

EU

인공지능법 투명성 의무 실천 강령 최종본 공개

서강대학교 법학전문대학원/교수
박준우

1. 개요: 투명성 의무 정의, 종류, 목적과 근거 규정

EU 인공지능법의¹⁾ 인공지능 시스템 ‘투명성(transparency)’은 ①적절한 추적과 설명이 가능하도록(traceability and explainability) 개발·이용됨, ②인공지능 시스템과 소통·상호작용함을 인간이 알도록 함(make humans aware), ③인공지능 이용사업자(deployer)에게²⁾ 인공지능 시스템의 역량과 한계(capabilities and limitations) 관련 정보 제공함, ④영향 받는 자(affected persons)에게 그들의 권리에 대한 정보 제공함을 모두 뜻한다(법 전문 27). 이 법의 투명성 의무는 ㉠‘고위험’ 인공지능 시스템 투명성(법 제13조), ㉡‘범용’ 인공지능 시스템 투명성(법 제53조 제1항), ㉢‘고위험’과 ‘범용’을 포함한 ‘모든’ 인공지능 시스템에 적용되는 일반 투명성(법 제50조)으로 나뉜다. 2026년 6월 10일 EU 집행위원회는 법 제50조 제7항에 따라³⁾ ㉣일반 투명성 의무에 관한 ‘인공지능 생성 콘텐츠의 투명성에 관한 실천 강령’을 공개하였다(이하 ‘투명성 강령’).⁴⁾ 그리고 투명성 강령은 8월 2일부터 적용되는데, 투명성 강령 준수에 서명한 개발사업자와 이용사업자(Signatories. 이하 ‘서명 기관’) 명단은 7월 31일 공개되었다.⁵⁾

인공지능법 제50조는 인공지능 시스템의 개발사업자와⁶⁾ 이용사업자에게 일반 투명성 의무를 부과하였다. 제2항은 오디오, 이미지 또는 텍스트 콘텐츠를 합성하여 생성하는 인공지능 시스템(범용 인공지능 시스템 포함함.) 개발사업자에게 인공지능 생성·조작물임을 그 산출물에 기계 판독 가능 형식으로 표시하고 탐지할 수 있도록 보장할 의무를 규정하였다.⁷⁾ 제4항은 ㉤딥페이크를⁸⁾ 구성하는 이미지, 오디오 또는 비디오 콘텐츠를

1) 이하에서는 ‘EU’는 원칙적으로 생략하며, ‘법’ 또는 ‘인공지능법’은 ‘EU 인공지능법’을 뜻한다.

2) deployer: 자기 권한 하에 인공지능 시스템을 이용하는 자연인, 법인, 공공기관, 대행 기관 등인데, 개인적 비직업적 활동 과정에서 이용하는 경우는 제외함 (법 제3조 제4호, 전문 13). 한국 인공지능기본법의 ‘인공지능이용사업자(인공지능개발사업자가 제공한 인공지능을 이용하여 인공지능제품 또는 인공지능 서비스를 제공하는 자. 제2조 제7호 나목)’ 정의와 유사하므로, 이하 ‘이용사업자’로 번역한다.

3) 법 제50조 제7항은 ‘인공지능 사무국의 실천 강령 작성 권장 추진 의무’와 ‘집행위원회의 실천 강령 승인을 위한 이행 법률 채택 권한’을 정하였다. 또한 법 제3조 제47호는 ‘인공지능 사무국(AI Office)’을 ‘집행위원회 결정(2024. 1. 24.)에 규정된, 인공지능 시스템과 범용 인공지능 모델의 이행·모니터링·감독, 그리고 인공지능 거버넌스에 기여하는 집행위원회 담당 부서’라고 정의하고, EU 인공지능법에서 모든 ‘인공지능 사무국’ 언급은 ‘집행위원회’ 언급으로 해석해야 함을 정하였다.

4) <https://ec.europa.eu/newsroom/dae/redirection/document/129555>

5) <https://digital-strategy.ec.europa.eu/en/news/strong-backing-code-practice-transparency-ai-generated-content>

6) provider: 유상 또는 무상으로, 인공지능 시스템이나 범용 인공지능 모델을 ㉠개발하거나 ㉡개발하게 한 것을 자기 이름·상표 하에 출시 또는 서비스 제공하는 자연인, 법인, 공공기관, 대행 기관 등임(법 제3조 제3호). 한국 인공지능기본법의 ‘인공지능개발사업자(인공지능을 개발하여 제공하는 자. 제2조 제7호 가목)’ 정의와 유사하므로, 이하 ‘개발사업자’로 번역한다.

7) “Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.” 법

생성·조작하는 인공지능 시스템 이용사업자에게 그 콘텐츠가 인공지능으로 생성·조작되었음을 공개할 의무와 ㉔ 공익 관련 정보를 공중에게 제공할 목적으로 텍스트를 생성·조작하는 인공지능 시스템 이용자에게 그 텍스트가 인공지능 생성·조작물임을 공개할 의무를⁹⁾ 부과하였다. 또한 제5항은 ㉕명확하고 구별 가능한 방식으로 ㉔늦어도 첫 번째 상호작용 또는 노출까지는 ㉔관련된 자연인에게 위 정보들이 제공될 것을 규정하였다.¹⁰⁾

투명성 강령은 법 제50조 제2항 및 제5항(개발사업자 의무)과 법 제50조 제4항 및 제5항(이용사업자 의무)의 구체적 실행 방법을 각각 섹션 1과 섹션 2에 정하였고, 각 섹션은 목적(objectives), 전문(recitals), 실행(commitments), 조치(measures)로 구성되었다. 투명성 강령의 **궁극적 목적(overarching objectives)**은 법 제1조 제1항의 ‘㉕EU 시장(internal market) 기능 개선, ㉕인간 중심적이고 신뢰할 수 있는 인공지능 도입(uptake) 촉진, ㉕혁신 지원, ㉕EU 기본권 헌장에 명시된 건강·안전·기본권(민주주의·법치주의·환경보호 포함)을 인공지능의 해로운 효과로부터 높은 수준으로 보호함을 보장함’이다. 그리고, 위 목적 달성을 위한 **세부 목적(specific objectives)**을 다음과 같이 열거하였다.

- a) 투명성 강령은 법 제50조 제2항, 제4항과 제5항 의무 준수 증명(demonstrating)을 위한 지침 문서(guiding document) 역할을 함. 다만, 이 강령 준수가 의무 이행 최종 증거(conclusive evidence)는 아님.
- b) 생성형 인공지능 시스템 개발사업자의 투명성 의무 이행과 이용사업자의 의무 준수를 보장함.
- c) 위 의무 준수를 이 강령으로 증명하려는 개발사업자를 권한 있는 시장 감시 당국이¹¹⁾ EU 전역에 걸쳐 일관되고 예측 가능하며 통일된(consistent, predictable and uniform) 방식으로 평가할 수 있게 함.

