딥페이크 뉴스 대처방안

- 미국과 EU의 입법 사례를 중심으로 -

성균관대학교 법학전문대학원
지성우 교수

(2024.09.26(목). 한반도선진화재단)

제1장 들어가는 말

- 인공지능(artificial intelligence: AI)이란 인간의 학습능력, 추론능력, 지각능력을 인공적으로 구현하려는 컴퓨터 과학의 하위 연구영역임. 정보기술에서 인프라 기술 중 하나이며, 인간이 가진 자연 지능(natural intelligence)을 모방하여 구상됨
- 인공지능을 구동하기 위한 핵심 기술은 데이터(기계학습)-클라우드-IoT-6G, 유선망-S/W, H/W 등임
- 미국 연방법률에서는 “인공지능”이란 인간이 정의한 일련의 주어진 목표에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 권고 또는 결정을 내릴 수 있는 기계 기반 시스템임. AI 시스템은 기계와 인간의 입력을 통해 현실 및 가상 환경을 인식하고 자동화된 분석을 통해 이러한 인식을 모델로 추상화하며 모델 추론을 사용하여 정보나 조치에 대한 선택지를 정립”한다(미국연방법전 제15편 제9401조(3))고 규정함
- 유럽연합 인공지능법안에서는 AI 시스템을 “[부속서 I]에 열거된 기술 및 접근법 중 하나 이상을 사용하여 개발되고 인간이 정한 목표로 설정되어 상호작용하는 환경에 영향을 미치는 콘텐츠, 예측, 추천 또는 결정과 같은 결과물을 생성할 수 있는 소프트웨어”로 정의함(제3조 제1호)

유튜브 공화국

조형래 칼럼

부국장 겸 경제에디터

사용자·시간 압도적 1위
혐오 편가르기 조장하고
광고·음악·쇼핑 등
시장 독점 갈수록 심화
국내 기업 역차별 불만도
거대 플랫폼 폐해 막을
최소한의 방어막은 있어야

한국인은 이제 유튜브로 뉴스를 본다. 글을 읽는 것보다 영상을 보는 게 편한 데다 마법 같은 유튜브 알고리즘은 내 구미에 맞는 콘텐츠만 골라서 보내준다. 내가 민주당 열성 지지자면 대통령과 여당을 공격하는 영상을 집중적으로 보여주는 식이다. 한번 빠져들면 점점 더 자극적인 콘텐츠를 찾게 되는 중독 현상, '토끼굴 효과'가 나타난다.

유튜브의 세계에서는 사실(fact)과 다양한 의견은 중요하지 않다. 불편한 진실보다는 누가 우리 편이냐가 더 중요하다. 여기에선 대통령을 "무식한 주정뱅이"라고 모욕해도, 야당 대표에게 "칼 맞은 김에 죽지"라고 저주해도 반박하는 사람이 없다. 같은 편끼리만 보는 영상이기 때문에 상대방을 더 자극적으로, 더 극단적으로 몰아세울수록 박수를 받는다. 이런 콘텐츠를 반복해서 접하다 보면 편향성은 갈수록 강해지고 때론 가짜 뉴스가 사실로 둔갑한다.

한국인의 유튜브 중독은 뉴스에만 국한되지 않는다. 출퇴근 지하철에서는 유튜브 음악을 듣고 점심때는 유튜브 맛집을 찾아다니고 유튜브로 주식·요리 공부를 하고 주말엔 15~20분짜리 영화 리뷰 영상을 몰아 보는 게 일상이 됐다. 학교 선생님들도 책이 아닌 유튜브 영상을 찾아보고 숙제를 하라고 한다. 한국의 유튜브 월간 사용자는 4547만명. 1인당 평균 사용 시간이 43시간으로 종주국인 미국(24시간)을 크게 앞선다. 토종 메신저 카카오톡(12시간), 네이버(9시간)도 이미 멀찌감치 따돌렸다. 전 국민 플랫폼인 유튜브는 수만 명의 유튜버들이 대박의 꿈을 좇아 쉴 새 없이 올리는 동영상을 기반으로 광고, 디지털 음원, 쇼핑으로 독점력을 확대하고 있다.

유튜브가 한국 인터넷 시장을 장악한 데에는 세계 최고의 한국 IT 인프라가 한몫 단단히 했다. 전 국민이 스마트폰을 보유하고 있고 통신 3사가 수십조 원을 투자해 세계 최초로 5G 전국 통신망을 깔아준 덕분에 스마트폰에서도 초고화질 영상을 즐길 수 있다. 유튜브의 국내 통신망 트래픽 비율은 28.6%로 넷플릭스(5.5%), 메타(페이스북·4.3%), 네이버(1.7%), 카카오(1.1%) 등 경쟁 업체들을 압도한다. 하지만 유튜브는 막강한 콘텐츠 파워를 무기로 국내외 인터넷 기업 중 유일하게 통신망 사용료를 한 푼도 내지 않는다.

게다가 서슬 퍼런 규제의 칼날도 유튜브 앞에서는 한없이 무뎌진다. 국내 기업들은 정부 가이드라인이나 지침 형태의 수많은 규제에 시달리지만, 유튜브는 해외 기업이라는 이유로 강제 규정이 아닌 가이드라인을 요리조리 피해간다. 국회 역시 CEO의 국감 증인 채택을 앞세워 국내 기업들은 쥐 잡듯이 잡으면서도 유튜브가 가짜 뉴스를 퍼뜨리는 숙주(宿主) 역할을 하는 것은 애써 외면한다. 또 유튜브가 광고 없이 동영상을 볼 수 있는 프리미엄 서비스에 음원 끼워 팔기를 해 단숨에 국내 디지털 음악 시장을 장악하고, 프리미엄 서비스 가격을 40% 넘게 올려도 아무런 제재를 받지 않는다. 한 인터넷 기업 임원은 "최근 방송통신위원회가 유튜브에 밀려 망해가는 국내 OTT(동영상 서비스) 4사 대표를 불러 가격 인상 자제를 당부했지만, 정작 가격 인상을 주도한 유튜브나 넷플릭스에는 아무 말도 못 하는 게 현실"이라고 했다. 국내 기업 역차별이라는 불만이 끊이지 않는 이유다.

