



해외 생성형 인공지능 관련 주요 규제 동향 및 시사점

AI법제도센터 채은선 수석 연구원

■ 개요

- 생성형 인공지능을 이용한 딥페이크 확산에 따라 공직선거에서의 부정확한 사용, 가짜뉴스, 허위 광고 등에 따른 유권자 기만 등 공정 선거 저해, 소비자 손해 등 피해 증가
 - 국내외 관련 규제 움직임과 딥페이크 영상 제작 서비스를 제공하는 기업의 자발적 규제도 등장하여 그 동향 및 시사점 분석

■ 국가별 규제 동향

① 미국

- 바이든-해리스(Biden-Harris) 행정부는 행정명령을 통해 전 분야에서 인공지능의 안전 및 보안을 위해 각 부처 등이 이행해야 할 사항을 특정 기간 내에 이행하도록 하고, 인공지능 선두기업이 인공지능의 '안전(safe), 보안(secure) 및 신뢰(trust)'의 자발적 이행 약속을 보증
- **(바이든-해리스 행정명령)** 인공지능 안전 및 보안을 위한 기준(표준) 수립, 미국인의 프라이버시 보호, 형평성 및 시민권 제고, 소비자·근로자 보호, 혁신 및 경쟁 촉진, 국제 사회에서 미국의 리더십 향상 등을 위한 행정명령 발표('23.10.30)¹⁾
 - 딥페이크를 포함한 인공지능을 이용한 사기 등으로부터 국민을 보호하기 위해 상무부에서 인공지능 생성 콘텐츠를 탐지하고 정부 공식 콘텐츠(official content)를 인증하는 모범관행 및 기준이 포함된 지침 마련(행정명령 Sec. 4.5, '안전 및 보안을 위한 새로운 기준(표준) 수립')
- **(인공지능 선두 기업의 자발적 약속)** Amazon, Anthropic, Google, Inflection, Meta, Microsoft, OpenAI 등 선도적 기업 7개 사(社)는 3가지 원칙인 '안전(safe), 보안(secure) 및 신뢰(trust)' 이행을 약속('23.7.21.)²⁾

1) <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>; (보도자료)<https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>

- 이후 Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, Stability 사(社)가 서명하여 '23.9. 기준 총 15개 사(社)가 본 '자발적 약속'에 서명3)
- 특히 '신뢰' 확보를 위해 콘텐츠가 언제 생성되었는지 사용자가 알 수 있도록 워터마크 시스템 등 강력한 기술체계를 개발할 것을 약속하여, 창작 활동에 인공지능을 사용할 수 있게 하면서도 사기 위험을 감소시키는 데 도움을 줄 것으로 기대

〈인공지능이 생성한 콘텐츠 관련 기준 수립의 주요 내용〉

- ▶ 상무부장은 인공지능으로 생성된 콘텐츠임을 파악하고 그러한 콘텐츠임을 명확히 표시할(label) 수 있도록 다음에 관한 기준 표준도구·관행 및 과학적으로 가능한 표준·기술을 파악한 보고서를 예산관리국(Office of Management and Budget, OMB) 및 대통령의 국가안보 보좌관에게 제출(행정명령을 발한 날로부터 240일 이내)
 - (i) 콘텐츠 인증 및 출처 추적
 - (ii) 워터마크 이용 등 합성 콘텐츠* 표시
 - * 합성 콘텐츠(synthetic content) : 인공지능을 포함하여 알고리즘으로 상당한 부분이 생성되거나 수정된 생성된 이미지, 영상, 오디오 및 텍스트와 같은 정보
 - (iii) 합성 콘텐츠 탐지
 - (iv) 생성형 인공지능이 아동성착취물을 생성하거나 실제 인물의 동의없이 해당 인물과 유사한 이미지(식별가능한 사람의 신체를 디지털로 묘사한 것 포함)를 생성하는 것 방지
 - (v) 상기 목적으로 이용되는 소프트웨어 테스트
 - (vi) 합성 콘텐츠 감사 및 관리(auditing and maintaining)
- ▶ 상무부장은 OMB 국장과 협력하여 '디지털 콘텐츠 인증 및 합성 콘텐츠 탐지를 위한 도구 및 관행에 관한 지침'을 개발(보고서 제출 후 180일 이내)
- ▶ 국방부장관, 법무장관, 상무장관(NIST가 수행), 국토안보부 장관, 국가정보국장 및 OMB 국장이 적절하다고 판단하는 기관의 장은 해당 기관이 생산하거나 게시하는 콘텐츠 표시(label) 및 인증(authentication) 지침 마련(지침 개발 후 180일 이내)