투명성 강령에서 정한 요건은 그 성격에 따라 실체, 절차, 기술 요건으로 나눌 수 있는데, 다음에서는 실체 요건을 중심으로 소개한다. 그리고 투명성 강령은 서명 기관의 조치(measure) 수준을

제50조 제2항 1문.

8) deepfake: ㉕인공지능이 생성·조작한 이미지, 오디오 또는 비디오 콘텐츠로서, ㉕기존의 사람(persons), 물건, 장소, 단체 또는 사건과 유사하고, ㉔사람이 진정하게 또는 사실로 오인할 수 있는 것. 법 제3조(정의) 제60호.

9) “Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, shall disclose that the content has been artificially generated or manipulated.” 법 제50조 제4항 1단락 1문. “Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated.” 같은 조항 2단락 1문.

10) “The information referred to in paragraphs 1 to 4 shall be provided to the natural persons concerned in a clear and distinguishable manner at the latest at the time of the first interaction or exposure.” 법 제50조 제5항 1문.

11) market surveillance authority(시장 감시 당국): EU 규정 2019/1020[시장 감시준수]에 따라 활동 수행하고 조치를 취하는 회원국 당국(national authority)(법 제3조 제26호).

▪ national competent authority(권한 있는 회원국 당국): 통지 당국(notifying authority) 또는 시장 감시 당국 뜻함. EU 산하 기구, 대행 기구, 사무국 및 기관들이 서비스를 개시하거나 사용하는 인공지능 시스템 관련, 이 법에서 ‘권한 있는 회원국 당국’ 또는 ‘시장 감시 당국’은 ‘유럽 데이터 보호 감독관(European Data Protection Supervisor)을 언급함으로 해석해야 함(법 제3조 제48호).

‘의무(will)’, ‘권장(encouraged)’, ‘선택(may)’으로 구별하였는데, 이 글에서는 각 ‘해야 함’, ‘권장됨’, ‘할 수 있음’ 등으로 번역하였다.¹²⁾ 또한 ‘shall’과 ‘must’도 ‘해야 함’으로 번역하고, ‘commit’는 ‘약속’으로 번역하였다.

2. 개발사업자의 인공지능 생성·조작 콘텐츠 표시 및 탐지 의무

투명성 강령 섹션1 전문은 ㉔정보 생태계 신뢰, ㉔표시 기술 솔루션, ㉔(표시·탐지를 위한 실효성·호환성·견고성·신뢰성 있는 기술 솔루션 개발을 위한) 가치사슬 전반에 걸친 협력, ㉔표시 기술과 탐지 메커니즘 혁신, ㉔이해관계자와 협력, ㉔표준화 촉진, ㉔중소·중견 기업을 위한 비례성 원칙(proportionality)을 설명했다.

법 제50조 제2항은 표시·탐지를 위한 기술 솔루션이¹³⁾ ㉔ 실효성, ㉔ 호환성, ㉔ 견고성, ㉔ 신뢰성 등 요건을 충족할 것을 규정하였는데,¹⁴⁾ 투명성 강령은 이 요건들의 의미를 구체적으로 설명하였다.

1) Measure 3.1 : 실효성(effectiveness)

서명 기관은 다음과 같은 표시(marking) 및 탐지(detection)를 위한 기술 솔루션을 구현해야 한다: ㉔위 두 솔루션을 함께 적용했을 때, ㉔‘인공지능 생성·조작 콘텐츠를 자연인이 구별할 수 있게 함’이라는 ㉔목적에 부합하고 ㉔(실제로) 그렇게 할 수 있음으로써, ㉔정보 생태계의 신뢰성과 무결성에 기여함(contributing to the trust and integrity of the information ecosystem).

그리고 표시 및 탐지 솔루션의 실효성 기준과 조치는 다음과 같다:

㉔ 자연인이 탐지 결과에 접근하여 그 의미를 이해할 수 있어야 실효성 인정됨;

㉔ 실효성 판단 관련 정량적 평가 지표는 없음. 대신 탐지 기능에 대한 사용자 기반 평가 필요함.

서명 기관은 자신들의 탐지 솔루션이 출력하는 탐지 정보를 표시하는 방법 안내를 위한 사양서(specifications)를

12) 투명성 강령 조치는 구체적으로 다음과 같이 나뉨. 투명성 강령, 7, 28쪽.

- will: 서명 기관이 인공지능법 제50조 제2항, 제4항 및 제5항 준수를 위해 반드시 충족해야 하는 이 강령의 강제(mandatory) 조치임. 권한 있는 시장 감시 당국(competent market surveillance authorities)이 그 준수(compliance) 여부를 감시함.
- encouraged: 선택적(optional) 조치임. 법적 강제가 아닌 순수한 자발적(not legally required and are purely voluntary) 조치지만, 권장됨.
- may: 선택적 조치임. 법적 강제가 아닌 순수한 자발적 조치임. 인공지능 사업자에게 조치 및 약속(measure or commitment) 이행방식에 대한 재량(flexibility)을 부여함.

13) marking(표시): 디지털 서명된 메타데이터나 사람이 인식할 수 없는(imperceptible) 워터마킹과 같이, 인공지능 생성·조작 콘텐츠에 표시(marker)를 추가하거나 내장(embedding)함. 투명성 강령, 23쪽(Glossary).

- detection mechanism for a marking technique(표시 기술 탐지 메커니즘): 워터마크 탐지 또는 디지털 서명된 메타데이터 표시 검증을 위한 메커니즘. 위 워터마크와 메타데이터 표시는 개발사업자 또는 제3자(예. 모델 개발사업자)가 의도적으로(purposefully) 추가한 것임. 서로 다른 개발사업자들이 같은 탐지 메커니즘을 공유할 수 있는데, 주로 이들이 표준화된 표시 기술을 사용하거나 공유 탐지 솔루션을 공동으로 개발하는 경우임. 그러나, 일반적으로는, 표시 기술마다 고유한 특정 탐지 메커니즘이 있음. 예를 들면, 워터마킹 기술이 다르다면 탐지 메커니즘도 다름. 투명성 강령, 22~23쪽(Glossary).

14) “Providers shall ensure their technical solutions are effective, interoperable, robust and reliable [생략]” 제50조 제2항 2문.

다운스트림 이해관계자들(stakeholders)에게 발행하도록 권장된다.

서명 기관은 자연인들의 탐지 결과 이해 정도 확인을 위해, 자신들(서명 기관)의 애플리케이션 사용 사례와 관련 있는 범위 내에서, 관련 패널 조사(panel studies)를 수행할 수 있다.

