KISDI Perspectives

초점 June 2024 No.4

딥페이크 관련 국내외 규제 현황 및 분석

노은정 전문연구원
정보통신정책연구원 미디어정책연구실 전문연구원

KISDI 정보통신정책연구원
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE

KISDI Perspectives

초점

2024. 6. 28

딥페이크 관련 국내외 규제 현황 및 분석

노은정 전문연구원
정보통신정책연구원 미디어정책연구실 전문연구원, eunroh@kisdi.re.kr

요 약

- 빠른 기술 발전으로 누구나 쉽게 딥페이크 영상을 생성할 수 있게 되면서 전 세계적으로 큰 문제가 되고 있음
- 구글과 유튜브에서는 딥페이크 탐지 기술을 개발하며 불법 딥페이크를 추적 및 식별하고 있음. 크리에이터가 AI를 활용했을 경우 해당 사실의 표기를 요구함
  - 구글은 ´24년 5월 30일부터 딥페이크를 활용하여 음란물 등을 변경/생성하는 서비스의 광고 금지
  - 유튜브는 ´24년 3월부터 크리에이이터에게 영상 업로드 시 AI 사용 여부 표시 의무화
- EU는 세계 최초로 AI를 포괄적으로 규제하는 「AI법」을 제정하였으며, 미국의 州 정부들은 선거 등과 관련한 별도의 딥페이크 법령을 제정
  - EU는 딥페이크 관련하여 AI 시스템의 공급자(provider)와 사용자(deployer)에게 투명성 의무 부과
  - 미국 캘리포니아 주는 선거 60일 전부터 후보자에 대해 기만적인 음성/영상 배포 금지
- 국내에서는 딥페이크로 인한 피해가 증가함에 따라 「성폭력범죄의처벌 등에 관한 특례법」에 처벌 규정을 신설하였으며, 「공직선거법」을 개정하여 딥페이크 영상을 활용한 선거운동을 금지함
- 딥페이크의 피해를 최소화하고 긍정적인 부분을 최대화할 수 있도록 정부의 효과적인 규제와 기술 개발 지원, 플랫폼의 빠른 대응, 그리고 이용자 교육이 병행되어야 함

01 개 요

- 딥페이크는 더이상 우리에게 낯선 기술이 아님. 우리는 생활 속 다양한 분야에서 쉽게 딥페이크 콘텐츠를 접할 수 있음
  - 딥페이크를 활용하여 유명 연예인의 목소리가 나오는 네비게이션 안내부터, 유명 가수 목소리로 재탄생한 다양한 노래의 커버까지, 딥페이크는 콘텐츠 생성의 지평을 획기적으로 넓히고 있음
- 딥페이크가 긍정적으로 사용된 예로, 영상 산업에서 딥페이크를 활용하여 현실적 제약을 뛰어넘는 혁신적인 제작이 가능해졌음
  - AI 합성 콘텐츠를 만드는 신서시아(Synthesia)에서 딥페이크를 이용해 데이비드 베컴이 말라리아 퇴치 캠페인 동영상을 중국어, 아랍어, 힌디어, 스와힐리어, 요루바어 등 9개 언어로 자막 또는 더빙하여 제작함[1]
  - 이밖에도 드라마 〈The Mandalorian〉에서 Luke Skywalker 역을 연기한 Mark Hamill의 젊은 시절을 딥페이크를 활용하여 제작하였으며[2], SBS 예능프로그램에서 故 김광석씨의 목소리를 구현함
- 하지만 딥페이크는 '성적 허위 영상물' 제작 등 사회적으로 심각한 문제를 야기하고 있음
  - 이전과 비교하여 일반인들도 쉽게 딥페이크 영상을 생성할 수 있게 되었으며, 이로 인해 허위정보 유포, 명예훼손 및 저작권 침해 등 발생하여 불법적인 딥페이크 악용 사례가 급증
  - 특히, 불법 딥페이크 허위 영상물의 피해자가 유명 연예인뿐만 아니라 청소년과 일반인으로 확대되고 있다는 점도 우려스러움

※ 방송통신심의위원회에 따르면, ´24년 1월부터 4월까지, 딥페이크 기술을 악용한 '성적 허위 영상물'은 모두 4,691 건으로 전년 대비 400% 폭증함[3]

※ ´19년부터 ´20년 사이에 딥페이크 온라인 콘텐츠의 수가 900% 증가하였으며, ´26년까지 온라인 콘텐츠의 최대 90%가 합성 콘텐츠(synthetically generated content)가 될 수 있다는 보고도 존재[4]

- 본 고에서는 딥페이크의 정의와 피해 유형을 살펴보고, 전 세계적으로 가장 많은 이용자를 보유한 구글과 유튜브를 중심으로 불법 딥페이크에 대한 대응 노력을 분석하며, 국내외 법제도를 검토하여 시사점을 도출하고자 함

1 한겨레 (2019.10.24.).
2 Ameena Qobrtay (2022.3.3.).
3 방송통신심의위원회 보도자료 (2024.5.2.).
4 World Economic Forum (2023.5.19.).

