

## AI 모델의 학습, 개발과 관련된 최근의 법적 쟁점들

EU는 2024년에 제정한 AI법에서 'AI'라는 단어가 아닌 'AI시스템'이라는 용어를 사용했고, 집행위원회 (Commission)는 엔지니어들을 위해 'AI시스템' 정의를 담은 가이드라인을 2025년 2월에 발표함으로써 역시 2월부터 금지되어지는 고위험 AI 활용영역에 대한 가이드라인 발표와 함께 EU AI법이 정해진 시간표에 따라 적용범위를 확대되고 있음을 잘 보여주고 있습니다.

범용 (general-purpose) AI모델이 AI법 제 5장에서 특별히 규율되고 있음을 AI시스템 가이드라인에서 밝힐 정도로 범용 AI모델은 AI법에서 주요한 규제 대상입니다. 범용 AI모델은 기술적으로는 광범위하게 데이터를 수집, 학습을 시키면서 개발, 개량되지만 모델 스스로도 계속 학습, 반응하면서 환각 (hallucination)이라 불리는 오류도 생성함에 따라 EU AI법은 범용 AI모델 제공자에게 학습, 테스트등에 관한 기술적 문서, 학습시키는데 사용된 내용물의 상세 요약본 등을 작성, 공개하도록 함은 물론, 특별히 EU 저작권법 준수를 요구하고 있습니다.

구체적으로 학습된 내용물의 상세한 요약본을 정해진 양식에 따라 공개하도록 강제함으로써 저작권자가 AI 활용에 따른 자신의 권리를 지키는데 실질적 도움이 되게 하였고 나아가 EU AI법 전문 (Recital)은 범용 AI가 EU밖에서 만들어졌을지라도 해당 AI모델을 EU내에서 사용하려고 하면 저작권 문제는 EU 저작권법을 따라야 한다고 설명하고 있습니다.

### Related Areas

TMT  
인공지능

### Contact

구태언 변호사

T. 02-3477-8695  
E. [tekoo@law-lin.com](mailto:tekoo@law-lin.com)

방석호 미국변호사

T. 02-3477-8695  
E. [shbang@law-lin.com](mailto:shbang@law-lin.com)

설기석 변호사

T. 02-3477-8695  
E. [ksseol@law-lin.com](mailto:ksseol@law-lin.com)

이런 EU의 AI법 관련 규정이 2024년 8월에 발효됨에 따라 최근에는 프랑스 출판협회, 창작자 등이 미국 빅테크인 Meta가 자사의 범용 AI를 훈련시키는데 EU 저작권법에 따르지 않고 불법으로 저작물을 사용했다는 이유를 들어 2025년 3월, 프랑스 법원에 소송을 제기하기도 했습니다.

데이터의 수집 단계에서부터 웹사이트 중심의 (web scraping) 수집, 저장부터 자율주행차, 스마트홈 기기, 공공기관이나 민간의 데이터베이스 등을 통한 방법 등 AI 기술발전에 비례하여 다양한 데이터 수집, 처리 수단들이 사용될수록 저작권, 인격권침해 등의 법적 논쟁 또한 더 불거지게 되기에 EU AI법은 특별히 AI Office가 2025년 5월2일까지 범용 AI 제공자가 준수할 행동준칙 (Code of Practice)을 마련하도록 하는 (제56조) 등 학습 데이터의 투명성을 확보하고 모델에 대한 지속적 평가 등을 통해 법적 논란을 최대한 관리할 수 있도록 노력하고 있습니다.

OECD 또한 2025년 2월에 발간한 보고서 (Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data)에서 데이터의 수집은 물론, 수집된 데이터의 1차 가공처리, 이를 바탕으로 한 AI모델 학습, 개발 및 데이터 보정 등 결과물 출력전 과정을 '데이터 스크래핑' (기술적으로는 웹스크래핑은 물론 하이퍼링크를 통한 웹크롤링, 웹이 아닌 스크린상의 데이터를 대상으로 하는 스크린 스크래핑 등의 다양한 수단을 모두 통칭)으로 정의를 내리면서 법적 쟁점들을 상기시키고 있습니다.