2) Measure 3.2 : 신뢰성(reliability)

서명 기관은 표시와 탐지 솔루션이 (함께 적용될 때) 예상할 수 있는 다양한 맥락과 사용 환경에서도 높은 신뢰성을 가지도록 해야 한다. 다만, 기술적으로 가능한 범위 내에서 최신 기술 수준(state of the art)에 부합하는 방식으로 솔루션을 이행한다.

‘신뢰성’이란 ‘표시 및 탐지 솔루션이 인공지능 생성·조작 콘텐츠 출처를 다른 콘텐츠로부터 정확히 식별하고 구별할 수 있음(accurately identify and distinguish)’을 뜻하는데, 정확도 측정방법은 다음과 같다:

㉠ 서명 기관은 탐지 정확도(accuracy)를 표시 유형 관련 지표(metrics, 예. 탐지 오류율¹⁵⁾)를 사용하여 측정해야 함;

㉡ 오류율 낮음은 자사 표시를 가진 인공지능 콘텐츠와 다른 콘텐츠(타사 인공지능 시스템 생성·조작 콘텐츠 포함 가능.) 모두 포함하여, 다양한 샘플을 통해 증명해야 함;

㉢ 신뢰성 평가를 위한 데이터는 가능한 한 해당 인공지능 시스템의 학습 및 개발 과정에서 사용되지 않은 것이어야 함(shall).

서명 기관은 표시 및 탐지 솔루션이 다양한 맥락과 사용 환경에서도 작동함을 증명해야 한다. 이를 위해 인공지능 시스템의 생성 또는 조작이 의도된 다양한(various) 길이, 크기, 다양성(diversity), 의미의 콘텐츠를 대상으로 정확도를 측정하고 보고함을 보장한다.

3) Measure 3.3 : 견고성(robustness)

서명 기관은 (함께 적용되어) ㉠ 일반적 변형(common alterations) 및 ㉡ 적대적 공격(adversarial attacks)을 포함한 변화하는 조건에서도 의도된 작동 수준을 유지할 수 있는(maintain intended performance levels) 표시 및 탐지 솔루션을 구현해야 한다. 다만, 기술적으로 가능한 범위 내에서 최신 기술 수준(state of the art)에 부합하는 방식으로 솔루션을 이행한다. 그리고 견고성 요건은 Measure 1.1의 단일 계층의 메타데이터 표시만을 적용받는 인공지능 시스템에는 적용되지 않는다. 또한 서명 기관은 신뢰성 평가와 동일한 수행 지표를 이용하여 견고성을 평가한다.

15) error rate(탐지 오류율): 탐지 메커니즘이 일부 콘텐츠를 ‘인공지능 생성·조작된 것’으로 잘못 분류할 개연성과 관련 있음. 오탐율은 해석 개선을 위해 더 세분화될 수 있음. 예. 오탐율(False positive rate)은 표시 기술 및 이와 연동된 탐지 메커니즘에 대해 부여할 수 있는 신뢰도에 부정적 영향을 미칠 수 있음. 투명성 강령, 23쪽(Glossary).

① ‘일반적 변형’ 또는 ‘전형적 처리 작업(typical processing operation)’은 다음 콘텐츠 처리 작업을 포함해야 한다:

- a) In-place 수정.¹⁶⁾ 예) 노이즈 추가/제거, 필터링, (재)압축, 스크린샷/화면 녹화(screenshot/screen-casting¹⁷⁾), 음성 향상, 어휘 대체, 유사 문자 교체(homoglyphs), 파일 포맷 변경 등;
 - b) 비동기화 메커니즘(Desynchronisation mechanisms).¹⁸⁾ 예) 좌우 반전(mirroring), 잘라내기(cropping), 해상도 증감(up/downscaling), 회전, 화면 비율 변경(aspect ratio change), 음고 변경 (pitch shifting), 시간 늘리기(time stretching), 문자 삽입/삭제, 문장 재구성(paraphrasing), 번역 사이클(translation cycles) 등;
 - c) 아날로그 홀(Analogue hole).¹⁹⁾ 예) 인쇄 후 재스캔 및 광학 문자 인식(print-and-scan and optical character recognition), 오디오 재생 및 재녹음 (audio playback and recording), 화면 캡처 촬영 등
- 또한 서명 기관은 탐지 기술이 콘텐츠 부분 변경도 탐지할 정도의 견고성을 가지도록 보장해야 한다.

예) 사용자가 서명 기관의 탐지 솔루션에 콘텐츠 제출 전, 프라이버시 보호를 위해 얼굴을 흐리게 처리하는 경우.

② 적대적 공격에 대한 견고성 (adversarial robustness)

- 서명 기관은 ‘적대적 공격에 대한 견고성’이 표시에 대한 복사, 제거, 재생성, 변조 공격 등 악의적 행위(malicious behavior)에 대한 회복력(resilience)으로 평가될 수 있도록 보장해야 함.
- 서명 기관은 ‘악의적 행위에 대한 견고성’을 표시 및 탐지 솔루션의 표시 무결성 검증 능력(ability to verify marking integrity)으로 측정해야 함. 여기서 ‘표지 무결성 여부’는 ‘콘텐츠 출처 왜곡(misrepresent the origin)을 위한 표시 변조(tamper) 또는 수정(modified) 여부’를 뜻함.
- 서명 기관은 콘텐츠 유형, 표시 유형 및 콘텐츠 배포 및 확산(deployment and dissemination) 맥락에 기반하여 실제 발생 가능한 위협(plausible real-world threat)을 악의적 행위로 선정해야 함.
- 서명 기관은 ㉠자사의 표시 및 탐지 솔루션에 대한 악의적 사용과 공격을 예방하고 대응하기 위해, 호출 횟수 제한(rate limits)과 같은 표준적 사이버 보안 관행을 적용함과 ㉡위협 평가(threat assess-

16) in-place modification: 이미지 픽셀과 같은 디지털 샘플 값이 수정되는 콘텐츠 처리 작업(processing operation)의 한 카테고리. 다만, 그 위치는 수정되지 않음. 예를 들면, 디지털 이미지의 대비를 높임. 이 정의는 비동기화 메커니즘 정의와 함께(side-by-side) 고려되어야 함. 투명성 강령, 23쪽(Glossary).

17) screencasting: 화면(컴퓨터, 태블릿, 스마트폰)에 게시되는(displayed) 영상 산출물을, 종종 음성 해설과 함께, 디지털로 기록하고, 파일로 저장하거나 라이브로 스트리밍하는 과정. 투명성 강령, 24쪽(Glossary).