해외에서는 빅테크를 규제하려는 움직임이 본격화하고 있다. 유럽연합이 유튜브·아마존·알리 등 20여 개 초거대 플랫폼 기업에 대해 불법 콘텐츠 유통 방지, 소비자 보호, 경쟁 환경 조성 의무를 부과하는 법안을 제정한 데 이어 영국·일본도 유사한 법안을 통과시켰다. 가짜 뉴스나 유해 제품의 온라인 유통에 대해 사용자·생산자 탓 말고 플랫폼 스스로가 책임지고 관리하라는 것이다.

한국에서도 공정거래위원회가 주도해 거대 플랫폼의 독점적 지위 남용을 제어하는 법안 제정에 착수했다가 총선 국면에 접어들면서 보류된 상태다. 하지만 시장 경쟁과 중소기업의 성장을 저해하는 독점, 민주주의 가치를 훼손하는 가짜 뉴스 유포에 대해서는 최소한의 방어막은 마련해야 한다. 이대로면 빅테크가 한국 대통령을 뽑을 날이 멀지 않았다.

朝鮮日報 2024년 9월 3일 화요일 A01면 종합

‘범죄 방조자’ 플랫폼, 책임도 안 진다

2일 유튜브에 한국 유명 여가수의 이름과 딥페이크(AI로 만든 진짜 같은 가짜 콘텐츠)를 영어로 함께 검색하자 해당 여가수의 얼굴을 합성한 영상이 수십 개 올라왔다. 목이 깊이 파인 옷을 입고 가슴을 절반쯤 드러낸 것도 있다. 여가수의 얼굴에 다른 여성의 몸을 합성한 것이다. 3년 전 올라온 이 영상의 조회 수는 153만회. 초상권을 침해한 명백한 불법 콘텐츠이지만, 계속 방치돼 있다. 조회 수 153만이면, 통상 발생하는 광고료는 약 205만원. 이를 제작자와 유튜브가 거의 절반씩 나눈다. 범죄 수익을 서로 공유한 것이다.

‘딥페이크 음란물’이 대거 유통된 텔레그램에는 지금도 비슷한 콘텐츠가 돌아다닌다고 한다. 한 플랫폼 기업 관계자는 “대화방 가입 조건만 조금 까다로워졌을 뿐, 크게 달라진 것이 없다”며 “텔레그램이 직접 조치를 취하지 않는 한, 이를 막는 건 불가능하다”고

‘범죄 방조자’ 거대 플랫폼

〈1〉 유해 콘텐츠의 진원지

유튜브·인스타그램·틱톡 등
딥페이크·마약 게시물 방치
자극적일수록 광고수익 많아
“부작용 막을 법적 장치 필요”

플랫폼

공급자가 콘텐츠·상품·서비스 등을 올려놓으면 이용자가 소비·구매하는 온라인 공간. 유튜브·페이스북·틱톡 등이 대표적이다.

램·틱톡 등 대부분의 플랫폼에서도 불법·유해 콘텐츠는 어렵지 않게 접할 수 있다. 유통뿐 아니라 단순 제작·소지도 처벌하는 방안을 정부가 추진하는 등 대책이 쏟아져 나오고 있지만, 플랫폼에서 불법 콘텐츠는 좀처럼 줄지 않고 있다. 플랫폼들이 불법·유해 콘텐츠를 적극적으로 추적·삭제하지 않고 방치하기 때문이다.

플랫폼 기업들은 “불법 콘텐츠는 제작적 이유가 더 크다는 것이 전문가들의 지적이다. 유해 콘텐츠와 가짜 뉴스 등 자극적 콘텐츠일수록 사람들을 더 많이 끌어들이고, 그로 인해 광고 수익이 늘기 때문이다. 최근 텔레그램에 ‘딥페이크 음란물’이 급증한 것도 지난 4월 텔레그램이 구독자 1000명 이상 채널(방) 개설자에게 광고 수익을 지급하기 시작한 것이 영향을 미쳤다는 지적이다. 특히 인공지능(AI)으로 콘텐츠 제작이 쉬워지면서, 불법·유해 콘텐츠를 매개로 한 제작자와 플랫폼의 공생관계는 더 강화되고 있다. 유현재 서강대 신문방송학과 교수는 “플랫폼은 최우선순위가 수익인 만큼, 수익을 내는 유해 콘텐츠를 방조해 온 것”이라고

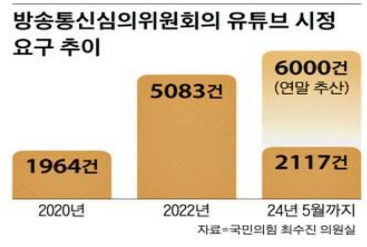
朝鮮日報 2024년 9월 3일

돈만 벌면 된다는 플랫폼… 가짜 뉴스·유해 콘텐츠

‘범죄 방조자’ 거대 플랫폼

〈1〉 유해 콘텐츠의 진원지

구독자 2만명을 보유하고 있는 한 연예 가십 유튜브 채널은 작년 5월 한 여자 아이돌 A씨가 다른 남자 아이돌 B씨와 사귄다는 3분짜리 허위 영상을 올렸다. 영상 게시 후 1년 3개월간 9만회에 가까운 조회 수를 기록 중이다. 두 사람이 등장하는 사진과 영상이 등장하고, 인공지능(AI) 음성이 허위 내용을 읽어준다. 역시 불법 콘텐츠다. 한 유튜브