02 딥페이크 개념 및 피해 유형

- (개념) 딥페이크(Deepfake)는 Deep learning과 fake의 합성어로, 딥페이크의 중요한 특징은 인공지능을 활용하는 것
  - 유럽의회(European Parliament)에서 ´21년 발간한 보고서에 의하면, 딥페이크를 "머신러닝과 딥러닝을 포함한 인공지능 기술을 사용하여 제작된, 실제처럼 보이지만 실제 사람이 말하거나 행동하지 않은 것처럼 보이는 조작 또는 합성된 오디오 또는 시각 미디어"로 정의함[5]
  - ´17년 레딧(Reddit) 커뮤니티 이용자 아이디 "deepfakes"가 유명인 얼굴을 합성한 음란물들을 게시하며 "deep fake"라는 용어가 사용되기 시작함
- (문제점) 딥페이크의 문제는 합성 콘텐츠를 진짜처럼 보이게 하여 허위 사실을 유포하는데 있음
  - 더구나 딥페이크가 빠른 속도로 발전하면서 점점 진위를 판별하기 어려워지고 있음
  - 딥페이크는 초기에 주로 음란물을 제작하는 데 사용되었으나, 기술이 진화함에 따라 개인, 조직 및 사회 전체가 상당한 위험에 노출되어 있음
- (음란물 제작 및 유포) ´24년 1월, 유명 가수 테일러 스위프트의 딥페이크 영상이 소셜 플랫폼 "X"에서 유포되면서 다시 한번 딥페이크 음란물에 대한 심각성이 대두되었으며, 이제는 유명인뿐만 아니라 일반인들도 딥페이크 음란물 대상이 되고 있음
  - ´23년 12월, 미국 플로리다에서 중학생 두 명이 AI로 반 친구들의 누드 사진을 만든 혐의로 체포되어 사회에 큰 충격을 주었음[6]
- (금융사기) 딥페이크를 활용한 금융사기도 증가하고 있음. 로맨스 스캠의 경우, 이전에는 전화나 문자를 통해 이루어졌으나 이제는 딥페이크로 영상 통화를 통해 범죄가 이루어지고 있음
  - 실제 가족이나 지인의 목소리를 합성한 보이스피싱 사례도 있으며, 유명인의 얼굴을 합성한 영상을 만들어 투자를 유도하는 등 다양한 금융사기가 발생하고 있음
  - 또한, 영국 기반의 세계적인 구조설계회사 에이럽(Arup)이 딥페이크로 생성된 가짜 화상회의 영상에 속아 약 2,500만 달러(약 340억 원)를 사기범의 계좌에 이체한 사례도 있음[7]

5 European Parliamentary Research Service (2021), p.2.

6 Gaby Del Valle (2024.5.4.).

7 서울경제 (2024.5.18)

- (저작권 침해) 원곡자의 동의를 받지 않고 다른 가수 목소리로 딥페이크 음원을 만들어 배포하거나, 기존 영화나 TV프로그램에 배우 얼굴을 교체하는 등 원작을 훼손하여 저작권을 침해하는 사례가 증가하고 있음
  - ´23년 10월, 미국 음악산업협회(Recording Industry Association of America)는 AI 음성 복제 서비스(AI vocal clone)가 폭발적으로 증가하여 목소리가 복제되는 아티스트의 권리뿐만 아니라 음원을 소유한 저작권자의 권리를 침해되고 있어 피해를 주고 있다고 발표함[8]
- (정치 악용) ´22년 3월, 우크라이나 대통령이 우크라이나 국민들에게 러시아 군인과 전쟁을 중단하고 무기를 반납하라는 내용의 딥페이크 영상이 유포되어 충격을 주었음
  - 조 바이든 대통령의 목소리로 예비 선거에서 투표하지 말라고 독려한 영상 등 딥페이크를 활용하여 정치에 개입할 수 있다는 점에서 민주주의를 위협이 되고 있음[9]

03 구글 및 유튜브의 딥페이크 정책

- ´23년 7월, 바이든 행정부는 아마존(Amazon), 앤트로픽(Anthropic), 구글(Google), 인플렉션(Inflection), 메타(Meta), 마이크로소프트(Microsoft), 오픈AI(Open AI) 등 7개 주요 AI 기업에게 안전하고 투명한 AI 기술 개발을 위한 자발적인 노력을 요청함
  - 이에 해당 기업들은 ① AI 시스템 발표 전 제품의 안전성 보장(내외부 보안 테스트 시행), ② 보안을 최우선으로 하는 시스템 구축 ③ 대중의 신뢰 확보(AI 생성물임을 표시, 투명성 강화 등)를 위해 노력하겠다고 약속함[10]
- 구글은 이전부터 이용자 신고와 자동화된 시스템을 활용하여 불법적인 딥페이크 콘텐츠를 추적 및 식별하고 있었음

8 Recording Industry Association of America (2023.10.6.), p.14.

9 Mack Degeurin (2024.2.8.).

10 The White House (2023.7.21.).

● ´23년 4월, 구글은 아래와 같이 AI 원칙을 발표[11]

구글의 AI 원칙

1. 사회적으로 유익해야 함
2. 불공정한 편견을 만들거나 강화하지 않아야 함
3. 안전성을 우선으로 설계하고 테스트되어야 함
4. 인간을 위해 책임을 다해야 함
5. 개인정보 보호 설계 원칙을 적용
6. 과학적 우수성에 대한 높은 수준을 유지
7. 구글 AI 원칙에 부합하는 용도로만 활용될 수 있어야 함

● ´23년 12월, 구글은 불법 딥페이크를 탐지할 수 있는 새로운 인공지능 기반 도구를 개발하고 있다고 발표함

- 구글은 이미지와 비디오를 정밀하게 조사하여 불일치 여부와 오디오를 분석하여 머신러닝을 통하여 딥페이크 패턴과 이상 징후를 식별[12]

● ´24년 2월, 구글은 메타(Meta), OpenAI와 함께 콘텐츠 출처에 대한 기술 표준인 Content Credentials를 수용하기 위해 C2PA(Content Provenance and Authenticity; 콘텐츠 출처 및 진위성 연합)에 가입할 것이라고 발표함[13]

- C2PA의 기술로 콘텐츠에 메타 데이터를 삽입하면 콘텐츠의 출처를 확인할 수 있기 때문에, 이를 이용하여 해당 콘텐츠가 AI로 제작된 것인지의 여부를 확인할 수 있음

● ´24년 5월, 구글은 AI를 이용하여 음란하거나 과도한 노출이 포함되도록 변경하거나 생성할 수 있는 서비스의 광고를 금지한다는 내용의 정책을 발표(5월 30일부터 시행)[14]

- 기존에는 음란 광고를 금지하였으나, 광고주가 딥페이크를 활용하여 음란물 등을 생성하는 서비스를 홍보하는 것은 금지하지 않음[15]

11 구글코리아 블로그 (23.4.7).
12 Medium (2023.12.1.).
13 Mike Kaput (2024.2.20).
14 Google Advertising Polices Help (2024.5.1.)
15 Gaby Del Valle (2024.5.4.).

