

이슈

생성형 AI 저작물 학습과 공정이용의 한계 : 학습용데이터로 이용된 뉴스콘텐츠를 중심으로

김태경

법무법인 케이씨엘 변호사

목차

- 1 들어가며
- 2 생성형 AI의 뉴스콘텐츠 학습에 대한 공정이용 규정의 적용 가능성
- 3 텍스트 및 데이터 마이닝(TDM) 면책 규정 도입의 문제점
- 4 생성형 AI 기업과 미디어 창작자의 상생 발전을 위한 제언
- 5 맺음말

요약

생성형 AI의 급격한 확산은 뉴스콘텐츠의 소비 방식을 ‘제로 클릭(zero click)’ 형태로 변화시키며 미디어 생태계의 조립을 위협하고 있다. 이 글은 이러한 기술적 전환기에 생성형 AI의 뉴스콘텐츠 학습이 우리 저작권법상 공정이용 법리에 부합하는지를 최근 미국 판례와 대법원 판례를 통해 비판적으로 검토한다. 상업적 목적의 AI 학습은 원저작물의 현재 및 잠재적 시장을 침해하는 ‘무임승차’ 행위로서 공정이용으로 정당화되기 어려우며, 성급한 TDM 면책 도입은 창작 생태계를 훼손하여 결과적으로 AI 산업의 지속가능성마저 저해할 우려가 있음을 지적한다. 지속 가능한 AI 혁신과 창작 생태계의 공존을 위해서는 학습용데이터의 투명한 공개 의무화와 뉴스 콘텐츠 가치에 부합하는 정당한 보상 체계의 확립이 필수적임을 제언한다.

1. 들어가며

바야흐로 인공지능(이하 ‘AI’)의 시대다. 2022년 11월, 오픈AI(OpenAI)가 별다른 홍보 없이 조용하게 출시한 챗GPT(ChatGPT)가 AI 역사 및 현대사에 유례없는 파급력을 가져온 이래, 생성형 AI는 언어, 이미지, 영상을 자유자재로 만들어내며 인간 고유의 영역이라 여겨졌던 ‘창작’의 성역까지 허물고 있다. 이제 AI는 단순한 기술적 도구를 넘어 전 세계 산업의 지형을 송두리째 재편하는 ‘게임 체인저(game changer)’로 확고히 자리 잡았다. 우리나라 정부 역시 ‘인공지능 3대 강국’ 도약을 목표로 달려가고 있다.

그러나 이 화려한 기술 혁신의一面에는 미디어 생태계의 존립을 위협하는 어두운 그림자가 드리워져 있다. AI를 개발하고 고도화하기 위해서는 수많은 콘텐츠를 수집·가공하여 데이터베이스를 구성한 다음 이를 인공 신경망에 전달하여 학습시키는 과정이 필수적인데, 이 과정에서 학습용데이터에 포함된 저작물에 대한 복제·전송 등의 행위가 필연적으로 수반되기 때문이다. 이로 인해 AI 학습과 관련한 저작권 침해 문제가 대두되고 있다.¹⁾

아울러, 이러한 AI 학습의 결과는 원작자의 생존 기반을 파괴하는 시장대체로 직결되고 있다. 뉴스콘텐츠의 상황은 더욱 심각하다. 과거 포털 사이트가 뉴스 유통의 관문(gateway) 역할을 했다면, 이제 생성형 AI는 뉴스 소비의 종착지(destination)가 되어 가고 있다. 이용자들은 더 이상 언론사 홈페이지나 포털 뉴스 페이지의 링크를 클릭하지 않는다. AI 챗봇이 뉴스 기사를 학습해 제공하는 정교한 요약 답변만으로 정보 욕구를 충족하는 이른 바 ‘제로 클릭(zero click)’ 시대가 도래한 것이다. 실제로 구글이 검색 결과 상단에 AI 요약을 노출하는 ‘AI 오버뷰(AI Overview)’를 도입한 이후, 미디어 업계의 트래픽은 최대 40%까지 급감했다는 충격적인 보고가 이어지고 있다.²⁾ 이는 AI가 뉴스의 유통을 보조하는 수단을 넘어, 뉴스 미디어의 수익 모델을 근본적으로 파괴하는 대체제로 기능하고 있음을 시사한다.

1) 이와 관련하여 2023년 12월 27일 미국 뉴욕타임스(NYT)가 오픈AI(ChatGPT)와 Microsoft(Bing Chat, 현재 Copilot)를 상대로 GPT 대규모언어모델(LLM)을 훈련시키는 데 뉴스저작물을 무단 이용하였다는 이유로 미국 뉴욕 남부지방법원(US District Court, Southern District of New York)에 제기한 저작권 침해소송이 많이 알려져 있으며, 이 소송은 뉴스 언론사가 인공지능사업자를 상대로 제기한 최초의 소송이다. 2025년 1월 13일 우리나라에서도 지상파3사가 네이버 등을 상대로 유사한 내용의 소송을 제기했으며, 현재 진행 중이다.

2) 한국저작권위원회, “구글 AI 오버뷰 도입 후 미디어업계 트래픽 감소와 대응 현황”, ‘저작권 이슈 브리프’, 2025-8-4호, 2025.

이러한 위기 속에서 2026년 1월 22일 시행을 앞둔 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(이하 “인공지능기본법”) 역시 깊은 우려를 안겨주고 있다. 해당 법률은 우리나라 AI 규제체계의 기본 프레임을 처음 설정한 법률이라는 점에서 의의가 있으나, AI 산업의 ‘진흥’에 방점이 찍혀 있을 뿐, AI 학습에 이용된 수많은 저작물에 대한 학습용데이터 공개나 보상 체계와 같은 권리자 보호 장치는 사실상 공백 상태로 남겨두었기 때문이다.³⁾

이 글은 학술적 엄밀성을 갖춘 심층 논문이라기보다는 급변하는 AI 시대에 우리가 놓치지 말아야 할 법적 문제를 환기하기 위한 시론적 성격의 글이다. 이에 AI 기업들이 무단 학습의 면죄부로 삼고 있는 ‘공정이용(Fair Use)’ 법리를 재검토하고, 왜 뉴스콘텐츠의 AI 학습이 법리적으로 공정이용으로 정당화될 수 없는지 개괄적으로 분석한다. 나아가 무분별 한 텍스트 및 데이터 마이닝(Text and Data Mining, 이하 “TDM”) 면책 도입의 위험성을 경고하고, AI 기업과 미디어 창작자가 공존하기 위한 법적 과제를 제언하고자 한다.