플랫폼과 불법 제작자들의 공생
빅테크들 “콘텐츠 제작은 자유”

글로벌 플랫폼에선 어린이와 청소년들이 유해 콘텐츠를 접할 수 있다는 점도 문제다. 유튜브, 인스타그램, 틱톡 등 10대가 많이 이용하는 플랫폼 중 단 한 군데도 성인 인증이나 연령 인증을 하지 않기 때문이다. 회원 가입 시 출생 연도를 요구하긴 하지만, 실제 태어난 해를 기입하지 않아도 가입이 가능하다.

◇규제 근거 없으니 더 활개

구글, 메타(페이스북·인스타그램), 바이트댄스(틱톡) 등은 마약이나 폭력, 성매매 등 불법 콘텐츠에 대해서는 모니터링을 하고, 계정 정지나 수익화 금지 등

(유사 or 상위개념 1) 허위조작정보(?)

I. 허위조작정보의 개념

- ‘허위조작정보’, 소위 ‘가짜뉴스’(Fake News)라 함은, “소셜 미디어 또는 가짜뉴스웹사이트 등과 같이 온라인상으로 혹은 전통적인 뉴스 매체에서 발생하여 사실에 근거하지 않지만 마치 사실상 정확한 것으로 믿을 수 있도록 제공되는, 비풍자적 뉴스 기사”를 일컫는 신조어임
- 다시 말해 허위조작정보는 “소셜 미디어 또는 전통적인 뉴스 기사 등을 통하여 금전적 혹은 정치적 이득을 얻기 위해 오인할 의도를 가지고 거짓 또는 의도적으로 유포되는 잘못된 정보”라고 할 수 있음

(유사 or 상위개념 1) 허위조작정보(?)

II. 확산 원인

(1) 허위조작정보가 진짜 현실보다 더 사람들의 이목을 집중시킬 수 있음

- 통계에 의하면 대부분의 SNS 독자들은 뉴스 기사의 헤드라인만 읽는 경향이 있음

(2) 경제적 이익을 극대화 하기 위함 (SNS를 활용한 수익 창출)

- (예) 가짜뉴스 언론사인 Disinfomedia의 사장으로 가짜뉴스를 작성하고 가짜뉴스에 자금을 투입했던 Jestin Coler는 이 일을 그만둘 때까지 매달 $10,000에서 $30,000의 광고 수익을 얻음

(3) 가짜뉴스는 민의를 왜곡하고 그릇된 여론을 형성하기 위한 목적으로 의도적으로 생성·배포됨

- SNS서비스 관계망상의 이용자들간의 신뢰관계를 악용하면 좀 더 쉽게 여론을 왜곡할 수 있음

(유사개념 2) 보복성 폭로콘텐츠 처벌과 규제

I. 개념

- 데이트 관계나 부부 관계 등 기존의 사회적 관계망에서 발생한 개인적인 Intimacy 영상이나 사진 등을 피해자의 동의 없이 온라인에 게시하여 피해자에게 심각한 정신적·사회적 피해를 주는 행위를 말함
- Henry & Powell은 이러한 유형의 범죄를 **"기술매개 성폭력"(Technology Facilitated Sexual Violence)**이라 명명함
- "리벤지포르노"(Revenge Pornography)도 이러한 유형의 범죄를 말하나, **"보복성 폭로콘텐츠"가 더 광의의 일반적인 개념**임

II. 유형

1. 성적 사진/영상 유포

- 데이트 관계나 부부 관계에서 촬영된 성적인 사진이나 영상을 동의 없이 온라인에 게시하여 피해자의 명예를 훼손하고 수치심을 느끼게 함

2. 사생활 침해

- 피해자의 사생활에 관한 정보를 게시하여 사회적 차별과 불이익을 받도록 유도함

3. 협박 및 위협

- 피해자에게 불리한 정보를 유포하겠다고 협박하여 피해자를 통제하고 복종하도록 강제하고, 나아가 금품을 갈취함(**예: 쯔양사건**)

- 최근 국내·외에서 허위조작정보와 보복성 폭로 콘텐츠가 심각한 사회문제가 되고 있어 생성·유포자 본인과 정보통신서비스 제공자의 책임을 대폭 강화해야 한다는 주장이 제기되고 있음
- "보복성 폭로 콘텐츠를 양산"한다 함은 좁은 의미로는 남의 불행이나 사고, 실수, 결점, 잘못 등을 인터넷 상에 주로 영상의 형태로 공론화해서 이익(조회수, 인지도, 광고 수익)을 챙기는 사람이나 행위를 말함
- 또한 넓은 의미로는 본인의 경제적 이익에 관계 없이 남의 불행이나 사고, 실수, 결점, 잘못 등을 인터넷 상에 공론화하고 이슈거리로 만드는 사람이나 행위를 말함
- 영미와 유럽의 경우 'Commentary YouTubers' 혹은 'Drama Channel'이라는 명칭으로 한국보다 일찍 2000년대 중~후반부터 하나의 유투브 장르로 자리 잡았음. 또한 일본의 경우에도 한국보다 훨씬 더 명예훼손적 내용을 방영하는 유투브 채널들이 다수 방영되고 있음
- 보복성 폭로 콘텐츠를 온라인상에서 생산·유포하는 행위는 비교적 최근의 현상임. 다만 사건·사고가 발생한 경우 이를 취재해서 선정적으로 보도하던 **황색언론과 파파라치의 온라인 버전**이라고도 볼 수 있음
- 그러나 황색언론이나 파파라치와 근본적으로 다른 점은 최근의 경향이 기존의 인간관계에서 생성된 동영상, 음성 등 개인에게 불리하거나 사생활 침해, 명예훼손적 표현물들을 주로 경제적 목적으로 온라인을 통해 유포시킨다는 것임

(유사개념) 사이버레커(?)

