국회입법조사처 | 2024. 10. 18. | 제2279호
이슈와 논점
Issues and Perspectives

딥페이크 식별을 위한 AI 생성물 표시 의무 입법 방안

정 준 화

딥페이크와 실제 사실을 구분하지 못해서 발생하는 위험으로부터 이용자를 보호하고 사회적 혼란을 줄이기 위해서 인공지능(AI)으로 생성되거나 조작된 정보는 별도로 표시하도록 의무화하는 방안이 필요하다. 표시 의무 대상자를 AI 모델 개발·제공자, AI 모델을 활용한 제품·서비스 운영자, AI 이용자로 나누어 적합한 의무를 부과하고, 온라인 플랫폼은 미표시 콘텐츠의 삭제 또는 임시조치 체계를 갖추며, 정부는 AI 생성물 표시에 관한 기술개발 및 국제협력을 지원하는 입법 방안을 고려해 볼 수 있다.

1 딥페이크 식별의 필요성

딥페이크(deep fake)는 딥러닝(deep learning)과 같은 인공지능(AI) 기술로 만들어 낸 가상의(fake) 정보로서, 사람들이 실제라고 생각할 수 있을 만큼 정교한 것을 말한다. 여기서 '페이크'가 주는 부정적 어감으로 인해 딥페이크 자체가 불법적으로 느껴질 수 있지만 실제로는 가치중립적인 기술이다. 사용 목적에 따라 프로필 사진 보정이나 방송·광고·교육용 영상 제작과 같이 유익하게 활용되기도 하지만, 합성음란물 제작·유포, 타인을 사칭한 사기와 같이 범죄 용도로 악용되기도 한다.

최근에는 청소년 딥페이크 합성음란물이 사회적 문제가 되면서 딥페이크의 위험성이 크게 부각되고 있다. 합성음란물 규제를 강화함으로써 발등의 불은 끌 수 있었지만, 다양한 영역에서 현실화되고 있는 딥페이크 위험에 대비하기에는 부족함이 있다.

보다 근본적인 대안이 필요한데, 기술 측면만 보자면 AI 발전으로 누구나 딥페이크 이미지·영상 콘텐츠를 만들 수 있는 상황에서, 많은 사람들이 실제와 가상을 구분하지 못해 발생하는 피해와 혼란을 막는 것이 중요하다. 대표적인 대안은 딥페이크 콘텐츠가 AI로 만든 것임을 표시하도록 하여 사람들이 표시 사항만 확인하면 쉽게 딥페이크임을 식별할 수 있도록 하는 것이다. 이를 통해 분야별 딥페이크 규제와 AI 관점의 일반 규제가 서로 맞물리는 '이중 안전장치'를 갖출 수 있다.

유럽연합과 미국에서는 AI 생성물 표시에 관한 법제를 완성하였고, 우리나라에도 다수의 법률안이 발의된 상태다. 이 글을 통해 국내외 입법례를 살펴보고 입법적 시사점을 모색하기로 한다.

2 유럽연합과 미국의 입법례

(1) 유럽연합 인공지능법

유럽연합 「인공지능법」[1] 제50조는 AI 시스템

1) 유럽연합 법률정보시스템, *Artificial Intelligence Act*(최종 검색일: 2024.9.30.), 〈https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689〉; 국회도서관, 「유럽연합 인공지능법(번

NARS www.nars.go.kr
발행처 국회입법조사처
발행인 박상철 국회입법조사처장
저 자 정준화 사회문화조사실 과학방송통신팀 입법조사관
02-6788-4715 joonhwa@assembly.go.kr
국회입법조사처 NATIONAL ASSEMBLY RESEARCH SERVICE

제공자(provider)와 배포자(deployer)[2]에게 AI 생성물임을 표시하는 의무를 부과하여 사람들이 딥페이크를 쉽게 식별하는 기반을 만들었다. 제50조 제2항에 따라 제공자는 AI 산출물임을 기계 판독 가능한(machine-readable) 형식으로 표시하고, AI로 생성 또는 조작되었다는 것을 탐지할 수 있도록(detectable) 해야 한다. 제4항은 배포자에게 해당 콘텐츠가 AI로 생성 또는 조작되었다는 것을 공개할 의무를 부과한다. 다만, 범죄 수사나 순수한 예술·표현 등에 사용되는 딥페이크에 대해서는 공개 의무가 적용되지 않는다.