이하에서는 AI법 적용대상인 'AI시스템'에 대한 상세한 정의, 또한 AI시스템 개발과 학습과정에서 일반적으로 사용되는 데이터 스크래핑, '문자 및 데이터 마이닝', 그리고 최근 제기되고 있는 지브리 효과 (Ghibli Effect)까지 관련 법적 이슈들을 살펴보고자 합니다.

## 1. EU AI법과 안면인식기술

### 1-1. AI시스템 정의

EU 집행위원회는 AI법 3조 ‘AI시스템’에 대한 구체적 가이드라인을 2025년 2월에 발표하면서 절대적 정의 기준은 없지만 현재의 기술수준 등을 고려, 판단을 위한 7가지 기준을 열거하고 있습니다. 특히 기계기반의 시스템 (machine-based system), 자율성 (autonomy), 입력데이터로부터 결과물을 어떻게 만들 것인가에 대한 추론 (inference), 그리고 인간의 관여 정도에 따라 달라지는 4가지 결과물, 즉 예측 (prediction), 내용물 (content), 추천 (recommendation) 및 의사결정 (decision)까지 EU AI법이 적용되기 위한 AI시스템은 4개가 필수요소임을 적시하고 있습니다.

구체적으로 AI시스템으로 정의되기 위한 첫 번째 필수요소인 ‘기계기반의 시스템’은 하드웨어적 기반에 해당되는 데이터처리, 메모리, 저장, 네트워크장비 등의 물리적 기계요소가 먼저 있어야만 되고 이것을 작동시키기 위한 컴퓨터코드, 운영시스템, 프로그램 등의 소프트웨어 요소가 결합되어야만 AI시스템이 된다는 점을 지적합니다. 따라서 인간이 관여하지 않아도 자동실행 되어지도록 프로그래밍 된 SW라 할지라도 SW만으로는 AI법 적용대상이 되지 않습니다. 2024년 12월에 통과된 우리의 인공지능기본법은 EU AI법과 비슷하게 인공지능시스템을 자율성, 적응성, 추론 등의 요소를 통해 정의하고 있지만 ‘기계기반의 시스템’에 국한됨을 법 자체가 명확히 하고 있지는 않습니다.

AI시스템 정의 가이드라인이 두 번째 필수요소로 적시한 ‘자율성’은 결과물 도출에도 연결되어지는 개념으로서 인간이 직간접으로 100% 통제하는 것을 배제하는 범주내에서의 탄력적 의미를 지니기 때문에 인간의 손을 떠난 완전 자율성도 AI시스템에 당연히 포함됩니다.

세 번째 요소인 ‘추론’ (inference)은 기계적 학습 (machine learning)기술에 기반하며, 신경망기술을

통한 심층학습기술을 통해 지원되며 데이터, 규칙, 모델 등을 기반으로 결과물을 도출해내는 추론과정 (reasoning)의 존재를 의미하기 때문에 자동 실행되는 응용 SW와 구별되는 AI시스템의 핵심 요소이기도 합니다. 데이터 기반의 기계적 학습 외에 환자의 증상에서 질병진단을 하게 만드는 '논리와 지식기반 학습'의 AI시스템도 AI법상 적용대상이 됩니다. 다만 수학, 과학 등에서 활용되어지고 있는 연산과 추론능력이 뛰어난 슈퍼컴퓨터는 인간이 사전에 만들어 놓은 프로그램 규칙으로 짜여진 '기본적 데이터처리' (basic data processing) 개념 범주를 넘지 못하고 있기 때문에 AI시스템으로 정의되어지지 못합니다. 마찬가지로 여론조사, 물건판매, 주식동향, 기후변화 등에서 기계적으로 학습된 다양한 데이터를 통계적으로 유의미하게 분석, 요약, 예측하여 주는 컴퓨터시스템은 기본적으로 데이터를 전달하는 수단을 향상시킨 것에 불과할 뿐 스스로 학습, 추론할 능력이 없기에 EU AI법 적용을 받는 AI시스템으로 분류되어지지 않습니다.