I. 정의

- 교통사고가 나면 달려오는 “렉카” 차량처럼 조회수가 되는 현안이 생기면 불을 켜고 달려드는 일부 인터넷방송 진행자들을 뜻하는 표현임 (☞ 용어의 적정성에 대해서는 의문임)
- 최근 온라인 상에서 타인의 명예를 훼손하거나 금전적 피해를 주는 범죄행위로 문제가 됨
- 사이버 위협, 사이버 스토킹, 온라인 명예훼손, 개인정보 유출, 금품갈취 등이 대표적인 유형임

II. 적용범위

- 소셜미디어, 온라인 게시판, 뉴스사이트 등 온라인 플랫폼 전반

III. 예외 사항

- 순수한 개인적 의견 표현, 예술 및 학술 등은 제외됨

IV. 일반적 인식

- 2024.2. 한국언론진흥재단의 조사에 의하면 일반인의 92%가 사이버레커가 사회적 문제라고 인식
- 응답자의 94.3%가 사이버레커 등 유명인의 명예훼손 가해자에 대한 처벌을 강화해야 한다고 답변
- 응답자의 93.4%가 피해자 구제제도 강화, 88.2%가 플랫폼 자율규제의 강화 필요성이 있다고 답변

제2장 인공지능과 딥페이크의 위험성/역기능

- 딥페이크 기술은 인공지능(AI)과 딥러닝을 이용해 실제 인물의 얼굴과 목소리를 조작해 허위 콘텐츠를 만드는 기술. 이 기술의 발전으로 인해 다음과 같은 위험과 문제점이 발생할 가능성이 있음

➢ **허위 정보 유포**: 딥페이크 영상은 현실과 구분하기 힘들 정도로 정교하게 제작될 수 있음. 이를 통해 허위 정보를 유포해 사회적 혼란을 야기하거나, 특정 개인이나 단체에 대한 비방, 명예 훼손 등이 이루어질 수 있음

➢ **개인 프라이버시 침해**: 허가 없이 타인의 얼굴을 이용해 딥페이크 콘텐츠를 만들 경우 개인의 사생활이 침해될 수 있는데, 특히 유명인들이나 정치인들이 자주 표적이 됨

➢ **금융 및 사기 범죄**: 딥페이크 기술을 이용해 금융 사기나 신분 도용이 발생할 가능성도 있음. 특히 음성 딥페이크를 이용한 CEO 사칭 사기 등은 이미 실제 사례로 보고된 바 있음

➢ **신뢰의 상실**: 영상과 오디오 증거물이 가짜일 수 있다는 인식이 퍼지면, 법적 분쟁이나 공적 논의에서 진실을 가리는 데 있어 혼란이 발생할 수 있음

➢ **국가안보 및 공공질서 혼란**: 정치적으로 악용될 경우 국가안보에 대한 위협이 발생할 가능성이 있음(정치인들의 발언을 조작해 선거 목전에 혼란을 야기하는 경우 선거 결과에 심각한 영향을 미칠 우려가 있음)

제3장 딥페이크 뉴스에 대한 해외 입법 사례

제1절 미국

- 요약 : 바이든 행정부에서는 “AI 행정명령”과 아울러 연방 각 주에서도 개별사항에 대해 규정

1. 바이든 대통령의 “AI 행정명령”(2023.10.30.)

가. 개관

➢ 바이든 대통령은 지난 2023년 10월 30일 ‘안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령(Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence 14110, 이하 “**AI 행정명령**”)을 발표함

➢ 이 AI 행정명령은 총13개 조문(Section)으로 구성되어 있음. 미국 연방 50여 개 기관에 대해 100개 이상의 조치 시행을 지시함

➢ 구체적으로 목적(제1조), 정책 및 원칙(제2조), 정의(제3조), 8개의 지도 원칙(Guiding Principles)과 각 원칙별 이행하여야 할 구체적 지침과 조치사항(제4조 내지 제11조), 이행(제12조), 일반조항(제13조)으로 구성됨

나. 조문별 구체적인 내용

•제1조(목적)

• 제2조(정책 및 원칙)

• 제3조(정의)

• 제4조(AI 기술의 기술의 안전 및 보안의 보장)

• 제5조(혁신 및 경쟁의 경쟁의 촉진)

• 제6조(근로자에 대한 지원)

• 제7조(형평성 및 시민권의 진전)

• 제8조(소비자, 환자, 승객 및 학생에 대한 보호)

• 제9조(개인정보 보호)

• 제10조(연방정부의 AI 사용의 발전)

• 제11조(해외에서 미국 선도자로서의 강화)

• 제12조(이행)

• 제13조(일반조항)

다. 주요 내용 요약

(1) AI 시스템의 개발 보장

➢ 이 명령 시행일부터 270일 이내에 안전하고 보안이 보장되며 신뢰할 수 있는 AI 시스템의 개발을 보장하기 위해 상무부장관은 국립표준기술원장을 통해 에너지부장관, 국토안보부장관 및 상무부장관이 적절하다고 판단하는 그 밖의 관련 기관장과 조정하여 다음을 수행함

