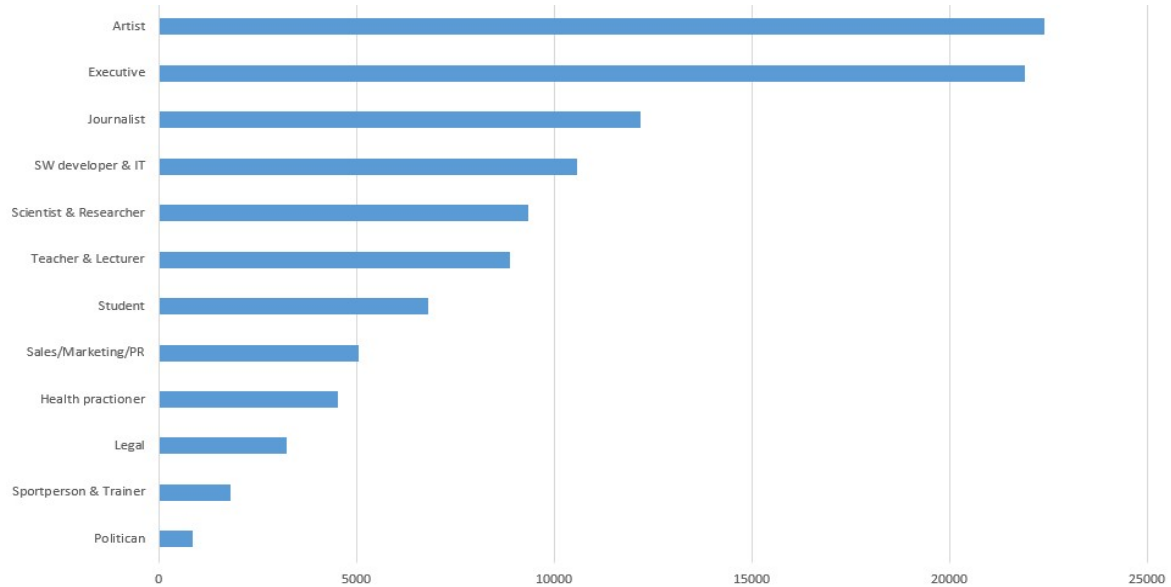
자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

17) www.zdnet.co.kr, 틱톡, 딥페이크 서비스 개발...美 외신 “악용·정보침탈 우려”

□ 예술, 경영관련 직군의 언급 비중이 높고, 분야별로는 영화·음악, 정치와 연관되어 주로 언급이 이루어지고 있는 상황

○ 예술, 경영, 언론, SW개발 직업군들의 언급 비중이 높은 것으로 분석

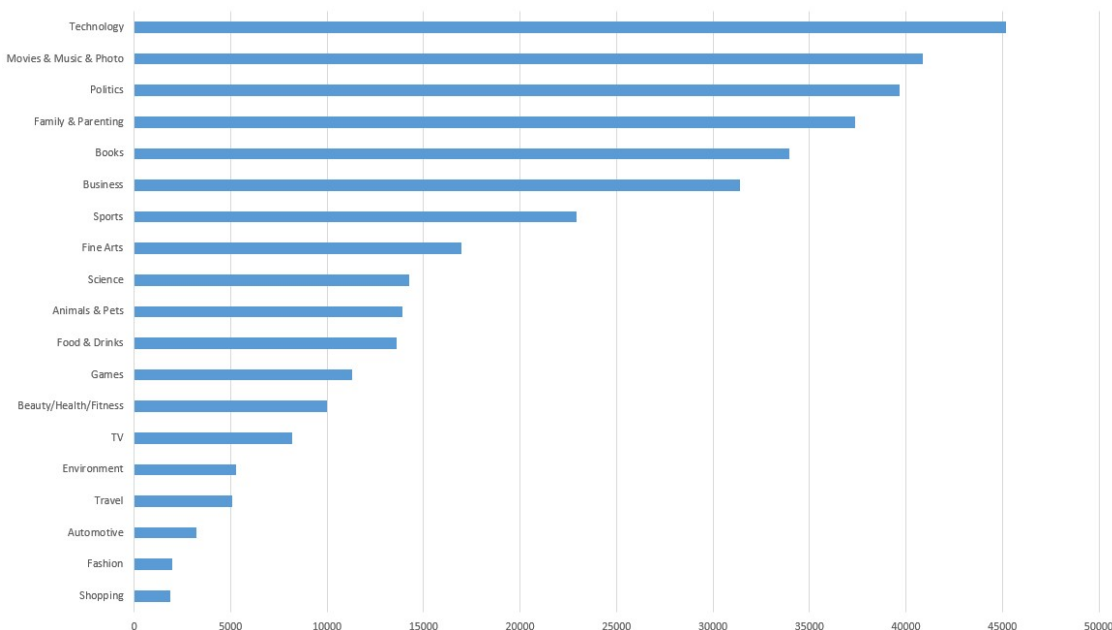
[그림] 직업에 따른 딥페이크 Mention Volume(2015.10~2020.1)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

