

AI 결과물 표시의무제, 해외 동향과 시사점

요약

- ▶ 각국이 추진 중인 AI 관련 법에서 규정하는 '결과물 표시의무'는 기술적 불완전성과 국가별 규제 차이로 인해, 기업들은 복잡한 상황에 놓임
- ▶ 한국의 AI 결과물 표시의무는 법(제31조)과 시행령(제22조)을 통해 가시적·비가시적 표시를 모두 허용하며, 콘텐츠 유형별로 이용자가 인식 가능한 수준의 고지를 요구
- ▶ EU, 미국 캘리포니아, 일본 등 주요국들도 각기 다른 방식으로 AI 결과물 표시 규제를 도입하고 있음
 - [EU] AI 개발자 및 AI 시스템을 자신의 책임 하에 사용하는 자가 수범 대상이며, EU 외 기업도 준수해야 함, 위반 시 최대 1,500만 유로 또는 전세계 매출 3%의 벌금이 부과되며 2026년부터 단계적으로 시행 예정
 - [미국 캘리포니아] 월 100만 명 이상 사용자를 보유한 생성형 AI 제공자(Covered Provider)에게 AI 탐지 도구 제공, 비가시적·명시적 표시(Disclosure) 의무를 부과
 - [일본] 강제성 없는 가이드라인을 통해 AI의 설명 가능성, 활용 사실 고지, 산출물 식별, 데이터 출처 투명성, AI 리터러시 제고를 명시, 워터마크·탐지·출처 추적 등 기술 개발과 국제 표준 참여를 권장

노재인 연구위원(jinoh@kinternet.org)

서론

- ▶ 최근 발표된 AI 기본법 시행령과 가이드라인은 이제 이해관계자들의 의견 수렴 과정을 거치며 구체화될 예정이며, 이 과정에서 정책 입안자들은 'AI 혁신 촉진'과 '이용자 신뢰 확보'라는 두 가지 핵심 가치 사이에서 최적의 균형점을 찾아야 하는 과제에 직면해 있음
- ▶ AI 기본법 및 시행령의 결과물 표시의무는 기업들에게 워터마크 등 기술적 조치와 규제 대응 전략을 요구하지만, 기술 개발의 불완전성과 혼합 창작물 같은 경계 사례로 인해 빠른 실무 적용이 어려운 상황
- ▶ 더욱이 여러 나라에서 동일한 서비스를 동시에 제공하는 디지털 산업의 특성상, 기업들은 국가별 상이한 규제에 동시에 대응해야 하는 복잡한 상황에 놓여 있음
- ▶ 따라서, 우리나라의 AI 기본법 및 시행령, 가이드라인 등에서 제시하고 있는 결과물 표시의무 내용을 살펴보고, 주요국의 AI 결과물 표시의무제도를 검토함으로써 제도의 보완 필요성 및 적용 한계를 검토하고자 함

AI 결과물 표시의무제란?

- ▶ AI 결과물 표시의무제는 생성형 AI로 만든 콘텐츠임을 명시적으로 표시하도록 법적으로 의무화하는 제도로, 이용자가 콘텐츠의 출처를 명확하게 인지하고 신뢰도를 스스로 판단할 수 있도록 함으로써 AI 생성 콘텐츠의 투명성 확보를 목적으로 함
- ▶ 즉, AI 생성 콘텐츠와 인간 창작물을 구별할 수 있는 기준을 마련하고 허위 정보 확산을 방지하며 이용자의 알 권리를 보장하기 위한 제도로, 기업은 가시적·비가시적 워터마크, 메타데이터 부착, 음성 안내 등 다양한 기술과 방식을 활용해 이를 이행해야 함