- AI 개발자, 특히 이중 용도 기초 모델의 개발자가 안전하고 보안이 보장되며 신뢰할 수 있는 시스템을 배포하기 위한 AI 레드티밍 시험을 수행할 수 있도록 적절한 절차와 절차를 포함한 적절한 지침(국가안보시스템의 구성요소로 사용되는 AI는 제외한다)을 수립할 것 (제4조 중 일부 번역)

(2) AI 시스템을 활용해 생성된 합성콘텐츠의 식별과 라벨링

➢ 합성 콘텐츠로 인한 위험의 축소 AI시스템을 활용하여 생성된 합성 콘텐츠를 식별하고 라벨링하는 기능을 육성하고 연방정부나 이를 대신하여 생성된 합성 및 비합성 디지털 콘텐츠의 진위성과 출처를 확립하기 위하여 다음을 수행함

- (a) 상무부장관이 이 명령 시행일부터 240 일 이내에 상무부장관이 적절하다고 판단하는 그 밖의 관련 기관장과 협의하여 행정관리예산국장과 국가안보보좌관에게 다음과 같은 기존 표준, 도구, 방법 및 관행 뿐만 아니라 과학적으로 뒷받침되는 추가 표준 및 기술의 잠재적 개발을 식별하는 보고서를 제출하는 것
  (i) 콘텐츠를 인증하고 출처를 추적하는 것
  (ii) 워터마킹 등을 사용하여 합성 콘텐츠에 라벨링을 하는 것
  (iii) 합성 콘텐츠를 발견하는 것
  (iv) 생성형 AI가 아동학대 자료를 생성하거나 실제 개인에 대해 합의되지 아니한 은밀한 이미지
  (v) 상기 목적상 사용되는 소프트웨어를 시험하는 것
  (vi) 합성 콘텐츠를 감사하고 유지하는 것

라. 평가

- 총론적으로 미국의 인공지능에 대한 대응은 낙관적이며 산업 진흥을 뒷받침하기 위한 지원적 성격, 미국 안보 위협 제거의 성격이 있는 것으로 평가됨
- 당해 행정명령은 연방정부와 기관들을 수범자로 하며 이들의 행위를 통해 시장에 영향력을 미치도록 설계되어 있음. 이 행정명령은 인공지능 관련 기업 등 민간을 직접적인 수범자로 하지 않고, 따라서 이를 위반했을 경우 제재 규정도 없음
- AI 행정명령은 마이크로소프트, 구글, 오픈 AI 등 AI 관련 15개 기업들이 자발적으로 합의한 AI의 안전한 개발과 사

2. 연방 주들에서의 딥페이크 규제

가. 선거운동에서 딥페이크 규제: 캘리포니아주, 택사스주, 하와이주, 애리조나주, 워싱턴주, 플로리다주, 뉴욕주

(1) 캘리포니아와 텍사스주

- 2019년 캘리포니아주와 텍사스주는 법안을 통과시킨 최초의 주임
- 새로운 캘리포니아주 법률(AB 730)하에서 선거 전 60일 기간 이내에 정치적 입후보자의 평판에 흠집을 내거나 유권자를 기만하여 후보자에 대한 찬반 투표를 유도할 의도로 해당 인물을 주인공으로 한 조작 콘텐츠를 배포하는 행위는 불법으로 간주됨
- 텍사스주 법률은 캘리포니아주와 매우 흡사하지만 30일 기간 이내의 배포를 금지할 뿐임
- 하지만 두 법률 모두 상당한 비판을 받았으며 특히 언론 자유에 대한 권리와의 양립가능성에 대해 의문이 제기되고 있음

(나) 하와이주, 애리조나주

- **하와이주**의 경우 H.B. 1766 및 S.B. 2396 법안이 의회에 제출되었는데, 하와이주에서 선거 전이나 선거 중에 딥페이크 또는 사기로 간주될 수 있는 허위 조작 정보의 유통과 매체 이용을 방지하는 내용을 담고 있음
- **애리조나주**는 S.B. 1078 및 S.B. 1336 법안에서 디지털화 된 오디오 녹음의 허위 사용 또는 금전적 이익 또는 악의적인 의도를 위해 딥페이크 비디오, 오디오 등의 매체를 무단으로 유포하는 것을 방지한다는 내용임
- **워싱턴주** 의회는 사기성 미디어와 선거의 공정성을 규제하는 S.B. 5152 법안를 통과시켰음. 이 법은 선거에 영향을 미칠 수 있는 조작된 미디어의 공개를 의무화하고 있음. 또한 정치 캠페인에 사용되는 AI 생성 콘텐츠를 명확하게 식별해야 함
- **플로리다주**도 S.B. 850 법안을 통하여 정치 광고 또는 선거 관련 매체가 생성형 AI로 만들어진 경우 라벨링을 의무화하고 있음
- **뉴욕주**도 2023년 10월부터 S. 1042 법률이 시행했는, 포르노와 선거 개입을 포함한 다양한 영역에서 딥페이크의 사용을 규제하는 것을 목표로 하면서 딥페이크 기술의 오용에 대한 명확한 지침과 행위에 상응하는 처벌 근거를 제공하고 있음