2. 생성형 AI의 뉴스콘텐츠 학습에 대한 공정이용 규정의 적용 가능성

1) 생성형 AI 개발·활용과 관련한 저작권 침해의 문제

생성형 AI 개발·활용 과정에서 발생할 수 있는 저작권 침해 문제는 다양한 형태로 나타날 수 있지만 주로 두 가지 측면에서 논의되고 있다. 첫째는 AI 학습 단계에서의 침해 가능성이 있고, 둘째는 AI 산출 단계에서의 침해 가능성이 있다. 이 둘은 밀접하게 연결되어 있지만, 법적 평가상 별개의 침해 행위로 다뤄질 수 있으며, 그 판단기준과 입증방식 역시 구별 될 수 있다.

AI 학습 단계에서의 저작권 침해는 기존 저작물에 대한 수집·가공을 통해 대규모 학습용데이터 세트를 구축하고 이를 인공 신경망에 전달하여 기계학습에 활용하는 일련의 행위에서 발생한다. 저작물을 포함한 학습용데이터 세트가 권리자의 동의 없이 사용될 경우, 이는 복제권 및 전송권 침해에 해당할 수 있다. 전통적인 저작권 침해 판단 구조는 의거

3) 과학기술정보통신부는 2025. 11. 12. 인공지능기본법 시행령 제정안을 입법예고한 바 있다.

성과 실질적 유사성 기준에 따라 구성되기 때문에 AI 학습 단계에서의 침해 문제를 원저작물과 AI 산출물과의 유사성의 관점에서 판단하고 이와 연결 짓는 견해도 제기되나, AI 학습 단계는 AI 산출 단계 이전에 원저작물의 이용을 완결하기 때문에, 예컨대 최종 AI 산출물이 원저작물과 유사하지 않다는 이유만으로 그 이용의 위법성이 자동적으로 해소된다고 보기 어렵다. AI 학습 그 자체가 침해의 본질이라는 점에 주목할 필요가 있다.

AI 산출 단계에서의 저작권 침해는 별도의 침해태양(침해유형)으로서 AI 산출물이 특정 저작물에 의거하고 있는지 여부와 표현의 실질적 유사성이 존재하는지 여부가 침해 판단의 핵심이 될 것이다. 문제는 AI가 수많은 저작물을 혼합하여 새로운 산출물을 생성하기 때문에, 그 표현이 단일 원저작물에 명백히 의거했다고 보기 어려운 경우가 있을 수 있다는 점이다. AI 산출물이 다수의 저작물을 변형·융합한 형태일지라도 그 핵심 표현을 그대로 모방했다면 독립적인 침해로 평가될 수 있을 것이다.

이 글에서 주로 다룬 부분은 AI 학습 단계에 관한 것으로, AI 기업들은 AI 학습의 저작권 문제에 대해 일정한 조건을 충족할 경우 저작권 보호의 예외를 인정하는 TDM 면책 적용을 주장하며(우리나라는 아직 TDM 면책 규정을 도입하지 않았다), 아울러 일반적·포괄적 저작재산권 제한 사유인 공정이용(Fair Use) 규정을 통해 면책된다고 주장하고 있다.

2) 생성형 AI의 저작물 학습에 대한 공정이용 규정의 적용 가능성

(1) 2025년, 3건의 미국 판결의 시사점

미국 저작권법 제107조에서는 저작물의 이용이 공정이용에 해당하는지를 판단하는 기준으로 ①제1요소 : 이용의 목적 및 성격(the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes)⁴⁾, ②제2요소 : 원저작물의 성질(the nature of the copyrighted work), ③제3요소 : 이용된 부분이 원저작물 전체에서 차지하는 양과 상당성(the amount and substantiality of the portion used in relation to the copyrighted work as a whole),

4) 변형적 이용(transformative use)에 관한 자세한 내용은 현서유. (2024). “저작권법상 공정이용 원칙의 판단 기준 및 인공지능 학습데이터의 변형적 이용 – 미국 연방대법원 Andy Warhol Foundation v. Goldsmith (2023) 판결을 중심으로”. 경영법률, 34(4), 207–235. 참조.

④ 제4요소 : 이용이 원저작물의 잠재적인 시장 또는 가치에 미치는 영향(the effect of the use upon the potential market for or value of the copyrighted work)을 규정하고 있다.

최근 미국에서는 AI의 개발·활용을 위해 저작물을 학습시키는 행위가 공정이용에 해당하는지 여부에 관한 판결들이 잇달아 선고되었다. 우리나라의 경우, 아직까지 TDM 면책 규정을 도입하지 않았고 저작권법에 공정이용에 관한 조항(제107조)을 두고 있는 미국과 유사한 체계를 가지고 있기 때문에, 최근 선고된 미국 판결들은 우리나라에서도 생성형 AI 저작물 학습에 대한 공정이용 규정의 적용 가능성에 대한 시사점을 줄 수 있다.⁵

① Thomson Reuters Enter. Ctr. GmbH v. Ross Intelligence Inc.⁶ : 공정이용 부정
로스 인텔리전스는 법률 데이터베이스를 검색해서 보여주는 ‘로스 AI(ROSS AI)’(참고로 생성형 AI라기보다는 이른바 ‘검색형 AI’ 또는 ‘판별형 AI’에 가깝다)를 개발하였다. 톰슨 로이터는 로스 인텔리전스가 위 AI를 학습시키는 과정에서 톰슨 로이터가 운영하는 ‘웨스트로(Westlaw)’의 헤드노트(판결 요지에 대한 요약문)와 키 넘버 시스템(자료 분류를 위한 색인 체계)을 무단 이용했다며 저작권 침해 소송을 제기하였다. 이에 대해 로스 인텔리전스가 공정이용이라며 항변했으나, 법원은 공정이용 항변을 배척하고 저작권 침해를 인정했다.