AI 기본법 및 하위 규정에 따른 결과물 표시의무

- ▶ 인공지능 기본법 제31조 (투명성 확보 의무)
 - 인공지능사업자는 고영향 또는 생성형 인공지능 기반 제품·서비스 제공 시 이용자에게 사전 고지를 해야 하고(제1항), 생성형 인공지능의 결과물임을 표시해야 하며(제2항), 나아가 실제와 구분하기 어려운 음향·이미지·영상 등 이른바 딥페이크 결과물의 경우 이용자가 명확히 인식할 수 있는 방식으로 고지·표시해야 함.
- ▶ 인공지능 기본법 시행령 제22조 제2항(결과물 표시의무)
 - 이미지, 음성 등이 생성형 AI를 통해 작성되었음을 명확히 하여 이용자들의 혼란을 예방하고, 결과물 악

용을 방지하기 위함

- 인공지능사업자는 생성형 인공지능 또는 이를 이용한 제품 등을 제공하는 경우 그 결과물에 사람 또는 기계가 판독할 수 있는 형식으로 표시할 수 있으며, 시각·청각적으로 직접 표시하는 방식과 SW를 통해 확인 가능하도록 비가시적으로 표시하는 방법 등이 있음

[인공지능 투명성 확보의무(제31조 제2항) 가이드라인 주요내용]

[이미지 콘텐츠]

- 로고나 텍스트 형태의 가시적 워터마크 삽입
- 전용 기술을 활용한 비가시적 워터마크 적용

[동영상 콘텐츠]

- 이미지와 동일하게 가시적 워터마크(로고, 텍스트 등)와 비가시적 워터마크 적용 가능
- 반드시 전체 영상에 표시할 필요는 없으며, 이용자가 생성형 AI 결과물임을 인식할 수 있는 수준이면 충분

[오디오 콘텐츠]

- 콘텐츠 재생 초반에 음성 안내를 통해 AI 생성물임을 고지
- 비가청 워터마크 기술 적용 가능
- 동영상과 마찬가지로 전체 구간이 아닌, 이용자가 인식 가능한 수준에서 표시하면 충분

[텍스트 콘텐츠]

- 챗봇 등은 대화창·메시지 화면에서 AI 생성물임을 표시
- 텍스트 결과물에는 비가시적 워터마크 부착 기술이 개발 중

해외 주요국의 AI 결과물 표시의무 규제 비교

[EU] AI Act Article 50 (투명성 의무)

▶ 표시의무 주체

- 제공자(Providers): AI 시스템을 개발하거나 개발을 의뢰하여 시장에 출시하는 자. 시스템의 설계 단계에서부터 투명성 의무가 확보되도록 해야 함.
- 배포자/운영자(Deployers): AI 시스템을 자신의 책임 하에 사용하는 자. 최종 사용자에게 고지·표시의무가 부여됨.

▶ 의무 내용 상세

- **챗봇·대화형 시스템**: 사람과 직접 상호작용하도록 설계된 AI 시스템은 사용자가 AI와 대화 중임을 명확히 인지할 수 있도록 “You are chatting with an AI assistant”와 같은 문구, 팝업, 화면 상단 고정 표시 등을 제공해야 하며, 다만 합리적이고 관찰력 있는 사용자가 명백히 인지할 수 있는 경우에는 예외가 적용 됨
- **합성 콘텐츠 생성 AI**: 텍스트, 오디오, 이미지, 비디오 등 합성 콘텐츠를 생성하는 경우 출력물에 AI 생성 또는 조작 사실을 기계 판독 가능한 형식(machine-readable format)으로 표시해야 하며, 메타데이터, 비가시적 디지털 워터마크, C2PA 표준 등이 활용되나 입력 데이터를 실질적으로 변경하지 않은 경우는 표시 대상에서 제외됨