AI시스템으로 정의되기 위한 네 번째 필수요소로 들고 있는 4개의 결과물 중 첫번째인 예측 (prediction)은 단순히 과거의 데이터 학습을 통한 확률적 예측이 아니라 자율주행자동차에서 보듯이 다양한 환경변수에 반응하면서 복잡하면서도 새로운 의사결정을 돕는 것을 의미합니다. 둘째의 결과물인 내용물 (content)은 기술적으로는 예측과 결정과정의 연결선에 있지만 생성형 AI를 통해 텍스트, 음성, 영상, 이미지, 음악 등 다양한 창작물을 만들어내는 것이며 현재 상업적으로 가장 일반화된 AI 상품입니다. 세번째의 결과물인 추천 (recommendation)은 미리 정해진 규칙과 데이터 입력에 따라 작동하는 것이 아니라 이용자 요구에 맞춰 그때 그때 필요한 것, 예를 들면 기업에 적합한 채용후보자를 추천하여 주는 것과 같은 기능을 말합니다. 마지막의 결정 (decision) 결과물은 완전자율주행자동차의 예처럼 인간의 판단을 대체하여 AI시스템이 최종 결정을 내리는 것을 의미합니다.

결국 AI시스템을 특징지을 수 있는 '결과물' 요소는 데이터 학습과정인 기계적 학습, 논리와 지식기반 학습을 거쳤을지라도 AI법상 AI시스템으로 정의되기 위해서는 반드시 데이터안에서의 패턴, 복잡한 상호관련성, 추론 등을 다루고 수행할 수 있어야만 된다는 점을 지적하고 있습니다.

## 1-2. 고위험 금지와 안면인식 (Facial Recognition)

2020년 1월 18일자 뉴욕 타임즈 기사는 스타트업 회사인 Clearview AI가 스캐닝 된 30억개 이상의 이미지를 데이터베이스로 활용, 얼굴인식 앱을 개발하였고, 수분 안에 다양한 신상정보를 제공할 수 있는 안면인식 (FR)기술을 FBI, 경찰 등에 판매하고 있다고 보도함으로써 충격을 준 바 있습니다. 페이스북, 인스타, 유튜브 등의 소셜미디어나 글로벌 플랫폼에 올려진 수많은 개인 이미지 파일들이 본인도 모른 채 무차별 스크래핑, AI와 결합된 FR를 통해 치안유지, 범죄예방, 국가안보 등 광범위한 목적으로 제공되는 것이 과연 허용될 수 있는가?에 대한 논쟁을 촉발시켰습니다.

FR기술 자체는 정확성이 높고 신속하게 사람을 식별할 수 있어 AI와 결합, 공권력에 의한 치안유지, 테러 및 범죄예방, 세관 및 출입국 통제등의 공공용도는 물론이고 기계적 학습을 통해 디지털 접근 및 사이버보안, 자율주행, 금융거래, 건강진료 등의 상업적 용도로 이미 널리 활용되고 있지만 2023년 1월, 우리의 국가인권위원회는 불특정 다수를 대상으로 공공장소에서 ‘실시간 원격 FR’를 사용하는 것은 기본권 침해 위험성이 매우 크므로, 원칙적으로 금지하여야 한다고 강조한 바 있습니다.

EU AI법에서 2025년 2월부터 전면 금지하고 있는 고위험 영역 중 하나인 안면인식정보의 상업적 활용은 개인정보침해 등을 이유로 ‘인터넷이나 CCTV를 통해 무차별적으로 안면인식정보를 스크래핑, 데이터베이스를 구축, 확장하는 것’으로 국한되고 있습니다. 흔히 ‘지능형 CCTV’라고 불리우는 AI기반의 CCTV를 기반으로 데이터베이스를 구축, 활용하는 경우도 물론 해당됩니다.

개인이 인스타나 페이스북 등과 같은 소셜미디어 플랫폼에 스스로 올린 이미지 사진을 외부 기관이 사용했는지라도 상업적 용도의 스크래핑과 데이터 추출 및 활용을 동의한 것은 아니기 때문에 AI 고위험 활용영역에 속하게 되면 수집금지대상이 됩니다.

