SHIN&KIM
법무법인(유) 세종

‘생성형 AI의 학습 데이터 사용 및 출력물에 관한 저작권 침해’를 인정한 최근 독일 판결(GEMA v. OpenAI)

2025.12.02

1. 들어가며

저희 법무법인(유) 세종의 2025. 7. 10.자 뉴스레터 「AI 개발을 위한 저작물 학습과 공정이용에 관한 최근 미국 판례 동향」에서 소개해 드린 바와 같이, 2025. 6.경 미국에서는 생성형 AI의 개발을 위하여 저작물을 학습시키는 행위가 공정이용(fair use)에 해당하는지 여부에 관한 판결이 잇따라 선고된 바 있습니다.

한편, 독일 뮌헨지방법원은 2025. 11. 11. GEMA v. OpenAI 사건에서, 「OpenAI가 운영하는 LLM의 학습·출력 과정에서의 노래 가사 이용은 저작권법상 복제권 및 공중이용 제공권 침해에 해당하고, 텍스트·데이터 마이닝(이하 ‘TDM’) 면책 규정의 적용 대상에도 포함되지 않는다」는 판결[1](이하 ‘OpenAI 판결’)을 선고하였습니다.

이번 뉴스레터에서는 위 ‘OpenAI 판결’의 주요 내용과 그 법적 의미를 살펴보고, 최근 선고된 미국 판결들과의 차이점도 살펴보고자 합니다.

2. OpenAI 판결의 주요 내용

이 소송에서 독일의 음악저작권 관리단체인 GEMA[2]는, GEMA가 관리하는 9개의 음악저작물(이하 ‘대상저작물’)의 가사를 OpenAI가 ChatGPT의 학습에 무단으로 사용하였고, GPT-4 및 4o 모델이 대상저작물의 가사를 기억하여 대상저작물과 거의 동일한 가사를 출력한다는 점을 지적하면서, 이러한 OpenAI의 대상저작물 이용 행위는, (i) 학습단계에서의 복제권(§ 16 UrhG[3]) 침해, (ii) 출력 단계에서의 공중이용 제공권(§ 19a UrhG)[4] 침해, (iii) 가사의 일부 왜곡 등에 따른 인격권 침해에 해당한다고 주장하였습니다.

이에 대하여 OpenAI는, LLM은 개별 학습 데이터를 저장, 복제하는 것이 아니라, 전체 데이터셋으로부터 통계적 패턴을 추출할 뿐이므로, 대상저작물 가사의 사본이 존재하지 않고, 출력 단계에서 침해가 있다고 보더라도 그 책임은 프롬프트를 입력한 사용자에게 있으며, 설령 침해 요소가 일부 존재하더라도 이는 TDM 면책 규정(§ 44b UrhG)에 따라 면책된다고 주장하였습니다.

독일 뮌헨지방법원(이하 ‘법원’)은 이러한 주장에 대하여 아래와 같이 판단하였습니다.

① 학습단계에서의 복제권 침해에 관한 판단: 먼저 학습단계에서의 복제권 침해와 관련하여, 법원은 대상저작물의 가사들은 OpenAI가 운영하는 Chatgpt-4 및 4o 모델 내부에서 **암기(memorisation)되어 재현 가능한 상태에 있다고 보았습니다**. 학습데이터가 암기되는 경우 모델에 포함되어 출력물로 다시 추출될 수 있는데, 이러한 암기 현상은 AI 모델이 학습 과정에서 단순히 학습데이터 세트로부터 일반적인 정보를 추출하는 데 그치지 않고, 학습 종료 후 설정된 파라미터에 개별 학습데이터를 전부 채택·고정하는 경우에 발생합니다.

법원은 이러한 암기 현상으로 인하여 대상저작물의 가사들이 AI 모델의 특정 파라미터 데이터에 체화(embodiment)되어 있다는 점에서 정보사회 저작권지침(InfoSoc Directive) 제2조[5] 및 저작권법상 복제 요건이 충족된다고 판단하였습니다.