유럽연합 「디지털서비스법」[3] 제35조는 초대형 온라인 플랫폼(very large online platforms, VLOP) 및 초대형 온라인 검색엔진(very large online search engines, VLOSE) 사업자가 실존 인물·물체·장소 등과 상당히 유사하고 진실인 것처럼 보이도록 생성 또는 조작된 콘텐츠를 온라인에 게시하려면 그것이 생성·조작된 것임을 눈에 잘 띄게 표시하는 의무를 부과한다. 다만, AI로 만들어진 콘텐츠만으로 대상을 한정하지는 않는다.

(2) 미국 인공지능 행정명령과 법률안

미국은 연방 차원의 AI 일반법은 존재하지 않는다. 대신, 정부·공공기관에 AI 관련 의무를 부과하기 위해 2023년 12월에 바이든 대통령이 서명한 「안전하고 보안이 보장되며 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령」[4]이 시행 중이다. 행정명령 제3조는 딥페이크를 합성콘텐츠(synthetic content)[5]로 규정하고 정부에 합성콘텐츠의 진위와 출처를 확인하는 방안을 마련하도록 했다. 이에 따라 정부는 상무부장관이 중심이 되어 콘텐츠 진위 여부를 확인하고 출처를 추적하는 방법, 합성콘텐츠임을 표시하는 방법, 합성콘텐츠를 탐지하는 방법 등 AI로 생성한 정보를 식별하는 기술을 개발하여야 한다.

연방의회에는 AI 생성물임을 표시하도록 의무화하는 법률안이 다수 발의되어 있다. 예를 들면, 딥페이크 제작자를 공개하고 딥페이크 탐지 기술을 적용하도록 규정한 「딥페이크 책임법안(DEEPFAKES Accountability Act)」, AI 개발자 및 이용자 등에게 AI로 만든 콘텐츠임을 표시하는 의무를 부과하는 「인공지능 표시법안(AI Labeling Act of 2023)」이 있다.

또한, 국가표준기술원에 생성형 AI로 만든 콘텐츠를 식별하는 기술 표준 및 지침을 개발하도록 하고, 생성형 AI 서비스 제공자에게 가상 정보를 생성한 애플리케이션의 이름, 콘텐츠를 생성하거나 수정한 날짜와 시간 등의 메타데이터를 제공하도록 의무화한 「기만적인 AI로부터의 소비자 보호법안(Protecting Consumers from Deceptive AI Act)」도 있다.

3 국내 입법 논의

(1) 인공지능법안

제22대 국회에 2024년 9월 30일 기준으로 11

역본)」, 2024.

2) "제공자"란 유상 또는 무상으로 인공지능시스템이나 범용인공지능모델을 개발하거나 인공지능시스템이나 범용인공지능모델을 개발하여 자신의 명의 또는 상표로 인공지능시스템을 시장에 출시하거나 서비스를 제공하는 자연인 또는 법인, 공공기관, 대행기관이나 그 밖의 기구를 말한다. "배포자"란 개인적인 비전문가 활동 과정에서 인공지능시스템을 사용하는 경우를 제외하고, 그 권한에 따라 인공지능시스템을 사용하는 자연인 또는 법인, 공공기관, 대행기관이나 그 밖의 기구를 말한다.

3) 유럽연합 법률정보시스템, *Digital Services Act*(최종 검색일: 2024.9.30.), 〈https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1727655566080〉; 국회도서관, 「디지털서비스법(규정 2022/2065호)」, 2022.

4) 미국 백악관 홈페이지, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, (최종 검색일: 2024.9.30.), 〈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/〉; 국회도서관, 「안전하고 신뢰할 수 있는 인공지능의 개발과 사용에 관한 행정명령」, 2023.

5) "합성콘텐츠"란 AI를 포함한 알고리즘을 활용하여 크게 수정하거나 생성하는 이미지, 비디오, 오디오 클립, 텍스트 등의 정보를 말한다.