② 중국

- '심층합성기술'(深層合性技術, 딥페이크)*은 「인터넷 정보 서비스의 심층합성 관리 규정」을 마련하고, '생성형 인공지능 서비스'은 별도의 규정인 「생성형 인공지능 서비스 규정」을 통해 규제
 - * 심층합성기술 : 심층학습(deep learning), 가상현실 등 합성 알고리즘을 사용, 텍스트, 이미지, 영상, 가상 인물·장면 등 네트워크 정보를 생성하는 기술

2) <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>

3) <https://fedscoop.com/eight-more-tech-companies-join-white-house-commitment-to-manage-ai-risks/>

- **(딥페이크 규제)** 「인터넷 정보 서비스의 심층합성 관리 규정」은 딥페이크 기술 이용은 허용하되, 허위정보 규제 뿐 아니라 딥페이크 서비스 제공자 및 플랫폼 인터넷 서비스 기술을 지원하는 자 등이 준수해야 할 사항을 규정(‘22.11.25. 공포, ‘23.1.10. 시행)⁴⁾
 - ‘딥페이크 서비스 공급자’는 5가지 유형*의 딥페이크 서비스 제공 시 딥페이크된 콘텐츠에 합리적 방식으로 표시하고 로그 정보 보관
 - * (5가지 유형) ①지능형 대화 지능형 글쓰기 등 자연인을 모방하여 텍스트 생성 또는 편집하는 서비스, ②음성 합성, 음성 모방 등 음성 생성 또는 개인 음성 특성을 현저하게 변경하는 서비스, ③얼굴 생성 얼굴 교체 얼굴 조작 자세 조작 등 인물 이미지 변경 또는 동영상의 생성 등 개인 영상 특성을 현저하게 변화시키는 편집 서비스, ④사실처럼 보이는 영상 등의 생성 또는 편집 서비스, ⑤콘텐츠를 생성하거나 현저하게 변경하는 기능을 가진 기타 서비스
 - 그 외의 딥페이크 서비스를 제공하는 경우 해당 서비스 이용자에게 딥페이크 표시 기능을 제공하고 이를 이용자에게 고지

- **(생성형 인공지능)** 「생성형 인공지능 규정」은 중국 내에서 생성형 인공지능 기술을 이용하여 텍스트, 이미지, 오디오, 영상 및 기타 콘텐츠 생성 서비스 제공을 규제(‘23.7.13.공포, ‘23.8.15.시행)⁵⁾
 - 서비스 유형별로 생성형 인공지능 서비스의 투명성 증진 및 생성된 콘텐츠의 정확성 및 신뢰성 향상을 위한 효과적 조치 수행
 - 생성형 인공지능 서비스 공급자들은 「인터넷 정보 서비스의 심층합성 관리 규정」에 따라 생성된 그림, 동영상 및 기타 콘텐츠에 표시
 - 생성형 인공지능 서비스 공급자는 이용자가 불법적인 활동에 자사의 서비스를 이용한다는 사실을 알게 된 경우, 해당 이용자에 대해 경고 및 해당 서비스 제한·정지·종료 조치

■ 국제적 규제 동향

① G7, 11개 지도원칙 및 그 이행을 위한 행동강령

- 고도의 인공지능(생성형 인공지능) 개발 기관이 준수해야 하는 11개 지도 원칙 중 다음의 두 가지 원칙에 근거하여 생성형 인공지능 생성물에 대한 표시의무 등 이행을 권고

4) https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm

5) http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

- (신뢰할 수 있는 콘텐츠 진본성 및 출처 확인 체계 개발 및 배치) 적절하고 기술적으로 가능한 경우 워터마크 또는 인공지능 생성 콘텐츠임을 알 수 있도록 하는 조치* 등 인공지능 생성 콘텐츠의 진본성 및 출처** 확인 체계를 개발 이행하고 관련 연구 협력 및 투자

