

기획특집

인공지능(AI) 기술 발전과 성인지적 대응 전략

• 인공지능(AI)의 젠더편향 완화를 위한 법제화 전략

김일우 | 한국청소년정책연구원 부연구위원

• 공정성을 넘어: AI 채용도구 형평성 제고전략

이지은 | 연세대 문화인류학과 부교수

신민정 | 연세대 문화인류학과 박사과정

신지연 | 연세대 문화인류학과 박사과정

• 딥페이크 성범죄 기술 대응 동향

유재흥 | 소프트웨어정책연구소 책임연구원

딥페이크 성범죄 기술 대응 동향

KOREAN WOMEN'S
DEVELOPMENT
INSTITUTE

유재흥 소프트웨어정책연구소 책임연구원

1. 들어가며

가. 딥페이크 문제 현황

2024년 8월 텔레그램발 청소년 딥페이크 이슈가 우리 사회를 휩쓸었다. 청소년들은 소셜미디어 등에서 쉽게 구한 친구나 선생님의 사진을 나체 사진과 합성해 음란 콘텐츠를 만든 후 텔레그램을 통해 공유했다. 2019년 우리 사회에 큰 충격을 주었던 텔레그램 N번방 사건을 떠올리게 하면서 사회적 이슈가 되었다. 사건 발생 직후 2024년 8월 22일 방송통신심의위원회가 딥페이크를 이용한 성적 허위 영상물을 강력히 대응하겠다고 나섰고, 같은 날 여성가족부도 디지털성범죄피해자지원센터를 통해 영상물 삭제 지원 등의 피해 대응 계획을 밝혔다. 8월 27일 국무회의에서 대통령은 ‘이번 사건은 명백한 범죄 행위’라며 디지털 성범죄를 뿌리 뽑아달라고 당부했다.

한편 시큐리티히어로(Security Hero)라는 보안

업체가 2023년 딥페이크 현황 보고서를 발간했다.¹⁾ 보고서는 95,820개의 딥페이크 동영상, 온라인 플랫폼의 85개 딥페이크 콘텐츠 전용 채널, 딥페이크 생태계와 연결된 100개 이상의 웹사이트를 종합적으로 분석한 결과를 담고 있다. 보고서에 따르면 2023년 딥페이크 동영상은 2019년 대비 550% 증가하였고 딥페이크 동영상의 98%는 딥페이크 포르노였다. 그리고 딥페이크 포르노의 표적이 되는 사람의 99%는 여성으로 나타났다. 놀라운 점은 딥페이크 포르노에 등장하는 인물 중 절반이 넘는 53%가 한국 배우와 가수였다는 사실이다. 선명한 얼굴 이미지 한 장만 있으면 60초 분량의 딥페이크 포르노 영상을 제작하는데 25분도 걸리지 않으며 비용도 제로다. 딥페이크 기술에 대한 접근성이 낮아지면서 국내의 딥페이크 성범죄 신고도 급증하고 있다. 딥페이크 성범죄와 관련된 경찰신고건수는 지난 2021년 156건에서 2024년 10월 기준 964건으로 5배 이상 늘었다.²⁾

1) Security Hero(2023.9). 2023 State of Deepfakes.

Top 10 individuals most frequently targeted by deepfake pornography

Nationality & Profession	Video Features	Views
South Korean Singer	1,595	5,611,500
South Korean Singer	1,238	3,865,000
South Korean Singer	923	2,305,500
South Korean Singer	844	3,395,500
South Korean Singer	803	1,365,200
South Korean Singer	761	3,294,000
South Korean Singer	751	1,995,800
Thai Singer	733	3,852,000
South Korean Singer	733	3,084,000
British Actress	713	3,190,100

자료: Security Hero(2023.9). '2023 State of Deepfakes'. (<https://www.securityhero.io/state-of-deepfakes/#targeted-individuals>)