나. 포르노그래피 딥페이크에 대한 규제: 버지니아주, 캘리포니아주, 뉴욕주

- 포르노그래피 딥페이크로부터의 보호에 일차적인 관심을 갖는 일부 주의 법률은 덜 비판적인 평가를 받음
- 예컨대, 2019년 7월 **버지니아주**에서 통과된 법률은 한 개인을 강압·학대·협박할 의도를 가진 경우 해당 콘텐츠의 배포를 범죄로 규정함
- **캘리포니아주 법률(AB602)**은 개별 진정 절차를 간소화하기 위해 포르노그래피 딥페이크에 등장하는 **개인들의 소송권(private right of action)**을 도입했음. 2020년 말 통과된 **뉴욕주**의 법률 또한 **포르노그래피 딥페이크에 대한 소송권을 도입**함. 사후 40년간 한 예술가와의 유사성을 승인되지 않은 상업적 이용으로부터 보호하기 위해 해당 법률에서 사후 퍼블리시티권(right of publicity)을 정립하고 있음

3. 연방 수준의 법안

- 연방 수준에서 채택된 기존 조치들은 정보 수집의 체계화 및 제도화에 국한되기 때문에, 이들 법률에 의해 가용해지는 정보를 바탕으로 추가 조치가 뒤따를 수 있을 것임
- **생성적 적대 신경망 산출물 식별법**(Identifying Outputs of Generative Adversarial Networks Act, IOGAN Act)은 딥페이크 제작 및 진위성 분석 연구에 대한 국립과학재단National Science Foundation(NSF)의 지원과 딥페이크 표준에 관한 국립표준기술연구소National Institute of Standards and Technology (NIST)의 연구 수행, 그리고 딥페이크 탐지 방안에 관한 양 기관과 민간부문의 공동작업과 협력을 규정하고 있음
- 나아가 딥페이크 대응 조치는 **미 국방수권법National Defense Authorization Act(NDAA)에도 2년 연속으로 포함**되었음. 2021 NDAA는 국토안보부(DHS)가 향후 5년간 '디지털 콘텐츠 위조'에 관한 연간 보고서를 발행할 것을 지시하고 있음. 해외 국가의 딥페이크 사용에 관한 보고에만 초점을 맞춘 2020 NDAA와 달리 DHS는 현재의 관점을 확장하여 해외 정부의 딥페이크 활용 방안뿐만 아니라 딥페이크가 일반 대중에게 가하는 광범위한 위협 또한 검토하라는 지시를 부여받았음. 또한 이 법은 DHS가 딥페이크를 생성·탐지하고 이에 대응하는 방안을 연구할 것을 요구하였음

➢ 딥페이크 책임법(deepfake accountability act)

- EU와 관련성이 있는 한 가지 논의는 2019년 제출된 “딥페이크 책임법안”(deepfake accountability act)을 중심으로 전개되었음
- 법안은 딥페이크에 레이블을 표시해야 하며 그렇지 않을 경우 제작자는 중한 처벌을 받을 것이라고 명기하고 있음
- 이 안은 유해한 딥페이크 동영상 배포의 방지라는 실제적 목표를 달성하지 못한다는 비판을 받았음. 즉 결국 악의적인 딥페이크 제작자는 선진적 기술을 사용함으로써 적발을 피할 수 있으므로 자신의 행동을 바꾸지 않을 것임. 반면 합법적 영상 제작자만 불필요한 부담을 지게 될 것임. 또한 일부 조치에는 플랫폼에 책임을 부과하는 요소가 포함될 위험성이 있음
- 전자상거래 준칙과 마찬가지로 콘텐츠 제공자의 면책을 규정하는 “통신품위법”(Communications Decency Act) 제230조의 개정안에서도 동일한 문제가 발생한 사례가 있음
- 한편으로 이러한 논쟁은 언제나 표현의 자유에 대한 매우 절대적인 미국적 이해방식을 배경으로 전개되며 다른 한편으로 플랫폼에 과중한 규제 부담을 주지 않으려 한다는 점에서 관련법들의 적용 범위가 매우 제한적인 성격을 갖고 있음을 암시함(한국의 입법에도 참고할 필요가 있음)

➢ 딥 페이크 사기 방지법(Preventing Deep Fakes Scams Act)

- 최근 미국 연방 의회는 딥페이크를 규제하고 감독하기 위한 다음과 같은 몇 가지 법안을 통과시켰음
- 딥 페이크 사기 방지법(Preventing Deep Fakes Scams Act). H.R. 5808은 금융 서비스 부문의 인공 지능에 대한 태스크 포스(Task Force on Artificial Intelligence in the Financial Services Sector)를 설립할 근거를 두고 있고, 이 태스크 포스는 금융 서비스 부문의 AI와 관련된 문제에 대해 의회에 보고할 것을 규정하고 있음

➢ 딥페이크 책임법 (Deepfakes Accountability Act)

- 이 법은 딥페이크 기술로 인한 위협으로부터 국가 안보 기관을 보호하는 내용의 규정을 두고 있고, 유해한 딥페이크의 피해자에게 법적 구제 수단을 제공하는 내용을 담고 있음

➢ 기만적인 AI로부터 소비자 보호법(Protecting Consumers From Deceptive AI Act)

- 이 법은 미국 국립표준기술연구소(National Institute of Standards and Technology)가 AI와 딥페이크에 대한 태스크 포스를 구성하도록 규정하고 있음. 태스크 포스는 생성형 AI에 의해 생성된 콘텐츠 식별과 관련된 기술 표준 및 지침의 개발을 촉진하고 관련 정보를 제공하는 것을 목표로 함. 이러한 표준에 따르면 AI에 의해 생성되거나 실질적으로 수정된 오디오 또는 시각적 콘텐츠에 해당 콘텐츠의 출처를 공개하도록 의무화 하고 있음

