

인공지능 생성물 고지 및 표시 제도

정 준 화*

Abstract

생성형 인공지능(AI) 기술 발전으로 실제와 AI 생성물의 구분이 어려워지면서, 주요국은 이용자의 진위 파악을 돕기 위해 AI 생성물임을 밝히는 고지·표시 제도를 도입하고 있다. EU는 「인공지능법」 제50조를 통해 제공자와 배포자 모두에게 다층적 투명성 의무를 부과하고, 미국 캘리포니아주는 「AI 투명성법」을 통해 잠재적·명시적 출처표시와 탐지도구 제공을 의무화하였다. 중국 역시 「인공지능 생성·합성 콘텐츠 표시 방법」을 통해 명시적·묵시적 표시와 메타데이터 기록 의무 등을 부과하고 있다. 한국도 「인공지능 기본법」 제31조에서 고영향·생성형 AI 기반 서비스에 대한 고지 의무, AI 생성물 표시 의무, 딥페이크 결과물의 고지·표시 의무 등 세 가지 투명성 체계를 규정하였다. 그러나 실제 현장에서 발생하는 다양한 피해 양상과 비교하면, 이용자·영향받는자 관계의 미비, 서비스 이용 중 실시간 인지 부족, 취약계층 고려의 한계, 표시 변조·삭제에 대한 대응 부재 등 보완해야 할 지점도 존재한다.

I. 서론

인공지능(AI) 기술은 발전 속도를 논하는 것조차 의미가 없을 만큼 빠르게 진화하고 있다. 활용 범위 역시 초기의 AI 챗봇 일변도를 지나 인간의 일상 전반으로 확장 중이다. 모바일 앱(app)이 어느 순간 PC 기반의 웹(web)을 대체했던 것처럼, AI도 조만간 결정적 전환점(tipping point)을 지나 디지털 사회의 보편적인 수단 중 하나가 되는 범용 AI(general purpose AI, GPAI)의 시대로 나아갈 것이다. 이 과정에서 우리는 다양한 AI 서비스를 경험해 왔고 앞으로 경험하겠지만, 그 중에서 가장 활발하게 사용된 것이 ‘생성(generation)’이다. 2022년 11월 OpenAI의 ChatGPT가 등장한 이후 불과 3년 만에, 전 세계 AI의 중심

* 국회입법조사처 과학방송통신팀 입법조사관, joonhwa@assembly.go.kr

축은 문장·문서 생성, 요약, 이미지·영상 생성 등 생성형 서비스에 있었다.

AI의 생성 기능으로 인해 인간은 사고와 창작에 필요한 시간과 비용을 혁신적으로 줄이는 편익을 얻었다. 그러나 동시에 실제 존재하는 것과 AI로 생성된 것 사이의 구분이 어려워지면서 사회적 혼란과 비용이 발생하기도 했다. 특히, AI의 심층학습(deep learning)을 통해 고도로 정밀한 이미지 합성·편집·생성이 가능한 심층조작, 즉 딥페이크(deepfake)가 등장하면서, 이용자의 눈과 귀만으로는 실제 존재하는 것과 AI로 생성된 것을 구분하는 것은 거의 불가능에 가까워졌다. 또한, 초기에는 AI를 이용한 생성 또는 딥페이크 제작이 일부 전문가의 전유물이었지만, 이제는 AI 기술 발전으로 쉽고 저렴한 비용으로 정밀한 결과물을 만들어 낼 수 있게 되면서 실제와 AI 생성물의 구분 문제는 누구나 직면할 수 있는 사회 전반의 문제가 되었다.

예를 들어서, 온라인쇼핑몰에서 이미지·영상·사용후기를 보고 물건을 샀는데, 이것이 딥페이크로 만들어진 것으로 밝혀져 소비자의 피해가 발생한 사례가 존재한다. 또한 이용자를 속여 실제 사람인 것처럼 가짜 신원을 생성, 도용하여 불법적인 목적으로 사용하는 신분 사기 위험도 있다.¹⁾ 딥페이크가 정치적으로 활용되어 선거에 영향을 미친 일도 있었는데, 2024년 1월 미국 대선의 뉴햄프셔주 프라이머리를 앞두고 바이든 대통령의 목소리를 흉내낸 딥페이크 음성으로 민주당 당원들에게 투표 거부를 독려하는 전화가 발송된 것이 대표적이다.²⁾ 이 외에도 국내에서는 2024년 8월을 전후로 하여 학생을 대상으로 한 딥페이크 합성음란물이 사회적 문제가 되기도 했다.

AI 생성물과 딥페이크가 빠른 속도로 확산되고 있는 상황에서 사후 점검과 처벌만으로는 이용자의 피해를 완전히 막기 어렵다. 따라서 주요국은 그 대안으로 AI 생성물임을 이용자에게 명확히 알리는 고지·표시 의무를 도입하고 있다. 대표적인 입법례로 유럽연합의 「인공지능법」, 미국 캘리포니아주의 「AI 투명성법」, 중국의 「인공지능 생성·합성 콘텐츠 표시방법」, 그리고 한국의 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」이 있다. 이에 이 글을 통해 주요국과 한국의 인공지능 생성물 고지 및 표시 제도를 살펴보고, 향후 개선 및 보완해야 할 과제를 모색해 보고자 한다.