법원은 로스 AI 서비스가 상업적 목적을 가지고 있고, 로스 인텔리전스의 이용이 톰슨 로이터의 이용과 ‘더 나은 목적이나 다른 성격’이 없으며, 결과적으로 톰슨 로이터의 헤드노트를 웨스트로와 경쟁할 법률 조사 도구를 만드는 데 이용했다고 판단했다(제1요소). 그리고 법원은 가장 중요한 공정이용 판단 요소가 저작물의 잠재적 시장에 미치는 영향이라는 판단 하에, 현재 시장뿐만 아니라 저작권자가 일반적으로 개발할 수 있거나, 다른 사람에게 라이선스를 줄 수 있는 잠재적 시장에 대한 영향까지 포괄적으로 고려하여야 한다고 하면서, 로스 인텔리전스의 이용이 웨스트로와 직접적으로 경쟁하는 법률 리서치 플랫폼

5) 다만, 2025년에 선고된 3건의 미국 판결은 모두 악식판결에 불과하다는 점을 감안하여야 하고, 그 내용이나 맥락도 모두 다르다는 점을 유의하여야 한다. 아울러, 공정이용 요소들은 개별적 평가가 아닌 종합적 고려라는 점에서 사안별로 그 특수성이 반영됨을 감안하여야 한다.

6) 미국 딜리웨어 연방지방법원 2025. 2. 11. 선고 : Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc., 765 F. Supp. 3d 382 (D. Del. 2025).

시장뿐만 아니라 항후 법률 AI 훈련 데이터 시장에까지 부정적인 영향을 미칠 수 있다고 보았다(제4요소).⁷⁾ 이는 AI가 원저작물을 대체할 우려가 있거나 경쟁관계에 있다면 공정이용의 핵심인 ‘변형적 이용’으로 인정받기 어렵다는 점, 그리고 원저작물 그 자체가 AI 학습의 원료로서 거래되는 ‘학습용데이터 시장’이라는 잠재적 시장을 인정할 수 있음을 시사한다.

② Bartz, et al. v. Anthropic PBC⁸⁾ : 공정이용 일부 인정

앤티로피는 ‘클로드(Claude)’라는 대규모언어모델(LLM)을 개발하였다. Andrea Bartz 등 작가 3명은 앤티로피이 클로드(Claude)를 훈련시키기 위해 자신들의 서적을 무단 이용했다며 저작권 침해 소송을 제기하였다. 하지만 법원은 앤티로피이 서적을 구입하여 라이브러리에 저장하고, 학습에 이용한 것을 모두 공정이용이라고 판단했다.

법원은 학습에 이용된 서적들은 비록 ‘표현적인 요소’를 포함하고 있긴 하지만(제2요소), ‘완전히 변형적인(exceedingly transformative)’ 이용이라면서(제1요소), 이를 “작가가 되기를 희망하는 독자가 책을 읽으며 글쓰기 능력을 연마하는 것”에 비유했다. 즉, 언어 간의 통계적 패턴을 학습하여 언어 모델을 만든 후 본질적으로 다른 목적의 새롭고 독창적인 텍스트를 생성하기 위한 것으로 보았다. 아울러, 법원은 저작물의 잠재적인 시장이나 가치에 미치는 영향 측면에서도 공정이용 인정에 문제가 없다고 보았다(제4요소). 법원은 작가들의 저작물 사본에 대한 수요를 대체하지 않았고 앞으로도 대체하지 않을 것이라고 하였고, AI가 경쟁 저작물의 폭발적 증가를 유발하더라도 경쟁 작가가 늘어난 것과 다를 바 없다고 보았다. 다만, 법원은 ‘원고들이 앤티로피의 대규모언어모델(LLM) 훈련이 원저작물이나 파생저작물의 시장에 심각한 위협이 된다는 것을 입증하지 못했다’는 전제 하에 위와 같이 판단한 것이라는 점에 주목할 필요가 있다.

7) 해당 판결에 대한 기타 내용은 한국저작권위원회, “[미국] AI 학습시 저작물 이용에 대한 미국 법원의 공정이용 판단 검토(박한빈) – Thomson Reuters, v. Ross Intelligence., 1:20-cv-00613”, 「저작권 동향 2025」 제2호, 2025. 2. 28. 참조.
8) 미국 캘리포니아 북부지방법원 2025. 6. 23. 선고 : Bartz v. Anthropic PBC, 2025 WL 1741691 (N.D. Cal. June 23, 2025).

한편, 법원은 불법사이트에서 서적을 다운로드하여 저장한 부분은 공정이용이 될 수 없다고 명확하게 선을 그었다. 정당한 방식으로 입수할 수 있는 서적을 불법 복제하여 전자도서관을 구축한 것은 그 자체로 저작권 침해에 해당한다는 것이다.⁹

③ Kadrey, et al. v. Meta Platforms, Inc.¹⁰ : 공정이용 인정

메타는 ‘라마(Llama)’라는 대규모언어모델(LLM)을 개발하였다. Richard Kadrey를 비롯한 13명의 작가들은 메타가 ‘라마(Llama)’를 훈련시키기 위해 자신들의 책을 무단 이용했다며 저작권 침해 소송을 제기하였다. 이 사건에서도 법원은 공정이용 항변을 받아들였다.

법원은 메타가 책을 복사한 목적이 사람들에게 읽히기 위한 것이 아니라 대규모언어모델(LLM)이 단어와 문장 사이의 통계적 관계를 학습하도록 훈련시키는 것으로서, 책의 목적과는 다른 목적과 다른 성격을 가진 것으로 ‘매우 변형적인(remely transformative)’ 이용이라고 하였다(제1요소). 법원은 시장에 미치는 영향에 대한 판단에 있어서도 메타의 손을 들어주었지만, 그 취지는 작가들이 라마(Llama)의 결과물이 그들의 책 시장에 어떤 식으로든 해를 끼칠 수 있다는 ‘실증적 증거를 제시하지 못했기 때문’이며, ‘만약 AI가 원작과 유사한 작품을 대량으로 생성하여 시장에 혼란을 초래한다는 증거를 제시했다면 소송결과가 달라졌을 것’이라고 하였다(제4요소). ‘해당 판결이 메타의 저작물 이용이 합법적이라는 명제를 지지하는 것이 아니라는 점’을 분명히 했다는 점에 주목할 필요가 있다. 원고의 입장 부족으로 인하여 해당 사건에서는 메타의 손을 들어주었지만, AI 학습 전반을 공정이용으로 일반화될 수 없다는 점을 시사한다.¹¹

9) 해당 판결에 대한 기타 내용은 한국저작권위원회, “[미국] Anthropic사의 AI 학습 목적의 저작물 이용에 대한 미국 법원의 공정이용 판단(총지회)”, 「저작권 동향 2025」 제7호, 2025. 7. 16. 참조.

10) 미국 캘리포니아 북부지방법원 2025. 6. 25. 선고 : Kadrey v. Meta Platforms, Inc., 2025 WL 1752484 (N.D. Cal. June 25, 2025).