○ 세부 분야별로는 기술, 영화·음악, 정치와 연관되어 언급

[그림] 분야별 딥페이크 Mention Volume(2015.10~2020.1)

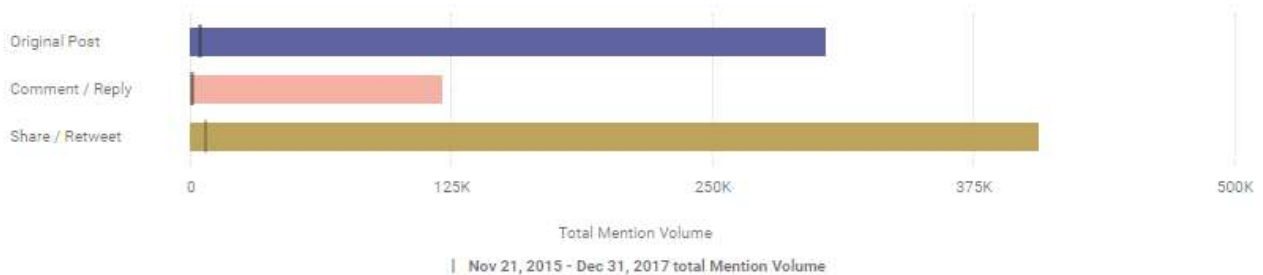


자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

□ 주로 트위터를 통해 언급된 내용에 답변을 달거나, 공유하는 형식으로 전파

- 직접 언급(Original Post)한 내용에 커멘트·답변(Comment·Reply), 공유·리트윗(Share·Retweet)하는 방식으로 전파 중

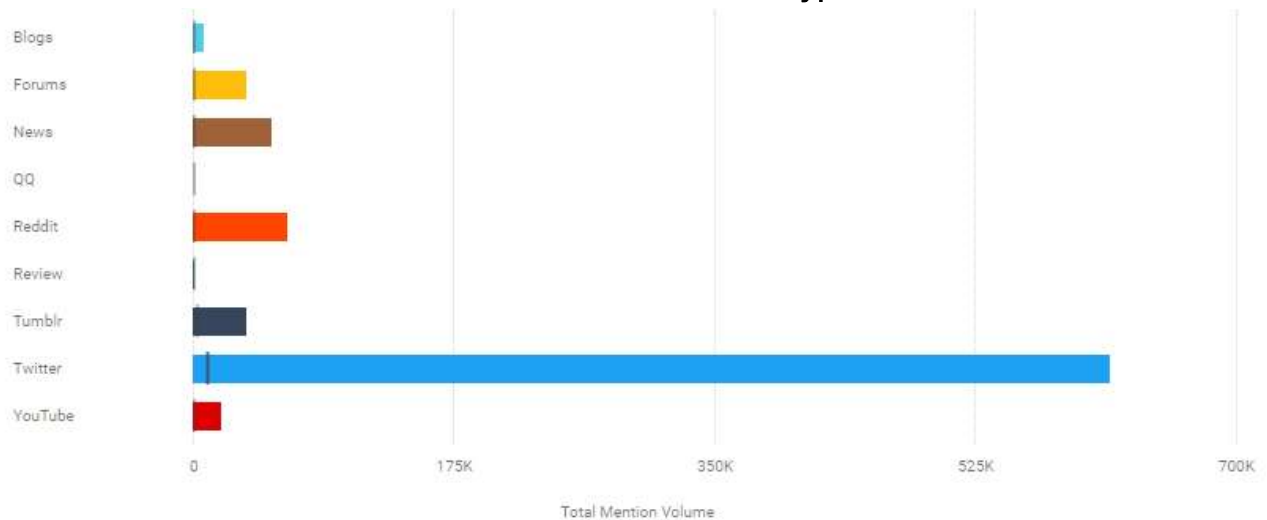
[그림] 딥페이크 언급형태(Mention Type)(2015.10~2020.1)