- 새로운 정책에 따라 5월부터는 딥페이크로 음란물을 생성하는 서비스를 제공하는 웹사이트 또는 앱, 딥페이크 음란물 생성 방법에 대한 방법을 알려주거나, 딥페이크 음란물을 홍보하는 서비스 등의 광고가 금지됨

● 유튜브도 ´23년 11월 AI를 통해 제작된 콘텐츠가 이용자에게 실제라는 오인을 줄 수 있는 경우, 크리에이터에게 AI 사용 여부를 공개하도록 요구할 것이라고 발표함[16]

※ AI 활용 여부를 공개해야 하는 콘텐츠의 예시로, 실제로 일어나지 않은 사건을 사실적으로 묘사하는 AI 제작 동영상이나 실제로 하지 않은 말이나 행동을 하는 사람을 보여주는 콘텐츠 등을 제시

- 크리에이터가 해당 정보를 공개하지 않을 경우 문제되는 콘텐츠를 삭제되거나, 유튜브 파트너 프로그램이 중단되거나 기타 처벌을 받을 수 있음
- 유튜브는 해당 표시를 두 가지 방법을 통하여 진행할 예정: ① 동영상 설명 부분에 콘텐츠 일부가 변형됐거나 합성됐다는 라벨을 추가 ② 민감한 주제를 다룬 특정 유형의 콘텐츠의 경우 더 눈에 띄는 라벨을 영상이 재생될 때 표시

● 유튜브는 콘텐츠의 불법성을 판단하기 위하여 2만 명 이상의 리뷰어(reviewer)와 머신러닝 기술을 활용하여 자사 가이드라인을 준수하는지 확인하고 있음

- AI를 통하여 잠재적 위반 가능성 있는 콘텐츠를 감지하고, 리뷰어가 해당 콘텐츠가 실제로 정책을 준수하지 않았는지 확인함[17]

● 유튜브는 해당 AI 라벨 정책을 ´24년 3월부터 시행하였으며, 이에 따라 크리에이이터가 영상 업로드 시 AI 사용 여부를 선택적으로 표시해야 함[18]

- 명백히 비현실적인 콘텐츠, 애니메이션, 특수 효과가 포함된 콘텐츠, 또는 제작 보조(production assistance)에 생성형 AI가 사용된 콘텐츠는 제외됨

※ 구체적으로, 애니메이션이나 유니콘을 타고 환상적인 세계를 여행하는 사람처럼 명백히 비현실적인 콘텐츠, 색상 조정 또는 조명 필터, 배경 흐림 또는 빈티지 효과와 같은 특수 효과, 뷰티 필터 또는 기타 시각적 향상 등이 이에 해당됨

16 YouTube Official Blog (2024.3.18.).
17 YouTube Official Blog (2023.11.14.).
18 YouTube Official Blog (2024.5.18.).

유튜브에서 콘텐츠 업로드 시 AI 활용 여부 표시

변경된 콘텐츠 (Altered Content)

다음 중 하나라도 해당되는 콘텐츠가 있나요?

- 실제 사람이 말하거나 하지 않은 것을 말하거나 행동하는 것처럼 보이게 하는 경우
- 실제 사건 또는 장소의 영상을 변경하는 경우
- 실제로 발생하지 않은 장면을 실제처럼 보이도록 생성하는 경우

- 예 (Yes) / 아니오 (No)

※ YouTube 정책을 준수하려면 콘텐츠가 변경되거나 합성되어 실제처럼 보이는지 여부를 알려야 합니다. 여기에는 AI 또는 기타 도구를 사용하여 만든 사실적인 음향 또는 비주얼이 포함됩니다. '예'를 선택하면 콘텐츠에 라벨이 추가됩니다.

자료: YouTube Official Blog (2024.5.18) 국문 번역.

04 유럽과 미국의 딥페이크 법제도 동향

Ⅰ 유럽연합(EU)

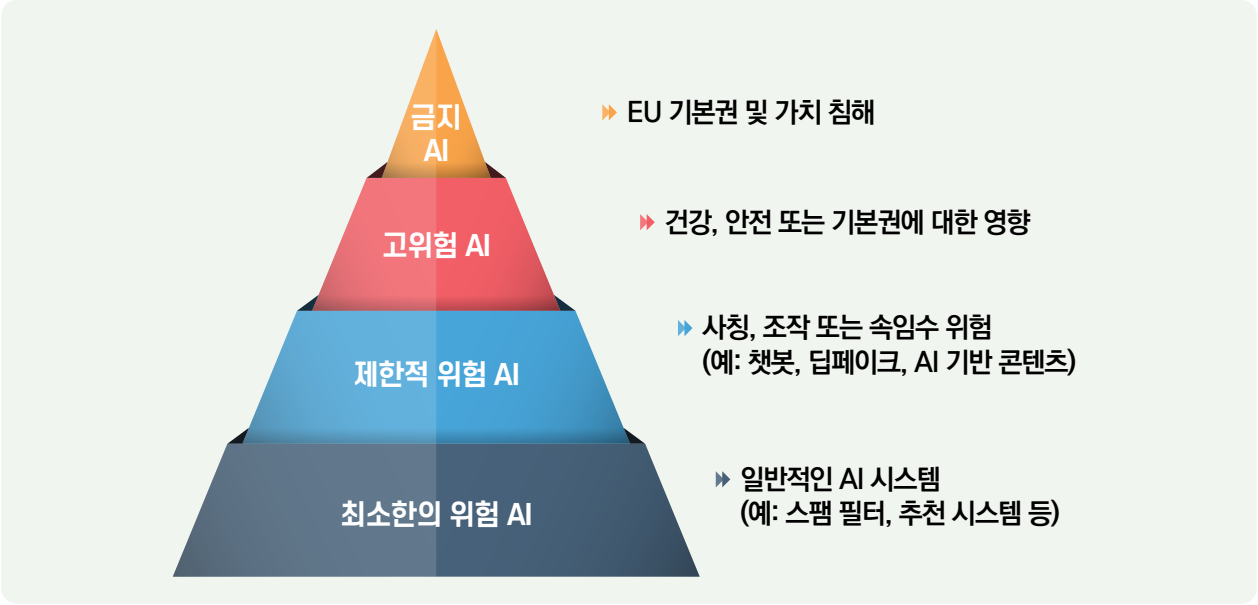
□ AI법

- ´24년 5월, EU 이사회(European Council)는 세계 최초로 AI를 포괄적으로 규제하는 「AI법(Artificial Intelligence Act)」를 최종 승인함
- (목적) AI법의 목적은 내부 시장의 기능을 개선하고, 인간 중심적이고 신뢰할 수 있는 인공지능(AI)의 채택을 촉진하는 동시에, 유럽연합 내 AI 시스템의 유해한 영향으로부터 건강·안전·민주주의, 법치 및 환경 보호를 포함하여 헌장에 명시된 기본권을 높은 수준으로 보호하고 혁신을 지원하는 데 있음(제1조)
  - AI법은 AI를 잠재적 위험에 따라 금지(unacceptable risk), 고위험(high risk), 제한적 위험(limited risk), 최소한의 위험(minimal risk) 등 네 가지 범주로 분류하였으며, 딥페이크는 '제한된 위험'에 해당[19]