* 애플리케이션 프로그래밍 인터페이스(Application Programming Interfaces, APIs) 등

** 콘텐츠를 생성한 서비스 또는 모델의 식별자는 포함하되 이용자 정보 불포함

- 또한 라벨링(labeling) 또는 면책조항(disclaimers) 도입 등 고려

- (국제기술표준 개발 추진 및 채택) 콘텐츠 진본 및 출처 입증 체계, 사이버보안 정책, 결과 공개 등을 개발하는 경우 워터마킹 등 국제기술표준 및 모범사례의 개발 이행에 기여하고, 표준개발 기구(Standards Development Organizations, SDOs)와 협력

- 특히 이용자들이 인공지능 생성 콘텐츠 여부를 구분할 수 있도록 상호운용 가능한 국제기술, 표준 및 체계 개발에 참여

② EU, 인공지능법안(‘24.3.13. 의회 채택안)

- 생성형 인공지능 결과물과 딥페이크를 구분하여 규제, 생성형 인공지능 결과물은 ‘생성형 인공지능 시스템 공급자’가, ‘딥페이크’는 ‘딥페이크 콘텐츠 생성·조작 인공지능 시스템’의 배포자*가 관련 의무 이행

* (개념) ①(딥페이크) 사람에게 거짓으로 진짜 또는 진실된 것으로 보이도록 인공지능이 ‘기존 사람·사물·장소, 기타 실제 및 장소를 생성·조작한 이미지·음성·영상 콘텐츠, ②(배포자, Deployer) 자신의 권한 하에 인공지능을 이용하는 자

- (생성형 인공지능 결과물) “오디오, 이미지, 영상, 텍스트 등 인위적인 콘텐츠를 생성하는 ‘인공지능 시스템’(범용 인공지능 시스템 포함) 공급자”(‘생성형 인공지능 공급자’)는 해당 시스템의 결과물이 인위적으로 생성·조작된 것임을 기계가 판독가능한 형식으로 표시

- 워터마크, 메타데이터 인식, 출처·내용 진본성 입증을 위한 암호화 방법, 로그인 방법, 핑거프린트 등과 같은 기술적 조치를 통해 표시하고, 이러한 기술적 조치를 인공지능에 내장

- (딥페이크) 딥페이크 콘텐츠 생성·조작 인공지능 시스템의 배포자는 해당 콘텐츠가 인위적으로 생성 또는 조작되었음을 라벨링 및 출처 공개 등의 방식으로 공개

· (예외) 명백히 예술적, 창의적, 풍자적, 허구적 유사 저작물 또는 프로그램의 일부를 구성하는 콘텐츠, 저작물의 전시·향유를 방해하지 않는 적절한 방식으로 생성 또는 조작된 것임을 공개

- 공익 목적으로 대중에게 정보를 전달하기 위해 텍스트를 생성하거나 조작한 경우, 해당 시스템의 배포자는 그 문자가 인위적으로 생성 또는 조작된 것임을 공개

- (예외) ①법에 따른 범죄 탐지·방지·조사·기소 목적, ②인공지능 생성 콘텐츠를 사람이 검토·편집, ③자연인 또는 법인이 출판에 대한 편집 책임이 있는 경우

○ (이행 지원) ‘AI Office’는 라벨링 의무 등의 효과적 이행을 위해 EU 차원에서의 행동강령*(code of practice) 마련을 장려, 집행위원회는 본 행동강령 승인을 위한 이행규칙을 마련

* 공급자·배치자·시민사회·학계 등이 마련

○ (처벌) 의무 위반시 15,000,000 유로(약 217억원) 또는 ‘전년도 전세계 매출액 3%’ 중 높은 금액을 부과
※ 중소기업 및 스타트업의 경우, 높은 금액이 아닌 낮은 금액을 부과

■ 사업자 자율규제

○ (뮌헨협약) 독일에서 개최된 뮌헨안보회의(Munich Security Conference, MSC)

