DOMESTIC REPORT

딥페이크 기술과 투자 사기 광고

함민정
(고려대학교 정보문화연구소 전임연구원)

인공지능(AI) 기술의 발전이 가져온 딥페이크 기술은 실제와 거의 구분이 불가능한 오디오, 사진, 동영상을 생성함으로써 현대사회에 다양한 영향을 미치고 있다. 딥페이크 기술 발전 초기에는 유명인 음란물 생성으로 유명인 명예 훼손 문제를 일으켰으며, 최근에는 딥페이크 가짜 투자 광고가 투자 사기 피해를 야기하고 있다. 본고는 최근 국내외 딥페이크 투자 사기 광고 관련 사례들을 검토하고, 이런 사기 딥페이크 콘텐츠에 대처하려는 여러 노력을 고찰하고자 한다.

1. 들어가며

인공지능(AI) 기술의 급속한 발전은 현대사회에 극적인 변화를 가져오고 있으며, 우리의 일상생활에 다양한 영향을 미치고 있다. 딥페이크(Deepfake)는 '딥러닝(Deep Learning)'과 '가짜(Fake)'의 합성어로, 인공지능 기술을 활용하여 실제와 구별하기 어려운 가짜 오디오, 사진, 동영상을 생성하는 기술이다. 딥페이크는 주로 딥러닝 기술을 사용하여, 기존의 사진, 영상, 음성 등을 합성하거나 변조, 새로운 콘텐츠를 생성한다.[1) 2)] 딥페이크 기술의 등장 이전, 가짜 콘텐츠는 주로 단순한 포토샵 작업이나 비전문적 비디오 편집에 의존했는데, 이러한 초기 가짜 콘텐츠는 대체로 품질이 낮아 피사체의 부자연스러운 움직임이나 피사체 주변의 흐릿한 경계 등으로 합성 콘텐츠임을 식별하기 쉬웠다.[2)] 그러나, 딥페이크 기술 발전으로 매우 정교하고 실제와 구분하기 어려운 콘텐츠가 생산되고 있다.[2)] 이러한 콘텐츠는 눈으로 봐서는 진짜와 가짜를 구별하기 어렵다. 딥페이크 기술은 유명 인사들의 얼굴을 다른 사람의 몸에 합성하는 등 가짜 음란물로 명예 훼손 사건을 일으켰으며, 최근에는 음란물 외 투자 사기 광고 등에 활용되어 투자 사기(Investment Scam) 등 문제를 유발하고 있다. 본고의 2장에서는 광고 조작(Advertising Manipulation)*의 발달 과정을 살펴보고, 3장에서는 딥페이크 기술을 활용한 국내외 투자 사기 광고 사례들을 살펴본 후, 4장에서 딥페이크 사기 광고 대응과 전망을 살펴보고자 한다.

> *광고 조작이라는 용어에서 조작이라는 용어가 부정적 의미로 받아들여질 수 있다. 그러나 최신 선행연구를 근거로, 본 고는 광고 조작을 광고 제작 전후에 광고 콘텐츠를 개선하기 위해 적용되는 모든 기술의 집합[1)]으로 정의하고자 한다. 즉, 본 고에서 광고 조작은 광고 콘텐츠의 질을 높이거나 광고 효과를 극대화하기 위한 창의적 활동과 기술적 활용을 의미한다. 광고 조작의 한 방식으로서 딥페이크 기술은 긍부정적 측면을 모두 갖고 있지만, 대체로 사기성 광고 제작에 활용된다는 점에서 부정적으로 평가된다.

2. 광고 조작(Manipulation in Advertising)의 진화

광고 조작이란 광고 제작 전후 적용되는 다양한 기술을 통해 광고 콘텐츠의 품질을 향상하는 포괄적 활동을 의미한다.[1)] 이러한 활동에는 사진 리터치, 특수 조명 사용, 특정 카메라 렌즈 사용, 메이크업 등이 포함되며, 이는 기술의 발전에 따라 1.0세대부터 3.0세대까지 진화해왔다.[1)] 1.0세대 광고 조작은 아날로그 조작(Analog Manipulation)으로, 광고 제작자가 콘텐츠를 완성한 후 흠을 자연스럽게 제거하기 위해 물리적 도구를 사용하는 기법이다. 이는 주로 메이크업, 조명, 에어브러시 사용 등을 포함하며, 특히 사진 광고에 주로 사용됐다. 광고 콘텐츠 제작을 위해 가능한 최상의 조건을 만드는 것을 목적으로 한다.[3)]

