

2026. 1.

## 미국 풀리처상 수상 작가 등 6인, 주요 6개 AI 기업 상대로 저작권 침해 소송 제기

충남대학교 법학전문대학원 교수  
이철남

### 1. 개요

2025년 12월 22일, 풀리처상을 수상한 탐사보도 전문 기자 존 카레이루(John Carreyrou)를 포함한 6명의 작가는 미국 캘리포니아 북부 연방지방법원에 앤스로픽(Anthropic), 메타(Meta), 구글(Google), 오픈AI(OpenAI), xAI, 퍼플렉시티(Perplexity) 등 미국의 주요 AI 기업 6곳을 상대로 저작권 침해 소송을 제기했다. 원고들은 소장을 통해 피고 기업들이 ‘립젠(LibGen)’, ‘Z-라이브러리(Z-Library)’, ‘비블리오틱(Bibliotik)’ 등으로 불리는 이른바 ‘새도우 라이브러리(Shadow Library)’—저작권자의 허락 없이 서적을 불법으로 공유하는 해적판 사이트—를 통해 자신들의 저작물을 조직적이고 고의적으로 탈취하여 거대언어 모델(LLM)을 학습시켰다고 주장한다. 이들은 미국 저작권법에 따른 법정 손해배상을 청구하며, 피고들의 행위가 고의적이었으므로 작품당 최대 15만 달러의 배상을 요구하고 있다.

### 2. 주요내용

#### 1) 원고 및 관련 저작물

이번 소송을 제기한 6명의 원고는 단순한 저작권자가 아니라, 오랜 기간 심층적인 조사와 연구, 창의적인 집필 과정을 통해 고품질의 텍스트를 생산해 온 전문 작가들이다. 이들의 작품은 논리적 구조, 긴 서사 호흡, 정확한 문장력을 갖추고 있어, AI 모델이 인간 수준의 텍스트 생성 능력을 갖추기 위해 필수적인 학습 데이터로 볼 수 있다. 예를 들면, 존 카레이루는 전 월스트리트저널(WSJ) 탐사보도 전문 기자이자, 풀리처상을 두 차례 수상한 저널리스트이다. 대표작으로는 『Bad Blood: Secrets and Lies in a Silicon Valley Startup』(2018) 등이 있으며, 그의 책은 LibGen 및 Z-Library 등의 해적판 사이트에서 불법 공유되고 있다.

#### 2) 피고 및 주요 혐의

본 소송은 단일 기업이 아닌, 현재 생성형 AI 시장을 장악하고 있는 6개의 핵심 기업을 모두 포함하고 있다. 원고들은 이들 기업이 서로 경쟁하면서도, 불법 데이터 활용에 있어서는 동일한 침해 카르텔을 형성하고 있다고 본다.

소장에는 각 기업이 불법 데이터를 확보한 구체적인 정황이 제시되어 있다. 예컨대, 앤스로픽이 사용한 것으로 의심받는 데이터셋 ‘The Pile’에는 ‘Books3’라는 하위 데이터셋이 포함되어 있는데, 이는 해적판 사

이트인 ‘비블리오틱(Bibliotik)’에서 추출한 약 19만 6천 권의 도서 텍스트로 구성되어 있다.

메타는 ‘Books3’ 데이터셋 사용을 공식 논문에서 인정한 것에 더해, 회사 내부망을 이용해 비트토렌트 프로토콜로 LibGen, Z-Library 등의 전체 사본을 직접 다운로드했다는 의심을 받는다. Kadrey v. Meta 소송<sup>1)</sup> 과정에서 공개된 내부 이메일 등에 따르면, 메타의 변호사들이 해적판 데이터 사용의 법적 위험성을 경고했음에도 불구하고, 엔지니어들과 경영진이 이를 무시하고 다운로드한 것으로 드러났다.

구글의 ‘Gemini’와 ‘Imagen’ 모델은 ‘C4(Colossal Clean Crawled Corpus)’ 데이터셋을 주요 학습원으로 사용했다. 구글은 이 데이터셋이 ‘정제된(Clean)’ 것이라고 주장하지만, 원고들은 C4 데이터셋 분석 결과 Z-Library와 같은 해적판 사이트에서 크롤링된 텍스트가 다수 포함되어 있음을 지적한다. 워싱턴포스트(Washington Post) 등의 분석 보도에 따르면, C4 데이터셋을 구성하는 도메인 중 하나가 미 법무부에 의해 폐쇄된 Z-Library 도메인임이 밝혀졌다. 구글은 이를 알면서도 필터링하지 않고 모델 학습에 투입했다는 혐의를 받는다.<sup>2)</sup>