11) 해당 판결에 대한 기타 내용은 한국저작권위원회, “[미국] Meta의 AI 학습 목적의 저작물 이용에 대한 미국 법원의 공정이용 판단(총지회)”, 「저작권 동향 2025」 제7호, 2025. 7. 16. 참조

(2) 우리나라 저작권법상 공정이용의 해석

우리나라 저작권법 제35조의5는 제1항에서 “제23조부터 제35조의4까지, 제101조의 3부터 제101조의5까지의 경우 외에 저작물의 일반적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다.”고 하면서¹², 제2항에서 “저작물 이용 행위가 제1항에 해당하는지를 판단할 때에는 다음 각 호의 사항 등을 고려하여야 한다.”고 규정하고 있다. 위 공정이용 규정은 개별적인 행위 유형을 정하고 있는 각 제한 규정과는 달리 행위유형을 특정하지 않고 일반적·포괄적인 형태로 규정하고 있다. 그리고 이를 위해 각 호에서 정한 ①이용의 목적 및 성격, ②저작물의 종류 및 용도, ③이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성, ④저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향을 고려해야 하며, 궁극적으로는 법원에서 사안에 따라서 개별적인 사실관계 등을 고려하여 판단하게 된다.

생성형 AI 학습에 대한 공정이용 규정 적용 여부에 대한 학계의 의견 대립이 있으며, 이를 직접적으로 판단한 국내 판례는 아직 존재하지 않는다. 다만, 위 공정이용 규정과 관련하여, 대법원 2024. 7. 11. 선고 2021다272001 판결¹³에서는 다음과 같은 법리를 설시하면서, 수학능력시험 등 전국적으로 시행되는 중요 시험의 문제를 인터넷에 공개·게시하는 행위에 대해 공정이용이 성립하지 않는다고 보았다.

▲저작물의 이용 행위가 구 저작권법 제35조의3 제1항에서 규정한 ‘저작물의 통상적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우’에 해당하는지 판단할 때에는, 같은 조 제2항 각 호에서 예시적으로 열거한 ‘이용의 목적 및 성격(제1호)’, ‘저작물의 종류 및 용도(제2호)’, ‘이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성(제3호)’, ‘저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향(제4호)’ 등을 종합적으로

12) 생성형 AI 저작물 학습의 경우 저작권법 제23조부터 제35조의4까지, 제101조의3부터 제101조의5까지의 경우에 해당하지 않으므로, 결국 제35조의5에 따른 공정이용 규정의 적용 여부가 핵심 쟁점이 될 것이다.

13) 원고는 저작권 신탁관리단체이면서 보상금 수령단체인 한국문학예술저작권협회이고, 피고는 한국교육과정평가원으로, “고등학교 이하 각급학교 교육과정의 연구·개발 및 교육평가의 연구 실시, 전국적으로 실시하는 학력평가시험의 출제, 시험 및 채점 등의 업무를 목적으로 설립된 공공기관”이다. 피고는 대학수학능력시험을 비롯하여 검정고시, 교육공무원 임용후보자 선정경쟁시험(이른바 ‘임용고시’) 등을 주관·관리하고 있다. 이들 중 일부 시험은 출제되고 응시된 후 피고의 홈페이지 등에 공개되고 있다. 시험의 지문에는 물론이고 문제에도 출제자가 직접 만들지 않은 ‘외부’ 저작물이 포함되어 있는데, 원고는 이 저작물을 인터넷에 게시해 누구나 접근할 수 있도록 한 피고의 행위가 저작권 침해·전송권 침해에 해당한다고 주장하며 소송을 제기했다.

고려하여야 하고, ▲이용의 경위나 방법 등과 같이 위 각 호에서 열거하지 않은 사항이라도 판단 요소로 고려할 수 있다. ▲‘이용의 목적 및 성격(제1호)’에 관하여는 그 이용이 원저작물을 단순히 대체하는 수준을 넘어 새로운 표현, 의미, 메시지 등을 나타내도록 변형한 것인지, 원저작물과는 구별되는 별개의 목적과 성격을 가지는지, 원저작물을 변형한 정도가 2차저작물 작성에 필요한 수준보다 더 높은 정도에 이르렀는지, 공익적이거나 비영리적인 이용인지 등을 고려할 수 있다. ▲‘저작물의 종류 및 용도(제2호)’에 관하여는 원저작물이 사실적·정보적 성격을 가진 저작물인지, 공표되거나 발행된 저작물인지 등이 고려되고, ▲‘이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성(제3호)’에 관하여는 원저작물 전체를 기준으로 그 이용된 부분이 차지하는 양적인 비중이나 질적인 중요성이 낮은지, 이용자가 반드시 필요한 범위 내에서 이용한 것인지 등이 고려될 수 있다. ▲‘저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향(제4호)’에 관하여는 저작물의 이용이 원저작물 또는 원저작물의 2차저작물에 대한 현재 시장의 수요나 장래 개발될 합리적인 개연성이 있는 통상적인 시장의 수요를 대체하거나 그 시장가치를 훼손할 우려가 없거나 적은지 등을 고려하여야 한다.

위 대법원 판결은 대법원에 의해 공정이용이 직접 검토된 첫 사례라는 의의 외에도 향후 공정이용에 관한 법원 실무에 중요한 지침을 제공했다는 데서도 매우 큰 의미를 가지는데, 무엇보다 미국의 공정이용 법리를 어느 정도 고려하면서도 그보다 엄격한 기준을 가지고 판단했다는 점에서 향후 공정이용이 적용되는 영역이 그만큼 줄어들 것임을 짐작하게 하는 판결이기도 하다.¹⁴⁾

(3) 뉴스콘텐츠의 특징과 공정이용 적용의 한계

① 생성형 AI의 학습 메커니즘과 뉴스콘텐츠의 고유한 가치

법리적 판단에 앞서, 생성형 AI 산업에서 뉴스콘텐츠가 갖는 기술적·산업적 중요성을 짚어볼 필요가 있다. 생성형 AI, 특히 대규모언어모델(LLM)은 방대한 텍스트의 통계적 패턴을 분석하여 확률적으로 단어를 예측하는 모델이다. 따라서 AI의 성능은 알고리즘뿐만 아니라 학습용데이터의 ‘양(quantity)’과 ‘질(quality)’에 크게 좌우된다.