※ 이밖에도 ChatGPT와 같은 범용(General-Purpose) AI 모델에 대하여 별도의 규제를 마련함

19 정재욱 (2024).

| 그림 1 | AI 잠재적 위험에 따른 분류



자료: Tambiama Madiega (2024), p.7 이미지 재구성 및 번역

- (적용 범위) EU 소재지와 관계없이 EU에서 AI 시스템을 제작·사용·수입 또는 배포하는 모든 사람에게 적용됨
  - 다른 곳에서 제작되었더라도 EU에서 사용되는 AI 시스템에도 적용(제2조)[20]
- (딥페이크의 정의) AI법은 '딥페이크'를 AI가 생성하거나 조작한 이미지, 오디오 또는 동영상 콘텐츠로서 실제 인물, 사물, 장소, 단체 또는 사건과 유사하여 사람에게 진짜인 것처럼 보이거나 진실인 것처럼 보이게 하는 것으로 정의

  ※ Artificial Intelligence Act Article 3 (60). 'Deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful
- (주요 내용) 딥페이크와 관련하여 AI 시스템 공급자(provider; AI 시스템 개발)*와 사용자(deployer: AI 시스템 사용)*에게 투명성 의무(Transparency Obligation)를 부과하고 있음

20 군사, 국방 또는 국가 안보 목적으로 사용되는 AI 시스템이나 외국 공공 기관 또는 국제기구가 법 집행 및 사법 협력을 위해 사용하는 AI 시스템에는 개인의 권리를 보호하는 한 이 규정이 적용되지 않음. 과학 연구 및 개발에 사용되는 AI 시스템이나 아직 시장에 출시되지 않은 AI 시스템에는 적용되지 않음. 데이터 보호, 개인정보 보호 및 기밀 유지에 관한 기존 EU 법률에도 영향을 미치지 않음. 개인적/비전문적 활동을 위해 AI 시스템을 사용하는 개인이나 무료 및 오픈 소스 라이센스에 따라 공개된 AI 시스템에는 고위험이거나 특정 조항에 해당하지 않는 한 적용되지 않음; EU Artificial Inteliigence Act.

※ 공급자(provider): AI 시스템 또는 범용 AI 모델(general-purpose AI model)을 개발하거나, AI 시스템 또는 범용 AI 모델을 개발하여 시장에 출시하거나 자신의 이름 또는 상표를 사용하여 유상 또는 무상으로 AI 시스템을 서비스하는 개인 또는 법인, 공공 기관, 대행사 또는 기타 단체

※ 사용자(deployer): AI 시스템을 사용하는 개인 또는 법인, 공공 기관, 대행사 또는 기타 단체 (사적인 비전문적 활동에서 사용되는 경우 제외)

- 표시 및 감지 의무: 합성된 오디오·이미지·비디오 또는 텍스트 콘텐츠를 생성하는 AI 시스템의 공급자(provider)는 해당 결과물이 기계가 판독할 수 있는 형식으로 표시하고, 인위적으로 생성 또는 조작된 것으로 감지할 수 있어야 함(제50조 제2항)

  ※ 이를 구현할 수 있는 기술적인 방법으로, 워터마크·메타데이터 식별·콘텐츠의 출처/진위를 증명하기 위한 암호화 등이 있음

- 공개 의무: 딥페이크에 해당하는 이미지·오디오 또는 비디오 콘텐츠를 생성하거나 조작하는 AI 시스템의 사용자(deployer)는 결과물이 인위적으로 생성 또는 조작되었음을 공개해야 함(제50조 제4항)

  ※ 해당 결과물에 라벨을 표시하고 출처(artificial origin)를 밝힘으로써 명확하게 구별할 수 있도록 공개해야 함

- (벌금) AI 시스템의 공급자와 사용자가 딥페이크 관련하여 투명성 의무를 준수하지 않을 경우 최대 1,500만 유로 또는 직전 회계 연도 전세계 연간 매출액의 3% 중 더 높은 금액이 벌금으로 부과될 수 있음 (제99조 제4항)

□ 디지털서비스법

- EU의 「디지털서비스법(Digital Service Act)」에도 딥페이크 관련 내용이 명시되어 있음[21]
- (적용 범위) ´23년 8월부터 선제적으로 19개 기업만을 대상으로 시행되었으며[22], ´24년 2월부터 규모와 상관없이 EU의 디지털 서비스 사업자*에게 적용

  ※ EU에서 4,500만 명 이상의 사용자를 보유한 온라인 서비스 제공자로, 위원회가 플랫폼을 초대형 플랫폼 사업자(VLOP) 혹은 초대형 온라인 검색 엔진(VLOSE)으로 지정하면 4개월 동안 DSA를 준수해야 함[23]

- (주요 내용) 초대형 온라인 플랫폼(VLOP)과 초대형 온라인 검색엔진(VLOSE)에 딥페이크를 사용된 영상이 업로드 된 경우 이를 표시해야 함

21 법제처 법제조정법제관실(2024), p.29.

22 대상 기업은 총 19개로, 월평균 이용자 4,500만 명 이상 보유
- 초대형 플랫폼 사업자(very large online platform) 17개: Alibaba AliExpress, Amazon Store, Apple AppStore, Booking.com, Facebook, Google Play, Google Maps, Google Shopping, Instagram, LinkedIn, Pinterest, Snapchat, TikTok, Twitter, Wikipedia, YouTube, Zalando
- 초대형 온라인 검색 엔진(Very Large Online Search Engines) 2개: Bing, Google Search

23 European Commision.

- 초대형 온라인 플랫폼(VLOP)과 초대형 온라인 검색엔진(VLOSE)의 서비스 제공자는 실존 인물·물체·장소 등 진실인 것처럼 보이도록 생성되었거나 조작된 이미지·오디오 또는 비디오로 구성된 정보 항목이 온라인 플랫폼에 노출된 경우 눈에 잘 띄는 표시로 구별할 수 있어야 하고, 서비스 이용자도 쉽게 표시할 수 있는 기능을 제공해야 함 (제35조 제1항)