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 초기 Reddit을 통해 전파되다 이후, 트위터가 주요 채널이 되었으며, 포럼과 뉴스 채널 비중도 증가

[그림] 채널 별 딥페이크 언급형태(Mention Type)(2015.10~2020.1)

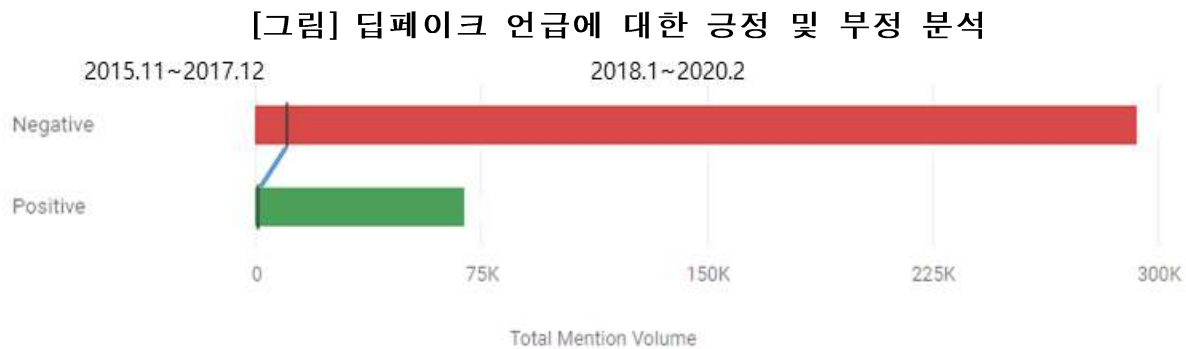


자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

2.2. 감성 분석(Sentiment Analysis)

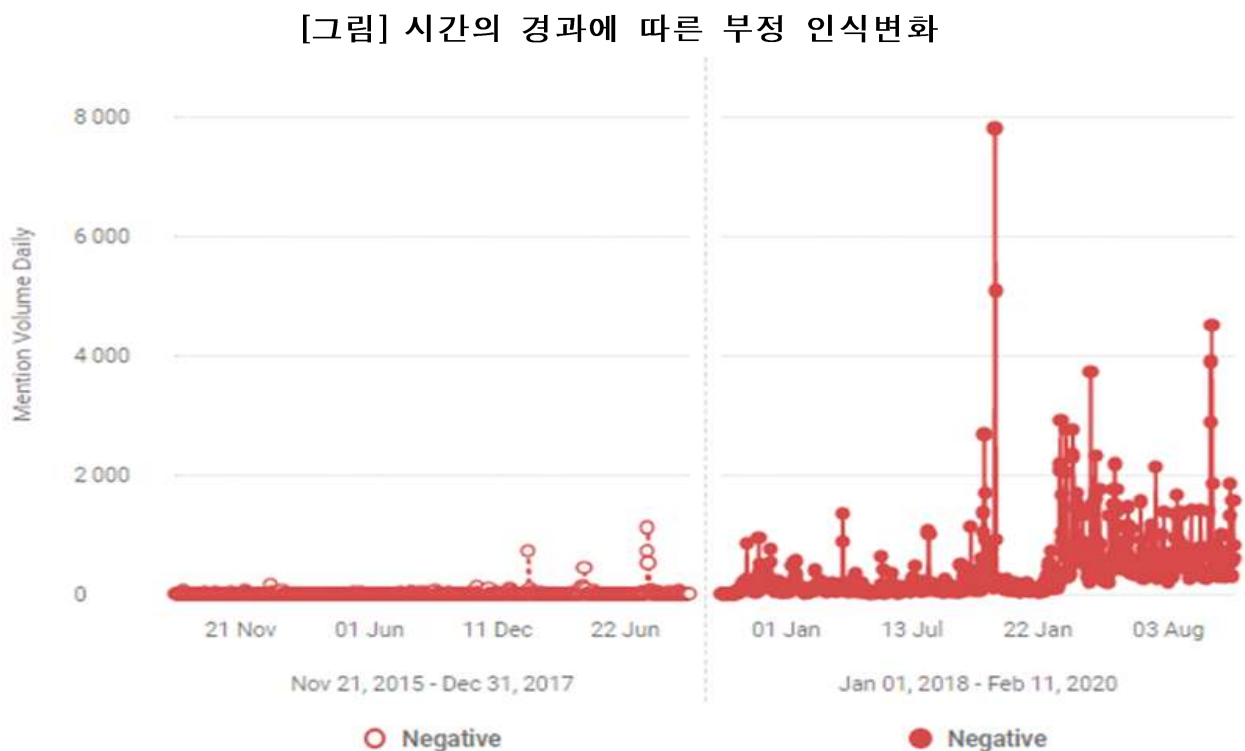
□ 딥페이크(Deepfake)에 대한 부정적인 인식이 강하고, 시간의 경과에 따라 증가 추세