【그림 1】 딥페이크 포르노 영상 상위 피해자의 국적과 직업

세계 각국은 딥페이크 문제에 대해 고심하고 있다. 특히 딥페이크의 주요 피해자가 여성이라는 점은 여러 보고서에도 공통으로 나타나고 있다. 미 국무부가 발간한 보고서³⁾에서는 여성과 소녀들이 인공지능을 이용한 학대의 흔한 표적이 된다고 지적한다. 인공지능으로 생성된 합성 콘텐츠 동영상의 대다수가 여성과 소녀들의 성적 묘사를 포함하고 있기 때문이다. 또한 2024년 UNESCO 보고서는 생성형 AI 도구가 젠더기반폭력(GBV, gender-based violence)을 조장하는 데 사용될 수 있다고 경고한다.⁴⁾ 기술매개 젠더기반폭력(TFGBV, Technology-facilitated GBV)은 정보통신기술이나 기타 디지털 도구를 사용하여 개인의 성별에 따라 성차별적 폭력을 저지르거나, 조장하거나, 악화시키거나, 증폭시키는 모든 행위를 말한다. 유엔 여성기구(UN Women)에서는 개발 연구소(the Institute of Development Studies)의

연구를 인용해 여성의 16~58%가 TFGBV를 경험한 적이 있다고 보고한다. 이코노미스트 인텔리전스 유닛(Economist Intelligence Unit)도 여성의 38%가 개인적으로 온라인 폭력을 경험한 적이 있으며, 온라인에서 시간을 보내는 여성의 85%가 다른 여성에 대한 디지털 폭력을 목격한 적이 있다는 사실을 보도했다.⁵⁾ 가장 흔하게 보고된 폭력 유형은 잘못된 정보와 명예훼손(67%), 사이버 괴롭힘(66%), 혐오 발언(65%), 사칭(63%), 해킹 및 스토킹(63%), 아스트로터핑⁶⁾(58%), 동영상 및 이미지에 기반한 폭력(57%), 폭행(55%), 폭력적인 위협(52%), 원하지 않는 이미지 또는 노골적인 성적 콘텐츠(43%)로 나타났다.

이러한 문제들은 전세계적으로 인공지능 사용이 증가함에 따라 더욱 심각해질 가능성이 있으며, 이는 성별 간 디지털 격차와도 맞물려 있다. AI 기술

2) 관계부처합동 보도자료(2024.11.). 딥페이크 성범죄 대응 강화 방안.

3) U.S. Department of State(2024.9.23.). The Global AI Research Agenda.

4) UNESCO(2023). "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI.

5) <https://www.unwomen.org/en/what-we-do/ending-violence-against-women/faqs/tech-facilitated-gender-based-violence>

6) 아스트로터핑(Astroturfing): 여러 플랫폼에서 동시에 해로운 콘텐츠를 공유하는 조직적인 활동.

의 발전과 비례하여 여성과 소녀들의 온라인 안전과 권리 보호에 대한 우려가 커지고 있으며, 이는 디지털 시대의 성평등 실현을 위한 중요한 과제가 되고 있다.

나. 딥페이크 기술의 악용

딥페이크 기술 자체는 디지털 아바타를 만들거나 이미지 또는 영상을 복원하는 유용한 목적으로 활용될 수 있다. 문제는 딥페이크 포르노와 같이 특히 여성을 타겟으로 한 음란물을 만들어 공유, 판매, 유포함으로써 피해자를 만들어 낼 수 있다는 점이다. 최근 딥페이크 생성 기술에 대한 접근성이 높아지면서 일반인들도 쉽게 이러한 피해의 대상이 되고 있다는 것이 문제를 더 심각하게 만들고 있다.

지난 2021년 딥페이크 봇(deepfake bot)이라 불리는 자동화된 프로그램이 텔레그램 메신저에 수십만 개의 여성 딥페이크 나체 이미지를 업로드했다.⁷⁾ 피해자들은 대부분 소셜미디어에서 올린 사진이나 온라인 개인 사진 보관함에 있는 사진이 도용당했다고 주장했다. 봇은 이렇게 불법 도용한 사진으로 딥페이크를 만들어 무료 이미지로 사용자를 유인하고 유료 서비스를 통해 수익을 올릴 목적으로 불법적인 방법을 통해 개인 이미지를 수집한 것이다.

