

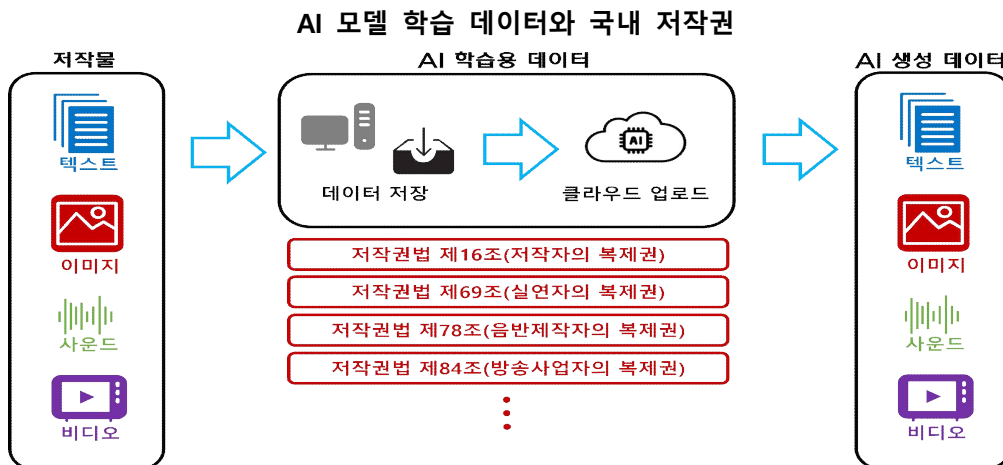
# AI 학습용 데이터의 저작권 침해 이슈 및 국내·외 대응 현황

KDB미래전략연구소 산업기술리서치센터  
박 성 수 (holly\_wp@kdb.co.kr)

- ◆ AI 기술의 발전으로 학습용 데이터의 범위가 다양한 종류로 확대됨에 따라, 저작권 침해 가능성이 AI 기업에 학습 데이터 구매비용 증가 및 법적 분쟁 리스크로 대두되며, AI 산업 발전을 저해할 우려 증가
- ◆ 해외는 일본을 중심으로 사전 이용 허락 없이도 AI 모델 학습 데이터로 저작물을 이용하는 것에 대한 법률적 허용이 논의되고 있으며, 국내에서도 구체적인 논의와 활용 가이드라인 필요