18) desynchronization mechanism(비동기화 메커니즘): 디지털 샘플의 위치(또는 그들의 값. value)가 수정되는 콘텐츠 처리 작업(processing operation)의 한 카테고리. 예를 들면, 디지털 이미지에 좌우 반전 효과(mirror effect)를 적용함. 이 정의는 in-place 수정 정의와 함께(side-by-side) 고려되어야 함. 투명성 강령, 22쪽(Glossary).

19) analogue hole: 디지털 콘텐츠를 재디지털화하기(re-digitised) 전에, 일반적으로 자연인에게 제시하기 위해 아날로그 형태로 변환(rendering)되는 콘텐츠 처리 작업(processing operation)의 한 카테고리. 예를 들면, 스크린에 상영되는 디지털 비디오 콘텐츠를 디지털 카메라로 녹화함. 투명성 강령, 22쪽(Glossary).

ment)를 자주 업데이트하여, 최신 위협 지형 변화에 대응함이 권장됨.

4) Measure 3.4 : 호환성(interoperability)

서명 기관이 이행하는 표시 및 탐지 솔루션은 ㉠기술적으로 가능한 범위 내에서, ㉡서로 다른 AI 시스템에 도입된 표시 기술과 관계없이, ㉢AI 생성·조작 콘텐츠를 탐지할 수 있도록 ㉣다수(multiple) 시스템, 행위자, 맥락 및 기술적 구현 전반에서(across) 원활하게(seamlessly) 작동해야 한다. 다만, 현재(투명성 강령 발행 시점) 관련 호환성 표준이나 모범 사례가(interoperability standards and/or best practices) 개발되지 않았으므로(디지털 서명된²⁰⁾ 메타데이터 제외), 호환성 요건을 단계별로 이행하여야 한다(staged implementation).

인공지능법 제50조 제2항 적용 시작 단계(initial stage)에서 서명 기관은 다음 의무를 부담해야 한다.

- (a) 확립된(established) 표준이나 모범 사례를 채택하여 메타데이터 표시 및 탐지를 실행함으로써, 누구나 탐지(open detection)할 수 있게 함.
- (b) 자기 인공지능 시스템에서 구현한 다른 유형의 표시 기술을 탐지하는 솔루션을 연동(integrate)하여 접근하는 방법에 관한 공개 정보(public information)를 EU 집행위원회(Commission)와 이해관계자들에게 제공함. 위 정보 확산을 위해 서명 기관은 ㉠자사 웹사이트에 관련 정보를 제공함과 ㉡원활한 연동을 위해 마련될 수 있는 공중 이용 가능한 레지스트리에 입력함(populate)이 권장됨.
- (c) ㉠서명 기관의 부담과 비용을 줄이면서도 ㉡최종 이용자(end-users)의²¹⁾ 원활한(seamless) 접근을 보장할 수 있는, 워터마크 탐지 솔루션들 사이의 최소 수준의 호환성을 위해 일함. 더 구체적으로 서명 기관은 2027년 2월 2일까지 자신의 탐지 메커니즘의 호환 솔루션에 다음 중 하나 이상을 구현해야 함. 다만, 호환 솔루션은 ㉠임베드된 다른 워터마크의 견고성을 훼손(detrimental)하지 않고, ㉡남용 위험 발생 방지를 위한 안전장치(safeguards)를 포함해야 함.

i) 공중 이용 가능하고 호환성 있는 업계 표준 접근 방식 (industry standard access method)

이 방식은 다음을 가능하게 해야 함: ㉠탐지 요청(queries)이 관련 탐지 메커니즘에 라우팅(전달)됨; ㉡탐지 결과는 다음 탐지 솔루션이 해석함: ㉢공유 또는 특정 개발사업자에게 종속되지 않은(shared or provider-agnostic) 솔루션; ㉣위와 같은 API(access method)를²²⁾ 구현하는 엔드포인트와 상호작용함; ㉤공개된

20) digital signature(디지털 서명): 인공지능 개발사업자 진위(authenticity) 검증과 서명된 콘텐츠 무결성(integrity) 검증을 확보(secure verification)할 수 있는 암호화(cryptographic) 서명. 투명성 강령, 23쪽(Glossary).

21) 최종 이용자(end-user): 인공지능 시스템 생성·조작 콘텐츠에 노출된(exposed to) 자연인(natural user). 투명성 강령, 23쪽(Glossary). 또한 '이용자(user)'는 법 제3조 제4항의 이용사업자 또는 개발사업자의 생성형 인공지능 시스템을 사용하는 자(person)를 뜻함. 위 강령, 24쪽.

사인포스트(public signposts)를²³⁾ 지원하지만 이를 반드시 요구하지는 않는 탐지 솔루션들을 포함함. 이 방식을 구현하는 서명 기관은 알려진 호환 솔루션(예: 사인포스트)을 직접 읽어내어 타사 서비스로의 불필요한 트래픽 발생을 피할 수 있으면, 불필요한 트래픽을 유발해서는 안 됨(shall not).

ii) 공중이 읽을 수 있는 사인포스트 (a publicly readable signpost)

위 사인포스트 기타 호환 메커니즘을 인공지능 생성·조작 콘텐츠에 구현하는데, 공중이 인공지능 시스템 개발사업자들의 모든 탐지 솔루션들을 모두 실행할 필요 없이, 어떤 탐지 솔루션을 사용해야 하는지 신호(signal)를 보냄. 서명 기관은 자사의 사인포스트 표시 기술 기타 호환 솔루션과 연계된 탐지 메커니즘을 EU 집행위원회(Commission) 및 관련 이해관계자들이 이용할 수 있도록 합리적인 조치(reasonable steps)를 함으로써, 자신들의 탐지 워크플로에 해당 기술을 통합(integrated)할 수 있게 해야 함.

iii) 서명 기관 컨소시엄의 개발사업자 공유 탐지 솔루션 (a shared provider detection solution by a consortium of Signatories)

이 탐지 솔루션은 다음 요건을 충족함: ㉠앞으로 가입할 다른 서명 기관(중소·중견기업 포함. SMEs and SMCs)도 접근할 수 있고 포용성 있음; ㉡기밀 기술 정보(confidential technical information) 보호를 위한 적절한 안전장치 마련함; ㉢이러한 통합(aggregated) 탐지 솔루션은 컨소시엄에 참여하는 특정 서명 기관에 종속되지 않아야 함(shall be agnostic); ㉣참여 기관의 생성형 인공지능 시스템이 생성·조작한 산출물을 탐지할 수 있음. ㉤다른 탐지 솔루션과 호환되는 방식으로 작동함.

iv) 서명 기관 사이의 기타 호환 솔루션으로, 위 솔루션들과 대등한(comparable) 수준인 것

(d) 완전한 호환성(full interoperability)으로 나아가기 위해, 서명 기관은,

㉠ 자원과 역량(resources and capacities) 허용 범위 내에서, 개선된 첨단(state of the art) 호환 솔루션 및 표준 개발을 위해, EU 집행위원회 및 관련 이해관계자들과 신의칙에 따라(in good faith) 협력해야 함.