- **딥페이크**: 실제 인물·사물·장소·사건과 유사하여 사실로 오인될 수 있는 이미지, 오디오, 비디오를 생성·조작한 경우 콘텐츠가 인위적으로 제작되었음을 가시적 워터마크, “AI-generated content” 라벨, 재생 전 경고 메시지 등을 통해 명시적으로 공개해야 하며, 다만 법 집행 목적이나 예술·창작·풍자·허구 등 명백히 예술적 맥락의 경우는 예외가 인정됨
- **공공정보 텍스트**: 뉴스 기사나 공공 이익과 관련된 정보 제공 목적의 텍스트는 AI로 생성·조작되었음을 반드시 공개해야 함

▶ 적용 범위 및 집행

- EU 시장에서 AI 시스템을 제공/배포 사용하는 모든 주체에 적용되며, EU 외 기업들도 EU 시장 준수시 필수로 해당 규정을 지켜야 함
- 투명성 의무를 위반할 경우, 최대 1,500만 유로 또는 전 세계 연매출 3% 중 더 높은 금액이 벌금으로 부과됨
- 2026년 8월 2일부터 집행되며, 고위험 AI는 2027년부터 적용됨

[미국] California SB-942 (AI 투명성법)

▶ 표시의무 주체

- Covered Provider : 생성형 AI 시스템을 생성·코딩·제작하는 자로, 월 100만 명 이상의 방문자 또는 사용자를 보유하고 캘리포니아 내에서 공개적으로 접근 가능한 서비스를 제공하는 자를 말하며, 여기에는 대규모 AI 모델 개발사(OpenAI, Google, Anthropic 등), 이미지·비디오·오디오 생성 AI 제공자, 그리고 캘리포니아에서 서비스를 운영하는 해외 기업이 모두 포함됨

▶ 의무 내용 상세

- **AI 탐지도구 제공 의무**: 업로드되거나 링크된 이미지·비디오·오디오 콘텐츠가 해당 covered provider의 생성형 AI 시스템에 의해 생성·변경되었는지를 판단할 수 있는 무료 AI 탐지도구를 사용자에게 제공해야 하며, 이 도구는 공개 접근이 가능해야 하고 covered provider는 사용자 피드백을 수집·활용할 의무가 있음
- **Latent Disclosure(잠재적 공개)**: 육안으로 보이지 않는 비가시적 표시로서 기계만 판독 가능한 임베디드 정보를 의미하며, 모든 AI 생성 이미지·비디오·오디오에는 기술적으로 실현 가능하고 합리적인 범위 내에서 반드시 포함되어야 하고, 여기에는 covered provider의 이름, AI 시스템 정보(사용된 모델 등), 생성 시간과 날짜, 고유 식별자(unique identifier), 콘텐츠 출처(provenance) 정보가 포함되며, 이 정보는 영구적이거나 제거가 극도로 어렵게 만들어져야 하고 탐지도구를 통해 탐지 가능해야 함
- * Covered Provider가 생성형 AI 시스템을 제3자에게 라이선스할 경우 계약상 Latent Disclosure 기능을 반드시 유지하도록 규정
- **Manifest Disclosure(명시적 공개)**: 육안으로 명백히 보이는 표시를 의미하며, 이미지·비디오·오디오 생성 및 변경 시 포함 여부를 사용자에게 선택권으로 제공하되, AI 생성 콘텐츠임을 명확하고 눈에 띄게(clear and conspicuous), 매체에 적합한 방식(appropriate for the medium), 합리적인 사람이 이해할 수 있는 형태로 표시해야 하고, 이 표시는 영구적이거나 제거가 극도로 어렵게 만들어져야 함

▶ 적용 범위 및 집행

- Covered Provider가 해당 규정을 위반할 경우 위반 건당 5,000달러의 벌금이 부과되며, 위반은 일별로 별도 위반으로 처리되어 누적 벌금이 산정
- 월 100만 사용자 미만의 서비스, 텍스트 생성 AI, 그리고 사용자가 직접 콘텐츠를 생성하지 않는 비사용자 생성 콘텐츠(스트리밍, TV, 게임 등)는 해당 규정 적용에서 제외