## □ 고성능 AI 모델의 발전으로 학습 데이터 이용이 증가하면서, 저작권 침해 이슈가 AI 산업의 리스크로 부상

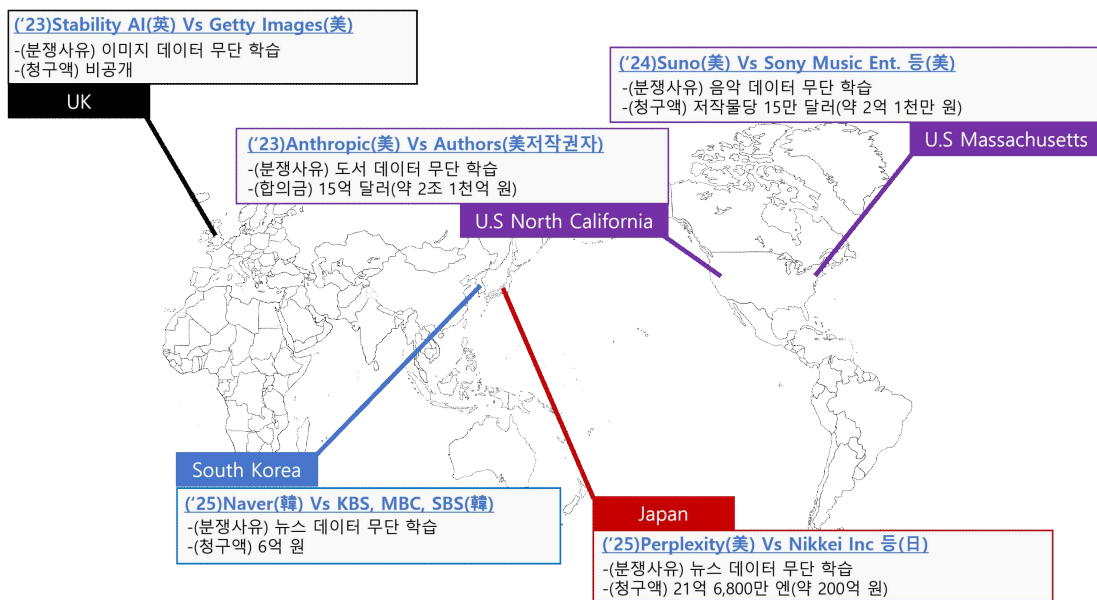
- AI 기술이 혁신\*을 거듭하며 학습 데이터의 종류가 텍스트부터 사운드, 이미지, 비디오에 이르기까지 다양화, 비정형화되면서 저작권법 위반 가능성 증가
  - \* 대형언어모델(LLM)이 다양한 종류의 비정형데이터(텍스트, 음성, 이미지 등 고차원 데이터)를 활용한 멀티모달모델(LMM)로 발전
- AI 모델 개발 과정에서 다양한 저작물을 학습용 데이터로 활용할 경우, 저장·복제·모델 파라미터 형태의 내재화 등 이용 방식에 따라 복제권\* 침해 가능성 존재
  - \* 저작권법 제16조, 69조, 78조, 84조 등



자료 : 한국산업은행

- 국내 저작권자들과 AI 업계 간 저작물 활용에 대한 견해차로 인해 AI 기업들의 학습용 데이터 구매비용 및 법적 분쟁 가능성 증가 우려
  - (저작권자) 음악, 영상, 웹툰 등 12개 단체가 참여한 ‘범창작자정책협의체’는 저작자 권리보호를 위한 TDM\* 법안 입법 반대 및 저작물 이용에 대한 적정 비용 산정 논의 요구
    - \* TDM(Text Data Mining) 허용법안 : 텍스트 등의 데이터를 AI 모델 학습에 활용 시, 사전에 권리자로부터 이용 허락을 받지 않더라도 저작물을 이용할 수 있도록 허용하는 규정
  - (AI 업계) 한국인공지능협회·소프트웨어산업협회는 “권리자와 매번 이용 허락 계약을 맺는 것은 불가능에 가깝고, 데이터 구매비용은 국내 AI 스타트업들의 성장과 AI 산업 발전을 저해한다”라는 입장이며, TDM 법안 입법 요구
  
- 글로벌 AI 기업들도 AI 학습용 데이터에 책, 이미지, 음악, 뉴스 기사 등의 저작물을 무단으로 이용한 사례가 확인되어 글로벌 법적 분쟁 발생
  - 저작권 분쟁은 AI 기업에 법적 불확실성과 대규모 손해배상 위험을 초래할 뿐만 아니라 산업 전반의 신뢰도 및 지속가능성에도 영향을 미치는 리스크로 부상
    - 美 캘리포니아 북부 연방법원은 AI 스타트업 엔트로픽을 상대로 한 작가들의 저작권 소송에서 15억 달러 규모\*의 합의안 승인
    - \* 역대 최대 규모의 저작권 침해 배상액

대표적인 저작권 침해 분쟁 사례



자료 : 언론보도 취합, 한국산업은행

□ 주요국은 저작물의 AI 모델 학습 활용에 대한 저작권법 개정, 판례를 통해 저작권 침해 리스크에 대응 중

- (일본) '19년 TDM 허용 법안\*의 시행을 통해 정보분석 및 데이터 학습 시 저작권 분쟁을 방지하고, 상업적 사용 허용을 통해 AI 산업 활성화 도모
  - \* 저작권법 제30조의4: 공표된 저작물을 '향유목적'이 아닌, '정보분석'을 위해 필요한 범위에서 복제하는 행위를 허용하는 규정. 단, 권리자 이익 부담 침해 시 예외
  - 저작물 복제의 허용 범위를 명확히 규정함으로써, AI 기업들의 학습용 데이터 이용에 대한 법적 불확실성 감소
  