딥페이크 도구들은 점점 정교해지고 있다. 단순히 얼굴을 교체하는 것뿐만 아니라 음성을 학습해 사람 목소리를 재생하기도 하고, 실시간 라이브 동영상 필터를 적용하여 바뀐 얼굴과 모습으로 동영상을 스트리밍할 수 있게 한다. 예를 들어 ‘딥누드 프

로’의 사용자가 옷을 입은 피사체의 사진을 입력하면 소프트웨어는 피사체가 옷을 벗은 것처럼 보이게 만든다. 사용자는 이미지에 나타나는 신체적 특성의 크기와 모양까지 선택할 수 있다. 연예인의 사진에서 추출한 얼굴로 3D 아바타를 생성한 다음 아바타에 녹음된 음성 또는 다른 음성 모델을 합성해 딥페이크 연예인 아바타도 만들 수 있다. 그리고 그 아바타 캐릭터로 동영상 플랫폼을 통해 라이브 방송을 할 수 있는 것이다. 딥페이크AI, 스왑페이스, 아바타AI 비디오콜스푸퍼(Deepfake AI, SwapFace, and AvatarAI VideoCallSpoofers)와 같은 애플리케이션이 대표적 예들이다.⁸⁾ ‘딥페이크 3D 프로’와 아바타 생성 기능은 금융 기관의 고객확인(KYC) 시스템을 우회하여 범죄자가 훔친 ID를 통해 계정에 액세스할 수 있도록 하는 데 사용될 수 있다. 또한 음성 복제 서비스와 함께 사용하면 사이버 범죄자가 유명인이나 기타 영향력 있는 사람의 아바타를 생성하여 새로운 사기 투자 프로젝트를 홍보하는 대본을 읽게 하고, 더 나아가 각 피해자의 이름을 언급하는 동영상을 생성하여 개인화할 수 있으므로 스크립트화된 비디오 피싱 캠페인을 만들 수 있다. 반면에 딥페이크AI는 딥페이크에 더 쉽게 접근하여 피해자의 얼굴을 침해 동영상에 붙여 피해자의 평판을 훼손하거나 가짜 뉴스를 퍼뜨리는 것을 가능하게 만들 수 있다. 이는 유명인뿐만 아니라 일반 시민을 포함한 모든 사람을 상대로 한 무기로 사용될 수 있음을 의미한다. 다크제미니(DarkGemini)의 사진 처리 기능이나 사용자 인터페이스(UI)에서 바로 이미지 생성 기능을 제공하는 TorGPT와 같은 인공지능모

7) CNET(2021.6.28.). Deepfake bot on Telegram is violating women by forging nudes from regular pics.

8) Trend Micro(2024.7.30.). Surging Hype An Update on the Rising Abuse of GenAI, <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/surging-hype-an-update-on-the-rising-abuse-of-genai>

텔(LLM)들도 악의적으로 활용되고 있다.

디지털 성범죄의 대응은 결국 딥페이크 포르노와 같은 불법 콘텐츠의 생성을 억제하고, 그러한 콘텐츠를 식별하여 유통을 차단할 수 있도록 돕는 것이다. 여기에서는 이미지를 생성하는 AI 기술의 흐름을 중심으로 최근 진행 중인 딥페이크 기술 동향에 대해 살펴보고자 한다.

2. 딥페이크 기술 동향

가. 이미지 생성 인공지능 기술의 흐름

딥페이크 생성 기술은 인공지능 분야에서 급속도로 발전하고 있다. 초기에는 적대적 생성 신경망(GAN)을 기반으로 한 기술이 주로 사용되었다. GAN(Generative Adversarial Network)은 2014년에 처음 제안된 혁신적인 인공지능 알고리즘으로, 생성자(Generator)와 식별자(Discriminator)라는 두 개의 주요 구성 요소로 이루어져 있다. GAN은 생성자와 식별자라는 두 신경망이 서로 경쟁하며 학습하는 방식으로, 더욱 사실적인 데이터를 생성하는 것을 목표로 한다. 최근에는 GAN을 넘어서 더욱 정교하고 사용하기 쉬운 기술들이 등장하고 있으며, 이는 딥페이크 생성의 접근성을 높이고 있다. GAN의 등장 이후 CycleGAN(2017)과 StyleGAN(2018) 등 다양한 후속 연구들이 진행되어 딥페이크 제작 도구인 FaceSwap 등에 널리 활용되며 이 분야의 발전을 이끌고 있다. 그러나 GAN 기술이 다방면으

로 연구되고 적용되었음에도 불구하고 생성된 이미지의 품질이 기대에 미치지 못하는 문제점도 존재한다. 이는 GAN 기술이 가진 한계점 중 하나로, 지속적인 개선과 연구가 필요한 부분이다.