그러나 특정인을 식별할 수 있는 음성의 수집이나 특정인의 이미지 파일을 AI 시스템이 다운로드 받아 인터넷 검색을 통해 사후 (소위 'reverse engineering') 특정인을 식별하는 것과 같은 활용은 EU AI법에서 금지되지 않습니다. 또한 특정인 식별이 되지 않는 다양한 얼굴 이미지 생성을 위한 AI 모델 개발이나 테스트용 안면인식 데이터베이스 구축은 해석상 당연히 허용됩니다.

## 2. AI와 TDM

### 2-1. EU의 저작권법 규정

2019년에 개정된 EU 저작권법 (Directive 2019/790)은 인터넷상에서 패턴, 트렌드 등의 정보를 얻기 위한 디지털 '문자 및 데이터 마이닝' (text and data mining: 이하 'TDM'이라 칭함)에 대한 저작권보호 예외를 인정하되 다만 권리자가 명시적으로 자신의 권리를 유보하겠다고 (opt-out) 웹사이트나 기술적 수단 등의 적절한 방법을 통해 밝히지 않는 조건으로만 (제3조와 제4조) 인정하였습니다.

AI모델 등에 따라 데이터 학습 및 훈련 방법은 달라질 수 있고 TDM의 구체적 방법에서도 저작권 침해 시비를 피하기 위해 데이터 자체를 조금씩 긁어오기도 하는 등 차이는 있지만 TDM은 대규모 데이터를 수집하고 처리하면서 AI모델 훈련, 개발에 사용되는 주요 수단임에 틀림없으며, 더욱이 EU AI법이 범용 AI 제공자는 EU 저작권법의 opt-out규정을 준수하도록 규정하고 있음에 비추어 EU의 TDM 예외규정은 AI모델 학습 등을 둘러싼 저작권 침해 논쟁에서 주목받을 수밖에 없습니다.

TDM 기법은 대체로 디지털 텍스트와 데이터를 자동 분석해 주로 패턴, 추세, 그리고 상관관계 등 관련 디지털정보를 생성하는 기술로 정의되고 원시 데이터를 추출, 수집하는 행위에 초점을 맞추는 데이터 스크래핑보다는 훨씬 광의의 과정을 의미합니다. EU 저작권법은 저작권 자료를 직접 사용하게 되는 TDM방식의 기술행위 논란에 대해 2019년에 그 답을 제시한 것이고, 특히 제 4조의 TDM 예외규정은 제 3조의 단순한 과학 연구목적에 뛰어넘어 AI를 포함한 디지털, 빅데이터 관련

기술개발, 혁신을 촉진시키려는 큰 의도를 가지고 스크래핑된 데이터 자료를 넓게 활용하는 것에 대해 EU가 저작권자의 의사를 존중하면서 만든 타협물입니다. 따라서 상업적 개발자는 물론이고 이미지, 오디오 등 다양한 디지털데이터도 모두 포함, 침해예외가 인정됩니다.

일본은 2018년에 저작권법 개정을 통해 '정보 해석' 목적의 복제나 이용 시 저작권자의 허락 없이 상업적 목적으로 하더라도 TDM이 허용된다고 예외를 인정했고, 영국은 EU에서 탈퇴한 뒤인 2022년에 EU의 TDM방식을 AI 기술발전에 따라 도입하는 것을 검토했다가 결국 저작권자들의 반발로 중단, 올 2월부터 다시 검토하고 있지만 상업적 목적의 TDM이라 할지라도 저작권자를 보호하기 위한 라이선스 계약이 필요하다는 반발에 부딪쳐서 갈등을 겪고 있습니다.

한편, AI개발과 사용에 대한 포괄적 규범을 다루기 위한 G7회의가 2023년 10월에 데이터 스크래핑을 포함한 데이터 입력 등과 관련된 지적재산권 보호이슈에 대해 '자발적 행동강령'(voluntary code of conduct)을 발표하기도 했지만 AI 학습 및 훈련방법은 저작권법과 관련, 여전히 논쟁거리입니다.