- Adobe, Amazon, Google, IBM, Meta, Microsoft, OpenAI, TikTok 등 주요 기업은 인공지능으로 인한 민주적 선거의 훼손을 방지하기 위해 ‘합리적 사전조치’(reasonable precautions)의 자발적 채택을 내용으로 하는 ‘2024년 선거에서 인공지능의 기만적 사용 방지 기술 협약’(Tech Accord to Combat Deceptive Use of AI in 2024 Elections)에 서명*(‘24.02.16.)⁶⁾

* (‘24.03.11. 기준 23개 기업 서명) Adobe, Amazon, Anthropic, Arm, IIElevenLabs, Gen, Google, IBM, Inflection, LG AI Research, LinkedIn, McAfee, Microsoft, Meta, NetApp, Nota, OpenAI, SNAPCHAT, Stability.AI, TikTok, Trendi, Truepic, X⁷⁾

- 7가지 중요 목표* 달성을 위해 ‘인공지능 생성 이미지 식별, 콘텐츠 및 그 출처·진위 인증’ 등을 포함하는 8가지 조치의** 이행 약속

* 예방(Prevention), 출처 표시(Provenance), 탐지(detection), 사고에 대한 적절하고 신속한 대응(Responsive Protection), 평가(Evaluation), 대중의 인식(Public Awareness), 회복력(Resilience)

〈**기술 협약에 따른 ‘8가지 조치’의 주요 내용〉

1. 다음과 같은 기만적 인공지능 선거 콘텐츠 관련 위험 경감을 위한 기술 개발·이행
 - a. 진짜 같은 인공지능 생성 이미지 식별과 콘텐츠 및 그 출처·진위 인증을 통해 기만적인 인공지능 선거 콘텐츠로 인한 위험을 감경할 수 있도록 기술적 혁신 지원. 다만, 워터마크 또는 서명된 메타데이터와 같은 분류체계 또는 출처인증방법에 한정되지 않음
 - b. 음향, 영상 및 이미지 관련 출처인증기술 발전에 계속적으로 투자
 - c. 인공지능이 생성한 것으로서 진짜 같은 음향, 영상 및 이미지에 기계 판독 가능한 정보를 포함(attach)

6) (원문) https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf; (MS 기사) <https://news.microsoft.com/2024/02/16/technology-industry-to-combat-deceptive-use-of-ai-in-2024-elections/>

7) <https://www.aielectionsaccord.com/commitments/>

2. 기만적인 인공지능 선거 콘텐츠와 관련된 위험을 이해하기 위해 본 협약 범위 내에 있는 모델을 평가
3. 의심 콘텐츠 신고, 탐지기술, 생성물 생산자가 콘텐츠 업로드 시 인공지능 사용했음을 공개하도록 하는 등의 방법 등을 통해 기만적 인공지능 선거 콘텐츠 유통 탐지
4. 진짜 같은 인공지능 생성물에 관한 정보 제공 및 관련 정책 채택·공개 등 표현의 자유 및 안전 원칙에 부합하는 방법으로 기만적인 인공지능 선거 콘텐츠에 적절히 대응
5. 모범관행을 공유하고 사고에 대응하기 위해 기술 신호를 공유하는 방법을 탐색하는 등 기만적인 인공지능 선거 콘텐츠에 대한 산업계의 회복력/복원력(resilience) 증진
6. 기만적 인공지능 콘텐츠에 대해 대응하는 방법을 대중에게 투명하게 공개
7. 위험에 대해 기업이 이해하는 바에 대한 정보를 제공하기 위해 글로벌 시민사회기관, 학계 및 기타 관련 전문가들과 협업 지속
8. 기만적인 인공지능 선거 콘텐츠에 대한 대중의 인식 제고 및 모든 사회의 회복력 증진 노력 지원

○ **(프론티어 모델 포럼)** Anthropic, Google, Microsoft, OpenAI 사(社)(이하, ‘회원사’)는 ‘프론티어 모델 포럼’ 창립(‘23.07.26.)⁸⁾

- 세계의 책임있는 인공지능 개발 및 배포에 기여하기 위해 출범, 모범관행 파악, 인공지능 안전 연구, 기업 및 정부 간 정보 공유 활성화 등 수행
- 포럼은 ‘안전, 보안 및 신뢰’ 약속(‘23.7.21.)의 후속조치로서 원활한 자발적 노력을 위해 창립되었으나, 아직까지 포럼 차원에서 딥페이크 또는 생성형 인공지능 결과물에 대한 구체적인 조치 방안 또는 입장을 발표한 바는 아직 없는 상황