2017년 구글이 소개한 트랜스포머(Transformer) 모델은 언어 처리 분야에 혁명을 일으켰으며, 이후 시각 인공지능 영역에서도 트랜스포머 기반 연구가 활발히 진행되고 있다. 최근 몇 년 동안 이미지 및 동영상 생성하는 AI 기술이 급속도로 발전하여 인간의 콘텐츠 제작 능력과 속도를 넘어서는 수준에 이르렀다. OpenAI의 DALL-E(2021)와 DALL-E2(2022), 구글의 Imagen(2022)과 Parti(2022), Stability AI의 Stable Diffusion(2022), 구글의 Muse(2023), 그리고 OpenAI의 Sora(2024) 등이 대표적인 예다. 이러한 모델들은 트랜스포머 아키텍처를 기반으로 하며, 텍스트와 이미지 간의 상호작용을 효과적으로 모델링하여 고품질의 이미지와 동영상을 생성한다.

최근 Stable Diffusion과 같은 고성능 이미지 및 동영상을 생성하는 AI 기술의 발전으로 일반 사용자도 쉽게 접근할 수 있는 웹 기반의 서비스가 대중화되고 있다. 2022년 시각 인공지능의 세계적 학회인 CVPR⁹⁾에서 제안된 Stable Diffusion 모델은 텍스트를 기반으로 하여 이미지를 생성하는 혁신적인 인공지능 모델이다.¹⁰⁾ 이 모델은 데이터의 라벨이 필요 없는 자기지도학습(self-supervised) 방식으로 이미지를 점진적으로 노이즈 벡터로 변환하는 확산(diffusion) 과정을 학습한다. 이미지 생성 시에는 노이즈 벡터에서 시작하여 점진적으로 개선된

9) Computer Vision and Pattern Recognition.

10) Rombach et al. (2022). High-Resolution Image Synthesis With Latent Diffusion Models, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10684-10695.

이미지를 만들어내는 역과정을 수행한다. Stable Diffusion의 특징은 생성하고자 하는 이미지에 대한 설명 문장을 프롬프트 형태로 입력하여 생성된 이미지를 세밀하게 제어할 수 있다는 점이다. 이후 Stable Diffusion 2.0(2022), Stable Diffusion 3.0(2024) 등 후속 모델이 개발되면서 기술적 진보가 이어지고 있다. Stable Diffusion 모델은 오픈 소스로 공개되면서 딥페이크 제작의 진입 장벽을 크게 낮추는 결과를 가져왔다.

또한 대규모 생성형 모델을 빠르고 효율적으로 특정 도메인에 적응시키는 기술도 발전하고 있다. 예를 들어 LoRA(Low-Rank Adaptation)와 같은 방법은 적은 양의 학습 데이터로도 모델을 효과적으로 미세 조정할 수 있게 한다. LoRA는 마치 전체 책을 다시 쓰는 대신 필요한 페이지만 골라 수정하는 것과 같이 인공지능 모델의 학습과 적용의 효율성을 높여준다. 이러한 기술의 발전으로 만화나 특정 스타일의 이미지 생성에 최적화된 모델 개발이 매우 용이해졌다. 이는 딥페이크 제작의 다양성과 효율성을 크게 향상시킬 수 있으나 오히려 적절한 조정이 되지 않을 경우 성능 저하로 이어질 수 있다.

나. 딥페이크 식별 탐지 및 대응 기술

딥페이크 기술이 급속도로 발전함에 따라 이를 탐지하고 대응하기 위한 기술 개발도 활발히 이루어지고 있다. 국내외 주요 기업과 연구기관들은 다양한 접근 방식으로 딥페이크 탐지 기술을 개발하고 있으며, 이는 디지털 미디어의 신뢰성을 확보하는데 중요한 역할을 하고 있다. 이러한 노력은 딥페이크로 인한 부정적 영향을 최소화하고 기술의 긍정적 활용을 촉진하는 데 기여하고 있다.

구글 답마인드의 신스ID(SynthID)는 AI 이미지 생성 플랫폼에서 만든 이미지에 육안으로는 보이지 않는 워터마크를 픽셀 단위로 삽입하는 기술을 개발했다. 이 기술은 AI로 합성된 이미지를 식별하는 데 효과적으로 사용되어 생성된 이미지의 출처를 추적하고 진위를 판별하는 데 도움을 준다.