건의 인공지능법안이 발의되어 있고, 이 중에서 5건 법안이 생성형 AI로 만든 콘텐츠는 해당 사실을 표시하도록 규정한다. 생성형 AI를 이용하여 제품 또는 서비스를 제공하려는 자는 해당 제품 또는 서비스의 결과물이 생성형 AI에 의하여 생성되었다는 사실을 표시하여야 한다.

다만, 실제 법률 집행 과정에서 무엇이 '생성형 AI'인지 판단하는 것이 쉽지 않고, 의무 불이행에 대한 처벌 조항이 없어서 이행력을 확보하기도 어려울 수 있다.

(2) 정보통신망법 개정안

제22대 국회에 발의된 「정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안」 중에도 AI 생성물임을 표시하도록 하는 법률안이 있다. 주요 내용은 AI 기술로 만든 실제와 구분하기 어려운 가상의 콘텐츠는 AI로 생성한 것임을 명확하게 인식할 수 있도록 표시하고, 표시 의무 위반 시 과태료를 부과하며, 정보통신서비스 제공자는 자신이 운영·관리하는 정보통신망에 적합한 표시를 하지 아니한 정보가 유통되는 것을 발견한 경우 지체 없이 해당 정보를 삭제하도록 한 것이다.

이 법률안은 피규제자의 다양성을 고려하지 않고 모든 사업자·이용자를 규제하며, 적용 대상을 '실제와 구분하기 어려운' 것으로 규정하여 주관적 판단이 개입할 우려가 크고, 제도 정착에 필요한 최소한의 정부 지원도 고려하지 않아서 건전한 딥페이크 생태계까지 위축시킬 우려가 있다.

[표 1] AI 생성물 표시 의무를 규정한 제22대 국회의 주요 법률안('24.9.30. 기준)

대표발의 (의안번호)	법률안 및 조항		내용	
이훈기의원 (2203960)	「인공지능 발전 진흥과 안전성 확보 등에 관한 법률안」 제27조		생성형 인공지능을 이용하여 제품 또는 서비스를 제공하려는 자는 해당 제품 또는 서비스의 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 함 (※5개 법률안 모두 표시 의무를 규정한 조항의 내용이 동일함)	
배준영의원 (2203297)	「인공지능 발전 진흥과 사회적 책임에 관한 법률안」 제27조제1항			
한민수의원 (2203072)	「인공지능 기본법안」 제28조제1항			
민형배의원 (2201158)	「인공지능기술 기본법안」 제30조			
정점식의원 (2200543)	「인공지능 발전과 신뢰 기반 조성 등에 관한 법률안」 제29조제1항			
안철수의원 (2200053)	「인공지능 산업 육성 및 신뢰 확보에 관한 법률안」 제29조			
조인철의원 (2203627)	「정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안」	제44조의11, 제76조	누구든지 인공지능 기술을 이용하여 만든 실제와 구분하기 어려운 가상의 음향·화상 또는 영상 등의 정보를 제작·편집·유포·상영 또는 게시하는 경우에는 해당 정보가 인공지능 기술 등을 이용하여 만든 가상의 정보라는 사실을 명확하게 인식할 수 있도록 전자적 표시를 하여야 하고, 위반시 1천만 원 이하 과태료를 부과함	정보통신서비스 제공자는 자신이 운영 · 관리하는 정보통신망에 적합한 전자적 표시를 하지 아니한 정보가 유통(게재 또는 전시)되는 경우에는 지체 없이 해당 정보를 삭제하여야 함
김승수의원 (2200713)		제43조의2, 제44조의2, 제76조	정보제공자는 인공지능 기술을 이용하여 만든 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 정보 중 대통령령으로 정하는 정보를 제공하려는 경우에는 해당 정보가 인공지능 기술 등을 이용하여 만든 가상의 정보라는 사실을 명확하게 인식할 수 있도록 표시하여야 하고, 위반시 1천만 원 이하 과태료를 부과함	

※ 자료: 국회 의안정보시스템 재정리

4 입법적 시사점

AI 발전 속도를 고려한다면 앞으로 딥페이크는 더욱 증가하고 정교해질 것이다. 지금처럼 피해자의 신고에 의존하거나 규제기관의 모니터링으로 차단하는 방식은 한계에 직면할 수 밖에 없다. 이미 유통된 정보가 AI로 만든 것인지 탐지하는 방법도 있지만 딥페이크 생산량이 급증하여 사후 탐지에 많은 비용이 소요되고, 완벽한 정확도를 보장하기도 어렵다. 따라서 딥페이크가 만들어지고 유통되는 단계에서 적절하게 표시될 수 있도록 의무화하는 것이 필요하다. 이를 위해 외국 입법례에서 도출한 시사점은 다음과 같다.