EU의 방식은 TDM방식이 예외적으로 허용되도록 한 입법이긴 하지만 미국과 우리나라 저작권법상의 '공정한 사용'(fair use) 규정에 비해 저작권자가 자신의 저작물을 지키겠다는 의사를 표명하지 않은 경우에만 허용된다는 조건이 붙었음에도 사례별로 침해여부에 대한 사후 판정을 다룰 필요가 없는 분명한 이점도 있습니다.

## 2-2. EU 데이터베이스법 적용문제

EU는 1996년에 미국주도의 디지털데이터 시장형성에 대응하기 위해 데이터베이스 제작자의 투자에 따른 재산적 권리를 보장해주는 DB법 (Directive 96/9/EC)를 만들었고 우리도 이에 영향을 받아 데이터베이스 제작자의 재산적 권리를 2003년에 처음으로 저작권법에 삽입한 바 있습니다.

저작권법에 따라 편집저작물로 보호되는 데이터베이스에 TDM방식으로 접근, 데이터의 1차 추출 후 재생, 번역, 수정, 정리 등의 기술적 과정을 통해 소재의 선택과 배열의 창작성을 훼손할 수 있기에 해당 사업자는 EU 저작권법 위반의 책임을 지게 되고 아울러 EU DB법에 따른 데이터베이스 제작자의 권리 또한 침해할 수 있습니다. 이에 따라 EU 저작권법 제 4조는 침해예외를 규정하면서 동시에 TDM을 통해 데이터베이스 내용물을 합법적으로 재생, 추출할 수 있도록 DB법상의 예외 내지 제한 규정을 국내 법으로 만들도록 EU국가들에게 요구함으로써 DB법과의 충돌도 해결하고 있습니다.

반면, 우리 저작권법은 편집저작물인 데이터베이스에 대한 침해행위를 복제·배포·방송 또는 전송으로 열거하고 있기 때문에 TDM을 통한 데이터 추출 자체가 금지되는지에 대해서는 다툼이 있을 수 있습니다. 또한 데이터베이스의 개별 소재 (데이터)는 저작권법에 따른 보호대상이 아니기 때문에 당연히 데이터베이스 제작자의 재산적 권리가 미치지 못하게 됩니다.

그러나 EU 저작권법 제4조가 TDM의 상업적 활용에 대해서도 예외를 인정하는 것과 달리 우리의 저작권법은 교육, 학술, 연구, 시사보도인 경우에만 데이터베이스 제작자 권리 예외를 인정하고 있기 때문에 영리목적의 TDM을 통한 AI학습 및 개발은 배려하지 않고 있습니다.

또한 경쟁사업자가 TDM을 통해 디지털 식별 표시된 데이터베이스의 데이터 전부 또는 상당부분을 무단으로 복제, 배포, 방송, 전송행위까지 할 경우에는 현행 콘텐츠산업진흥법에 따른 법적 책임도 추가로 물을 수 있습니다. 다만 콘텐츠산업진흥법은 미국의 부정경쟁 (unfair competition) 법리에 따라

콘텐츠제작자의 투자재산으로 식별될 수 있는 것을 부정한 경쟁수단으로 왜곡, 침해한 책임을 묻고자 한 것이기에 우리 저작권법이 데이터베이스 제작자를 보호하기 위해 특별한 (sui generis) 재산권을 5년간 인정하고 갱신을 위한 상당한 투자 시 투자분에 대한 5년 보호를 다시 연장하는 것과는 달리 최초의 콘텐츠제작일로부터 5년간만 영업적 이익침해로부터 보호하도록 하고 있습니다.

### 2-3. 미국에서의 소송

미국에서 진행되고 있는 AI관련 저작권 침해소송 가운데 특히 주목을 받고 있는 NY Times v. Open AI, MS간 1심 소송에서도 Open AI가 챗 GPT를 훈련시키기 위해 사용한 TDM기법이 미국 저작권법상의 공정한 사용에 해당되는지에 대한 공방이 벌어지고 있습니다.