● (벌금) 디지털서비스법을 준수하지 않을 경우 전 세계 연간 매출액의 최대 6%에 해당하는 벌금이 부과될 수 있음(제52조 제3항)

미국

● 미국은 현재 연방 차원에서 딥페이크를 직접적으로 규제하는 법은 없지만, 매년 많은 법안들이 의회에 제출되고 있음

- No AI FRAUD Act(No Artificial Intelligence Fake Replicas And Unauthorized Duplications Act of 2024): 개인의 초상(Likeness)과 음성(Voice)에 대한 권리를 보호하고, 무단 복제 및 유포 시 소송 권한을 부여[24]
- Deepfakes Accountability Act(Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2023) : 딥페이크(advanced technological false personation record)로 제작된 경우 콘텐츠가 변경되었음을 식별할 수 있도록 공개해야 함
- DEFIANCE Act(Disrupt Explicit Forged Images And Non-Consensual Edits Act of 2024): 당사자 동의 없이 '디지털 위조'의 대상이 된 개인에게 민사 소송 권한을 부여[25]

● 주 정부 차원에서는 선거 혹은 음란물과 관련한 딥페이크 규제가 존재, '24년 2월 기준 미국에서 40개 이상의 주에서 총 407건의 AI 관련 법안이 발의됨[26]

- (캘리포니아) 불법 딥페이크 방지 법안을 통과시킨 최초의 주 중 하나로, '19년 10월 선거법 개정(A.B. 730 법안)을 통해 60일 이내 후보자에 대해 기만적인 음성 혹은 영상을 배포하여 후보자의 명예를 의도적으로 훼손하거나 유권자를 속여 상대 후보자에게 찬성 또는 반대투표를 하도록 유도하는 것을 금지함[27]

24 Resemble AI.

25 Ocasio Cortez (2024.3.7.); US Senate Committee on the Judiciary (2024.6.12.).

26 Ryan Heath (2024.2.14.).

27 California State Legislature. (2019). Assembly Bill No. 730.

- (텍사스) ´19년 9월, 선거법 개정을 통해 후보자의 평판을 훼손하거나 선거 결과에 영향을 미치려는 의도로 고의로 선거 30일 이내에 딥페이크 영상 게시 및 배포를 금지함[28]
- (뉴욕) ´23년 9월, 형법을 개정하여 당사자의 동의 없이 딥페이크 포함하여 디지털로 제작된 성적인 이미지를 의도적으로 유포하거나 게시하는 행위를 금지함[29]
- (사우스디코타) ´24년 2월, AI가 생성한 아동 성적 학대 이미지를 제작, 배포 또는 소지한 혐의로 유죄 판결을 받은 사람에 대해 징역형을 선고하도록 관련 법률을 개정함[30]

● 한편, ´23년 10월, 바이든 행정부에서 안전한 AI 개발 및 이용에 관한 행정명령(Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)을 발표하는 등,[31] 향후 딥페이크와 관련 규제도 강화될 것으로 예측

- 안전한 AI 개발 및 이용에 관하여 8가지 원칙과 우선순위에 따라 AI를 개발 및 관리할 예정

안전한 AI 개발 및 이용에 관한 행정명령의 8가지 원칙

1. 인공지능은 안전하고 보안이 유지되어야 함
2. 책임 있는 혁신, 경쟁, 협력을 촉진해야 함
3. 인공지능의 책임 있는 개발과 사용에 있어 미국 노동자를 지원하겠다는 약속을 해야 함
4. 인공지능 정책은 형평성과 시민권 증진을 위해 헌신하는 행정부의 정책과 일치해야 함
5. 일상생활에서 인공지능 및 인공지능 기반 제품을 사용하거나 상호작용 혹은 구입하는 미국인의 이익이 보호되어야 함
6. 인공지능이 계속 발전함에 따라 미국인의 프라이버시와 시민의 자유(civil liberties)가 보호되어야 함
7. 연방 정부의 인공지능 사용에서 발생하는 위험을 관리하고, 인공지능의 책임 있는 사용을 규제, 관리 및 지원하는 내부 역량을 강화하여 미국인들에게 더 나은 결과를 제공하는 것이 중요함
8. 연방 정부는 미국이 이전의 파괴적 혁신과 변화의 시대에 그랬던 것처럼 글로벌 사회, 경제, 기술 진보를 선도해야 함

28 Kenneth Artz (2019.10.22.).
29 One Trust Data Guidance (2023.10.30.).
30 Pymnts (2024.2.14.).
31 The White House (2023.10.30.).

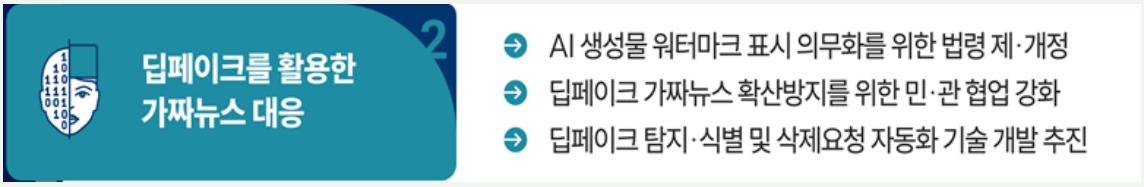
05 국내 딥페이크 정책 현황

관련 정책 현황

- 각 부처에서 딥페이크를 포함한 생성형 AI 관련 다양한 정책을 추진하고 있음
- (과학기술정보통신부) ´20년 「인공지능(AI) 윤리기준」 마련하였으며, 딥페이크 탐지 등 공공분야 디지털 한계극복 기술 개발 착수[32]하는 등 다양한 노력을 기울이고 있음
  - ´24년 5월 과기부는 관계부처 합동으로 '새로운 디지털 질서 정립 추진계획' 마련[33]
  - 해당 추진계획에 디지털 심화 쟁점을 해결하기 위해 20대 정책과제가 포함되어 있으며, 이 중 8대 핵심 과제에 "딥페이크를 활용한 가짜뉴스 대응"이 포함되어 있음

※ 지난해 발표한 「디지털 권리장전」을 구체적인 정책으로 구현하기 위한 범부처 계획으로, 디지털 심화시대의 새로운 질서를 정립하고 신속히 쟁점을 해결하기 위해 마련

| 그림 2 | 8대 핵심 과제 중 "딥페이크를 활용한 가짜뉴스 대응"



자료: 과학기술정보통신부 보도자료(24.5.21) 발췌

- (방송통신위원회) ´24년 3월, 대통령 업무보고를 통해 「인공지능서비스 이용자 보호에 관한 법률」 제정 추진을 발표
  - AI 서비스의 신뢰성을 보장하고 역기능으로부터 이용자를 보호하기 위하여 AI 생성한 콘텐츠를 게시할 경우 AI 생성물 표시 의무화[34]

32 연합뉴스 (2024.5.26.)
33 과학기술정보통신부 보도자료 (2024.5.21.).
34 연합뉴스 (2024.3.21.)