㉡ 특히, 조치(Measure) 3.5조에 기술된 태스크포스 참여가 권장됨.

㉢ 본 강령 Section 1(생성형 인공지능 시스템 산출물의 표시·탐지)이 마련한 조치들을 실행할 수 있게 하는 데(operationalise) 도움 되는 개방형(open) 콘텐츠 표시 및 탐지 표준 개발에 초점을 맞춘 컨소시엄 이니셔티브(consortia initiatives)뿐만 아니라 국제 및 유럽 표준화 기구에 가입 및 지원함이 권장됨.

이러한 후속 단계에서, 인공지능법 제50조 제2항의 호환성 의무를 계속 준수하기 위해, 서명 기관은 앞으로

22) API: Application Programming Interface. 인공지능 시스템 또는 다른 기타 소프트웨어 서비스에 대한 기계 이용 가능(machine-usable) 인터페이스. 투명성 강령, 22쪽(Glossary).

23) signpost solution: 어떤 (워터마크) 탐지 메커니즘을 사용해야 하는지 안내하기 위해, 콘텐츠 내에 공개적으로 드러난 표시 기술 및 이와 연계된 탐지 메커니즘임. 이는 메타데이터를 이용할 수 없을 때마다 각 콘텐츠를 생성형 인공지능 시스템 개발사업자들의 모든 탐지 메커니즘에 대하여 검증해야 하는 비효율을 피하고, 탐지 워크플로를 원활하게 하기 위한 호환 메커니즘임. 이는 예를 들어, 사람이 인지할 수 없지만 기계는 읽을 수 있는(imperceptible machine-readable) 표시로 구현될 수 있음. 투명성 강령, 24쪽(Glossary).

나타날 표준들을, 기술적으로 가능하고 자사 인공지능 시스템 적용 사례(application use case)에 필요한 범위에서, 시기적절하게 구현할 것을 약속하였다.

3. 이용사업자의 딥페이크 및 인공지능 생성·조작 공개 텍스트 라벨링 의무

법 제50조 제4항이 인공지능 시스템 이용사업자에게 의무로서 부과한 공개 내용은 ㉠ ‘딥페이크를 구성하는 이미지, 오디오 또는 비디오 콘텐츠가 인공지능 생성·조작물임’과 ㉡ ‘공중에게 제공되는 공익 정보 텍스트가 인공지능 생성·조작물임’이다. 다만, 두 의무 모두 범죄의 발견, 예방, 수사 또는 기소를 위해 법이 허용하는 경우에는 적용되지 않는다. 또한, ㉢ 딥페이크 콘텐츠가 명백하게 예술·창작·풍자·허구·비유적 작품·프로그램의 일부인 경우, 위 작품의 게시나 향유(display or enjoyment)를 저해하지 않는 적절한 방식으로 공개해야 한다. 그리고 ㉣ 공익 정보 텍스트가 ㉤인간의 검토 또는 편집상의 통제 과정을 거치고 ㉥자연인 또는 법인이 그 콘텐츠 발행에 대한 편집상의 책임을 가지는 경우에는, 공개 의무를 부담하지 않는다. 투명성 강령 섹션2 전문은 ㉦명확하고 구별 가능한 공개, ㉧일관되고 효과적인 이행, ㉨접근성, ㉩인식과 책임성, ㉪투명성과 혁신 생태계 장려, ㉫다른 EU 및 회원국 법에서의 추가 안전장치를 설명했다. 투명성 강령 섹션2 주요 내용은 다음과 같다(참고로 섹션 2에서 ‘딥페이크’, ‘콘텐츠’, ‘공개된 텍스트’ 등은 모두 ‘인공지능 생성·조작 콘텐츠’를 뜻한다.²⁴⁾).

1) Commitment 1: 딥페이크 및 공익 정보 텍스트 공개

딥페이크 및 공익 정보 텍스트 공개 관련 서명 기관 약속 등은 다음과 같다.

- 서명 기관은 법 제50조 제4항과 제5항 의무 이행을 위해, 딥페이크와 공익 정보 텍스트 기원이 인공지능임(artificial origin)이 일관되고 효과적으로 공개되도록 보장함을 약속한다.
- 서명 기관은 ㉠EU 아이콘(부록1) 또는 ㉡이에 상당하는(equivalent) 투명성 강령 디자인 사양과 배치 사양을 준수한 ㉢아이콘이나 ㉣라벨을 통해 공개함을 약속한다.
- 서명 기관은 아이콘 또는 이에 상당하는 라벨(이하 ‘아이콘·라벨’)을 적용 가능한 부문별(applicable sectoral) EU와 회원국 법률 또는 업계 강령의 기존 공개 관행(practices)에 통합할 수 있다. 다만, 위 관행이 투명성 강령 섹션 2와 법 제50조 제4항과 제5항을 준수하는 범위 내에서 통합할 수 있다.
- 서명 기관은 라벨링이 제3자 권리와 자유 보호에 관한 EU 및 회원국 법에서 면제되지 않음을 인정한다(recognise). 위 보호는 묘사된 사람(depicted person)이나 권리자 동의에 적용되는 법적 요건을

24) 투명성 강령, 28쪽.

포함한다.

투명성 강령은 아이콘·라벨 디자인·배치 사양에 관한 서명 기관의 의무 등을, 시각적 공개(visual disclosure)가 가능한 경우와 불가능한 경우로 나누어 정하였다.

(1) Measure 1.1: 디자인 사양

① 시각적 공개 가능한 경우



[그림] 기본 EU 아이콘. 투명성 강령, 부록1, Figure 3, 38쪽.

- a) 의무 사항: 영어 대문자 약어 “AI”를 아이콘·라벨의 주된 시각적 요소로서 구성해야 함. 다만, 영어 사용이 회원국법과 상충하면 자국어로 공개할 수 있음. 약어를 구성하는 각 글자 높이가 같아야 하며, 크기 조정할 때 글자 비율 유지해야 함.
- b) 권장 사항: ㉠인공지능 생성·조작되었다는 정보를 ㉡기술적으로 가능한 경우 (상호작용하는) 2차 레이어에 또는 ㉢아이콘 안이나 옆에 보완함(예. ‘수정됨’ 또는 ‘생성됨’을 나타내는 텍스트). ㉣인공지능 시스템 수정 부분을 (상호작용하는) 2차 레이어에 공개함(예. 얼굴 변형을 기술하는 텍스트나 픽토그램).