○ 부정적인 인식이 긍정적인 인식의 3배 이상



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

○ 부정적인 인식은 2018년 초부터 점차 증가하다 2019년 이후 증가폭이 확대



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 긍정적인 인식도 증가하고 있으나, 부정적 인식대비 양적 측면에서는 열위

[그림] 시간의 경과에 따른 긍정 인식변화

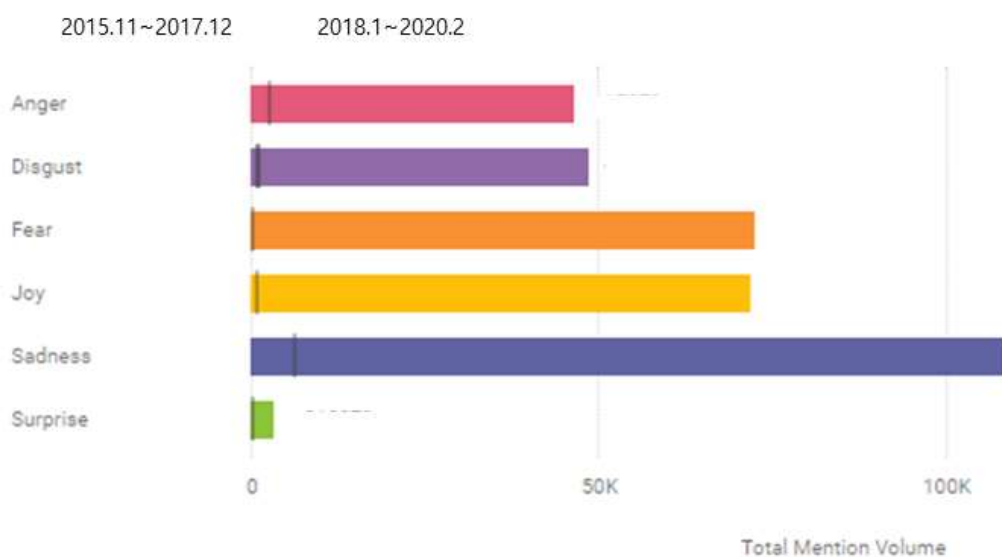


자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 슬픔과 두려움에 대한 감정이 강하며, 즐거움에 대한 감성도 내재