〈‘프론티어 모델 포럼’⁹⁾〉

- ▶ **(배경 및 역할)** ‘프론티어 모델 포럼’은 인공지능이 세계에 도움이 되는 한편 위험성도 존재하므로 적절한 안전장치가 필요하다는 공감대가 형성됨에 따라, 책임있는 인공지능 개발 및 배포를 위해 관련 기관들 간 안전·책임에 관한 논의를 할 수 있도록 하는 하나의 수단으로서 역할 수행 예정
- ▶ **(운영 원칙)**
 - ① **(인공지능 안전 연구 Advancing AI safety research)** 연구를 통해 프론티어 모델을 책임감 있게 개발하고 위험을 최소화하며, 독립적이고 표준화된 성능·안전 평가 추진
 - ② **(모범 관행 파악 Identifying best practices)** 모범 관행은 프론티어 모델을 책임있게 개발·배포하는데 필수적이며, 대중이 기술의 특성·성능·한계·영향을 이해하는 데에도 도움이 됨
 - ③ **(분야간 협력 Collaborating across sectors)** 정책입안자, 학계, 시민단체 및 기업은 서로 협력하고 신뢰 및 안전 관련 위험에 대한 지식을 공유할 필요
 - ④ **(인공지능이 사회 최대의 현안에 도움이 되도록 지원 Help AI meet society’s greatest challenges)** 사회 최대의 현안*에 도움이 될 수 있는 애플리케이션 개발 노력 지원

* 기후변화 완화 및 적응, 암 발견 및 예방, 사이버 위협에의 대응 등

8) Microsoft, “Microsoft, Anthropic, Google, and OpenAI launch Frontier Model Forum”, 2023.07.26., <https://blogs.microsoft.com/on-the-issues/2023/07/26/anthropic-google-microsoft-openai-launch-frontier-model-forum/>

- ▶ **(회원)** ‘프론티어 모델’*을 개발 및 배포하는 기관이면 회원으로 가입 가능
 - * ‘프론티어 모델’ : 현재의 가장 고도화된 모델의 현재 성능을 초과하고 광범위하고 다양한 임무를 수행할 수 있는 ‘거대 기계 학습 모델’(large-scale machine-learning models)
- ▶ **(역할)** 세 가지 사항에 주안점을 두고 역할 수행
 - ① **(모범관행 파악)** 잠재적 위험 경감을 위해 안전 표준 및 관련 관행에 관하여 산업계, 정부, 시민단체, 학계 간 모범관행 및 지식공유를 장려
 - ② **(인공지능 안전 연구 추진)** 인공지능 안전 관련 중요 연구 과제를 파악함으로써 인공지능 안전 생태계를 지원하고, 특히 적대적 견고성(강건성), 기술적 해석가능성, 창발적 행동, 확장가능한 감독* 등 연구에 협력
 - * 적대적 견고성 : 모델이 해커공격 등 적대적 공격을 받았을 때 속지 않고 저항하는 능력
 - 기술적 해석가능성 : 기계학습모델 (심층신경망의)층별 작업의 이해, 설명을 위한 뉴럴 네트워크 연구
 - 창발적 행동 : 명시적 프로그래밍 없이 단순한 규칙 또는 요소 간 상호작용으로 인해 발생하는 복잡한 행동
 - 확장가능한 감독 : 사람의 인공지능 감독을 지원하는 방법과 감독에 필요한 시간·노력 감경 방법 연구
 - ③ **(기업 및 정부 간 정보 공유 원활화)** 기업, 정부 및 이해관계자간 정보를 공유하는 신뢰할 수 있고 안전한 체계를 수립, 사이버안보 등에서의 책임있는 공개에 관한 모범관행을 따름
- ▶ **(‘인공지능 안전성 기금’ 조성)** 현 회원, 재단 등의 자금 지원을 통해 천만 달러 기금 조성
 - ※ 본 기금은 외부 전문가로 구성된 자문위원회의 지원을 받아 Meridian Institute가 관리¹⁰⁾