이러한 맥락에서 뉴스콘텐츠는 대체 불가능한 가치를 지닌다. 첫째, 정제되지 않은 블로그, 소셜미디어상의 데이터와 달리, 뉴스콘텐츠는 게이트키핑(gatekeeping)을 거쳐

14) 이일호 · 박윤석. (2024). 「공정이용 판단 기준의 구체화 – 대법원 2024. 7. 11. 선고 2021다272001 판결에 관한 평석 –」, *정보법학*, 28(3), 79–119.

검증된 사실을 기초로 하고 정확한 문법을 갖추고 있어 AI의 ‘환각(hallucination)’ 현상을 완화하는 데 필수적이다. 둘째, 육하원칙에 기반한 논리적 서술 구조는 AI의 문맥 파악 및 추론 능력 고도화에 기여한다. 또한 정치, 경제, 사회, 문화 등 여러 분야에 걸친 뉴스콘텐츠는 AI 모델이 다양한 문체와 어휘를 구사할 수 있도록 풍부한 소스를 제공한다.셋째, 실시간으로 업데이트되는 뉴스는 AI가 최신성을 유지하도록 돋는 유일한 소스다. 뉴스 데이터의 결여는 곧 AI 모델의 시의성과 정확성 저하로 직결된다.

따라서, AI 기업들의 뉴스콘텐츠 이용은 단순한 정보 습득을 넘어, 고비용이 투입된 고품질 데이터를 통해 자사 모델의 상업적 가치를 제고하는 과정으로 이해해야 한다. 이는 기술 혁신의 일환으로 볼 수도 있으나, 정당한 대가 없는 이용이 계속될 경우 타인의 성과에 무단 편승한다는 비판에서 자유롭기 어렵다. 앞서 살펴본 대법원 판결에서 “이용의 경위나 방법 등과 같이 위 각 호에서 열거하지 않은 사항이라도 판단 요소로 고려할 수 있다.”고 판시한 점을 고려한다면 더욱 그러하다.

② 제1요소 : 상업적 목적과 원저작물 대체

생성형 AI의 뉴스콘텐츠 학습은 순수한 학술·연구나 교육 등을 위한 것이 아니다. 그 목적은 명백히 상업적 서비스 개발을 위한 필수적 과정에 가깝다. 영리 목적상 AI 기업들의 서비스 제공은 유료로 이루어질 가능성이 높고, 초기에는 무료라 하더라도 장기적으로는 유료 전환을 고려하는 것이 대부분이다.

한편, 대규모언어모델(LLM)은 뉴스콘텐츠의 독창적인 표현(문체, 서술구조 등) 자체를 학습하여 인간처럼 유려한 문장을 구사하는 능력을 획득한다. 이는 저작권법이 보호하는 창작적 표현 형식을 그대로 이용하는 것이며, 단순한 정보 분석을 넘어서선다. 따라서 학습 단계에서부터 저작물의 본질적 가치를 침해한다고 볼 여지가 많다.¹⁵

15) AI는 수십만 개 토큰에 달하는 예시들을 학습하면서 단어, 문장, 문단, 문서 수준에서 언어가 어떻게 선택되고 배열되는지, 즉 언어 표현의 본질까지 흡수한다. 이를려, AI 학습 단계에서 ‘언어들 간의 패턴’만을 이용하는 것이 아니며, 원저작물 전체를 복제하되 그 속에서 언어들 간의 패턴을 기술적으로 추출해내는 것이다. 기계학습 후 저장장치에 남아있는 결과물(예컨대 언어들 간의 패턴)을 기준으로 ‘변형적 이용’에 해당하거나 ‘비표현적 이용’에 해당한다는 시각은 다소 형식적인 접근으로 생각되며, 실질적으로는 원저작물의 창작적 표현 구조를 그대로 내재화하여 이를 대체하거나 모방할 목적을 지니고 있으므로, 원저작물에 새로운 의미나 메시지를 부기하는 ‘변형적 이용’에 해당한다고 보기 어렵다.

아울러 생성형 AI가 뉴스콘텐츠를 학습하여 산출하는 결과물은 뉴스의 핵심 기능인 ‘정보의 전달’, ‘사실 확인 및 검증’, 그리고 ‘비판과 제언’ 등의 역할을 그대로 수행할 수 있다. 결과적으로 뉴스콘텐츠를 새로운 가치로 변형하기보다, 뉴스콘텐츠의 기능을 그대로 대체할 가능성이 높다는 점에서 공정이용 인정에 신중을 기할 필요가 있다.

③ 제2요소 : 고품질 및 고신뢰도의 창작물

일각에서는 뉴스콘텐츠가 주로 사실(fact)을 다루므로 다른 저작물에 비해 보호 범위가 좁다고 주장하나, 이는 뉴스콘텐츠 생산의 특수성을 간과한 시각이다. 뉴스콘텐츠는 단순한 사실의 나열이 아니다. 수많은 정보 중 무엇을 보도할지 선택하고(gatekeeping), 어떤 관점에서 분석하며, 어떻게 배열할지 결정하는 과정에는 언론사의 고도화된 기획력과 막대한 자본, 기자의 사상과 감정이 투입된다. 즉, 뉴스콘텐츠는 막대한 인적·물적 자원을 투입하여 생산한 고도의 창작물이다.

한편, 뉴스콘텐츠는 다른 저작물과 달리 공표를 당연한 전제로 하고, 언론사가 대중에게 공표하여 경제적 가치를 창출하는 것을 본래 목적으로 하는 창작물이므로, 공표된 저작물이라는 점이 AI 기업들에게 유리하게 작용한다고 보기는 어렵다.

④ 제3요소 : 대량 이용, 전체 복제와 핵심 가치 추출

생성형 AI의 뉴스콘텐츠 학습은 기술적으로 대량의 기사를 이용하게 되고 기사 전체를 복제하여 저장하고 가공하는 과정을 수반한다. 양적인 측면에서 일반적인 이용의 범위를 한참 넘어선다.

양적인 측면뿐만 아니라 질적인 측면에서도, 생성형 AI는 기사의 핵심인 육하원칙에 따른 사실관계나 인과관계 분석 등을 추출하여 학습한다. 이는 저작물의 본질적 가치를 상당 부분 차용하는 것으로, 단순히 기술적 필요성만으로 정당화하기에는 무리가 있다.