- 슬픔에 대한 감정이 가장 크고, 두려움과 즐거움에 대한 비중이 유사하게 나타남

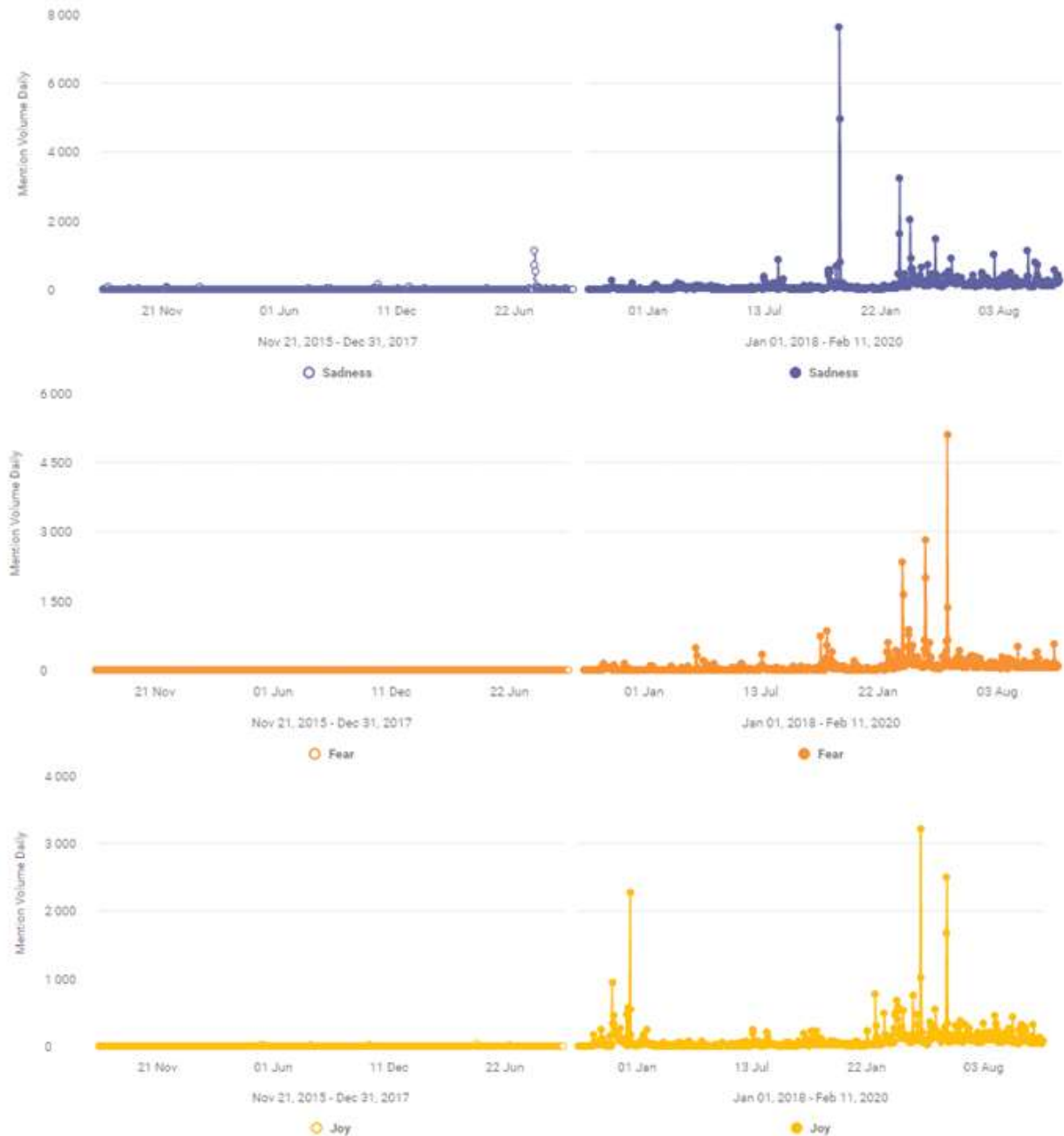
[그림] 딥페이크에 대한 세부 감성 분석



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 딥페이크 발생 초기에는 즐거움에 대한 감정이 높았으나, 시간의 경과에 따라 슬픔과 두려움의 감정이 증가

[그림] 시간 경과에 따른 감성 변화



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

3. 시사점

- 딥페이크의 확산으로 가짜 콘텐츠와의 전쟁이 본격화되고, 적용대상, 분야 등도 다변화 전망
 - 딥페이크 콘텐츠수와 언급 양(Volume)증가 추세이며, 향후 다양한 SNS와 딥페이크 기술이 결합되면 확산 폭은 더욱 확대
 - 키워드 분석에 등장한 TikTok, Snaphcat등 다양한 SNS와의 결합으로 이용자가 급증 전망
 - 영화 등 영상에 사용자 얼굴 합성해주는 App '자오'는 2019년 9월 출시와 동시에 중국 iOS 앱스토어 다운로드 1위를 차지
 - * 중국 채팅 App인 모모(陌陌)가 투자한 것으로 알려진 'ZAO'는 사용자의 정면 얼굴 사진 한 장 만 있으면 영화TV작품이나 동영상 속의 인물에 자신의 얼굴을 붙여넣어 자신이 주인공인 동영상을 만들 수 있도록 만들어진 App
 - 현재 유명인사 중심으로 딥페이크가 제작되고 있으나, 기술의 보편화와 접근 용이성으로 인해 일반인 및 사물로 확대될 전망
 - 적용 분야도 현재는 성인물과 정치 이슈이나, 향후 영화, 음반, 게임, 스포츠 등 다양한 분야로 확대
 - 최근 주요 키워드로 정책이슈가 부각되면서 향후 규제강도가 더욱 높아질 가능성
 - 빅 데이터 주요 키워드가 특정 인물과 사건 위주에서 최근 「Policy」 비중이 높아졌고, 채널측면에서도 뉴스와 포럼 비중이 확대