1) 노재원·유재홍·장진철·조지연.(2024). 『AI위험 유형 및 사례 분석』, SPRi 이슈리포트 IS-183, 소프트웨어정책연구소.

2) 정준화.(2024). 정보통신망 이용촉진 및 정보보호등에 관한 법률 일부개정법률안(의안번호 2126048) 입법영향분석 - 인공지능으로 만든 가상 정보임을 표시하도록 하는 의무, NARS 입법영향분석 기획보고서 제7호, 국회입법조사처.

II. 외국 입법례

1. 유럽연합의 「인공지능법」

유럽연합은 AI의 위험 수준에 따라 차등적인 규제를 적용하는 일반적이고 포괄적인 규제 법률인 「인공지능법(AI Act)」을 제정하였다. 이 법은 2024년 8월 1일 발효되어 순차적으로 시행될 예정이며, 특히 AI 생성물에 대한 표시를 규정한 제50조 투명성 의무 조항(Article 50 Transparency obligations for providers and deployers of certain AI systems)은 「인공지능법」의 핵심인 ‘고위험 AI’에 대한 규제와 함께 2026년 8월 1일 시행될 예정이다. 다만, EU 집행위원회가 최근 디지털 기업의 규제를 간소화하는 ‘디지털 옴니버스(Digital Omnibus)’ 패키지를 제안했는데, 이 제안이 받아들여진다면 제50조의 시행일은 2027년 12월로 기존보다 16개월 연기될 가능성도 있다.³⁾

「인공지능법」 제50조는 AI 제공자 및 배포자가 늦어도 최초의 상호작용 또는 노출 시에 이용자가 AI 생성물임을 명확하고 구별할 수 있는 방식으로 표시하도록 규정하고 있다. 주요 내용은 자연인과 직접 상호작용하는 AI 제공자(provider)의 의무, 콘텐츠를 생성하는 AI 제공자의 의무, 딥페이크를 만드는 AI 배포자(deployer)의 의무 등이며, 그 내용은 다음과 같다.⁴⁾

첫째, AI 제공자는 자연인과 직접 상호작용하기 위한 AI를 제공할 경우, 그것을 이용하는 자연인 스스로 AI와 상호작용하고 있음을 알 수 있는 방식으로 설계 및 개발되도록 해야 한다. 다만, 사용 상황 및 맥락을 고려할 때 합리적으로 정보에 정통하고 관찰력이 있으며 신중한 자연인의 관점에서 AI와 상호작용하고 있음이 명백한 경우에는 이러한 의무를 면제할 수 있다. 또한, 이 의무는 제3자의 권리 및 자유에 대한 적절한 안전장치를 갖춘 상황에서 범죄를 탐지·예방·수사 또는 기소하기 위하여 법에서 승인한 AI, 일반 대중이 범죄를 신고하기 위하여 이용하는 AI 등에는 적용되지 아니한다.

둘째, 범용 AI를 포함하여 합성 오디오·이미지·비디오 또는 텍스트 콘텐츠를 생성하는 AI 제공자는 AI 산출물이 기계 판독 가능한 형식으로 표시되고 인위적으로 생성 또는 조작되었다는 것을 감지할 수 있도록 보장해야 한다. 제공자는 관련 기술표준에 반영되는 다양한 콘텐츠 종류의 특수성 및 제한, 구현비용, 일반적으로 알려진 최신기술을 고려하여 기술

3) European Commission, “Digital Omnibus Regulation Proposal”, (최종검색일: 2025. 11. 30.), <<https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal>>

4) 국회도서관 외국법률번역DB, “유럽연합 인공지능법”, (최종검색일: 2025. 11. 20.), <<https://law.nanet.go.kr/foreignlaw/foreignIndex/foreignIndexView.do?countPerPage=10&cn=TLAW1202400312>>

적으로 실현 가능한 범위에서 그 기술적 해법이 유효하고 상호운용 가능하고 견고하며 신뢰할 수 있도록 해야 한다. 이 의무는 AI가 표준 편집을 위한 보조 기능을 수행하거나 배포자가 제공한 입력데이터나 그 의미를 실질적으로 변경하지 아니하는 경우, 또는 범죄를 탐지·예방·수사 또는 기소하기 위하여 법에서 승인하는 경우에는 적용되지 아니한다.

셋째, 딥페이크를 구성하는 이미지, 음성 또는 비디오 콘텐츠를 생성하거나 조작하는 AI 배포자는 해당 콘텐츠가 인위적으로 생성 또는 조작되었다는 것을 공개해야 한다. 이 의무는 범죄를 탐지·예방·수사 또는 기소하기 위하여 법에서 사용을 승인하는 경우에는 적용되지 아니한다. 또한, 해당 콘텐츠가 명백히 예술적, 창의적, 풍자적, 허구·유사 저작물 또는 프로그램의 일부를 구성하는 경우에는 해당 저작물의 전시 또는 향유를 방해하지 아니하는 적절한 방식으로 생성 또는 조작된 콘텐츠임을 공개할 수 있다.