- (문화체육관광부) ´23.12월, '저작권 강국 실현, 4대 전략'을 발표하였으며, 인공지능(AI) 기술 상용화로 인한 시장 혼란을 최소화하기 위해 "인공지능(AI)-저작권 안내서"를 발표함[35]
  - 이밖에도 ´24.2월 한국저작권위원회와 함께 '2024 인공지능-저작권 제도개선 워킹그룹'을 발족하고, 11월까지 학계·법조계·권리자·사업자·산업기술계 의견 청취, 쟁점별 분과 회의로 분야별 전문가와 함께 합리적 대안 모색 예정[36]
- (경찰청) ´24년 3월, 경찰청에서도 딥페이크 관련 범죄가 증가함에 따라 딥페이크 의심 영상의 진위를 판별하기 위하여 신규 소프트웨어 개발[37]
  - 해당 소프트웨어는 페이스 스왑(Face Swap) 등 딥페이크 영상을 5분~10분 내 분석하여 진위를 판단할 수 있으며, 결과보고서를 즉각 창출하여 수사에 곧바로 활용할 수 있음

관련 법제도 현황

- 현재, 우리나라는 EU처럼 AI를 전반적으로 포괄하는 통합법은 없으며, 딥페이크를 악용한 범죄행위 등에 대해서는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」이나 「형법」이 적용될 수 있음
- 또한, 음란물·선거와 관련된 딥페이크 피해 증가에 대해서는 관련 법령의 개정을 통해 대응
- (딥페이크 음란물) ´20년 3월, 딥페이크로 인한 피해가 증가함에 따라 별도의 처벌 규정을 마련할 필요성이 증가하여 「성폭력범죄의처벌 등에 관한 특례법」을 개정
  - 반포할 목적으로 사람의 신체 등을 촬영하고, 대상자의 의사에 반하여 성적 욕망 또는 수치심을 유발할 수 있는 형태로 편집·합성 또는 가공하거나 이를 반포 혹은 판매할 경우 처벌받음 (제14조의2)
  - 해당 범죄는 미수범 규정의 적용을 받기 때문에 결과가 발생하지 않더라도 처벌받을 수 있음 (제15조)
  - 다만, "개인 소장 목적"과 "딥페이크 음란 행위" 자체는 처벌 대상이 아닌 점, 그리고 "반포 목적"이 없을 경우 제3자가 제작한 음란물을 유포한 행위에 대해서는 법적 처벌 공백이 존재함[38]

35 문화체육관광부 보도자료 (2023.12.27.).
36 문화체육관광부 보도자료 (2024.2.19.).
37 경찰청 보도자료 (2024.3.5.).
38 법무법우 화우 (2024), p.2.

● (선거) ´23년 12월, 중앙선거관리위원회는 제22대 국회의원선거 당시 딥페이크 영상을 활용한 선거운동을 금지하고자 「공직선거법」을 개정함[39]

- 선거일 90일 전부터 선거일까지 선거운동을 위하여 인공지능 기술 등을 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등을 제작·편집·유포·상영 또는 게시하는 행위 금지 (제82조의8 제1항)
- 선거 기간이 아닌 경우 선거운동을 위하여 딥페이크 영상 등을 제작·편집·유포·상영또는 게시하는 경우 가상의 정보라는 사실을 딥페이크 영상 등에 표시해야 함 (제82조의8 제2항)

관련 입법 추진 동향

● (입법 추진 동향) 제22대 (2024~2028) 의안 중 의안명에 “인공지능”이 포함된 법안은 총 4건이며, 이 중 원문이 공개된 법안은 2건임[40]

의안명	제안자	제안일자	심사진행 상태	목적
인공지능산업 육성 및 신뢰 확보에 관한 법률안	김성원의원 등 11인	2024-06-19	접수	의안 원문 없음
인공지능산업 육성 및 신뢰 확보에 관한 법률안	조인철의원 등 19인	2024-06-19	접수	의안 원문 없음
인공지능 발전과 신뢰 기반 조성 등에 관한 법률안	정점식의원 등 108인	2024-06-17	소관위접수	인공지능의 발전 지원 및 신뢰 기반 조성하여 국민의 권익과 존엄성을 보호하며, 삶의 질 향상과 국가경쟁력 강화
인공지능 산업 육성 및 신뢰 확보에 관한 법률안	안철수의원 등 12인	2024-05-31	소관위접수	안전하고 신뢰할 수 있는 인공지능 개발·이용 기반 조성, 인공지능 관련 정책 수립·추진에 필요한 사항 규정하여 국민의 권익과 삶의 질 보호

● 「인공지능 발전과 신뢰 기반 조성 등에 관한 법률안」 및 「인공지능 산업 육성 및 신뢰 확보에 관한 법률안」 공통적으로 생성형 인공지능에 대하여 표시 의무를 부과하고 있음

- (공통) 생성형 인공지능을 이용하여 제품 또는 서비스를 제공하려는 자는 해당 제품 또는 서비스가 생성형 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하고, 해당 제품 또는 서비스의 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 함

39 중앙선거관리위원회 보도자료 (2024.1.9.).

40 의안정보시스템 2024.6.22 검색 결과.