다음으로, 2.0세대 광고 조작은 디지털 조작(Digital Manipulation)이다. 이는 포토샵이나 파이널 컷 프로, 인스타그램 필터 등 디지털 도구를 사용한 사진이나 비디오 수정을 의미한다. 고차원적으로는 컴퓨터 생성 이미지(Computer Generated Image) 사용도 포함한다. 아날로그 조작이 광고 제작 시 여전히 흔하게 사용되지만, 디지털 도구 활용으로 점점 전환되는 추세이기도 하다.[1)]

마지막으로, 3.0세대 광고 조작은 합성 조작(Synthetic Manipulation)이다. 인공지능 알고리즘을 사용해 콘텐츠를 자동으로 편집하거나 생성하는 기술을 포함한다. 딥페이크는 이러한 합성 조작의 한 예이며, 합성 조작은 소비자에게 초개인화된 광고를 제공함으로써 소비자의 관심과 반응을 극대화하는 것을 목적으로 한다.[1)]

1.0~3.0세대의 광고 조작은 서로 대체하기보다는 보완하며 공존한다.[1)] 아날로그 조작은 다른 조작들에 비해 상대적으로 간단한 도구를 통해 이뤄지고, 디지털 조작은 콘텐츠 편집과 개선을 더욱 손쉽게 만들며, 합성 조작은 고품질 광고를 적은 비용으로 생산할 새로운 기회를 제공하고 있다.[1)]

표 1 광고 조작의 진화(Generations of Manipulation in Advertising)

세대	1세대: 아날로그(Analog)	2세대: 디지털(Digital)	3세대: 합성(Synthetic)
사용 도구 예시	메이크업, 조명, 카메라 렌즈, 물리적 편집	컴퓨터 생성 이미지, 포토샵, 인스타그램 필터 등	딥페이크, 생성적 적대 신경망(GANs)
주체	인간(수동)	인간-컴퓨터 상호작용	인공지능(AI와 머신러닝 기법 자동화)
대상	대중(Mass Focused)	개인, 기업 대 기업(B2B)	초개인화(Hyper-personalization)
채널	TV, 라디오, 인쇄 매체	TV, 인쇄 매체, 온라인	온라인

출처: Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2022). Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising, 51(1), 22-38.

3. 딥페이크 기술을 활용한 투자 사기 광고 사례

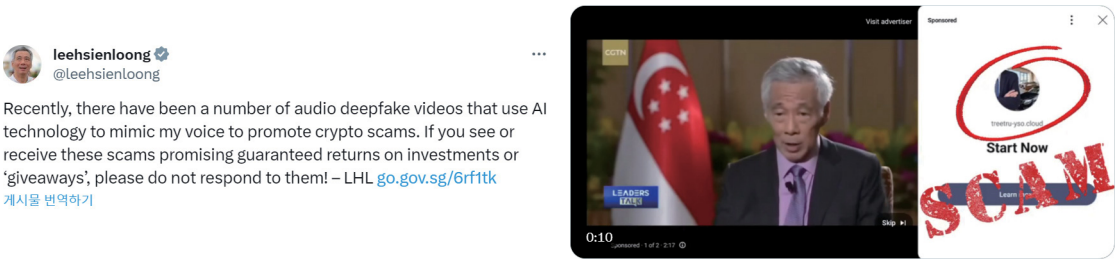
딥페이크 기술이 광고 분야에서 활용되고 있지만, 그 결과가 대부분 긍정적이라고 보긴 어렵다. 특히, 최근에 소비자들의 투자를 유도하는 사기성 광고에 주로 이 기술이 사용되고 있어, 그 위험성이 높게 보고되고 있다. 즉, 딥페이크 기술의 부정적 사례들이 긍정적 사례들보다 훨씬 많이 알려지고 있어서 기술의 부적절한 사용이 불러올 위험성에 대한 우려가 존재한다는 것이다.