한편, 딥페이크를 공익적 목적으로 사용하는 경우, 즉 일반 대중에게 공익 문제를 알릴 목적으로 공표되는 텍스트를 생성 또는 조작하는 경우에도 AI 배포자는 해당 텍스트가 인위적으로 생성 또는 조작되었다는 것을 공개해야 한다. 다만, 이 의무는 범죄를 탐지·예방·수사 또는 기소하기 위하여 법에서 사용을 승인하는 경우 또는 AI 생성 콘텐츠가 인적 검토 또는 편집통제절차를 거치고 자연인 또는 법인이 콘텐츠의 공표를 위한 편집 책임을 지는 경우에는 적용되지 아니한다.

2. 미국 캘리포니아주의 「AI 투명성법」

미국 캘리포니아주의 「AI 투명성법(California AI Transparency Act)」은 2024년 9월 19일 제정되어 2026년 1월 1일부터 시행될 예정이다. 이 법률의 목적은 생성형 AI가 만든 디지털 콘텐츠에 대해 출처표시와 탐지 가능성을 의무화하여 이용자가 진위 여부를 확인할 수 있도록 투명성과 책임성을 확보하는 데 있다. 다만, 사용자가 직접 생성하는 것이 아니라(non-user-generated) 전문 콘텐츠 기업이 만드는 비디오게임·TV·스트리밍·영화 등과 같은 콘텐츠 서비스는 이 법의 적용 대상에서 제외된다.

핵심 규제 대상은 ‘커버드 제공자(covered provider)’이다. 생성형 AI를 개발·제작하는 자로서, 월간 방문자·사용자 수 100만 명 이상이며, 캘리포니아주에서 공개적으로 접근 가능한 서비스를 제공하는 자를 의미한다. 이와 관련한 「AI 투명성법」의 주요 내용은 다음과 같다.⁵⁾

5) 미국 캘리포니아주 의회 홈페이지. “SB-942 California AI Transparency Act”.(최종검색일: 2025. 11. 20.).
(https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942)

첫째, 커버드 제공자는 AI 탐지 도구(AI detection tool)를 무료로 제공해야 한다. 자사의 생성형 AI가 만든 이미지·영상·음성 등 콘텐츠를 탐지할 수 있는 무료 AI 탐지 도구를 제공해야 하는데, 이 도구는 시스템 출처 데이터를 식별·출력할 수 있어야 하고, 개인 출처 데이터는 절대 출력해서는 안 되며, 웹·API 형태로 누구나 접근 가능하도록 공개해야 한다. 또한 사용자가 제공하는 콘텐츠나 정보는 필요한 범위 외에는 보관할 수 없으며, 개인 정보는 수집·저장해서는 안 된다.

둘째, 커버드 제공자는 AI 생성물임을 표시(이를 ‘공개(disclosure)’라고 한다)해야 한다. 커버드 제공자는 자사의 생성형 AI가 만들어 내는 모든 콘텐츠에 대해 잠재적(latent, 또는 비가시적) 출처표시를 반드시 삽입해야 하고, 사용자가 요청하는 경우에는 사람이 볼 수 있는 명시적(manifest, 또는 가시적) 표시까지 포함할 수 있도록 옵션을 제공해야 한다. 명시적 표시는 AI 생성물임을 명확히 표시하고, 명료하게 눈에 띄어서 매체 특성에 적합하고 일반인이 이해할 수 있어야 하며, 기술적으로 가능한 범위에서 영구적 또는 제거가 극히 어려워야 한다. 잠재적 표시에는 제공자 이름, 모델 이름·버전, 생성·수정 시간, 고유 식별자 등이 포함되어야 하고, 탐지 도구로 감지 가능해야 하며, 업계 표준과 일치해야 하고, 기술적으로 가능한 범위에서 영구적 또는 제거가 극히 어려워야 한다.

셋째, 커버드 제공자가 생성형 AI 시스템을 제3자에게 라이선스하는 경우, 제3자가 이 출처표시 기능을 유지하도록 계약적으로 강제해야 한다. 만약 제3자가 이를 삭제·변조하면 96시간 이내 라이선스를 철회할 법적 의무를 갖는다.

3. 중국의 「인공지능 생성·합성 콘텐츠 표시 방법」

중국 국가인터넷정보판공실, 공업정보화부,公安部, 국가라디오텔레비전총국 등 4개 부처가 공동으로 「인공지능 생성·합성 콘텐츠 표시 방법(人工智能生成合成内容标识办法)」을 제정하여 2025년 9월 1일부터 시행하고 있다. 이 규정의 적용을 받는 ‘AI 생성·합성 콘텐츠’는 AI 기술로 생성 및 합성된 텍스트, 이미지, 오디오, 비디오, 가상 장면 등의 콘텐츠를 의미한다. 표시의 유형은 명시적 표시와 묵시적 표시로 구분된다. 명시적 표시는 텍스트·음성·이미지 등으로 사용자에게 명확히 인식 가능한 표시를 의미하며, 콘텐츠 유형별로 구체적인 방식이 다르다. 묵시적 표시의 경우, 파일의 메타데이터에 콘텐츠 속성 정보, 서비스제공자 명칭 또는 코드, 콘텐츠 번호 등과 같은 정보를 삽입해야 한다. 또한 디지털 워터마크 등을 추가하는 것을 권장한다. 이와 관련하여 사업자가 준수해야 할 의무는 다음과 같다.⁶⁾