[그림] EU 아이콘-인공지능 완전 생성 콘텐츠.

투명성 강령, 부록1, Figure 1, 37쪽.



[그림] EU 아이콘-인공지능 부분 수정 콘텐츠.

투명성 강령, 부록1, Figure 2, 37쪽.

- c) 선택 사항: 아이콘·라벨은 ㉠자연인에게 명확하고 구별 가능하게 공개됨을 서명 기관이 보장하는 한, 맥락에 맞는 다른 크기로 표시할 수 있으며, ㉣명확하고, 접근 가능하며, 구별 가능함, 즉 자연인이 읽고 인식할 수 있음(readable and recognisable)을 유지하는 한, 다른 스타일(예. 대비율, 색상 또는 서체)로 표시(express)할 수 있음.

② 시각적 공개 불가능한 경우 (예. 오디오만으로 된 콘텐츠)

- a) 의무 사항-면책 고지: 오디오 딥페이크의 기원이 인공지능임을 공개하는 짧은 음성 면책 고지(short audible disclaimer)를 ㉠딥페이크 자체의 시작 부분에 ㉣평이하고 간결한(plain and simple) 자연어로 ㉢콘텐츠와 같은 언어 또는 영어로 ㉡인지할 수 있는 방식으로

공개해야 함. 적절한 경우, 위 음성 면책 고지는 인공지능 생성·조작 여부 관련 정보로 보완되어야 하며, 인공지능 시스템 수정 부분 공개를 권장함. 다만, EU 전역에 적용되는 오디오 솔루션 개발 전에는 일정 조건 하에 다른 형태로 음성 공개할 수 있음.

b) 의무 사항-접근성: 서명 기관은 다양한 콘텐츠 유형(modalities)과 맥락에서도, EU 지침 2019/882(유럽 접근성 법)과 2016/2102(웹 접근성 지침) 등 EU법에 따라 공개에 접근할 수 있도록 보장해야 하며, 다음 조치를 포함해야 함:

- ㉠ 시각적 공개 요소에 대한 오디오 설명 또는 대체 신호;
- ㉡ 청각 장애 최종 이용자 필요를 고려한, 오디오만으로 된 콘텐츠에 대한 촉각 신호(tactile and haptic cues) (예: 오디오 재생 전에 딥페이크 콘텐츠 포함을 알리는 진동);
- ㉢ 고대비 아이콘·라벨 및 화면 낭독기(screen-reader) 호환성으로써, 색맹·색약 최종 이용자를 위한 것을 포함함;
- ㉣ 보조 기술(assistive technologies)로 아이콘·라벨 탐지 가능함(detectability).

c) 권장 사항-접근성: 서명 기관은 유럽 표준화 기구의 조화 표준(harmonised standard) ETSI EN 301 549 'ICT 제품 및 서비스를 위한 접근성 요건' 또는 W3C 웹 콘텐츠 접근성 가이드 2.1을 포함하여, 이용 가능한 모든 관련 접근성 표준이나 가이드를 구현함. 다만, 서명 기관 서비스나 제품이 관련 접근성 요건과 표준·지침 범위 안일 것.

d) 권장 사항-초대형 플랫폼/검색엔진: 초대형(very large) 온라인 플랫폼 또는 초대형 온라인 검색엔진인 서명 기관은 ㉠이용사업자가 투명성 강령 섹션 2 및 법 제50조 제4항을 준수하는 방식으로 라벨링 의무를 이행할 수 있도록, ㉢EU 아이콘이나 이에 상당하는 다른 준수 솔루션(equivalently compliant solution)을 도입함.(예. 이용사업자를 위해 플랫폼 업로드 인터페이스에서 공개 솔루션을 이용할 수 있게 함.)

(2) Measure 1.2: 배치 사양

투명성 강령은 배치 사양 관련 모든 경우에 준수해야 할 의무 사항을 일반 원칙으로 정하고, 추가 의무 사항을 시각적 공개가 가능한 경우와 아닌 경우로 나누어 정하였다.

㉠ Sub-measure 1.2.1: 일반 원칙 (overarching principles)

서명 기관은 다음 원칙을 모든 콘텐츠 유형에 적용해야 한다. 다만, 딥페이크가 예술·창작·풍자·허구·비유적 작품의 일부로서 별도 공개 기준이 적용되는 경우는 예외다(Commitment 3):

a) 콘텐츠 형식(format)과 유포 맥락을 고려하여, 이용자 (상호)작용이나 지속적인 주의(sustained attention) 필요 없이, 자연인이 즉시 인식할(recognition) 수 있도록,

적절하고 인지 가능한(perceivable) 방식으로, 아이콘·라벨을 배치하여야 함;

- b) 아이콘·라벨은 인지 가능성을 보장할 수 있도록, 적어도 통상적인 노출 조건에서 알아차리는데(noticed) 충분한 시간(duration) 동안 보여야 함(remain visible);
- c) 서명 기관들이 ㉠이 조치의 배치 사양을 준수하고, ㉡답페이지 또는 공익 정보 텍스트를 공개할 때 콘텐츠 배포 및 유통 체인(distribution and dissemination chain)을 고려함을 보장하기 위해 최선의 노력을 하는 한(as long as), 아이콘·라벨은 콘텐츠 내에 직접 임베드되어야 함. 다만, 직접 임베드된 아이콘에 상응하는 대체 방법(예. 자연인에게 콘텐츠 위에 있는 것으로 보이는 사용자 인터페이스 오버레이)을 이용할 수 있음;
- d) 아이콘·라벨은 늦어도 자연인이 답페이지 또는 공익 정보 텍스트에 최초 노출되는 시점(first exposure)에는 명확히 인지되고 구별될 수 있어야 함(clearly perceivable and distinguishable)(예. 다른 오버레이 요소, 면책 고지, 스크린/디스플레이 상의 청각/시각 요소들과 충분한 공간을 유지하고, 어떤 배경에서도 가시성 유지함.).

서명 기관은, ㉡답페이지·공익 정보 텍스트가 그 유통 전반(온라인/오프라인, 콘텐츠 작성, 배포·유통 체인 전반)에 걸쳐 공개 요건을 충족함, 즉 유통 전반에 걸쳐 공개가 수반됨(accompanies)을 보장할 수 있도록, ㉡답페이지·공익 정보 텍스트를 계속 배포·유통하는 데 이용하는 서비스·제품 공급자(actor. 예. 출판업자·온라인 플랫폼·소매업자)와 ㉢아이콘·라벨 보존(preserve)을 위해 최선의 노력을 다해 협력함이 권장된다. 이 권장 사항은 투명성 강령 섹션 2, 조치 3과 4에도 적용된다.