⑤ 제4요소 : 현재 시장 및 잠재적인 시장의 저해

공정이용 적용 여부 판단에서 가장 중요하게 고려되는 요소 중 하나인 제4요소는 저작물의 이용이 원저작물 또는 원저작물의 2차적 저작물에 대한 현재 시장의 수요나 장래 개발될 합리적인 개연성이 있는 통상적인 시장의 수요를 대체하거나 그 시장가치를 훼손할

우려가 없거나 적은지를 기준으로 한다. 이처럼 공정이용 판단의 핵심인 ‘시장에 미치는 영향’ 측면에서 AI의 뉴스콘텐츠 학습은 심각한 우려를 낳고 있다.

우선 AI가 뉴스콘텐츠를 학습하는 행위는 단순히 내부 성능 향상에 그치지 않고, 결과적으로 원저작물인 뉴스콘텐츠의 수요를 직접적으로 잠식하는 결과를 초래하고 있다. 대표적인 현상이 바로 ‘제로 클릭(zero Click)’이다. AI는 뉴스콘텐츠를 학습하여 사실관계와 맥락을 완벽하게 파악한 뒤, 사용자 질문에 대해 완성된 답변을 생성한다. 이렇게 AI가 뉴스콘텐츠의 내용을 학습하여 직접 제공하게 되면, 사용자는 굳이 원본 뉴스 사이트를 방문하여 기사를 읽을 필요가 없어진다. 즉, 뉴스콘텐츠 학습이 AI로 하여금 뉴스콘텐츠의 ‘대체재’ 역할을 수행하게 만들어, 언론사의 트래픽을 급감시키고 핵심 수익원인 광고 또는 구독 모델 등을 붕괴시키는 원인으로 작용하는 것이다. 이는 저작물을 학습하여 원저작물의 현재 시장 수요를 직접적으로 대체하는 것으로, 공정이용이 허용될 수 없는 요인으로 작용한다.

나아가 잠재적인 시장의 관점에서도 문제가 제기된다. 이미 뉴욕타임스(NYT), AP통신, 뉴스 코퍼레이션 등 글로벌 미디어 그룹은 주요 AI 기업들과 연간 수천만 달러 규모의 뉴스콘텐츠 학습용 라이선스 계약을 체결하고 있다. 이는 뉴스콘텐츠가 AI 학습용데이터로서 거래될 수 있는 ‘통상적인 시장’이 이미 형성되었음을 시사한다. AI 기업이 대가 없이 뉴스를 학습하는 것은, 저작권자가 정당하게 누려야 할 이 거대한 잠재적 시장의 기회를 원천 봉쇄하는 것이다. 앞서 살펴본 대법원 판결이 ‘장래 개발될 합리적인 개연성이 있는 시장’을 고려 요소로 명시한 점을 감안할 때, AI 기업의 무단 이용은 저작권자가 정당하게 향유해야 할 잠재적 라이선스 시장의 기회를 저해할 소지가 크다.

3) 소결

요컨대 생성형 AI의 뉴스콘텐츠 학습은 기술적 혁신이라는 명분에도 불구하고 현행법상 공정이용 법리를 일률적으로 적용하기에는 무리가 있어 보인다. 뉴스콘텐츠는 AI 모델의 성능을 좌우하는 핵심 자산임과 동시에 막대한 비용이 투입된 창작물이며, AI 서비스는 결과적으로 뉴스의 기능을 대체하여 기존 미디어 시장의 수익 구조를 위협하고 있기 때문이다. 아울러 최근 데이터 라이선스 시장이 구체화되고 있는 현실을 고려할 때, 정당한 대가 없는 무단 이용은 저작권자의 잠재적 시장 이익을 침해하는 행위로 판단될 가능성이

높다. 따라서 생성형 AI 시대의 저작권 균형점은 일방적인 면책이 아니라, 뉴스콘텐츠의 기여도를 정당하게 평가하고 그에 상응하는 대가를 지급하고 이용 관계를 명확하게 정립하는 방향에서 모색되어야 할 것이다.

3. 텍스트 및 데이터 마이닝(TDM) 면책 규정 도입의 문제점

1) TDM 면책의 입법 취지와 생성형 AI의 현실적 고리

TDM 면책 규정은 당초 연구 목적의 데이터 분석을 장려하기 위해 고안된 것으로, 원저작물의 표현을 재현하거나 대체하는 생성형 AI의 학습을 염두에 둔 것이 아니다. 학계에서는 세밀한 입법적 검토 없이 TDM 면책 조항을 조급하게 도입할 경우, 모든 저작물이 AI 학습을 위한 단순 재료로 전락하여 창작 의지를 꺾고 문화 산업의 기반을 훼손할 수 있다는 우려를 제기한다.

특히 뉴스콘텐츠의 경우, 단순한 데이터의 집합이 아니라 시의성과 사실 겹중이 결합된 고부가가치 저작물이다. EU의 DSM 지침(Directive on Copyright in the Digital Single Market)조차 상업적 목적의 TDM에 대해서는 권리자의 거부권(opt-out)을 보장하고 있음을 상기할 때, 상업적 AI 모델이 뉴스콘텐츠를 통째로 학습하여 원저작물과 경쟁하거나 이를 대체하는 서비스까지 면책하려는 시도는 TDM의 본래 취지를 벗어난 것이다.

2) TDM 면책 규정의 한계와 부작용

① 학술연구와 상업의 경계 모호성 및 법적 불확실성

TDM 면책 입법은 명확성을 제고하고 예측가능성을 높인다는 명분을 내세우나, 실제로는 법적 불확실성을 가중시킬 소지가 크다. ‘비상업적·학술연구 목적’, ‘적법한 접근’, ‘옵트아웃’ 등 면책과 관련된 여러 조건들은 그 자체로 새로운 해석 분쟁의 씨앗이 될 수 있다.

TDM 면책 논의에서 가장 대표적으로 거론되는 ‘비상업적 목적’ 내지 ‘학술연구 목적’이라는 면책 요건 자체가 현대 AI 산업의 생태계를 전혀 반영하지 못하는 것일 수 있다. 생성형 AI 개발에 투입되는 막대한 비용은 순수한 학술 예산만으로는 감당하기 어려워, 대

학·연구소의 기초 연구 단계에서부터 영리 기업의 자금 지원이나 산학협력이 관여하는 것이 일반적이다. 더욱이 초기에 비상업적 목적으로 개발된 AI 모델이나 기술이 추후 영리 기업에 판매되거나 라이선스되는 기술 이전은 AI 산업의 핵심적인 성장 동력이기도 하다.