6) 백지연.(2025). 정부부처 공동으로 「인공지능 생성·합성 콘텐츠 표시 방법」 발표, 저작권 동향 제9호, 한국저작권위

첫째, AI 서비스제공자는 AI 생성·합성 콘텐츠를 생성하는 파일의 메타데이터에 묵시적 표시를 추가해야 하며, AI 생성·합성 콘텐츠의 식별 방식 및 표시 관리 정책을 이용약관에 명확히 기재하고, 이용자가 이 내용을 알 수 있도록 고지해야 한다. 아울러, 서비스 이용자가 명시적 표시가 없는 AI 생성·합성 콘텐츠를 제공하는 경우, 이용약관을 통해 이용자의 표시 의무와 콘텐츠 사용에 대한 책임을 명확히 고지하고, 이용자 관련 로그를 6개월 이상 보관해야 한다.

둘째, 온라인 콘텐츠 전송 서비스 제공자는 AI 생성·합성 콘텐츠의 전송에 있어 유형별로 조치를 취해야 한다. 먼저, 파일의 메타데이터에 묵시적 표시가 포함되어 있는 경우, 게시물 주변에 명확한 표시를 추가해 해당 콘텐츠가 AI 생성·합성 콘텐츠임을 알려야 한다. 다음으로, 메타데이터에 묵시적 표시가 없더라도 사용자가 AI 생성·합성 콘텐츠임을 밝힌 경우, 동일하게 명시적 표시를 부착해야 한다. 마지막으로, 메타데이터의 묵시적 표시와 이용자의 소명 모두 없지만, 명시적 표시나 합성 흔적이 발견될 경우, 이를 ‘생성·합성 콘텐츠 의심’으로 분류하고, 게시물에 경고성 표시를 부착해야 한다. 또한, 이용자가 자발적으로 AI 생성·합성 콘텐츠의 생성 여부를 신고할 수 있는 기능을 제공해야 하며, 위 세 가지 상황에서는 메타데이터에 생성·합성 콘텐츠 속성 정보, 전송 플랫폼 명칭 또는 코드, 콘텐츠 번호 등 전송 요소 정보를 기록해야 한다.

셋째, 온라인 응용프로그램 제공 플랫폼은 앱을 출시하거나 심사할 때, 해당 앱이 AI 생성·합성 기능을 제공하는지를 서비스제공자에게 확인해야 한다. 또한 AI 생성·합성 기능이 있는 경우, 플랫폼은 생성된 생성·합성 콘텐츠의 표시 의무 이행과 관련된 자료를 검증해야 한다. 이용자는 생성·합성 콘텐츠의 생성 및 게시 시, 자발적으로 이를 밝히고 적절한 표시기능을 사용해야 한다. 표시의 삭제·변조·위조·은폐 행위, 위 행위를 위한 도구·서비스를 제공하는 행위 또는 부정한 표시 수단을 이용해 타인의 권익 침해하는 행위는 금지된다.

원회.

Ⅲ. 한국의 인공지능 기본법과 하위 법령

1. 개 요

AI 산업 진흥과 ‘고영향 AI’를 중심으로 AI 규제 사항을 규정한 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “인공지능 기본법”이라 한다)이 2025년 1월 22일 제정되어 2026년 1월 21일부터 시행된다. 이 중에서 AI 생성물 고지 및 표시 제도는 인공지능 기본법 제31조 및 동법 시행령⁷⁾ 제22조에 AI 사업자의 3가지 의무로 규정되어 있다. 첫째, 고영향 및 생성형 인공지능에 기반하여 운영된다는 사실의 고지 의무다. 고지는 사전에 미리 알리는 것으로 서비스 가입 과정에서 약관 등에 생성형 AI 기반임을 명시하는 방법이 고지의 사례다. 둘째, 생성형 인공지능이 생성한 결과물임을 표시해야 한다. 표시는 AI 제품·서비스 제공 과정 중에 해당 생성된 결과물이 생성형 AI를 통해 생성된 것임을 나타내는 것이다. 셋째, 딥페이크는 이용자가 명확하게 인식할 수 있도록 고지 또는 표시해야 한다. 이에 대하여 현재 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령안」이 입법예고 중이다.

표 1 AI 생성물 고지 및 표시에 관한 인공지능 기본법과 시행령안 내용

구분	법률	시행령안
고지	제31조(인공지능 투명성 확보 의무) ① 인공지능사업자는 고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스를 제공하려는 경우 제품 또는 서비스가 해당 인공지능에 기반하여 운영된다는 사실을 이용자에게 사전에 고지하여야 한다.	제22조(인공지능 투명성 확보 의무) ① 인공지능사업자는 고영향 인공지능이나 생성형 인공지능을 이용한 제품 또는 서비스(이하 “제품등”이라 한다)를 제공하기 전에 다음 각 호의 어느 하나에 해당하는 방법으로 법 제31조제1항에 따른 고지를 하여야 한다. 1. 제품등에 직접 기재하거나, 계약서, 사용 설명서, 이용약관 등에 기재 2. 이용자의 화면 또는 단말기 등에 표시 3. 제품등을 제공하는 장소(해당 장소와 합리적으로 관련된 범위의 장소를 포함한다)에 인식하기 쉬운 방법으로 게시 4. 그 밖에 제품등의 특성을 고려하여 과학기술정보통신부장관이 인정하는 방법