[일본] AI 사업자 가이드라인 1.0

▶ 표시의무 주체

- AI 개발자(Developer): AI 시스템을 설계·개발하는 조직이나 개인으로, 시스템 단계에서부터 투명성 확보 기능을 구현할 책임이 있음
- AI 제공자(Provider): AI 시스템이나 서비스를 시장에 제공하는 자로, 이용자에게 AI 활용 사실을 명확히 개시할 책임이 있음
- AI 사용자(User): 업무 목적 등으로 AI를 활용하는 자로, 최종 이용자에게 적절한 정보를 제공할 책임이 있음

▶ 의무 내용 상세

- 설명 가능성 확보: AI 시스템은 입력과 출력 및 동작 방식을 이해할 수 있도록 설계되어야 하며, 알고리즘의 작동 원리를 투명하게 공개하고 의사결정 과정을 추적할 수 있어야 한다.
- AI 활용 사실 개시: 서비스 제공자는 이용자에게 AI를 활용하고 있다는 사실을 적절히 고지하고, 서비스나 제품 설명에 이를 명시해야 한다.
- AI 산출물 식별: AI가 생성한 콘텐츠는 사람이 식별할 수 있도록 표시되어야 하며, 이를 위해 워터마크 등 기술적 수단을 개발·적용하는 것이 권장된다.
- 데이터 출처 투명성: AI 학습에 사용된 데이터의 출처와 수집·관리 절차는 명확히 공개되어야 한다.
- AI 리터러시 향상: 이용자에게 교육과 정보를 제공하고, AI의 한계와 위험에 대해 명확히 안내함으로써 AI 리터러시를 높여야 한다.
- 워터마크 기술의 개발·배포, AI 생성물 탐지 메커니즘 마련, 출처 추적 시스템 구축, C2PA 등 국제 표준 참여 등이 권장됨

▶ 적용 범위 및 집행

- 명시적인 제재는 없으며, AI 관련 분쟁 발생 시 법원이 가이드라인 준수 여부를 기준으로 과실 판단을 하며, 손해배상 책임 인정에 영향
- 일본은 AI 전용 규제법을 제정하지 않고 기존 법률(저작권법, 개인정보보호법, 부정경쟁방지법 등)을 통해 규제하고 있기 때문에, 위반 시에는 해당 법률에 따른 제재가 부과

이행 준비와 직면한 한계

- ▶ 기업 유형에는 차이가 있으나, 모든 AI 서비스 제공자는 표시 체계를 구축해야 하며, 규제 리스크를 최소화하기 위한 준비가 필요
 - 대중에게 노출되는 이미지, 영상 등을 만드는데 사용되는 AI 서비스를 제공하는 기업은 사전 고지 및 표시의무를 필수적으로 짐
 - 경계선상에 있는 활용 사례에 대해서는 기업의 리스크가 한층 가중될 수 있으므로 사전적 대비 뿐만 아니라 사후적 분쟁 가능성을 고려한 거버넌스 설계가 불가피함
- ▶ 워터마크 등 결과물 표시 기술의 불완전성과 부분적·혼합적 생성물 같은 경계 사례는 인공지능 결과물 표시의무 이행을 어렵게 만드는 핵심 기술적 한계로 지적
 - AI 결과물 표시 기술은 콘텐츠 유형(텍스트·이미지·음악·영상 등)에 따라 각기 다른 기술이 요구되며, 워터마크 제거 기술 등 반(反)기술이 발전하고 있어 실효성에 대한 의문이 큰 상황
 - AI 생성물과 인간 창작물이 혼합된 경우처럼, 부분적 AI 활용이나 복합적 창작물이 늘어나면서 식별이 곤란한 사례가 늘어나고 있음
- ▶ 글로벌 비즈니스의 복잡성은 국가별 규제 차이와 가장 엄격한 규제가 사실상 국제 기준처럼 작동하는 현상으로 인해 심화되며, 이는 기업의 비용 부담과 혁신 제약으로 이어질 수 있음
 - 국가마다 규제 요구사항이 달라 기업의 컴플라이언스 비용은 지속적으로 증가할 전망(법무법인 화우, 2025)
 - 일부 예외 규정이 기업 규모나 서비스 유형에 따라 인정되기도 하지만, 다국적 서비스를 운영하는 기업은 결국 가장 엄격한 규제를 기준으로 삼아야 하는 ‘브뤼셀 효과’의 영향을 받게 될 것이며, 이러한 환경은 신규 서비스 개발이나 혁신적 비즈니스 모델을 구현하는 데 추가적인 제약과 부담으로 작용할 가능성이 큼