㉢ Sub-measure 1.2.2: 시각적 공개 가능한 경우

- a) 아이콘·라벨은 방해되는(intervening) 오버레이 요소 없는 적절한 위치에 보여야 한다(예. 이미지·영상 답페이지의 오른쪽 위 모서리).
- b) 서로 다른 자연인이 서로 다른 순간에 답페이지 영상에 노출될 수 있는 상황(예. 라이브 콘텐츠)과 하류 이용(예. 스크린샷과 클립)에 대응하기 위해, 서명 기관은 아이콘·라벨을 ㉣영상 시작 부분뿐만 아니라, 가능한 경우, ㉤영상 재생 동안 일정 간격(regular intervals throughout the video)으로, 그리고 ㉥최소한 중단(interruptions. 예. 상업 광고)이 발생한 후에는 표시(display)해야 한다. 맥락에 따라 가능하고 적절한 경우, 답페이지 부분(segment)이 영상 전체에 걸쳐 있지(span) 않아도, 서명 기관은 아이콘·라벨을 답페이지 영상 전체(throughout) 또는 영상의 답페이지 부분 재생 동안 표시(display)함이 권장된다.
- c) 직장 내부의 폐쇄적 맥락(closed internal professional context. 예. 직원 교육·정보 제공 목적)의 일환으로만 사용되는 답페이지: ㉦사용자 인터페이스, 물리적 환경(physical setting) 기타 답페이지에 노출되는 자연인이 쉽게 이용할 수 있는 적절한 매체에, ㉧자연인의 답페이지 노출 전에 정보 제공할 수

있도록, ©그 직장 환경 관련 자연인에게 명확하고 구별 가능하게(clear and distinguishable), 공개를 배치할 수 있다.

d) 시각적 딥페이크: 음성 공개(audible disclosures)는 오로지 ‘추가적 공개 수단’으로만 이용될 수 있으며(may only be implemented), 언제나 시각적 공개를 동반해야 한다(예. 오디오가 사용자 경험의 일부거나 접근성 목적상 적절한 경우).

e) 오디오 딥페이크: 스크린 이용 가능하면, 음성 면책 고지(audible disclaimer)에 추가하여, 아이콘·라벨에 기반한 시각적 공개가 이용 가능해야 한다(예. 차량 내 이용 가능한 스크린 또는 스마트폰 디스플레이.).

f) 공개 텍스트: 서명 기관은 최종 이용자가 명확하고 일관되게 구별할 수 있도록 아이콘·라벨을 배치해야 한다. 예. 텍스트 위나 꼭대기, 텍스트 헤드라인 주변 또는 텍스트 시작 부분의 판권란(colophon). 적절한 경우, 서명 기관은 인공지능 생성·조작 부분만을 라벨링할 수 있다. 짧은 텍스트(단어 하나 또는 간단한 구문) 산출물에 라벨링 하면 가독성과 유용성이 저하되므로(degrade), 서명 기관은 여전히 라벨링 해야 하지만(must) 사용자 인터페이스에 맥락에 맞게 공개함으로써 공개를 보장할 수 있다(예. 산출물 옆 표시(indicator) 또는 사용자 노출 세션 시작 부분이나 콘텐츠와 처음 상호작용할 때 인공지능 사용 공개함).

③ 시각적 공개 불가능한 경우 Sub-measure 1.2.3

a) 최초 고지 의무: 서명 기관은 늦어도 딥페이크에 최초로 노출되는 시점에는 음성 면책 고지(audible disclaimer)를 포함해야 한다. 인식 지원을 위한 음성 신호(audible cues to support recognition. 예. 신호음 또는 이어콘(earcon))을 추가할 수 있다.

b) 중간 고지 의무: 하류(downstream) 이용(예. 오디오 클립)과 롱폼(long-form) 또는 라이브 노출 대응을 위해, 서명 기관은, 적절한 경우, 콘텐츠 시작 부분의 최초 음성 면책 고지를 일정 간격의 알림(reminders at regular intervals. 예. 면책 고지 문구, 신호음 또는 이어콘)으로 보완해야 한다. 이 보완은 지속적 인지(ongoing awareness) 보장을 위해, 오디오 딥페이크 전체 재생(entire duration) 동안, 최소한 중단(interruptions. 예. 광고)이 발생한 후에는 이루어져야 한다.

2) Commitment 3: 예술·창작적 작품과 딥페이크 공개 제한

딥페이크가 명백히 예술·창작·풍자·허구·비유적 작품 또는 프로그램(evidently artistic, creative, satirical, fictional or analogous work or programmes)의 일부인 경우, 투명성 의무는 그 작품의 게시나 향유(display or enjoyment)를 저해하지 않는 적절한 방식의 공개(disclosure)로 제한된다(법 제50조 제4항 1단락 3문). 이는 그 작품의 통상적 이용과 사용(normal exploitation and use)이 방해받지 않음과 작품의 유용성(utility)과 품질이 유지됨을 포함한다.²⁵⁾

이를 위해 서명 기관은 다음을 이행해야 한다:²⁶⁾

a) 조치(Measure) 1.1의 디자인 사양을 따르는 아이콘·라벨 사용함. 그리고 이를 예술·창작·풍자·허구·비유적 작품 유형 및 그 작품이 제시되는(presented) 맥락에 적절한 방식으로 배치함(place);

b) 그러한 공개 및 배치(disclosure and placement)가 ㉠모든 자연인에게 명확식별 및 접근 가능함(clear, distinguishable, and accessible)과 ㉡늦어도 답페이지 포함 콘텐츠에 [자연인이] 최초로 노출되는 시점에는 제공됨(예. 사용자에게 함께 제공되는 안내나 설명, 처음/엔드 크레딧 등)을 보장함(ensure);

c) 아이콘·라벨을 통상적 관람 또는 노출 조건(normal viewing or exposure conditions) 하에서 쉽게 알아차려 인지 가능성을 보장할 수 있도록, 이를 충분한 기간 동안 인지 가능하게 함(perceivable for a sufficient duration)을 보장함.

또한, 투명성 강령은 답페이지 이용 방식을 ‘디지털 또는 상호작용 방식’과 아닌 것으로 나누어, 아이콘·라벨 공개 방식을 규정했다.

① 디지털 또는 상호작용 방식 (예. 웹사이트, 앱 등 사용자 인터페이스)

㉠서명 기관 통제하에, ㉢비디오·이미지 프레임에 인접한 외부, 또는 오디오 콘텐츠에 인접한 곳에 아이콘·라벨을 배치하고, ㉣사용자 인터페이스 요소나 오버레이로 통합할 수 있음.