7) 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령안」은 2025. 11. 12. ~ 2025. 12. 22.까지 입법예고 중이다.(자료: 법제처 국민참여입법센터, “인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령 제정안 입법예고”, (최종검색일: 2025. 11. 20.), <<https://opinion.lawmaking.go.kr/gcom/ogLmPp/84360>>)

구분	법률	시행령안
표시	② 인공지능사업자는 생성형 인공지능 또는 이를 이용한 제품 또는 서비스를 제공하는 경우 그 결과물이 생성형 인공지능에 의하여 생성되었다는 사실을 표시하여야 한다.	② 인공지능사업자가 법 제31조제2항에 따른 표시(법 제31조제3항에 따른 실제와 구분하기 어려운 결과물을 제공하는 경우로서 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하는 경우에는 적용하지 아니한다)를 할 때에는 다음 각호 중 하나의 방법으로 할 수 있다. 1. 사람이 인식할 수 있는 방법 2. 기계가 판독할 수 있는 방법. 다만, 이 경우에는 생성형 인공지능에 의하여 생성되었다는 사실을 1회 이상 안내 문구·음성 등으로 제공하여야 한다.
디스플레이 고지 및 표시	③ 인공지능사업자는 인공지능시스템을 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 결과물을 제공하는 경우 해당 결과물이 인공지능시스템에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하여야 한다. 이 경우 해당 결과물이 예술적·창의적 표현물에 해당하거나 그 일부를 구성하는 경우에는 전시 또는 향유 등을 저해하지 아니하는 방식으로 고지 또는 표시할 수 있다.	③ 법 제31조제3항에 따른 고지 또는 표시는 인공지능사업자가 다음 각 호의 사항을 고려하여 이용자가 명확하게 인식할 수 있는 방식으로 해야 한다. 1. 이용자가 시각, 청각 등을 통하거나 소프트웨어 등을 이용하여 쉽게 내용을 확인할 수 있는 방법으로 고지 또는 표시할 것 2. 주된 이용자의 연령, 신체적·사회적 조건 등을 고려하여 고지 또는 표시할 것
예외	④ 그 밖에 제1항에 따른 사전고지, 제2항에 따른 표시, 제3항에 따른 고지 또는 표시의 방법 및 그 예외 등에 관하여 필요한 사항은 대통령령으로 정한다.	④ 법 제31조제1항부터 제3항까지는 다음 각 호에 해당하는 경우에는 적용하지 아니한다. 다만, 제3호의 경우에는 법 제31조제1항부터 제3항까지 중 전부 또는 일부를 적용하지 아니할 수 있다. 1. 제품·서비스명, 이용자 화면이나 제품 겉면 및 결과물에 표시된 문구 등을 고려할 때 고영향 인공지능 또는 생성형 인공지능을 활용한 사실이 명백한 경우 2. 인공지능사업자의 내부 업무 용도로만 사용되는 경우 3. 그 외 제품 등의 유형·특성이나 결과물의 내용, 이용형태 및 기술 수준 등을 고려하여 법 제31조제1항부터 제3항까지 중 전부 또는 일부에 대한 적용 예외가 필요한 사항으로 과학기술정보통신부장관이 정하여 고시하는 경우

법률상 투명성 확보 의무는 이용자에게 최종적으로 AI 제품 및 서비스를 제공하는 ‘인공지능사업자’에게 부과된다. 따라서 파운데이션 모델을 개발·공급하는 AI개발사업자, AI개발사업자로부터 API 등을 받아서 특화 AI 서비스를 만들어 이용자에게 제공하는 AI이용사업자에게 투명성 확보 의무가 발생한다. 한편, 단순히 AI 제품·서비스를 이용한 결과물을 자신의 서비스 등에 활용하는 ‘이용자’는 인공지능 기본법상 사업자에 해당하지 않기 때문에 투명성 확보 의무가 없다. 예를 들어, AI를 이용하여 컴퓨터그래픽을 생성한 후 영화에

삽입한 영화제작자는 단순히 AI 서비스를 이용한 결과물을 자신의 콘텐츠에 활용한 것이므로 인공지능 기본법상 사업자가 아니며 이용자에 해당하고, 투명성 의무가 발생하지 않는다.

2. 주요 의무

첫째, 고지 의무다. 인공지능 기본법 제31조제1항에 따라 인공지능사업자는 고영향 AI나 생성형 AI를 이용한 제품 또는 서비스를 제공하기 전에, 해당 제품·서비스가 고영향 및 생성형 AI에 기반하여 운영된다는 사실을 고지하여야 한다. 고지의 방법은 문서(제품·서비스에 직접 기재하거나, 계약서, 사용 설명서, 이용약관 등에 기재), 표시(이용자의 화면 또는 단말기 등에 표시), 게시(제품·서비스를 제공하는 장소에 인식하기 쉬운 방법으로 게시), 그 밖에 제품·서비스의 특성을 고려하여 과학기술정보통신부장관이 인정하는 방법이다.