3-1. 해외 사례

유명 정치인을 사칭한 딥페이크 투자 사기 광고

한 동영상에서 싱가포르의 리셴룽 총리(Lee Hsien Loong)가 싱가포르 정부가 지원하고 테슬라 CEO인 일론 머스크(Elon Musk)가 주관하는 투자 플랫폼을 홍보하며 암호화폐 투자를 조장한다.[5)] 그러나 해당 영상은 리 총리의 얼굴과 음성을 딥페이크 기술로 재창조한 투자 사기 광고였다. 그림 1에서 보는 바와 같이, 리 총리는 자신의 소셜 미디어에 해당 광고를 업로드하며 "투자 수익 보장을 약속하는 광고를 보면 절대 대응하지 말라(If you see or receive these scams promising guaranteed returns on investments or 'giveaways', please do not respond to them!)"는 메세지를 남겼고[5)] 딥페이크 광고에 대한 위험을 공개적으로 경고한 바 있다.

그림 1 싱가포르 리셴룽 총리의 딥페이크 투자 사기 광고 경고



출처: 리 총리 소셜 미디어

유명 사업가를 사칭한 딥페이크 투자 사기 광고

호주의 유명 사업가 딕 스미스(Dick Smith)는 페이스북과 인스타그램을 통해 유포된 딥페이크 투자 사기 광고의 위험성을 경고했다.[5] 해당 광고는 호주 시사 프로그램인 <A Current Affair>의 한 부분처럼 조작되어, 진행자 앨리슨 랭던(Allison Langdon)이 딕 스미스와 또 다른 호주 유명 사업가인 지나 라인하트(Gina Rinehart)를 인터뷰하는 형태로 투자 기회를 홍보한다.[6] 딕 스미스는 자신의 웹사이트를 통해 이 비디오가 자신의 말을 조작하여 만든 완전한 사기임을 밝혔으며, 대중에게 이러한 동영상을 믿지 말고, 페이스북이나 인스타그램 광고를 통해 투자를 판단하는 행위를 자제할 것을 촉구했다.[5] 실제로 퀸즐랜드에 사는 파울라(Paula)와 론(Ron) 부부는 이 딥페이크 광고에 속아 모든 연금을 잃었다.[6]

3-2. 국내 사례

유명 작가와 연예인을 사칭한 딥페이크 투자 사기 광고

일명 '배터리 아저씨'로 불리는 박순혁 작가가 유튜브 광고에 등장해 고수익 투자 기회를 제시하며 자신의 소셜 미디어에 친구 추가를 할 것을 유도하고 이후 네이버 밴드를 통해 주식 투자 정보를 제공할 것이라 안내한다.[7] 이어서 이은주 비서가 등장해 박순혁 작가가 투자자들의 자산을 불려줄 것이라며 적극적으로 투자를 권유한다.[7] 또 다른 광고에는 배우 조인성과 송혜교가 출연해 박순혁 작가의 자선사업을 응원하며 동참하겠다는 식으로 투자를 유도한다.[7] 그러나 팩트 체크 결과, 광고에 등장한 이은주 비서는 한 방송사 기상 캐스터의 얼굴을 무단으로 사용한 것이었고, 박순혁 작가 본인은 이러한 광고와의 연관성을 전면 부인하며 소셜 미디어를 사용하지 않는다고 밝혔다.[7] 배우 조인성과 송혜교의 모습 역시 딥페이크 기술로 조작된 가짜였다.[7] 안타깝게도, 이 광고의 사기성을 인지하지 못한 투자자들은 유명인의 추천에 속아 대포 통장으로 추정되는 계좌에 약 6,600여 만 원을 이체했으며, 이 투자금을 되찾기는 매우 어려울 것으로 전망되고 있다.[7]

유명 방송사 사장을 사칭한 딥페이크 투자 사기 광고

손석희 전 JTBC 보도 담당 사장이 페이스북 광고에 등장해 "한국인을 위한 혁신적 플랫폼을 개발해 AI 기반 투자를 통해 경제적 자유를 얻었으며, 500원을 투자하면 매월 최대 1만 5,000원의 수익을 얻을 수 있다"고 말한다.[8] 이어서 손석희 사장은 "인생을 바꿀 기회를 놓치지 말라"며 광고 링크를 클릭하라고 유도한다.[8] 이 광고는 손 전 사장의 얼굴과 음성을 합성한 딥페이크 광고였다.[8] 이와 같은 사기성 광고는 특정 전문가나 유명인의 권위를 남용하여 리딩방 가입을 유도하며, 가짜 시스템을 만들어 높은 수익을 보장하는 것처럼 속여 투자금을 가로채는 사기로 이어질 위험이 크다. 딥페이크 기술의 고도화로 콘텐츠 진위 구별이 점점 더 어려워지는 데 대한 우려가 제기되고 있다.