■ 시사점

- 미 대선을 앞두고 바이든 대통령의 투표 불참을 권하는 '가짜 전화', 악담을 하는 기시다 총리 가짜 영상, 영국·슬로바키아 등에서도 정치 관련 음성 조작 사례 등이 지속적으로 발생
 - 중국은 딥페이크가 적용된 영상통화를 하여 금전을 요구하는 피싱사기가 발생, 여러 국가에서 딥페이크를 이용하여 안면인식 통한 신원인증 사기 또한 빈번
 - 생성형 인공지능 기술의 고도화 및 해당 기술을 적용한 서비스의 이용 확산에 따라 사기 등 각종 범죄에서 딥페이크 이용에 따라 범죄의 용이성이 증가
- 일본은 별도의 법제 마련보다는 기존 법령 해석 및 지침 마련 등으로 관련 문제(저작권, 개인 정보 등)에 대응하는 것이 적절하다는 입장을 유지해 왔으나,
 - 일본 자민당은 “인공지능의 진화와 구현에 관한 프로젝트팀”(AIの進化と実装に関するプロジェクトチーム)을 구성하여 인공지능법 입법을 추진 중

9) <https://www.frontiermodelforum.org/>

10) Google, “Anthropic, Google, Microsoft and OpenAI announce Executive Director of the Frontier Model Forum and over \$10 million for a new AI Safety Fund”, 2023. 10. 25., <https://blog.google/outreach-in-initiatives/public-policy/google-microsoft-anthropic-open-ai-frontier-model-forum-executive-director/>

- 최근 프로젝트팀은 '이용자에게 생성형 인공지능 사용을 고지' 내용을 포함하는 「책임있는 인공지능 추진 기본법안」(責任あるAI推進基本法(仮))을 마련('24.2.16. 공개안 기준)¹¹⁾

○ 우리나라는 딥페이크 또는 생성형 인공지능 생성물을 규제하는 일반법령은 없으나,

- 「공직선거법」 개정('23.12.28.)으로 선거 전 90일부터 선거일까지 딥페이크영상등*의 제작·편집·유포·상영·게시를 금지(제82조의8제1항)

* 인공지능 기술 등을 이용하여 만든 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등
 ※ (부정선거운동죄) 7년 이하의 징역 또는 1천만원 이상 5천만원 이하의 벌금(제255조제5항)

- 해당 기간 외에 딥페이크영상등을 제작·편집·유포·상영·게시하는 것은 허용되나, '실체가 아닌 인공지능 기술 등을 이용하여 만든 가상의 정보'임을 표시*(법 제82조의8제2항)

* 딥페이크영상등의 형태(영상, 음향, 이미지 등)별 표시는 「공직선거관리규칙」 제45조의6에 따른 별표 1의3 참조
 ※ 중앙선거관리위원회규칙으로 정하는 사항(「공직선거관리규칙」 제45조의6)을 딥페이크영상등에 표시하지 아니한 자에게는 1천만원 이하 과태료 부과(법 제261조제3항제4호)

- ① 당선되거나 되게 할 목적인 경우, '허위사실공표죄'에 해당하여 5년 이하의 징역 또는 5천만원 이하의 벌금,
 ② 당선되지 않게 할 목적인 경우, 7년 이하의 징역 또는 1천만원 이상 5천만원 이하의 벌금(법 제250조제4항)

- 이용자의 혼선 방지 및 인공지능 콘텐츠의 신뢰성·책임성 강화를 위해 인공지능 기술로 콘텐츠를 제작한 경우 이를 표시하도록 하는 내용을 포함하는 「콘텐츠산업 진흥법」 개정 추진 중(이상헌의원안, '23.5.22. 발의)

- 「정보통신망법 일부개정법률안」, 「지능정보화 기본법 일부개정법률안」, 「민법 일부개정법률안」 등 또한 딥페이크로 인한 개인의 권익 피해에 대응하여 딥페이크 규제 입법 추진 중¹²⁾

○ 생성형 인공지능 콘텐츠가 '인공지능이 생성한 것임'을 알리도록 하여 '가상의 것'이 '진실되거나 실제하는 것'으로 오인되지 않도록 하여 그로 인한 피해(저작권, 개인정보, 인격권, 초상권 침해, 금융범죄로 인한 재산상 피해 등) 예방 필요