(1) 표시 의무 적용 범위

AI로 생성한 콘텐츠는 모두 AI로 만든 것임을 표시하도록 할 것인지, 딥페이크만 표시할 것인지 고려해야 한다. 표시 대상을 딥페이크로 정할 경우 그것이 '실제와 구분하기 어려운' 것인지 판단해야 하는데, 판단 주체에 따라 결과가 달라서 법률 집행에 혼란이 발생할 우려가 있다. 참고로, 외국 입법례는 딥페이크가 아니라 AI로 생성된 콘텐츠를 표시하도록 하여 적용 범위를 명확하게 한다.

이러한 측면을 고려하여 표시 의무 대상을 딥페이크 또는 생성형 AI로 만든 콘텐츠로 정해서 법률 해석과 적용 과정에 혼란을 주기보다는, AI로 생성한 콘텐츠로 정하는 것이 효과적일 수 있다. 대신 과도한 표시 부담을 줄이기 위해 피해 가능성이 낮은 분야는 예외로 정할 필요가 있다.

(2) 표시 의무 부과 대상

사업자의 경우, 먼저 AI 모델을 개발·판매하는 자(유럽연합 「인공지능법」의 '제공자'에 해당)는 식별가능성 확보 의무가 필요하다. 해당 모델의 산출물은 기계 판독 가능한 정보를 포함하고, AI로 생성·조작된 것임을 탐지할 수 있도록 해야 한다. 다음으로, AI 모델을 자신의 제품·서비스에 활용하는 자(유럽연합 「인공지능법」의 '배포자'에 해당)는 표시 의무가 적절하다. 자신의 제품·서비스가 생산한 정보가 AI로 생성·조작되었음을 표시해야 한다. 두 경우 모두 의무 위반 행위, 거짓 표시 행위에 대한 제재조치를 마련해야 한다. 딥페이크의 위험은 사업자의 규모가 작다고 해서 사소한 것은 아니므로 이용자·매출 등으로 사업자의 적용 예외를 두는 것은 신중하게 고려할 필요가 있다.

이용자의 경우, 대부분은 앞서 살펴본 제공자·배포자가 표시 의무를 이행하기 때문에 별도로 이용자가 AI로 생성·조작된 정보임을 표시할 필요성은 낮다. 다만, 이용자가 오픈소스 등을 이용하여 직접 콘텐츠를 생성·조작하는 경우라면 타인에게 제공하는 목적 등에 대해서는 이용자가 직접 AI로 생성·조작하였음을 표시하도록 해야 한다.

딥페이크 콘텐츠를 유통하는 온라인 플랫폼은 적법한 표시 의무를 이행하지 않은 콘텐츠에 대한 처리 방침을 공개하고, 부적절한 표시가 되었다고 발견·신고된 콘텐츠에 대해서는 삭제 또는 임시조치하도록 할 필요가 있다.

(3) 제도 정착을 위한 정부 지원

AI 생성물 표시 기술의 신뢰성을 높이고, 일부 기업이 기술을 독점하여 표시 기술 자체가 시장의 진입장벽이 되지 않도록 정부는 기술개발에 대한 지원을 강화해야 한다. 이 과정에서 국제표준과 조화를 이루도록 긴밀한 국제협력도 필요하다.

이용자가 다양한 정보와 표시를 확인하여 딥페이크임을 알고 비판적으로 활용할 수 있도록 정부는 개인의 디지털 리터러시(digital literacy)를 함양하는 조치도 병행해야 한다.

『이슈와 논점』은 국회의원의 입법활동을 지원하기 위해 최신 국내외 동향 및 현안에 대해 수시로 발간하는 보고서입니다.

07233 서울시 영등포구 의사당대로 1 국회입법조사처
Tel. 02-6788-4510
발간등록번호 31-9735044-001340-14
ISSN 2005-744X