- (영국) '14년 저작권법 29A\* 시행 및 정부의 가이드라인\*\*을 통해 저작권 분쟁 방지 및 학술 과학 연구 효율성 향상에 기여
  - \* Copyright, Designs and Patents Act 1988, 29A: 합법적인 접근이 가능한 저작물을 '비상업적'이고 '연구 목적'으로 이용할 수 있도록 하는 영국 저작권법상 예외 규정
  - \*\* 영국 정부 IPO, "Exception to copyright: Text and data mining for non-commercial"
  - '비상업적', '연구 목적'만을 저작권 침해 예외 조항으로 두어, 저작물의 AI 학습 활용에 대한 허용 범위를 명확히 규정
  
- (미국) '70년대 제정된 Fair Use 원칙\*은 AI 학습 데이터 관련 저작권 분쟁 발생 시, AI 기업들이 저작권 침해 주장에 대한 방어 규정으로 활용
  - \* 저작권법 제107조: 저작자의 허락 없이도 '공익적 목적, 변형적 이용'에 한해 저작물을 제한적으로 이용할 수 있도록 허용하는 미국 저작권법상 예외 규정
  - Fair Use 원칙은 저작물 이용의 허용 범위를 '공익적 목적', '변형적 이용'과 같이 추상적으로 규정하여 저작권 분쟁의 소지 존재
  - 다만, 판례법주의 국가 특성상, 수십 년간 누적된 Fair Use 원칙에 대한 다양한 기술 판례를 바탕으로 AI 기업들은 저작권 분쟁 리스크에 대응 가능

주요국의 AI 학습 관련 저작권법 규정

국 가	규 정	저작권 허용 범위	비 고
일 본	저작권법 제30조의4	'향유목적이 아닌 경우', '정보분석 목적 이용'	권리자 이익 부담 침해 시 예외
영 국	저작권법 29A	'합법적 접근일 것', '비상업적 이용', '연구 목적 이용'	정부 가이드라인 ① 제3자에게 공유 금지 ② 출처 표기할 것 등
미 국	저작권법 제107조	'공익적 목적 이용', '변형적 이용'	법원이 구체적 사실 관계를 기반으로 판단

자료 : OECD(2025.2), "Intellectual property issues in artificial intelligence trained on scrapped data", 당행 재구성

□ 국내에서도 저작권 침해 분쟁이 AI 산업 발전을 저해하지 않도록, 저작물의 AI 학습 활용에 대한 법률 개정 등 명확한 가이드라인 필요

- 베른협약\*과 미국의 Fair Use 원칙 판례를 참조하여 국내 저작권법 내 공정이용 규정\*\*을 신설('11년)하였으나 AI 학습용 데이터와 관련된 대법원 판례 부족과 가이드라인 부재로 AI 업계와 저작권자 간 혼란 증가

\* 1886년 스위스 베른에서 체결된 저작권 국제 조약

\*\* 저작권법 제35조의5, 제87조(제35조의5 준용): 저작자의 허락 없이도 저작물의 '일반적인 이용 방법'과 충돌하지 않고, '저작자의 정당한 이익을 부당하게 해치지 않는 경우'에만 저작물 이용을 허용하는 규정

- AI 기업들의 법적 분쟁 리스크 최소화 및 창작자 권리보호를 위한 합리적이고 명확한 절충 방안 모색 필요

- 등록 저작물에 한하여 사전 이용 허락을 구하고, 미등록 저작물은 사후 보상금을 청구하는 방안, 옵트 아웃\* 방안 등 다양한 논의 필요

\* Opt Out: 권리자가 이용 반대 의사를 표시한 경우에만 학습에 이용 불가