표 2 생성형 AI 생성물에 대한 표시 방식

대상	방식	비가시적/비가청적 방식 사용시 조치 사항
이미지 콘텐츠	이미지 위에 로고, 텍스트 등의 가시적 워터마크를 표시하거나, 특정 기술을 활용한 비가시적 워터마크 적용	해당 콘텐츠가 생성형 인공지능을 통해 생성된 것임을 인식할 수 있도록 생성 결과물에 관련 문구 등을 포함하여 안내 ※ (예시) 생성 결과 이미지에 “이 그림은 AI로 생성되었습니다.”와 같이 생성형 AI를 활용 사실 포함하여 최종 결과로 사용자에게 전달
동영상 콘텐츠	이미지와 유사하게 로고나 텍스트 등의 가시 가능한 워터마크 및 특정 기술을 활용한 가시 불가능한 워터마크 적용	반드시 동영상 전체에 표시할 필요는 없으며, 이용자가 해당 콘텐츠가 생성형 인공지능을 통해 생성된 것임을 인식할 수 있도록 사전에 안내 ※ (예시) 영상 시작 부분에 “이 애니메이션은 AI로 생성되었습니다.”와 같이 생성형 AI를 활용 사실을 멘트로 간단히 안내
오디오 콘텐츠	콘텐츠 재생 초기 생성형 AI 결과물이란 사실을 음성으로 알리거나, 비가청적 방식의 기술적 워터마크 활용 가능	동영상과 동일하게 오디오 전체에 표시할 필요는 없으며, 이용자가 해당 콘텐츠가 생성형 인공지능을 통해 생성된 것임을 인식할 수 있도록 사전에 안내 ※ (예시) 오디오 콘텐츠 시작 부분에 AI로 만든 음성 결과물 시작 시 “이 음성은 AI로 생성되었습니다.”와 같이 생성형 AI를 활용 사실을 멘트로 간단히 안내

자료: 과학기술정보통신부·한국정보통신기술협회.(2025). 인공지능 투명성 확보 가이드라인.

둘째, 표시 의무다. 인공지능 기본법 제31조제2항에 따라 인공지능사업자는 생성형 AI 또는 이를 이용한 제품·서비스를 제공하는 경우 그 결과물이 생성형 AI에 의하여 생성되었다는 사실을 표시하여야 한다. 표시는 사람이 인식할 수 있는 방법 또는 기계가 판독할 수

있는 방법 중 하나의 방법으로 할 수 있다. 이 중에서 기계가 판독할 수 있는 방법을 적용하더라도 생성형 AI에 의하여 생성되었다는 사실을 1회 이상 안내 문구·음성 등으로 제공하여야 한다.

셋째, 딥페이크에 대한 고지 또는 표시 의무다. 인공지능 기본법 제31조제3항에 따라 인공지능사업자는 AI를 이용하여 실제와 구분하기 어려운 가상의 음향, 이미지 또는 영상 등의 결과물을 제공하는 경우 해당 결과물이 AI에 의하여 생성되었다는 사실을 이용자가 명확하게 인식할 수 있는 방식으로 고지 또는 표시하여야 한다. 이때 이용자가 시각, 청각 등을 통하거나 소프트웨어 등을 이용하여 쉽게 내용을 확인할 수 있는 방법으로 고지 또는 표시해야 하고, 주된 이용자의 연령, 신체적·사회적 조건 등을 고려하여 고지 또는 표시해야 한다.

표 3 딥페이크에 대한 고지 또는 표시 방식

구분	내용
고지 또는 표시 대상	• 사람의 신체·목소리를 실제와 구별하기 어려울 정도로 구조와 특성을 모방한 콘텐츠를 생성하거나, 인공물·자연물·자연현상 등을 이용자로 하여금 실재(實在)하는 것으로 오인하게 할 소지가 있을 정도로 모방한 콘텐츠 • (예외) 실제와 다른 2D·3D 애니메이션, 플레이어가 직접 조작하는 게임, 누가 보더라도 가상임이 명확한 이미지 등은 딥페이크에 해당하지 않음
고지 또는 표시 방식	• 주된 이용자의 연령, 신체적, 사회적 조건 등을 고려하여 표시: 정보 접근성과 이해도를 높이고 오인·혼동을 방지하기 위해 사용자 특성을 고려하여 사전 고지 및 표시함 • (예시 1) 아동 또는 노인의 경우, 성인과 비교해 이해할 수 있는 어휘, 단어에 한계가 있어 사용자 연령을 고려해야 함 • (예시 2) 장애인인 신체적 제약으로 발생할 수 있는 한계를 고려해야 함. 그 예로는 신체 크기, 신체능력, 인지능력이 있음 • (예시 3) 초보자의 경우 관련 서비스의 경험 여부와 사전 배경지식의 차이로 지식수준이 다를 수 고려해야 함
고지 또는 표시 예외	• 해당 결과물이 예술적·창의적 표현물에 해당하거나 그 일부를 구성하는 경우에는 전시 또는 향유 등을 저해하지 아니하는 방식으로 고지 또는 표시할 수 있음 • (예시) 제품 패키지에 표시하여 안내하는 방법, 회원가입 시 서비스 이용약관 내 고지하는 방법, 서비스 온보딩 상황에서 서비스 이용 방법 및 시 기반 운용 안내 방법, 세션 연결 상황에서 AI 생성 결과물 및 오류 가능성 안내 방법, 캡션·주석·설명 추가하여 실제 콘텐츠 향유 과정에서는 방해가 되지 않는 방식 등이 가능

자료: 과학기술정보통신부·한국정보통신기술협회.(2025). 인공지능 투명성 확보 가이드라인.