➢ AI 사기 방지법(No AI Fraud At): H.R. 6943은 초상과 음성에 대한 개인의 재산적 권리를 규정하고 있음

1. 유럽연합(EU)의 인공지능법

가. 개관

- 유럽연합은 오랜 논의 끝에 지난 2021년 4월 21일 인공지능 규칙안(Proposal for a Regulation laying down harmonised rules on artificial intelligence)을 제안했고, 2023년 12월 28일 이 규칙안의 통과에 합의했음

나. 주요내용

1) 이 규칙안에서는 인공지능 시스템을 당해 인공지능 시스템이 가진 위험 수준에 따라 허용할 수 없는 위험(unacceptable risk) 수준, 고위험(high risk) 수준, 저위험(low risk) 수준, 최소위험(minimal risk) 수준 이렇게 4단계로 분류하고, 그 위험 수준에 비례하여 관리를 제안하고 있음. 이를 위험 수준에 기반한 규율(Risk-based regulation)이라고 함
2) 여기서 위험 수준의 평가는 당해 인공지능 시스템의 사용 목적, 사용 정도, 건강·안전·기본권에 미치는 영향, 피해의 강도와 범위, 피해 복원 가능성 정도 등을 고려하여 이루어짐
3) 이러한 분류의 효과는 허용할 수 없는 위험(unacceptable risk)을 초래할 수 있는 인공지능 시스템은 활용 금지, 고위험(high risk)을 초래할 수 있는 인공지능 시스템은 법안에서 언급한 필수요건을 충족하고 사전에 적합성평가를 수행한 경우에 제한적으로 허용, 저위험(low risk) 수준과 최소위험(minimal risk) 수준을 초래할 수 있는 인공지능 시스템은 고위험 인공지능 시스템에 부과된 필수요건을 충족하는 것이 강제되지는 않지만 자발적인 준수를 위한 행동강령(code of conducts)의 수립을 권장한다는 것임

4) 여기서 허용할 수 없는 위험 수준의 인공지능 시스템은 다음 4가지임

- ✓ 첫째, 신체적.심리적 피해를 유발하거나 유발가능성이 있는 방식으로 사람의 행동을 중대하게 왜곡하기 위한 잠재의식에 영향을 미치는 기술이 탑재된 인공지능 시스템의 시장출시, 서비스제공, 사용

- ✓ 둘째, 특정 집단에 속하는 사람의 행동을 중대하게 왜곡하기 위해 해당 집단의 연령, 신체 또는 정신장애로 인한 취약성을 이용하는 인공지능 시스템의 시장출시, 서비스제공, 사용

- ✓ 셋째, 공공기관이 자연인의 행동, 성격, 특성을 토대로 일정 기간에 걸쳐 그의 신뢰성을 평가 또는 분류하고 정당하지 못한이유와 방식으로 그 당사자 및 그와 비슷한 집단에게 해롭거나 불리하게 대우하는데 사용되는 인공지능 시스템의 시장출시, 서비스제공, 사용

- ✓ 넷째, 법집행을 위한 목적으로 공개적으로 접근가능한 공간에서‘실시간’ 그리고 원격의 생체인식정보를 활용한 신원확인 시스템의 사. 다만 이 경우 실종아동 또는 범죄 피해자 표적 수색, 테러 방지나 중범죄자 색출의 경우 예외적으로 허용한다(이상 법안 제5조)

5) 여기서 고위험 수준의 인공지능 시스템은 다음과 같음

- ✓ 첫째, 부록(Annex) II에 언급된 유럽연합 규범을 적용받는 제품 또는 안전부품으로 사용되는 AI 시스템
- ✓ 둘째, Annex III에 열거된 생체인식, 교육, 고용, 필수적인 서비스, 법집행, 출입국관리, 사법행정 및 민주적 절차 등 기본권에 영향을 미칠 수 있는 환경에서 독자적으로 활용되는 인공지능 시스템
- ✓ 셋째, 이와 관련하여 유럽연합 집행위원회는 건강/안전에 피해를 주거나 기본권에 악영향을 미칠 수 있는 고위험 AI 시스템을 추가하여Annex III에 열거된 AI 시스템을 업데이트할 권한을 가짐(이상 법안 제6조-제7조).
- ✓ 여기서 정치운동에서 유권자에게 영향을 미치는 시스템은 고위험 수준, 딥페이크 시스템은 저위험 수준으로 평가됨

다. 평가

- ➢ 유럽연합 인공지능법안은 리스크 기반으로 인공지능을 유형화하고 사전주의 원칙(precautionary principle)에 따라 상대적으로 엄격한 일괄 규제를 하는 것임. 이것은 인공지능 기술과 산업 육성보다는 소비자 권리 보호와 유럽연합 역내 통상 보호라는 관점이 전제되어 있다고 해석되고 있음