4. 마치며: 딥페이크 광고 대응 및 전망

앞서 딥페이크 광고 콘텐츠가 어떻게 투자 사기를 유도하는지 검토하였다. 그러나 이에 대응해 딥페이크를 탐지하고 방지하는 기술이 개발돼 디지털 콘텐츠의 신뢰도를 지키는 데 핵심적인 역할을 할 것으로 기대되고 있다.

딥페이크 탐지 기술은 AI를 활용하여 사람 얼굴의 혈류 변화를 추적하고 영상을 픽셀 단위로 분석함으로써 딥페이크 적용 여부를 정밀하게 파악한다.[9] 예를 들어, 인텔(Intel)의 페이크캐처(Fakecatcher)는 실시간으로 딥페이크 유무를 분석하며 높은 정확도를 자랑하고, DARPA가 진행하는 세마포(SemaFor)는 딥페이크 콘텐츠 생성 시 발생하는 'AI의 실수'를 근거로 딥페이크 유무를 감지한다. 한편, 딥페이크 방지 기술로 워터마크를 통한 구분 방법도 고려되고 있다.[9][10] 구글 딥마인드의 신스ID(SynthID)는 AI 이미지 생성 플랫폼에서 만든 이미지에 육안으로 보이지 않는 워터마크를 넣어 실제가 아님을 식별할 수 있게 한다.[9] 궁극적으로, 이러한 탐지·방지 기술들의 진보가 딥페이크로 인한 사회적, 경제적 손해를 줄이는 데 기여할 것으로 보인다.

딥페이크 기술의 발전과 이를 탐지하는 기술 사이에는 지속적 경쟁이 존재하며, 마치 '고양이와 쥐 게임'[9]처럼 한쪽이 발전할 때마다 다른 한쪽도 그에 맞춰 발전한다. 과학 학술지 네이처에 따르면 현재 잘 알려진 딥페이크 생성기들에 대한 탐지 성공률은 높지만, 새로운 생성기에 대한 탐지 성공률은 상대적으로 낮은 편이다.[9] 즉, 딥페이크 기술이 더욱 정교해지면서 딥페이크 탐지 기술 역시 더 발전된 AI 알고리즘과 탐지 기술을 갖추기 위해 노력할 필요가 있다는 것이다.

한편, 디지털 플랫폼들도 딥페이크 피해를 줄이려 노력하고 있다. 메타와 구글과 같은 대기업들은 사용자들이 AI 생성 콘텐츠를 인식할 수 있도록 정치 광고에 AI 사용을 명시하는 정책을 시행하고 있으며, 이러한 정책은 가짜 뉴스와의 싸움에 중요한 전략이 될 것이다.[11] 이런 정책들이 정치 광고뿐만 아니라 투자 광고를 포함해 일반 콘텐츠에도 확장 적용되어 가짜 정보의 확산을 막는 데 기여할 것으로 기대된다.

주석 Commentary

1) Campbell, C., Plangger, K., Sands, S., & Kietzmann, J.(2022), Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising, 51(1), 22-38.

2) 이민석(2023.08.15.), 대선 뒤덮는 딥페이크...미국 정부, AI 가짜뉴스에 말 뽑았다.

3) McDonald, C., & Scott, J.(2007), A brief history of advertising. The Sage Handbook of Advertising, London: Sage, 17-34.

4) Omelchenko, D.(2023.12.29.), Singapore's prime minister issues warning following deepfake crypto video with him

5) Bucci, N.(2023.11.27.), Dick Smith criticises Facebook after scammers circulate deepfake video ad

6) NCA NewsWire(2023.12.19.), Dick Smith warns about growing threat of artifical intelligence scmas

7) 박세용(2023.12.27.), 조인성, 송혜교 영상 보고 투자했는데 실체는?

8) 금준경(2023.11.23.), 손석희 "재정적 자유를 위한 길을 열었습니다" 사칭 광고 딥페이크까지 등장

9) 황규락(2023.10.17.), 딥페이크 교활해질수록 탐지 기술도 정교해진다

10) 이도경(2023.11.06.), 음란물, 보이스피싱, 가짜뉴스...AI 딥페이크 점입가경

11) 노정연(2023.11.09.), 메타, 내년부터 페이스북 정치 광고에 AI 사용 여부 공개