인텔이 개발한 ‘페이크캐처(FakeCatcher)’ 기술은 사람 얼굴의 혈류 변화를 추적하여 실시간으로 딥페이크를 분석한다. 이 기술은 96%의 높은 정확도로 실제 영상과 딥페이크 영상을 구분할 수 있다.¹¹⁾ 페이크캐처는 비디오 픽셀에서 나타나는 미세한 혈류 변화를 감지하고, 이를 알고리즘을 통해 시공간 지도로 변환하여 딥러닝으로 영상의 진위를 즉시 판단한다.

11) Intel(2022.11.). Intel Introduces Real-Time Deepfake Detector.



[그림 2] 딥페이크 식별탐지 및 대응 기술 사례

미국 국방고등연구계획국(DARPA)의 세마포(SemaFor, Semantic Forensics) 프로그램은 딥페이크 생성 과정에서 발생하는 ‘AI의 실수’를 단서로 활용하여 딥페이크를 탐지하는 기술을 개발하고 있다.¹²⁾ 예를 들어 생성된 얼굴에서 양쪽 귀의 귀걸이가 서로 다르게 나타나는 등의 공통적인 특징을 추출하여 딥페이크를 식별한다. 또한 이러한 기술을 통해 딥페이크의 생성 이유까지 추론하는 것을 목표로 하고 있다.

센티넬(Sentinel)의 센티넬AI는 디지털 미디어를 프레임 단위로 분석하여 조작 여부를 판별하는 기술

을 제공한다. 이 기술은 조작된 부분을 시각적으로 표시하여 사용자가 쉽게 이해할 수 있도록 한다. 마이크로소프트(MS)는 비디오 인증 도구를 통해 AI 알고리즘을 활용하여 스틸 사진이나 비디오를 분석하고 미디어의 조작 여부를 나타내는 신뢰 점수(confidence score)를 제공한다.¹³⁾ 이를 통해 사용자는 미디어의 진실성을 빠르게 평가할 수 있다. 센시티(Sensity)는 세계적인 딥페이크 탐지 솔루션 제공업체로 Dall-E, Stable Diffusion, FaceSwap, Midjourney 등 다양한 동영상 생성 엔진을 통해 제작된 영상을 95% 이상의 정확도로 탐지할 수 있

12) DARPA, Semantic Forensics (SemaFor), <https://www.darpa.mil/program/semantic-forensics>

13) Microsoft Blogs(2021.9.1.). New Steps to Combat Disinformation.

는 뛰어난 성능을 자랑한다. 또한 ChatGPT와 같은 대규모 언어 모델(LLM)로 생성된 텍스트를 식별할 수 있으며, 사람이 AI로 생성된 콘텐츠를 편집한 경우에도 이를 감지할 수 있다.

스탠포드 대학교와 캘리포니아 대학교 연구팀은 입의 움직임과 구어를 비교해 불일치를 탐지하는 기술을 개발했다. 이 기술은 AI가 입의 움직임과 음성을 완벽히 일치시키는 데 한계가 있다는 점을 활용하여 불일치가 발견되면 해당 동영상을 딥페이크로 간주한다.¹⁴⁾

다. 딥페이크 탐지 기술 연구 개발 동향

국내 딥페이크 탐지 기술 개발은 정부 주도의 연구 프로젝트를 중심으로 이루어지고 있으며 학계와 산업계의 협력을 통해 실용적이고 효과적인 기술 개발에 주력하고 있다. 이러한 노력은 딥페이크로 인한 사회적 문제를 해결하고 디지털 환경의 신뢰성을 강화하기 위한 중요한 시도로 평가된다. 몇 가지 주요 연구 개발 과제는 다음과 같다.

2021년부터 2022년까지 진행된 과기정통부의 차세대 인공지능 핵심원천기술개발 프로젝트에서는 두 개의 주요 연구과제가 수행되었다. 하나는 "인물의 행동 양식을 모방하는 극사실적 실사 인물 동영상 합성 기술 및 판별 기술 개발"이며 다른 하나는 "실제와 지각 역치 이하 수준까지 동일한 특성을 갖는 인물 영상 합성 기술 및 판별 기술 개발" 과제다. 이 두 과제는 딥페이크 기술의 발전과 함께 그에 대응하는 탐지 기술을 동시에 개발하는 것을 목표로 했다.