국내 사업자 대응

- 국내 플랫폼 사업자들도 딥페이크 악용을 막기 위하여 다양한 노력을 기울이고 있음
- ´24년 3월, 사단법인 한국인터넷기업협회 회원사인 네이버, 카카오, SK커뮤니케이션즈는 22대 국회의원 선거를 앞두고 선거 과정에서의 공정성과 신뢰성을 위하여 '악의적 선거 딥페이크 사용 방지를 위한 공동선언'을 채택함[41]
  - 협회는 악의적 선거 딥페이크가 기술적 문제에 국한되지 않고 정치적·사회적·윤리적 문제임을 인식하고 있기에 사회 전반적으로 대응하고자 다음과 같이 선언함

악의적 선거 딥페이크 사용 방지를 위한 공동선언

1. 우리는 악의적 선거 딥페이크의 위험을 완화하기 위한 탐지와 신속한 조치에 노력한다.
2. 우리는 악의적 선거 딥페이크 대응 정책 공개 등을 통해 대응 투명성을 높인다.
3. 우리는 악의적 선거 딥페이크 확산 방지를 위해 적극적으로 논의하며, 정보와 의견 교류를 활성화한다.
4. 우리는 악의적 선거 딥페이크에 대한 대중의 인식을 높이기 위해 노력한다.
5. 우리는 다양한 시민단체, 학계 등 외부 전문가와 지속적으로 교류하며 논의한다.
6. 각 기업의 서비스 특성에 따라 추가적 조치 방안을 모색한다.

  - 추가적으로, 네이버는 모니터링을 강화하고, 검색결과 상단에 딥페이크로 인한 문제와 주의를 환기하는 안내 문구를 노출하는 등의 조치를 취하였음. 또한, 선거 관련 허위 정보 신고를 위한 채널도 운영함[42]
  - 카카오는 이미지 생성형 모델인 '칼로(Karlo)'에 비가시성 워터마크를 도입하였으며, AI를 이용해 생성한 기사에도 사용자가 쉽게 인지할 수 있도록 상단에 해당 사실 표기함[43]
  - SK커뮤니케이션즈는 AI를 이용한 이미지를 생성할 경우 주요 정치인 이름 검색어 제한 조치로 딥페이크를 방지하고자 노력[44]

41 한국인터넷기업협회 보도자료 (2024.3.8.).
42 네이버 공식 블로그 (2024.2.28.).
43 카카오 그룹 공식 홈페이지 참고자료 (2024.3.13.).
44 AI타임스 (2024.3.20.).

● 이밖에도 ´24.5월 삼성전자, SK텔레콤, KT, LG AI 연구원, 네이버, 카카오 등 국내 기업과 어도비(Adobe), 오픈AI, 구글(Google), 마이크로소프트(Microsoft), 앤스로픽(Anthropic), IBM, 세일즈포스(Salesforce), 코히어(Cohere) 등 해외 기업 총 14사에서 주요 인공지능(AI) 관련 기업들은 안전한 AI 사용을 위한 '서울 기업 서약(Seoul AI Business Pledge)'을 발표함[45]

〈서울 기업 서약〉 주요 내용

- 책임감 있는 AI 개발 및 사용 보장
- AI의 지속 가능한 발전과 혁신 추구
- 모두를 위한 AI의 공평한 혜택 보장

자료: AI 서울 정상회의 공식 홈페이지.

- 해당 서약에 의하면, 딥페이크와 관련하여 ① 이용자가 AI에 의해 생성된 콘텐츠를 식별할 수 있도록 워터마킹과 같은 적절한 방법론을 개발하기 위해 노력할 것이며, ② 민주주의를 위협하는 허위 정보 및 잘못된 정보에 대한 AI 생성 콘텐츠의 사용되는 것을 완화하기 위한 조치를 추구할 예정

06 소결 및 시사점

● 기술이 발전함에 따라 딥페이크 악용 사례가 증가하여 심각한 피해를 야기하고 있음
- 이러한 현실 속에서 정부와 플랫폼 사업자들은 딥페이크의 피해를 최소화하기 위하여 각기 다양한 노력을 기울이고 있음

● 구글의 경우 여러 가지 방법으로 딥페이크 악용 문제를 막기 위하여 노력
- 딥페이크를 감지할 수 있는 기술 개발, C2PA의 가입, 딥페이크를 활용하여 음란물 등을 변경/생성하는 서비스의 광고 금지 등

45 한국경제신문 (2024.5.22.); AI 서울 정상회의 공식 홈페이지.

- 유튜브도 콘텐츠의 투명성을 강화하고자 크리에이터가 콘텐츠를 업로드할 때 AI 사용 여부를 공개해야 한다는 자사 가이드라인을 발표함
  - 콘텐츠의 투명성을 높이기 위한 유튜브의 노력은 긍정적이지만, 애니메이션이 AI 라벨 대상에서 제외됨
  - 어린이는 성인보다 인지능력이 떨어지기 때문에 딥페이크 콘텐츠가 조작되거나 합성된 것을 인식하지 못하고, 허위정보를 무분별하게 수용할 수 있음
  - 유튜브의 AI 라벨 정책에 애니메이션도 포함되어야 딥페이크 애니메이션를 필터링할 수 있어 어린이를 불법 딥페이크로부터 최소한 보호할 수 있을 것
- 각 국가에서 추진/시행 중인 딥페이크 관련 법률을 살펴보면 공통으로 딥페이크 콘텐츠에 'AI 생성물' 표기를 요구하고 있으나, 이러한 조치가 얼마나 실효적인지 의문
  - 딥페이크 영상에 AI로 생성하였음을 표기하면 투명성이 강화되어 일차적으로 허위정보의 확산을 막을 수 있으며, 제작자와 유포자에게 해당 콘텐츠에 대한 책임을 물을 수 있음
  - 하지만 딥페이크 제작/유포자가 'AI 생성물' 표기를 제거하거나 딥페이크를 활용하지 않았다고 거짓으로 고지할 가능성 존재
  - 또한, 이용자가 해당 표기를 인식하지 못하거나 이를 무시하고 딥페이크 콘텐츠를 그대로 수용할 수 있음
  - 딥페이크 기술이 빠른 속도로 정교해지고 있는 현실에서 플랫폼에서 불법 딥페이크 콘텐츠를 바로 감지하고 처리하는 것이 점점 더 어려워질 수 있음
  - 'AI 생성물' 표기는 딥페이크의 피해를 줄이기 위한 최소한의 노력일 뿐, 이것만으로 딥페이크 악용 문제를 해결할 수 없음
- 따라서 불법 딥페이크 피해를 방지하기 위하여 'AI 생성물' 표기뿐만 아니라, 딥페이크 탐지 기술 개발, 이용자의 디지털 리터러시 교육도 같이 병행되어야 함
  - 이용자가 딥페이크 콘테츠를 판별할 수 있는 능력과 딥페이크의 불법적인 사용은 범죄라는 인식이 반드시 교육되어야 함
  - 이밖에도 제도적으로 불법 딥페이크를 규제하기 위하여 제작자 및 유포자에 대한 강력한 처벌과 피해자를 보호할 수 있는 보상제도가 마련되어야 함

참고문헌

경찰청 보도자료 (2024.3.5.), “경찰청, 「딥페이크 탐지 소프트웨어」 개발”,

과학기술정보통신부 보도자료 (2024.5.21), “대한민국이 새로운 디지털 질서 정립의 마스터플랜을 공개합니다”.