넷째, 고지 또는 표시 의무가 완화되는 상황이다. 인공지능 기본법 제31조제4항에 따라 AI 사용이 명백한 경우나 사업자가 자신의 내부업무를 처리하기 위해 AI를 활용하는 등 사전고지 및 표시가 필요 없는 예외적인 경우에 투명성 확보 의무를 완화한다. 구체적으로

먼저, 제품·서비스명, 이용자 화면이나 제품 겉면 및 결과물에 표시된 문구 등을 고려할 때 고영향 AI 또는 생성형 AI를 활용한 사실이 명백한 경우이다. 다음으로, 인공지능사업자가 직접 개발한 AI 서비스를 사내 직원들만 활용하는 경우 또는 내부적으로 연구개발을 위해 AI가 사용된 경우와 같이 내부 업무 용도로만 사용되는 경우이다. 마지막으로, 제품 등의 유형·특성이나 결과물의 내용, 이용형태 및 기술 수준 등을 고려하여 법 제31조제1항부터 제3항까지 중 전부 또는 일부에 대한 적용 예외가 필요한 사항으로 과학기술정보통신부장관이 정하여 고시하는 경우이다.

IV. 결 론

지금까지 주요 국가와 한국의 AI 생성물 고지·표시 제도를 살펴보았다. 우선 외국의 경우, EU는 「인공지능법」 제50조를 통해 제공자와 배포자 모두에게 투명성 의무를 부과하며, 합성 콘텐츠·딥페이크에 대해 기계판독과 가시적 표시를 병행하는 다층적 체계를 마련하였다. 미국 캘리포니아주는 「AI 투명성법」을 통해 ‘커버드 제공자’에게 잠재적·명시적 출처표시 삽입과 무료 탐지도구 제공 의무를 부여하여, 생성물 검증의 기술적 기반을 마련하였다. 중국은 「인공지능 생성·합성 콘텐츠 표시 방법」에서 명시적·묵시적 표시에 더해 메타데이터 기록, 경고표시, 신고 기능, 플랫폼의 전송 단계 관리 등 가장 포괄적인 의무 체계를 도입하고 있다.

한국은 「인공지능 기본법」 제31조와 시행령안에서 고지·표시·딥페이크 고지의 세 가지 투명성 의무를 규정하고 있다. 고영향·생성형 AI 기반 서비스임을 사전에 고지하고, 생성형 AI가 만든 결과물임을 표시하며, 실제와 구분하기 어려운 딥페이크 결과물은 이용자가 명확히 인식할 수 있는 방식으로 고지 또는 표시하도록 하고 있다. 또한 적용 주체를 ‘인공지능사업자’로 한정하며, AI 사용이 명백한 경우나 내부업무 목적 등은 예외로 규정한다. 이와 같은 틀은 이용자가 합리적 판단을 통해 AI 생성물을 스스로 인지할 수 있는 최소한의 기반을 제공한다.

다만, 외국 입법례와 국내에서 발생한 다양한 활용 사례를 종합해 보면, 현행 제도의 보완 필요성도 확인된다. 첫째, 인공지능 기본법은 주된 규율 대상을 ‘인공지능사업자 → 이용자’ 관계로 설정하고 있어, 실제 피해가 빈번히 발생하는 ‘이용자 → 영향을 받는자’ 관계에 대한 규정이 미비하다. 최근 문제가 된 딥페이크 사건들은 대부분 이용자가 타인을

대상으로 피해를 야기하는 구조였으며, 형법상 사기죄, 공직선거법상 뱃페이크 금지, 성폭력처벌법상 허위영상물 처벌 등 사후적 규율은 존재하지만, 사전적 안전장치를 마련하기 위해서는 이용자와 영향을 받는자 관계를 별도로 설정하는 방안도 검토할 필요가 있다. 또한 이 과정에서 ‘영향을 받는 자’를 AI 제품 또는 서비스에 의하여 자신의 생명, 신체의 안전 및 기본권에 중대한 영향을 받는 자로 정의한 인공지능 기본법 제2조제9호의 타당성도 검토해 볼 필요가 있다. 실제로 뱃페이크로 발생하는 문제는 생명·신체·기본권 수준보다 낮은 재산적 손실 문제가 빈번할 수 있기 때문이다.