또한, 과기정통부의 실감콘텐츠 핵심기술개발 사업의 일환으로 2023년부터 3년간 성균관대의 주관 하에 "악의적 변조 콘텐츠에 대응하기 위한 딥페이크 탐지 고도화, 생성억제, 유포 방지 플랫폼 개발" 과제가 진행 중이다. 이 프로젝트는 딥페이크 탐지 기술의 고도화와 생성 억제, 유포 방지까지 포괄적인 접근을 시도하고 있다. 2024년부터 2027년까지는 딥페이크 역기능 대응을 위해 "생성형 인공지능의 사회적 부작용을 방지하기 위한 자가 진화형 딥페이크 탐지 기술 개발" 과제가 한국전자기술연구원 주관하에 진행된다. 이 프로젝트는 딥페이크 기술의 빠른 진화에 대응할 수 있는 자가 진화형 탐지 기술 개발에 초점을 맞추고 있다.

한편 최근에는 딥페이크 탐지와 관련된 국내외 학술 연구가 활발히 진행되고 있으나 실제 적용에 있어 여전히 큰 격차가 존재한다. 학술적 연구에서 높은 성능을 보이는 방법들이 실제 환경의 딥페이크 탐지에서는 낮은 성능을 보이는 경우가 많기 때문이다.

3. 나가며

딥러닝(Deep Learning)과 페이크(Fake)의 합성어인 딥페이크는 2017년 말 소셜미디어 커뮤니티 레딧(Reddit)에서 한 사용자(아이디 deepfakes)가 유명인의 얼굴을 포르노 영상에 합성한 게시물을 올리면서 대중의 주목을 받았다. 7년이 지난 지금 생성형 인공지능 기술의 발전으로 인해 10대 청소년들조차 손쉽게 유명인뿐만 아니라 일반인들의 딥페이크 영상을 제작하고, 이를 소셜미디어를 통해 은

14) Bohacek & Farid(2024). Lost in Translation: Lip-Sync Deepfake Detection from Audio-Video Mismatch, CVPR Workshop.

밀히 공유하는 상황이 벌어지고 있다. 특히 딥페이크가 음란물 제작에 주로 활용되면서 여성들이 주된 피해자로 나타나고 있다.

이러한 문제를 해결하기 위해 딥페이크 식별과 유통 억제를 위한 기술 개발이 활발히 진행되고 있지만 학술적 연구 성과와 실제 적용 간의 격차가 있는 것이 사실이다. 현재는 고품질 딥페이크나 새로운 유형의 딥페이크에 대한 탐지 능력을 강화하는 기술 개발에 주력하고 있다. 딥페이크 성범죄 문제

는 기술적, 법적, 윤리적, 교육적 차원의 다각적 접근이 필요하다. 딥페이크 탐지 기술의 개발, 관련 법규의 정비, 대중의 인식 제고를 위한 교육 프로그램 등이 종합적으로 추진되어야 할 것이다. 딥페이크 기술은 창의적 표현의 새로운 도구로 활용될 수 있는 동시에 심각한 사회적 위험 요소로 작용하고 있으므로 기술 발전과 사회적 합의를 통해 딥페이크 기술이 건전하고 생산적으로 활용될 수 있는 방안을 모색해야 할 것이다.

• 참고문헌 •

- 관계부처합동 보도자료(2024.11.). 딥페이크 성범죄 대응 강화 방안.
- Bohacek & Farid(2024). Lost in Translation: Lip-Sync Deepfake Detection from Audio-Video Mismatch, CVPR Workshop.
- CNET(2021.6.28.). Deepfake bot on Telegram is violating women by forging nudes from regular pics.
- Intel(2022.11.). Intel Introduces Real-Time Deepfake Detector.
- Microsoft Blogs(2021.9.1.). New Steps to Combat Disinformation.
- Rombach et al. (2022). High-Resolution Image Synthesis With Latent Diffusion Models, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10684-10695.
- Security Hero(2023.9). 2023 State of Deepfakes.
- Trend Micro(2024.7.30.). Surging Hype An Update on the Rising Abuse of GenAI.
- UNESCO(2023). "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI.
- U.S. Department of Sate(2024.9.23.). The Global AI Research Agenda.