구글코리아 블로그 (2023.4.7), “구글 AI 원칙”.

네이버 공식 블로그 (2024.2.28), “네이버는 생성형 AI, 딥페이크 등으로 인한 피해 방지를 위해 노력하고, 허위 정보 확산을 막기 위해 앞장섭니다.”

문화체육관광부 보도자료 (2023.12.27.), “인공지능(AI)-저작권 안내서 발표로 시장의 불확실성 해소하고, 안무·건축 등 ‘저작권 사각지대’ 없앤다“

문화체육관광부 보도자료 (2024.2.19.), “인공지능과 저작권 쟁점별 구체적 정책 방안 마련한다”.

방송통신심의위원회 보도자료 (2024.5.2), “딥페이크 악용·유명 연예인 합성 ‘성적 허위영상물’ 400% 폭증“.

법무법인 화우(2024), “AI가 생성하는 맞춤형 음란물 시대의 도래와 시사점”, Legal Update.

법제처 법제조정법제관실 (2024), “딥페이크 관련 해외 입법동향”, 2024년 3월호.

서울경제 (2024.5.18), ““나 사장인데…돈 좀 보내”…딥페이크에 글로벌 기업도 당해”.

연합뉴스 (2024.3.21.), “방통위, ‘인공지능서비스 이용자 보호법’ 제정 추진(종합)”.

연합뉴스 (2024.5.26.), “딥페이크 탐지 등 공공분야 디지털 한계극복 기술 개발 착수”.

의안정보시스템, https://likms.assembly.go.kr/bill/main.do.

정재욱 (2024), “세계 최초로 통과된 EU 「AI법」, 우리 기업의 대응 방향은?”, 나라경제 2024.6월호, KDI 경제정보센터.

중앙선거관리위원회 보도자료 (2024.1.9), “국선 D-90, 딥페이크 영상 등을 이용한 선거운동 제한된다”.

카카오 그룹 공식 홈페이지 참고자료 (2024.3.13.), “딥페이크 허위 조작 정보 근절을 위한 카카오의 노력”.

한겨레 (2019.10.24), “‘딥페이크’ 활용해 영화도 자동더빙”.

한국경제신문 (2024.5.22.); AI 서울 정상회의 공식 홈페이지.

한국인터넷기업협회 보도자료 (2024.3.8.), “악의적인 선거 딥페이크 사용 방지를 위한 자율협의체 공동선언 채택”

AI 서울 정상회의 공식 홈페이지, https://aiseoulsummit.kr.

AI타임스 (2024.3.20.), “SK컴즈, 네이트 총선 페이지에 딥페이크 방지 위해 검색어 제한”.

Ameena Qobrtay (2022.3.3.), “’Star Wars’ fans love CGI Luke Skywalker, but deepfake implications are dangerous”.

California State Legislature. (2019). Assembly Bill No. 730.

EU Artificial Intelligence Act, https://artificialintelligenceact.eu.

European Commision, “DSA: Very large online platforms and search engines”.

European Parliamentary Research Service (2021), “Tackling deepfakes in European policy”.

Gaby Del Valle (2024.5.4), “Google bans advertisers from promoting deepfake porn services”, THE VERGE.

Google Advertising Polices Help (2024.5.1.), “Update to Inappropriate Content Policy (May 2024)”

K.C. Halm, Ambika Kumar, Jonathan Segal, and Caesar Kalinowski IV(2019.10.14), “Two New California Laws Tackle Deepfake Videos in Politics and Porn”.

Kenneth Artz (2019.10.22), “Texas Outlaws ‘Deepfakes’—but the Legal System May Not Be Able to Stop Them”.

MACK DEGEURIN (2024.2.8), “Google wants to fight deepfakes with a special badge”, Popular Science

Medium (2023.12.1), “Demystifying Deepfakes: Google’s AI Weapon in the Fight Against Misinformation”.

MIKE KAPUT (2024.2.20), “AI Rivals Band Together to Fight Deepfakes”.

Ocasio Cortez (2024.3.7), “Rep. Ocasio-Cortez Leads Bipartisan, Bicameral Introduction of DEFIANCE Act to Combat Use of Non-Consensual, Sexually-Explicit “Deepfake” Media”.

Pymnts (2024.2.14), “State-Level AI Legislation Balloons Amid Rise of Deepfakes”.

Recording Industry Association of America (2023.10.6.), “RIAA Submission to Comment Request for the 2023 Review of Notorious Markets for Counterfeiting and Piracy”.

Resemble AI, “What is the No AI FRAUD Act?”.

Ryan Heath (2024.2.14), “Exclusive: States are introducing 50 AI-related bills per week”.

Tambiama Madiega (2024), “Artificial Intelligence Act.” European Parliamentary Research Service.

The White House (2023.7.21), “FACT SHEET: Biden- Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI”.

The White House (2023.10.30), “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence”.

US Senate Committee on the Judiciary (2024.6.12), “Senate Republican Blocks Durbin’s Attempt to Tackle Nonconsensual, Sexually-Explicit Deepfakes”.

World Economic Forum (2023.5.19), “How can we combat the worrying rise in the use of deepfakes in cybercrime?”

YouTube Official Blog (2023.11.14), “Our approach to responsible AI innovation”.

YouTube Official Blog (2024.3.18), “How we’re helping creators disclose altered or synthetic content”.

YouTube Official Blog (2024.5.18), “How we’re helping creators disclose altered or synthetic content”.

KISDI Perspectives 발간 내역

KISDI PERSPECTIVES는 국내 외 정보통신방송 관련 주요 정책 및 시장 동향을 분석한 리포트입니다.

문의 : 노희윤 전문연구원 (정보통신정책연구원 방송미디어연구본부, hyoooon@kisdi.re.kr, 043-531-4042)