사실 ‘비상업적 목적’이라도 언론사의 뉴스콘텐츠를 무료로 이용하는 것이 당연하다고 보기는 어렵다. 그런데 초기 연구 목적으로 개발된 모델이 추후 상업화되거나 기술이 전되는 경우가 빈번한 현실에서, TDM 면책 규정이 도입된다면 ‘학술연구 목적’이라는 명분으로 뉴스콘텐츠를 무상 확보한 후 그 성과가 영리적으로 귀결되는 우회로가 열릴 수 있다. 이는 결국 ‘비상업성’의 범위를 둘러싼 소모적인 법적 분쟁을 야기할 것이다.

② 기술 중립성과 유연한 대처의 어려움

우리 저작권법은 이미 일반적·포괄적 저작재산권 제한 사유인 공정이용(Fair Use) 규정을 통해 기술 변화에 유연하게 대응할 수 있는 법적 토대를 갖추고 있다. 생성형 AI 기술은 대규모언어모델(LLM)을 넘어 예측 불가능한 속도로 진화하고 있다. 현 시점의 기술적 이해에 기반하여 구체적인 TDM 면책 요건을 입법화하는 것은, 급변하는 기술 환경을 반영하지 못하고 도리어 미래의 법적 판단을 경직시키는 결과를 초래할 수 있다.

③ 뉴스 미디어의 투자 유인 저해 및 시장 실패 우려

TDM을 지지하는 논리는 종종 학습용데이터를 마치 공기나 물처럼 누구나 채굴할 수 있는 ‘천연자원’으로 취급하는 오류를 범한다. 그러나 AI 학습에 사용되는, 특히 뉴스콘텐츠와 같은 데이터는 자연 상태의 것이 아니다. 이는 언론사가 막대한 자본, 인력, 시간을 투입하여 취재, 검증, 편집, 가공한 창작물이다.

저작권 제도의 핵심은 창작자에게 배타적 권리를 부여하여 창작 활동에 대한 투자를 유도하는 인센티브를 제공하는 데 있다. 그러나 TDM 면책은 이러한 인센티브 구조를 근본적으로 약화시켜 장기적으로 시장 실패를 야기할 위험이 있다. 특히 뉴스콘텐츠가 AI 학습용데이터로서 거래되는 새로운 시장이 형성되고 있는 시점에서, 면책 규정 도입은 저작권자가 정당하게 누려야 할 신규 시장의 이익을 원천적으로 배제하여 고품질 저작물 생산을 위축시키는 시장 실패를 초래할 수 있다.

④ 데이터 주권 및 글로벌 협상력 약화

만약 우리나라가 성급하게 포괄적인 TDM 면책을 도입한다면, 이는 국내의 고품질 한국어 뉴스콘텐츠를 글로벌 빅테크 기업들에게 대가 없이 제공하게 되는 결과를 낳을 수 있다. 글로벌 기업들이 해외에서는 뉴스 사용료를 지불하면서 국내에서는 면책 규정을 근거로 무임승차한다면, 이는 국내 콘텐츠 산업의 역차별 문제뿐만 아니라 데이터 주권과 글로벌 협상력을 스스로 포기하는 결과로 이어질 수 있다.

3) 소결

종합하면, 현재 상황에서 별도의 TDM 면책 규정을 신설하는 것은 AI 기업들에게 과도한 특혜를 부여하는 것이자 입법의 실효성 측면에서도 의문이 제기된다. 문화체육관광부 주관 협의체 등에서 TDM 면책 도입에 신중한 입장은 보이는 것 또한 이러한 우려를 반영한 결과로 이해된다.

AI 기업들의 예측가능성 확보라는 명분이 타인의 지식재산권을 무상으로 이용하는 것을 정당화하는 근거가 될 수는 없다. 현재 AI 개발·활용 등과 관련하여 제기되고 있는 여러 문제들이나 분쟁들은 근본적으로 상업적 이익을 위해 타인의 저작물을 무상으로 이용하려는 사업자가 감수해야 하는 법적·경영적 위험일 뿐이다. 이러한 위험부담을 창작 생태계에 전가하여 면책의 근거로 삼으려는 주장은 본말이 전도된 것이라 할 수 있다.

뉴스콘텐츠는 AI 산업의 핵심 원료이자 보호받아야 할 지적 자산이다. 따라서 향후의 논의는 일방적인 면책 입법보다는, 투명한 학습용데이터 공개를 전제로 뉴스콘텐츠의 가치에 부합하는 합리적인 보상 체계와 상생 모델을 구축하는 방향으로 나아가야 할 것이다.

4. 생성형 AI 기업과 미디어 창작자의 상생 발전을 위한 제언

1) 대가 지급의 당위성과 필요성

생성형 AI 학습과 관련하여 논의되고 있는 ‘공정이용’이나 ‘TDM 면책’ 논리는 복잡한 법리적 수사를 동원하고 있지만, 그 본질은 결국 “고비용이 투입된 고품질 뉴스콘텐츠를

대가 없이 AI 학습에 이용하겠다”는 경제적 유인으로 귀결된다. 그러나 AI와 미디어의 견전한 공존을 위해 정당한 대가 지급은 선택이 아닌 필수적인 선결조건임을 이해하여야 한다.

첫째, 경제적 정의의 실현이다. AI 기업이 천문학적인 가치를 지닌 AI 모델을 개발하면서, 그 핵심 원료인 뉴스콘텐츠를 무료로 사용하는 것은 타인의 투자와 노력에 무임승차하는 행위일 수 있다. 혁신이라는 미명 하에 정당한 대가를 지불하지 않는 것은 자본주의 시장 경제의 기본 원칙을 훼손하는 것이다.

둘째, AI 생태계 자체의 지속가능성 담보다. 뉴스콘텐츠는 끊임없이 변화하는 세상의 정보를 검증하여 공급하는 ‘마르지 않는 샘’과 같다. 만약 언론사가 수익 악화로 취재 역량을 상실하여 양질의 기사 공급이 중단된다면, AI는 학습할 원시 데이터를 그만큼 잃게 된다. 즉, 뉴스콘텐츠에 대한 정당한 대가 지급은 언론사 돋기가 아니라, AI 기업 자신의 생존과 미래를 위한 필수적인 재투자 비용으로 인식될 필요가 있다.