즉 NY Times 신문사는 Open AI가 웹크롤링, TDM 기법을 통해 자사 기사 전문을 그대로 복제하거나 요약하였기에 복제권이 침해되었다고 주장하였고 이에 대해 Open AI는 TDM 과정에서 데이터를 일시적으로 복제하는 것은 AI학습의 필수 단계이며, 더욱이 원본을 그대로 재사용하는 것이 아니고 통계적 패턴을 분석해 새로운 AI모델을 만드는 과정이기 때문에 저작권법상 허용되어지는 ‘변형적 사용’ (transformative use)에 해당된다고 반박을 하고 있습니다. 현재는 증거개시 (discovery) 절차 과정에서 NY Times가 AI 학습 데이터셋과 TDM 프로세스에 대한 상세한 정보를 요구하고 있고 이를 Open AI는 비공개 기술정보라고 거부하면서 소기각을 요청했지만 받아들여지지 않음에 따라 최종 판결이 나기까지는 시간이 많이 걸릴 것으로 보입니다.

Open AI를 둘러싼 이런 비슷한 논쟁은 3월 25일, Open AI의 챗 GPT-4o 최신 버전에 일본 애니메이션 제작사인 지브리 (Ghibli)가 만든 이미지 생성기능을 추가, 출시하자마자 또 벌어지고 있을 정도로 AI모델 개발을 둘러싼 상수가 되고 있습니다. 지브리 (Ghibli) 특유의 화풍으로 각종 사진, 그림들을 챗 Gpt-4o가 재미있게 만들어줌에 따라 ‘화풍 (스타일)은 저작권 보호대상이 아니다

v. 챗 GPT-4o가 기존 저작물들을 무단으로 학습했기에 그런 결과물이 나왔다'는 싸움이 그것입니다.

한편, 학습, 개발과정에서 어떤 저작물을 어떻게 사용했는지에 대한 논쟁이 AI모델을 둘러싼 저작권 소송의 핵심이었다면 약간의 변화를 보인 것이 유니버설뮤직 등과 같은 음반제작자들이 Claude AI가 노래가사 저작물을 대량으로 수집, 불법으로 학습시키고 사용하였다는 이유로 Anthropic을 상대로 2023년 10월에 제기한 소송 건입니다. 연방지방법원이 텍스트 생성능력에 특화된 Claude AI가 학습된 가사 저작물을 출력하거나 변형된 형태로도 보여주지 못하게끔 차단한 Claude AI모델의 기술적 안전장치 (guardrail)를 계속 유지하고 향후에도 지속하도록 합의한 양 당사간의 타협을 올 1월에 받아들였기 때문입니다. 비록 핵심 쟁점인 학습, 개발과정에서의 저작권 다툼은 여전히 진행되지만 적어도 프롬프트 결과물을 이용자에게 보여줄 때에는 저작권이 침해되지 않도록 AI모델 스스로 기술적 조치를 갖춰야 된다는 점이 새롭게 채택되었기 때문입니다.

과거 구글이 2004년에 전세계 도서관과 협력하여 수백만권의 책을 디지털화 하면서 처음 선보인 TDM과 이를 기반으로 하는 구글북스 프로젝트가 10년에 걸친 소송 끝에 미국 연방대법원으로부터 공정한 사용으로 2016년에 최종 결론이 날 당시와 현재 AI와의 관계, 시장상황 등을 단순 비교할 수는 없습니다. 당시 구글이 비영리목적으로 TDM을 사용한 것은 원본 아날로그 책에 대한 접근을 손쉽게 하기 위한 것이 아니라 키워드 검색을 통해 디지털화된 책의 일부 (snippet)만을 미리 보여주려 한 것이라는 등의 이유로 공정한 사용이라고 법적 결론을 낸 상황과 달리 지금의 AI 모델개발 경쟁은 글로벌 시장차원에서 벌어지고 있고 결과물을 생성하면서 새로운 영리시장을 폭넓게 열고 있기 때문에 다양한 데이터 학습방법을 둘러싼 갈등이 자연히 커질 수밖에 없기 때문입니다.

### 3. 약관을 통한 해결방식

저작권법상, 또한 미국처럼 판례를 통해 설사 TDM이 합법적으로 허용된다 하더라도 이용약관을 통해 TDM의 전 단계인 데이터 수집을 위한 스크래핑을 하지 못하도록 막아버린다면 AI의 학습 및 모델개발은 처음부터 불가능하게 될 수 있습니다. 이 점을 고려, EU 저작권법은 특별히 제3조, 즉 연구목적의 TDM 허용 규정은 계약을 통해서도 배제할 수 없게끔 규정하고 있습니다.