OpenAI는 GPT-3 논문에서 학습 데이터로 ‘Books1’과 ‘Books2’를 사용했다고 밝혔다. ‘Books1’은 프로젝트 구텐베르크(Project Gutenberg)와 같은 공개 도서로 추정되지만, ‘Books2’의 출처에 대해서는 철저히 함구해왔다. 원고들은 데이터의 규모와 특성을 분석할 때, ‘Books2’가 LibGen이나 Z-Library 등에서 수집된 것으로 의심한다.

xAI가 개발한 ‘Grok’ 모델은 학습 데이터를 전혀 공개하고 있지 않다. 그러나 원고들은 Grok 모델 자체가 자신의 학습 데이터 출처를 실토하는 현상을 포착했다. 소장에는 Grok 모델 스스로가 사용자 대화에서 “LibGen과 같은 해적판 사이트에서 데이터를 진공청소기처럼 빨아들였다”고 자백한 사례가 증거로 제시되었다.

퍼플렉시티는 AI 검색 서비스를 제공하면서, 웹사이트 소유자가 설정한 ‘Robots.txt’를 무시하고 무단으로 콘텐츠를 스크레이핑했다는 혐의를 받는다. 또한, RAG 기술을 통해 원고의 책 내용을 요약해 달라는 요청에 대해, 책을 구매하지 않은 사용자에게도 상세한 장(Chapter)별 요약을 제공함으로써 원작의 시장 수요를 대체하고 있다고 비판받고 있다.

### 3) 청구 내용

원고들은 피고들이 원고들의 저작권을 침해했다는 점, 해당 저작권 침해 행위가 고의적(Willful)이었음을 법원이 선언해 줄 것을 청구했다. 아울러 피고들이 저작권 침해 행위를 지속하지 못하도록 하는 영구적인 금지 명령을 내려 줄 것을 요청했다.

1) Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1041 (N.D. Cal. 2025) 참조.

2) Kevin Schaul et al., Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart, THE WASHINGTON POST (Apr. 19, 2023), <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>

손해배상액과 관련하여, 원고들은 배심원 재판을 통해 저작권법상 고의적 침해에 대한 법정 최고액인 작품당 15만 달러의 손해배상을 청구하고 있다. 2024년 하반기, 앤스로픽은 작가들과의 집단소송(Bartz v. Anthropic)에서 약 15억 달러 규모의 잠정 합의안을 도출했었다. 이 합의안에 따르면, 실제 작가들에게 돌아가는 보상금은 작품당 약 3,000달러 수준이다. 원고들은 이를 두고 빅테크 기업들이 수천억 달러 가치의 모델을 만드는 데 기여한 저작물의 대가로는 터무니없이 적은 액수로 보고 개별 소송을 진행한 것이다.

### 3. 결론 및 시사점

이번 소송은 학습에 사용된 데이터 자체가 해적판 사이트에서 유출된 불법 복제물이라는 점을 정면으로 겨냥하고 있으며, 특히 거대언어모델 저작권 분쟁의 초점이 '출력물의 변형성'에서 '학습 데이터 취득 경로의 불법성'으로 이동했음을 보여주고 있다. 아무리 AI 기술이 혁신적이라 하더라도, 그 기초가 되는 데이터를 LibGen, Z-Library 등의 해적판 사이트에서 확보했다면 이는 저작권 침해로 볼 가능성이 크다. 또한 AI 기업들이 해당 사이트의 불법성을 인지하고 있었거나, 내부적으로 위험성을 경고하는 목소리가 있었음에도 불구하고 학습을 강행했다는 정황은 '고의적 침해(Willful infringement)'를 입증하는 강력한 근거가 될 수 있다.

AI 개발사들로서는 데이터 수집 관행에 대한 전면적인 재검토가 요구된다. 단순히 인터넷에 공개된 데이터를 긁어오는 방식에서 벗어나, 데이터 수집의 투명성과 적법성을 우선 과제로 삼아야 한다. 기업 내에서 데이터 소싱에 대한 엄격한 기준을 세우고, 불법성이 의심되는 데이터셋은 학습에서 제외하는 리스크 관리가 필수적이다. 아울러 법적 분쟁을 피하기 위해 출판사, 작가 단체, 뉴스 매체 등과 정당한 라이선스 계약을 맺는 것을 고려할 필요도 있다.

### 참고자료

- BloombergLaw, "OpenAI, Anthropic, xAI Hit With Copyright Suit from Writers".
- Publishers Weekly, "Authors File New Lawsuit Against AI Companies Seeking More Money".
- Carreyrou v. Anthropic PBC et al., Complaint (N.D. Cal. Dec. 22, 2025).
- Bartz v. Anthropic PBC, 3:24-cv-05417 (N.D. Cal. Mar. 27, 2025).
- Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1041 (N.D. Cal. 2025).