참고문헌

- 국회. (2024). 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법.
- 광장 법률사무소 Tech & AI팀. (2025, 9). AI 기본법 시행령 등 하위법령 제정방향 공개 [PDF]. Leeko 뉴스레터. Retrieved from <https://www.leeko.com/news/techai/202509/202509.pdf>
- 김현경. (2025). AI 기본법의 발전적 시행을 위한 제언. 스타트업얼라이언스 이슈페이퍼.
- 법무법인 화우. (2025, 9월 9일). 인공지능기본법 하위법령 제정 방향 공개. 화우 인사이트. https://www.hwawoo.com/kor/insights/newsletter/13226?currentPage=1&utm_source=chatgpt.com
- 일본 경제산업성·총무성. (2024). AI 사업자 가이드라인 Ver.1.0. 도쿄: 일본 정부. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf
- California State Legislature. (2024). SB 942 — California AI Transparency Act. Sacramento, CA: California Legislative Information. Retrieved from https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942
- European Parliament. (2024). Regulation (EU) 2024/1689 — Regulation on Artificial Intelligence. Official Journal of the European Union.

부록: 해외 주요국 및 국내 AI 결과물 표시의무 규제 비교

구분	한국	EU	미국(캘리포니아)	일본
규제 성격	법적 강제 규정	법적 강제 규정	법적 강제 규정	자율 가이드라인 (법적 구속력 제한)
법적 근거	• 인공지능 기본법 제31조 • 시행령 제22조 제2항	• AI Act Article 50	• California SB-942	• AI 사업자 가이드라인 1.0
의무 주체	• 인공지능사업자 (고영향 또는 생성형AI 기반 제품·서비스 제공자)	• 제공자(Providers): AI 시스템 개발·출시자 • 배포자/운영자(Deployers): AI 시스템 사용자	• Covered Provider - 월 100만 명 이상 사용자 보유 - 캘리포니아 내 공개 서비스 제공자 - 대규모 AI 모델 개발사 포함	• AI 개발자(Developer) • AI 제공자(Provider) • AI 사용자(User)
핵심 의무	• 사전 고지 • 결과물 표시 • 딥페이크 강화 표시	• 챗봇 대화 중 고지 • 합성 콘텐츠 표시 • 딥페이크 명시적 공개 • 공공정보 텍스트 공개	• AI 탐지 도구 제공 • Latent Disclosure • Manifest Disclosure	• 설명 가능성 확보 • AI 활용 사실 개시 • AI 산출물 식별 • 데이터 출처 투명성 • AI 리터러시 향상
챗봇·대화형 시스템	• 대화창·메시지 화면에 AI 생성 표시	• "You are chatting with an AI assistant" 문구 • 팝업/화면 상단 고정 표시 • 예외: 합리적 사용자가 명백히 인지 가능한 경우	(규정 없음)	• 이용자에게 AI 활용 사실 적절히 고지 • 서비스·제품 설명에 명시