-서명 기관은 최종 이용자가 맥락에 맞는 공개(contextual disclosure. 이하 ‘맥락 공개’) 솔루션을 별도 행위나²⁷⁾ 추가 조작(dedicated actions or additional engagement) 필요 없이 즉시 인지할 수 있도록(perceivable) 보장함을 약속함. 맥락 공개 예시: 사용자 인터페이스 일부(예: 웹사이트 또는 스마트 글래스 등 디지털 기기의 앱 인터페이스 일부), 짙은 안내문, 마우스를 클릭하거나 커서를 올리면(hovering) 추가 정보를 제공하는 방해 되지 않는(non-obtrusive) 아이콘·라벨, 콘텐츠에 인접하여 보이는 면책 고지(disclaimers) 등

② **비디지털 또는 비상호작용 방식** (예. 전시(exhibition), 미술 갤러리, 영화관, 축제 기타 유사 맥락, 물리적 매체(physical carrier)의 오디오 또는 비디오)

공개(disclosure)는 ㉠온라인 또는 물리적 입구나 판매처(physical point of entry or sale)에서 소개 또는 동반 정보의 일부(예: 전시회 리플릿 또는 입장권)로서, 또는 ㉡물리적 매체를 통해 제공되는 정보(예: 패키징) 형태로 제공될 수 있음.

25) 투명성 강령은 이를 서명 기관의 ‘약속(commit)’으로 기재했다.

26) 투명성 강령 1:2차 초안은 실시간 비디오, 비디오, 기타 멀티모달 콘텐츠, 이미지, 오디오 등으로 나누어 규정했으나(각 29쪽과 34쪽), 최종본은 위 분류를 하지 않았다.

27) 2차 초안은 ‘dedicated action’ 자리에 ‘scrolling’을 기재했다. 2차 초안, 34쪽.

3) Commitment 4: 공개된 공익 정보 텍스트에 대한 인간 검토·편집 통제

법 제50조 제4항 4문은 ‘공중에게 제공되는 공익 정보 텍스트가 인공지능 생성·조작물임’을 공개할 의무를 인공지능 시스템 이용사업자에게 부과하였다. 이 의무에도 딥페이크 콘텐츠 공개 의무와 같이 법 제50조 제5항이 적용되어 명확하고, 구별가능하며, 늦어도 첫 번째 상호작용 또는 노출할 때 실행되어야 한다. 다만, 범죄의 탐지, 예방, 수사 또는 기소를 위해 법이 허용하는 경우에는 적용되지 않는다(2단락 2문 전단). 그리고, 공익 정보 텍스트가 ㉠인간의 검토 또는 편집상 통제 과정을 거치고 ㉡자연인 또는 법인이 그 콘텐츠의 발행에 대한 편집상의 책임을 가지는 경우에는, 공개 의무를 부담하지 않는다(2단락 2문 후단). 투명성 강령은 서명 기관이 언론·방송사인 경우와 아닌 경우로 구별하여, 위 예외 적용을 위한 내부 절차 수립과 최소한의 문서 유지를 정하였다.

① 서명 기관이 미디어 서비스 제공자인 경우

㉠서명 기관이 EU 규정 2024/1083 제2조 제2항에 따른 미디어 서비스 제공자고, ㉡EU와 회원국 미디어 법체계(framework)에 따라 미디어 서비스 제공자에게 적용되는 대로, 편집 기준(editorial standards) 적용과 규제공동 규제-자율 규제 체계 적용을 이미 받고 있으면, 이 강령 확약 사항 준수를 위해 기존 검토·편집 절차와 확립된 업계 기준(established professional standards)을 적용함으로써 법 제50조 제4항 2단락 예외를 인정받을 수 있다.

② 서명 기관이 非미디어 서비스 제공자인 경우

서명 기관은 인간 검토·편집 통제를 위한 적절한 정책을 수립·조정·유지함을 약속한다. 인간 검토·편집은 텍스트 공개(publication) 전에 이루어지며, 자연인 또는 법인이 편집상 책임을 가진다. 이 내부 정책은 ㉠기존 프로세스를 이용할 수 있고, ㉡인공지능 이용사업자 규모와 자원에 비례해야 하며, ㉢최소한 다음 요소를 포함해야 한다:

a) 편집 책임 있는 자연인 또는 법인의 신원 정보(identification. 이름, 역할 그리고 상세 연락처 등.);

b) 인적 자원뿐만 아니라 구체적인 조직상 조치에 관한 개요(overview). 위 인적 자원과 조치는 텍스트 공개 전에 적당한(adequate) 자연인 검토 또는 편집상 통제가 실행되고 편집 책임을 가지도록 보장하기 위해 할당된 것임. 이는 개별 텍스트 공개에 대한 개별 인간 검토·편집 통제까지 문서화해야 함을 뜻하는 것 아님.

서명 기관은 책임성(accountability) 보장을 위해 편집 책임 부서(function), 자연인 또는 법인 연락처 정보를 공개함을 약속한다. 서명 기관은 검토 성격이나 공개된 텍스트에 인공지능 시스템이 관여된 유형에 대한 추가 정보를 기록할 수 있다. 이 확약 사항 이행은 미디어 자유, 편집 독립성, 그리고 언론의 취재원 정보 보호에 어떠한 영향도 미쳐서는 안 된다(shall).

4. 결론 및 시사점

지금까지 EU 인공지능법 제50조 제2항, 제4항, 제5항이 정한 투명성 의무 실천을 위한 투명성 강령의 주요 내용을 실제 규정 중심으로 소개하였다. 투명성 강령에는 이밖에 이용사업자와 이용자들이 준수하고 협력해야 할 기술적, 절차적 요건, 그리고 장애인을 위한 요건도 함께 정하고 있다. 최종본은 1차 및 2차 초안과 비교하여 상당히 구체화 되었고 체계 변화도 많았다. 저작권법 측면에서는, 인공지능 생성·조작물의 표시·탐지와 관련하여, 저작물 유통 및 편집 과정에서 정보의 왜곡이나 손실이 일어나지 않도록 하는 견고성(reliability) 보장, 이력 정보(provenance information)²⁸⁾ 관리와 기존 저작물 분류 표시와의 호환성 확보가 중요한 의미를 가진다.

참고 자료

- Code of Practice on Transparency of AI-Generated Content. <https://ec.europa.eu/newsroom/dae/redirection/document/129555>

28) provenance information(이력 정보): 인공지능 시스템 생성·조작 콘텐츠의 기원(origin), 생성·조작 방법과 시점, 그리고 처리 과정을 보여주는 디지털 기록. 투명성 강령, 24쪽(Glossary). 다만, 이력 정보 기록과 제공은 아직 의무가 아닌 권장 사항이다. 투명성 강령, Measure 1.3.