[표] 딥페이크 이슈 진화 방향

구분	AS-IS	TO-BE
대상	· 사람(유명인)	· 사람(유명인+일반인), 사물, 동물 등
적용 분야	· 성인물, 정치	· 게임, Entertainment 등 다변화
플랫폼	· 독립제작 및 배포	· 독립제작+SNS 결합
규제강도	· 低 : 자율규제, 가이드라인	· 高 : 입법을 통한 규제 강화 등

- 딥페이크 기술은 명과 암이 존재하나, 현재 부정적인 인식이 더 강해 이를 고려한 기업전략 및 정책 수립이 요구되는 시점
 - 딥페이크에 대한 감정변화가 확산초기 즐거움(Joy)에서 시간의 경과와 함께 슬픔(Sadness), 두려움(Fear)이 강해지는 추세
 - 하지만, 여전히 즐거움(Joy)에 대한 감정은 존재하며, 향후 산업적 활용 측면에서 이를 반영하는 사업모델 발굴과 위험을 줄이는 정책수립이 필요
- 기업은 딥페이크를 활용한 다양한 사업모델 발굴에 주력하고 위험을 최소화
 - 기업은 디지털 트윈 가수, 과거재현, 광고 등 다양한 분야에 딥페이크를 긍정적으로 활용방안을 검토
 - 딥페이크 적용 시, 우려되는 프라이버시 등 다양한 위험을 고려하고 활용 가이드라인 수립 등 대책을 강구
 - 페이스북은 비디오의 인물이 실제로 말하지 않은 것을 말한 것처럼 편집·합성한 경우와 동영상에 다른 콘텐츠를 합성·교체·추가해 실제처럼 조작한 경우 규제하는 방안을 발표¹⁸⁾
 - * 딥페이크 영상의 식별은 사용자 제보와 사실 확인기관(fact checker)¹⁹⁾ 활동을 통해 진행되며, 콘텐츠 사실 확인 후 단계적 조치 실행
 - * 사용자의 자발적 제보와 국제 사실 확인 네트워크*에서 인증한 제3자 사실 확인기관의 모니터링 활동을 통해 딥페이크 가능 영상을 식별
 - 구글은 2019년 11월 정치광고를 포함한 기타 광고에서 딥페이크 사용을 금지²⁰⁾했고 트위터도 동참²¹⁾
 - ZAO를 개발한 中 기업 모모는 약관에 “사용자가 업로드 한 데이터 중 얼굴 사진이나 영상은 앱 성능 개발 외의 목적으로는 절대 사용하지 않겠다.” 고 적시

18) The Newyork times, Facebook Says It Will Ban Deepfakes, 2020.1.7.

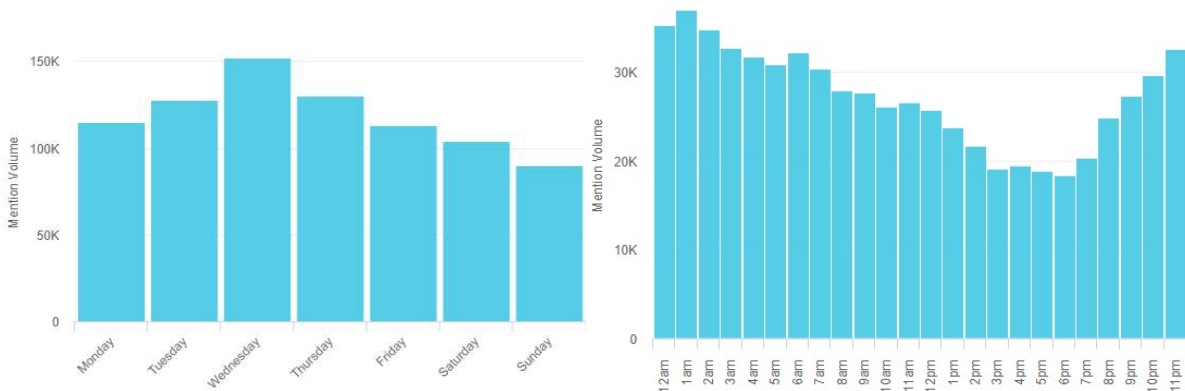
19) 국제 사실 확인 네트워크(International Fact-Checking Network): 2015년 미국 미디어연구 교육기관 포인터(Poynter)에 의해 설립된 전 세계 팩트 체크기관 포럼