- 인공지능을 통한 다양한 결과물을 만들어내는 것은 기술의 활용·발전 및 문화적 측면에서 허용 하되, 타인의 법익 또는 사회질서 침해의 방지를 위하여 주요 분야별 위반 행위는 규제 필요

11) https://www.aplawjapan.com/application/files/3417/0919/1282/Newsletter_Policy_Research_Institute_004.pdf; (법안 관련 자료 확인) https://note.com/masanao_ozaki/n/nbd4dd013a5cb

12) 우리 법상 딥페이크 규제 및 입법 추진 현황에 대해서는 다음 참조: 채은선, “중국 「인터넷 정보 서비스의 심층합성 관리 규정」 시행”, 디지털 법제 Brief, NIA, 2023.1.3., pp.4~5.

- 현재 입법 추진 중인 인공지능 기본법에서는 생성형 인공지능 결과물 표시 및 식별제도를 도입, 생성형 인공지능 서비스를 제공하려는 사업자가 결과물에 인공지능 결과물 표시를 할 수 있도록 하고,
 - 결과물에 서비스제공자가 기계 판독이 가능한 기술적 조치 등을 하도록 하여 제3자가 알 수 있도록 하여 책임있고 신뢰할 수 있는 인공지능 개발 및 이용 환경 조성을 위한 제도적 장치를 마련할 필요
- 인공지능 기술 발전, 문화적 활용, 다양한 콘텐츠 산업 진흥을 위하여 딥페이크 기술의 완전 금지는 지양하되, 딥페이크 기술의 건전한 이용을 위한 규제장치 마련 필요
- 인공지능 기본법제에서는 생성형 인공지능의 이용 및 결과물을 식별할 수 있도록 하는 기본적인 제도(표시 등) 방안을 마련하고,
 - 선거, 명예훼손, 저작권, 문화활동(영화제작 등), 교육 등 개별 분야에서 기술의 오남용을 방지하고 국민의 디지털 권리 보장 및 사회질서 보호를 위한 안전장치 마련

■ 참 고

■ 공직선거관리규칙 [별표 1의3] (규칙 제45조의6 관련) <신설 2024. 1. 19.>

딥페이크영상등 표시사항 및 표시방법

구 분		내 용
1. 음향	표시 사항	이 음향은 실체가 아닌 인공지능 기술 등을 이용하여 만든 가상의 정보입니다.
	표시 방법	누구든지 쉽게 인식할 수 있도록 시작과 끝부분에 음성으로 각각 표시한다. 이 경우 재생 시간이 5분을 초과하는 때에는 5분마다 1회씩 전단의 표시를 추가 하되 음향 중간에 적절히 표시하여야 한다.
2. 이미지	표시 사항	이 이미지는 실체가 아닌 인공지능 기술 등을 이용하여 만든 가상의 정보입니다.
	표시 방법	누구든지 쉽게 인식할 수 있도록 전체크기의 100분의 10 이상의 테두리 안에 배경과 구분되도록 표시한다. 이 경우 테두리 안에는 표시사항 이외에는 표시할 수 없으며, 음향을 포함한 경우 제1호의 표시를 추가하여야 한다(이하 제3호에서 같다).
3. 영상	표시 사항	이 영상은 실체가 아닌 인공지능 기술 등을 이용하여 만든 가상의 정보입니다.
	표시 방법	누구든지 쉽게 인식할 수 있도록 전체크기의 100분의 10 이상의 테두리 안에 배경과 구분되도록 상시 표시한다.
4. 기 타		해당 정보가 인공지능 기술 등을 이용하여 만든 가상의 정보라는 사실을 누구든지 쉽게 인식할 수 있도록 위에 준하는 방법으로 표시한다.

디지털 법제 Brief

기 획 : 한국지능정보사회진흥원
인공지능정책본부
AI법제도센터
작 성 : 채은선 수석 연구원
발 행 일 : 2024년 3월

1. 본 보고서의 무단전재를 금하며, 가공·인용할 때는 반드시 출처를 「한국지능정보사회진흥원 (NIA)」이라고 밝혀 주시기 바랍니다.
2. 본 보고서의 내용은 한국지능정보사회진흥원(NIA)의 공식 견해와 다를 수 있습니다.