2. 유럽연합의 딥페이크 규제

가. 유럽의회 결의안

- 유럽의회는 조작 개입과 허위정보에 맞서 민주적 선거를 보호하기 위한 전 유럽적인 활동에 지속적으로 참여해 왔음. 유럽의회는 몇 차례의 결의안과 보고서를 통해 인공지능의 바람직하지 않은 결과를 다루는 구체적 조치를 함
- 딥페이크에 관한 특별한 언급은 딥페이크 자료 또는 합성 동영상의 제작자에게 레이블 표시 요건을 도입할 것을 집행위원회에 촉구한 2019년 2월의 결의안('Report on the Implementation of the Action Plan Against Disinformation) 등 의회가 밝힌 여러 입장에서 찾아볼 수 있음. 선거 상황에서 딥페이크 사용에 대해 엄격한 제한을 도입하라는 요구는 이 문서의 대부분에서 찾아볼 수 있음
- 인공지능 개발의 지식재산권에 관한 2020년 보고서에서 의회는 딥페이크를 통한 대중 조작의 가능성과 싸우기 위해 인식 및 미디어 리터러시 제고를 촉구하고 있음
- 딥페이크 쟁점의 논의와 관련해 가장 포괄적인 최신의 문서는 '교육, 문화, 시청각 부문의 인공지능'에 관한 2021년 5월 19일의 결의안임. 이 결의안에는 다양한 제안이 담겨 있다. 여기에는 딥페이크의 위험에 대한 인식 제고 및 디지털 리터러시 개선의 중요성(90번), 기술적 수단을 통한 허위조작 콘텐츠 탐지 및 레이블 표시 애로의 해소(91번), 집행위원회를 향한 유해 목적의 딥페이크 창작·제작·배포를 관할하는 적정한 법적 준거틀 도입 촉구(91번), 탐지 역량의 추가 개발 장려(92번), 플랫폼 사용자에게 노출되는 콘텐츠 관련 투명성 개선 및 정보의 수령 여부와 수령 정보의 내용에 관한 사용자의 결정권 제고(93번) 등이 포함되어 있음

나. 디지털서비스법(Digital Sevice Act: DSA) 개정

1) 개관

- 유럽연합(EU) 집행위원회는 '디지털서비스법'(DSA)을 개정하여 딥페이크 정보에 대응하고 있음
- 이 법은 구글, 트위터, 페이스북, 틱톡, 인스타그램 등 유럽연합 내 이용자가 월 4500만 명 이상인 19개 글로벌 플랫폼을 규제함
- 딥페이크 정보를 비롯한 불법정보를 방치하거나 방지 시스템을 갖추지 않으면 연간 글로벌 매출의 최대 6%를 과징금으로 부과한다는 내용임
- 이 법은 2022년 10월 19일에 채택되어 2022년 10월 27일에 유럽 연합 관보에 게재되었으며, 2022년 11월 16일에 발효되었음 . 매우 큰 온라인 플랫폼과 검색 엔진은 2023년 8월 23일까지 시행이 유예되었으며 현재 시행되고 있음
- 이 법에 따라 온라인 플랫폼 사업자는 유해·불법 콘텐츠를 신속히 제거하고 예방할 수 있는 시스템을 마련해야 하며, 위반 시 연간 글로벌 수익의 최대 6%를 과징금으로 부과 받을 수 있음. 즉 규제 대상은 온라인 중개자와 플랫폼으로서 시장, 사회 네트워크, 콘텐츠 공유 플랫폼, 앱스토어 등을 포함함
- EU 집행부는 시민들의 기본적 권리를 더 잘 보호하고, 통제권과 선택권을 부여하며, 불법적 콘텐츠 노출을 줄여서 온라인 상에서 어린이를 강하게 보호하고자 한다고 이 법의 취지를 밝혔음. 즉 사용자의 안전, 기본적 인권의 보호, 공정하고 개방적인 플랫폼 환경을 만드는 것임
- 디지털 서비스 제공자에게 법적 명확성과 EU 전역에 적용되는 통일적인 규칙을 제공하고, 디지털 서비스를 사용하는 비즈니스 업계에게는 불법적 콘텐츠 제공자에게 맞설 수 있는 동등한 장을 마련해주는 것이 당해 법률의 목표임
- 이는 사회 전반에서 시스템 플랫폼에 대한 민주적 통제와 감시를 가능하게 하고, 허위 조작 정보와 같은 리스크를 줄일 수 있게 하기 위함임

2) 주요 내용

- 규제의 핵심은 EU 4억5천만 인구의 10% 이상인 4천5백만 유럽인을 한 달 평균 사용자 숫자로 가지고 있는 업체를 대규모 온라인 플랫폼과 검색엔진(Very large online platforms, VLOPs)이나 대규모 온라인 검색엔진(Very large search engines, VLOSEs)으로 지정하고 규제하는 것임
- 이렇게 지정된 업체는 4개월 이내로 다음의 사항을 준수해야 함
  - 첫째, 당국과 사용자들에게 연락처를 제공하고, 위법행위가 발생했을 시에는 보도해야 하며, 사용자 친화적 여건을 제공하고, 선전이나 추천 시스템 또는 콘텐츠 조정 결정에 관해 투명해야 함
  - 둘째, 지정된 업체들은 사회적 영향력을 고려하여 자신들의 서비스에 관련된 시스템 리스크를 규명하고 분석하고 평가해야 함
  - 셋째, 업체의 리스크가 규명되어 유럽연합 집행부에 감독을 위해 보고되면 이들 업체들은 이러한 리스크들을 줄이기 위한 조치들을 취해야 함
  - 넷째, VLOPs나 VLOSEs로 지정된 업체들은 확실하게 리스크를 규명하고 저감시키기 위한 내부통제기능을 확립해야 한다 최소 일 년에 한번 이상 독립적 회계감독을 실시함. 이러한 의무를 지키지 않을 시 2024년 2월 17일부터 전 세계에서 얻은 글로벌 수익의 6%까지 과징금이 부과될 수 있음

다. 유럽연합 개인정보보호법(General Data Protection Regulation: GDPR)

- 유럽에서 '딥페이크 허위조작정보'로 인한 피해자는 EU 회원국에 직접 적용되는 일반 개인정보보호법(General Data Protection Regulation, GDPR) 제17조 '잊힐 권리(right to erasure)'에 근거해 해당 콘텐츠의 삭제를 요청할 권리가 있으며, 동법 제21조에 의해 데이터 처리에 이의를 제기할 수 있음