20) The Wall Street journal, Facebook Bans Deepfakes but Permits Some Altered Content, 2020.1.7.

21) ZDnet, Twitter: We'll kill deepfakes but only if they're harmful. 2020.1.5.

- 정부는 딥페이크 산업 활성화를 고려한 정책조합(Policy Mix)구상 및 모니터링을 강화하고 선거관련 가짜 뉴스에 대비
 - 자율규제, 가이드라인, 입법규제, 과징금 부과 등 다양한 규제 강도 수준을 고려
 - 딥페이크 발생 요일, 시간 등을 고려한 데이터 기반의 정책 모니터링 강화

[그림] 요일, 시간대별 딥페이크 언급 양



자료: Web DB 분석 기반(전 세계, 영어 기준) 소프트웨어정책연구소 분석

- 주요 선거에 대비하여 딥페이크 기반의 가짜뉴스 이슈에 대응
 - 2020년은 한국(총선, 4월), 미국(대선, 11월) 등 주요국 선거가 예정

[참고문헌]

- [1] Bloomberg(2019.09.17.) Italy's Renzi Seeks Kingmaker Role in Bid to Start New Party.
- [2] Deeptrace(2019) "The state of Deepfakes, Landscape, Threats and Impact"
- [3] FINANCIAL TIMES(2019.10.10.)"Deepfakes: Hollywood's quest to create the perfect digital human"
- [4] Fast Company(2020.2.13.)"Meet Travis Bott, the Travis Scott twin whose music and lyrics were created with AI"
- [5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, 『Generative Adversarial Networks』
- [6] Ruben Tolosana et al, Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. JOURNAL OF LATEX CLASS FILES, VOL. 13, NO. 9, MARCH 2016;
- [7] Palacode Narayana Iyer Anantharaman(29 Oct 2018), Generative Adversarial Networks
- [8] Techcrunch.com(April 25, 2019). "The startup behind that deep-fake David Beckham video just raised \$3M"
- [9] Techcrunch.com(April 25, 2019)."The startup behind that deep-fake David Beckham video just raised \$3M"
- [10] The Newyork times, Facebook Says It Will Ban Deepfakes, 2020.1.7.
- [11] The Wall Street journal, Facebook Bans Deepfakes but Permits Some Altered Content, 2020.1.7.
- [12] Youtube Channel Shift Face, Bill Hader channels Tom Cruise [DeepFake]
- [13] www.prolificlondon.co.uk, Malaria Must Die campaign uses video AI to make David Beckham speak nine languages(9 April 2019)

- [14] www.cnn.com, “Doctored videos shared to make Pelosi sound drunk viewed millions of times on social
- [15] www.cnn.com, “deepfake video of Mark Zuckerberg presents a new challenge for Facebook“
- [16] www.rollingstones.com; Watch Bill Hader Become Tom Cruise, Seth Rogen in Eerie Deepfake Video
- [17] www.techcrunch.com: ByteDance & TikTok have secretly built a deepfakes maker
- [18] www.zdnet.co.kr, 틱톡, 딥페이크 서비스 개발...美 외신 "악용·정보침탈 우려"
- [19] ZDnet, Twitter: We'll kill deepfakes but only if they're harmful. 2020.1.5.
- [20] 최순욱, 오세욱, 이소은(2019). 딥페이크 이미지 조작: 심층적 자동화에 따른 사실의 위기와 폰크툼의 생성. 미디어, 젠더&문화, 34(3), 339-380

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.