셋째, 글로벌 스탠다드에 부합하는 길이다. 이미 오픈AI(OpenAI), 구글(Google) 등 글로벌 빅테크 기업들은 주요 언론사들과 수천만 달러 규모의 라이선스 계약을 체결하며 ‘뉴스콘텐츠에 대한 대가 지급의 당위성과 필요성’을 인정하고 있다. 법적 불확실성을 이유로 대가 지급을 거부하는 것은 글로벌 흐름에逆行하는 것이며, 장기적으로는 국내 콘텐츠 시장의 고립을 자초할 수 있다.

2) 학습용데이터 공개 의무 도입과 계약·분쟁 실무상 합의

2026년 1월 22일 시행을 앞둔 인공지능기본법은 고영향 AI에 대한 영향평가와 안전성 확보 의무, 투명성 고지 의무 등을 규정하고 있으나, 학습용데이터의 출처나 구성에 대한 공개 의무는 별도로 규정하지 않는다. 이는 EU AI Act 또는 미국 Generative AI Copyright Disclosure Act of 2024에 명시된 범용 AI 모델에 대한 학습용데이터 공개 및 저작권 준수 의무와 비교할 때 상당히 미흡한 수준이라 할 수 있다.

AI 학습용데이터의 ‘블랙박스(Black-box)’ 상태를 방지하는 것은 다음과 같은 심각한 법적·산업적 문제를 야기하므로, 법적 강제를 통한 정보 확보가 시급하다.

① 정보 비대칭에 따른 계약 협상력의 무력화

협상의 대전제는 정보의 대등함이다. 그러나 현재는 AI 사업자가 어떤 데이터를 열

마나 학습했는지 철저히 비공개로 일관하고 있어, 저작권자는 자신의 권리가 침해되었는지 조차 인지하기 어렵다. 이러한 극심한 정보 비대칭 상황에서는 권리자는 라이선스 계약 체결 또는 이용료 산정을 제안, 요구, 청구할 수 있는 유인이나 실익을 상당 부분 상실하게 되며, 이는 구조적으로 권리자의 협상력을 무력화시켜 공정한 시장가격 형성과 데이터 라이선스 시장의 지속가능성을 저해한다.

또한, 학습용데이터의 이용내역이 불투명할 경우, 저작권 관련 단체·기관들이 통합적 권리 관리 또는 협력적 라이선스 규제 체계를 개발하는 데 실질적 제약이 발생한다. 이는 향후 AI 산업계와 콘텐츠 산업계 간의 산업적 협력 생태계를 구축하는 데에도 큰 장애 요인이 될 수 있다.

② 입증책임의 부담과 사법적 구제의 형해화

AI 학습과 관련된 저작권 분쟁에서 나타나게 될 실무상 문제는 입증책임의 문제다. 학습용데이터가 공개되지 않으면, 실제 소송 단계에서 저작권자가 AI의 권리침해 사실을 입증하거나 ‘의거성’과 ‘실질적 유사성’을 입증하는 것 등이 매우 어려워진다. 이는 결과적으로 AI 기업에게는 손쉬운 면책을 부여하고, 권리자에게는 침해 입증이라는 넘을 수 없는 장벽을 세우는 꼴이다. 학습용데이터 공개 의무는 기울어진 운동장을 바로잡고 실효성이 있는 사법적 구제를 가능케 하는 최소한의 장치다.

③ 불법적 데이터 수집 및 독과점 구조의 고착화

학습용데이터의 공개가 전제되지 않으면 크롤링 등 부정한 방법으로 수집된 데이터가 세탁되어 AI 학습에 활용되더라도 이를 외부에서 감시하거나 통제할 수 없다. 이는 AI 기업에 의한 불법적인 대규모 데이터 수집·이용을 사실상 합법화시키는 결과로 이어질 수 있다.

또한, 자본력을 갖춘 소수 AI 기업들이 데이터를 독점적으로 구축하고 은폐할 경우, 신규 진입자는 배제되고 시장지배력은 더욱 공고해지는 독과점 구조가 심화될 우려가 크다. 이는 시장 감시 기능을 마비시키고 콘텐츠 생태계를 왜곡시키는 결과를 낳는다. 특히, 플랫폼 산업이 급속도로 성장하고 소수의 플랫폼 사업자가 강력한 시장지배력을 지니고 있는 우리나라 시장의 특성상 그와 같은 독과점적 구조가 쉽게 고착화될 가능성이 높다. 결국 미디어 기타 콘텐츠 산업의 생태계 자체를 왜곡시킬 수 있다.

따라서 학습용데이터의 출처와 구성 내역을 일정 요건 하에서 상세히 공개하도록 의무화하는 제도의 도입은 단지 정보의 공개를 넘어서 계약의 투명성과 협상의 공정성을 확보하고, 사후 분쟁 발생 시 입증부담의 형평성을 회복하기 위한 필수적 수단으로 생각된다.

5. 맷음말

생성형 AI의 뉴스콘텐츠 학습에 대해서는 공정이용 인정에 신중을 기할 필요가 있다. 뉴스콘텐츠는 AI가 대체할 수 없는 고유한 사실 검증 가치를 지닌 창작물이며, 이를 대가 없이 이용하여 결과적으로 원저작물의 현재 및 잠재적 시장을 잠식하는 행위는 법적으로 정당화되기 어렵다.

AI 산업의 지속 가능한 발전은 창작 생태계와의 공존을 전제로 한다. 뉴스 미디어의 수익성 악화로 양질의 기사 생산이 중단된다면, AI 역시 고품질 학습용데이터를 확보하지 못해 성능 저하에 직면하게 될 것이다. 따라서 뉴스콘텐츠에 대한 정당한 가치 인정과 보상은 미디어 산업의 보호를 넘어 AI 산업 자체의 안정적인 성장을 위한 필수 조건이다.

문화체육관광부는 최근 생성형 AI 학습 관련 공정이용 안내서(가칭)를 연내 발간할 예정임을 알렸다. 이번 발표가 ▲공정이용 규정의 엄격한 해석 ▲학습용데이터의 투명한 공개 ▲뉴스콘텐츠 이용에 대한 합리적 보상 기준 마련 등 현행법의 원칙이나 정책 방향성을 재확인하고, 기울어진 운동장을 바로잡는 실질적인 계기가 되었으면 하는 바람이다.

기술의 진보가 타인의 권리를 침해하는 방식으로 이루어져서는 안 된다. 진정한 혁신은 미디어 산업계의 콘텐츠를 단순한 데이터가 아닌 보호받아야 할 저작물로 존중하고, 투명한 이용 관계 속에서 정당한 대가를 지급하는 공정한 법적 질서 위에서만 가능할 것이다.