표 4 인공지능 기본법 적용 대상자와 미적용 대상자 예시

사례	인공지능 기본법 적용 대상자			인공지능 기본법 미적용 대상자
	인공지능사업자		이용자	
	개발사업자	이용사업자		
의료 영상 판독	의료 이미지 판독 AI 개발 기업	진단 참고용 AI솔루션 기업	AI솔루션 활용하여 진료에 참고하는 의사	AI 진료 결과에 따라 처방을 받은 환자
영화 제작	범용 AI 개발 기업	영화 제작용 AI영상 기업	AI영상 활용하여 영화를 만드는 제작자	AI 영상물의 가상 설정을 신뢰한 관객
음성 생성	범용 AI 개발 기업	음성 안내 AI챗봇 기업	AI 챗봇 이용하여 고객 콜센터 운영 기업	딥페이크 음성을 듣고 금전을 송금한 개인
이미지 생성	범용 AI 개발 기업	사진편집 AI이미지 기업	AI이미지를 사용하여 온라인 쇼핑몰에 업로드하는 이용자	딥페이크 이미지를 보고 상품을 구매한 소비자

자료: 과학기술정보통신부·한국정보통신기술협회(2025), 인공지능 투명성 확보 가이드라인, 일부 수정

둘째, EU의 「인공지능법」은 이용자가 AI와 상호작용(interacting with an AI system)하고 있음을 실시간으로 인지할 수 있도록 설계할 것을 명시하고 있다. 반면 한국의 현행 제도는 서비스 가입 단계의 사전 고지와 결과물에 대한 표시 중심으로 구성되어 있어, 실제 AI 사용 도중에는 AI와 상호작용하고 있다는 사실을 이용자가 잊거나 인지하지 못할 가능성이 있다. 서비스 이용 과정에서도 AI와의 상호작용을 인지할 수 있도록 안내하는 장치를 도입하는 방안도 고려할 수 있다.

셋째, 인공지능 기본법은 뱃페이크 고지 또는 표시의 방식으로, ‘주된 이용자의 연령, 신체적, 사회적 조건 등을 고려하여 표시’하도록 규정하고 있다. 그러나 현실에서 뱃페이크 문제에 더 쉽게 노출되는 계층은 이른바 디지털취약계층이다. 또한 앞서 살펴본 바와 같이 ‘이용자’보다는 ‘영향을 받는자’가 뱃페이크 문제를 직면하게 될 가능성이 더 높다. 따라서

이러한 상황에서 ‘주된 이용자’를 기준으로 딥페이크 고지 또는 표시 방식을 설정할 경우 취약 계층을 더욱 취약하게 만들 우려가 있다. 따라서 딥페이크로부터 보호할 대상자를 보다 강하게 보호할 수 있도록 관련 규정을 정비하는 방안을 고려해 볼 수 있을 것이다.

넷째, 고지·표시 제도의 변조·삭제 등 훼손에 대한 대응 방안도 보완이 요구된다. 생성형 AI가 만든 콘텐츠임에도 불구하고 고지·표시가 삭제되면, 그 콘텐츠가 오히려 ‘AI 생성물이 아닌 것’으로 오인되는 문제가 발생한다. 미국 캘리포니아주의 「AI 투명성법」은 표시의 영구성·비가역성 확보, 무료 탐지도구 제공 등을 통해 이러한 공백을 보완하고 있으나, 한국 법령은 아직 이에 대한 규정을 두지 않고 있다. 기술적 표시가 위조·삭제될 가능성을 고려하면, 제도 자체의 신뢰성을 확보하기 위한 기술적·절차적 대응 장치 마련이 필요하다.

결국, AI 생성물 고지·표시 제도는 생성형 AI의 빠른 확산 속에서 최소한의 투명성과 신뢰를 확보하기 위한 핵심 기반이다. 한국의 인공지능 기본법은 이러한 기반을 마련했다는 점에서 의미가 있지만, 이용자-영향받는자 관계, 사용 과정에서의 인지 가능성, 취약계층 보호, 표시 변조 방지와 같은 현실적 과제를 반영하여 제도를 정교화할 필요가 있다. 국제적 흐름 역시 투명성을 중심으로 한 안전장치 강화를 공통된 방향으로 추진하고 있는 만큼, 한국도 단계적·기술적·사회적 보완을 통해 AI 생성물 규율의 실효성을 강화해 나가야 할 것이다.

참고문헌

과학기술정보통신부·한국정보통신기술협회.(2025). 인공지능 투명성 확보 가이드라인.

국회도서관 외국법률번역DB. “유럽연합 인공지능법”.(최종검색일: 2025. 11. 20.).

〈<https://law.nanet.go.kr/foreignlaw/foreignIndex/foreignIndexView.do?countPerPage=10&cn=TLAW1202400312>〉

노재원·유재홍·장진철·조지연.(2024). AI위험 유형 및 사례 분석』, SPRi 이슈리포트 IS-183, 소프트웨어정책연구소.

미국 캘리포니아주 의회 홈페이지. “SB-942 California AI Transparency Act”.(최종검색일: 2025.11.20.). 〈https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942〉

백지연.(2025). 정부부처 공동으로 「인공지능 생성·합성 콘텐츠 표시 방법」 발표, 저작권 동향 제9호, 한국저작권위원회.

법제처 국민참여입법센터. “인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령 제정안 입법에

고”. (최종검색일: 2025. 11. 20.). <<https://opinion.lawmaking.go.kr/gcom/ogLmPp/84360>>
정준화. (2024). 정보통신망 이용촉진 및 정보보호등에 관한 법률 일부개정법률안(의안번호 2126048)
입법영향분석 - 인공지능으로 만든 가상 정보임을 표시하도록 하는 의무. NARS 입법영향분
석 기획보고서 제7호. 국회입법조사처.
European Commission. “Digital Omnibus Regulation Proposal”. (최종검색일: 2025. 11. 30.).
<<https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal>>.