이미지 콘텐츠	• 가시적 워터마크(로고, 텍스트) • 비가시적 워터마크	• 기계 판독 가능 형식 표시 • 메타데이터 • 비가시적 디지털 워터마크 • C2PA 표준	• Latent Disclosure - 비가시적 기계 판독 정보 - 제공자명, 모델 정보, 생성 시간, 고유 ID, 출처 정보 • Manifest Disclosure - 육안 식별 가능 표시(사용자 선택)	• 워터마크 등 기술적 수단 권장 • 사람이 식별 가능하도록 표시
동영상 콘텐츠	• 가시적/비가시적 워터마크 • 전체 영상 표시 불필요(이용자 인식 가능 수준)	• 기계 판독 가능 형식 표시 • 메타데이터, 비가시적 워터마크 • C2PA 표준	• Latent Disclosure 필수 • Manifest Disclosure 선택 제공 • 영구적·제거 극난 방식	• 워터마크 권장 • 출처 추적 시스템 구축 권장
오디오 콘텐츠	• 재생 초반 음성 안내 • 비가시적 워터마크 • 전체 구간 표시 불필요	• 기계 판독 가능 형식 표시 • 비가시적 워터마크	• Latent Disclosure 필수 • Manifest Disclosure 선택 제공	• 워터마크 권장 • 탐지 메커니즘 마련 권장
텍스트 콘텐츠	• 대화창·메시지 화면 표시 • 비가시적 워터마크 기술 개발 중	• 공공정보·뉴스는 AI 생성 사실 필수 공개	• 텍스트 생성 AI는 적용 제외	• AI 활용 사실 고지 • 투명한 정보 제공 권장
딥페이크 강화 규정	• 실제와 구분 어려운 콘텐츠 • 명확히 인식 가능한 방식으로 표시	• 가시적 워터마크 • "AI-generated content" 라벨 • 재생 전 경고 메시지 • 예외: 법 집행, 예술·창작·기자·허구 목적	(합성 콘텐츠 규정에 포함)	• 명확한 고지 권장 • 사람이 식별 가능하도록 표시
표시 방법	• 사람 또는 기계 판독 가능 형식 • 시각·청각적 직접 표시 • SW 통한 비가시적 표시	• 기계 판독 가능 형식(machine-readable format) • 메타데이터, 워터마크, C2PA	• 이중 공개 체계 - Latent (비가시적, 필수) - Manifest (가시적, 선택)	• 워터마크 • 출처 추적 시스템 • C2PA 등 국제 표준 참여 권장
적용 제외	(명시 규정 없음)	• 입력 데이터 미변경 콘텐츠 • 명백한 예술 창작 목적	• 월100만 미만 사용자 서비스 • 텍스트 생성 AI • 비사용자 생성 콘텐츠	(가이드라인이므로 적용 제외 개념 없음)
제재	(시행령에서 추후 규정 예정)	• 최대 1,500만 유로 또는 (br/>) 전세계 연매출3% 중 高額	• 위반 건당 5,000달러 • 일별 누적 벌금	• 명시적 제재 없음 • AI 분쟁 시 법원의 과실 판단 기준 • 기존 법률(저작권법 등) 제재 적용
시행 시기	• 2026년 1월(예정)	• 2026년8월2일 • 고위험 AI는 2027년	• 2026년 1월(예정)	(현재 적용 중)
적용 범위	• 국내 인공지능사업자 • 고영향 또는 생성형 AI 제공자	• EU 시장 내 모든 제공·배포 사용 주체 • EU 외 기업도 EU 시장 진출 시 의무 준수	• 주 내 서비스 제공자 • 해외 기업 포함	• 일본 내 AI 사업자 전반(권고 사항)
특징	• 딥페이크 강화 규정 • 콘텐츠 유형별 구체적 가이드 • 이용자 인식 가능 수준 기준	• 가장 포괄적이고 엄격 • 높은 벌금 • 기계 판독 중심 • 공공정보 특별 규정	• 탐지도구 제공 의무 특징적 • 이중 공개 체계 • 라이선스 계약 규정	• 자율 규제 방식 • 권장 사항 중심 • 기존 법률 체계 활용 • AI 리터러시 강조