한편, AI 대기업이 아닌 중소, 스타트업들은 저작권자와의 합법적 라이선스 계약을 체결하는 것부터 어려움을 겪을 수 있다는 점도 고려, 계약자유를 허용하지 않는 EU 저작권법 제3조와 달리 제 4조, 즉 ‘상업적 목적’의 데이터 스크래핑과 TDM에 대해 실제 표준약관 방식을 통해 AI 기술개발과 확산을 도와주면서 동시에 저작권자를 지키는 경제적, 기술적 타협이 행해지기도 합니다. 다만 나라마다 저작권에 대한 이해방식이 서로 상이할 수 있기 때문에 표준약관 채택만으로는 해결되기 어려운 점이 남아있게 되고 전문적 법률자문이 필요하게 됩니다.

예를 들면, 표준화된 이용약관이나 CCL (Creative Commons License)방식을 통해 스크래핑된 데이터에 대한 TDM기반의 사용허락까지 받았다 하더라도 AI가 창작자의 스타일을 모방하여 내용물을 만들어낼 경우 우리나라, 프랑스, 독일 등 대륙법계 국가들이 일반적으로 인정하는 저작인격권 (moral rights) 침해문제도 제기될 수 있습니다. 특히 거대언어모델 (LLM)로 학습한 범용 AI가 만들어내는 결과물이 학습된 수집 데이터, 창작물에 대한 저작권을 침해하지는 않았지만 스타일을 모방함으로써 창작물에 대한 유사성 (likelihood)에 대한 불만이 고조되고 있고, 더욱이 AI 비서 (agent) 모델이 확산되면서 창작자의 목소리 등을 모방한 AI 서비스모델은 향후 인격권침해 문제를 제기할 수 있기 때문에 저작인격권의 포기여부 등에 대한 법률자문이 요구됩니다.

TDM을 포함한 AI모델의 학습, 개발방법 등에 대해서는  
미국에서 진행중인 많은 소송에서도 쟁점으로 다루기 때문에  
글로벌 창작물 시장에서 약진하고 있는 우리 역시 관심을 가질 수밖에 없습니다.

최근 챗 GPT-4o에 탑재된 새로운 이미지 생성기능은 지브리 효과 (Ghibli Effect)라고 불리울 정도로  
전 세계적인 흥미거리를 제공하면서 저작권, 개인정보보호 문제를 다시 들춰냈습니다.

특히 수억명이 자발적으로 자신의 사진, 이미지를 올리면서 놀이처럼 참가한 덕분에  
데이터 스크래핑부터 걸리게 된 '개인정보의 자발적 동의' 규정을 너무 쉽게, 현명하게 우회했고,  
나중에 본인이 Open AI에게 탈퇴 (opt-out)의사를 표시하지 않는 한  
Open AI는 합법적으로, 비용도 들이지 않고 AI모델 추가 학습, 개발도 할 수 있게 되었습니다.

소셜미디어보다 더 많고 다양한 원본을 사실상 독점하게 될 Open AI가  
다른 소셜미디어 내용물을 지브리 이미지로 점령하려는 (Ghiblified) 영업전략의 성패 포함,  
최종 비즈니스 정산은 향후 저작권 소송 결말을 지켜봐야 될 듯합니다.  
또한 AI의 최신 이미지 생성기능으로 인한 서류 위변조와 사기범죄 등  
AI발전에 따른 부작용 법 문제도 계속 나올 것입니다.

매월 발간하는 법무법인 린 AI산업센터의 뉴스레터인 AID 내용에 대한 질문이나 조언 등은  
구태연 TMT 전문그룹장 ([tekoo@law-lin.com](mailto:tekoo@law-lin.com)), 방석호 AI산업센터장 ([shbang@law-lin.com](mailto:shbang@law-lin.com)),  
설기석 구성원 변호사 ([ksseol@law-lin.com](mailto:ksseol@law-lin.com))에게 보내주십시오.