② TDM 면책 규정(§ 44b[6] UrhG)의 적용 여부: OpenAI는 이러한 학습단계에서의 복제가 TDM 면책 규정의 보호 범위에 해당한다고 주장하였으나, 법원은 이 주장을 받아들이지 않았습니다. 법원은 TDM 면책 규정이 허용하는 것은 학습 데이터셋을 구축하는 과정에서의 기술적·준비적 복제에 한정되는데, 이 사건처럼 AI 모델이 학습과정에서 단순히 정보만 추출하는 것이 아니라 저작물을 그대로 복제하는 경우는 TDM 면책 규정이 예정한 범위를 벗어나며, 오히려 권리자의 저작재산권을 침해하는 행위를 구성한다고 판단하였습니다.

③ 부수적 이용 또는 묵시적 동의 인정 여부: OpenAI는 학습 데이터의 이용이 부수적 이용에 해당한다거나 묵시적 동의가 있었다는 주장도 하였으나, 법원은 훈련데이터 전체를 하나의 주저작물로 볼 수 없고, 그 안에 포함된 개별 가사를 단순한 부수적 요소로 평가할 수도 없다고 보아, **독일 저작권법 제57조(부수적 이용 예외)의 적용을 부정하였습니다**. 또한, AI 학습을 위한 저작물 이용은 권리자가 통상적으로 예상하거나 전제하는 정상적이고 예측 가능한 이용 형태(normal and expected use)에 해당하지 않으므로, 묵시적 동의 역시 인정되지 않는다고 보았습니다.

④ 출력단계에서의 저작권 침해 여부: 출력단계에서의 저작권 침해와 관련하여서는, 법원은 **사용자의 프롬프트에 대한 응답으로 사실상 동일한 가사가 출력되는 행위를 복제권 및 공중이용 제공권[7]에 대한 침해 행위로 평가하였습니다**.

⑤ 저작권 침해의 주체: 나아가 법원은 저작권 침해의 주체는 사용자가 아니라 OpenAI라고 보았는데, 그 근거로 (i) 문제된 출력이 단순한 프롬프트에도 반복적으로 생성된 점, (ii) OpenAI가 학습데이터를 선택하고 언어모델 아키텍처를 설계하였기 때문에 이로 인하여 발생한 암기 현상에 대해서도 이를 통제·관리할 지위에 있다는 점 등을 들었습니다. 법원은 이와 같은 사정을 종합하여, **출력물의 구체적 내용은 언어모델에 의해 실질적으로 결정되므로, 출력단계에서의 저작권 침해에 대한 책임은 사용자가 아니라 OpenAI에게 귀속된다고 보았습니다**.

⑥ 인격권 침해 여부: 한편, 가사 왜곡 등에 따른 인격권 침해 주장에 대하여, 법원은 2025. 9. 29. 이루어진 예비판정에서 이 사건에서 가사의 변형은 기술적 확률모델의 작동 과정에서 발생한 부수적 결과에 불과하고, 저작자를 비방하거나 명예를 훼손하려는 의도적 왜곡으로 보기 어렵다는 입장을 보였는데,[8] 본안 판결에서도 인격권 관련 주장은 기각되었습니다.

결론적으로 법원은 인격권 침해 주장을 제외한 GEMA의 대부분의 주장을 받아들이면서 (i) 언어모델 내부에서의 복제 및 출력 단계에서의 가사 재현을 하지 말라는 내용의 침해금지 청구, (ii) 침해 범위 및 손해 산정을 위하여 필요한 범위 내에서 학습데이터 이용 내역에 관한 정보를 제공하라는 정보제공 청구 및 (iii) 저작권 침해에 따른 손해배상 청구를 인용하는 판결을 선고하였습니다.[9]

3. 이번 판결의 의의

① 판단 기준: 공정이용 vs TDM 면책 규정

미국은 AI 개발과 관련하여 별도의 TDM 면책 규정을 도입하지 않았습니다. 이 때문에 최근 미국에서 선고된 앤트로픽[10], 메타[11], 웨스트로[12]판결(이하 '미국 판결')에서는 모두 공정이용(fair use)의 4요소를 종합적으로 고려하여 저작권 침해 여부를 판단하고 있습니다. 미국 판결에서는 LLM 학습을 위한 저작물 이용이 목적·성격 측면에서 변형적(transformative) 이용으로 평가될 여지가 있다는 점을 인정하면서도, (i) 불법 복제물의 사용, (ii) AI 학습용 라이선스 시장 침해, (iii) 원저작물과 실질적으로 유사한 출력물에 의한 간접적 시장 대체 효과 등이 있는 경우에는 공정이용에 해당하지 않을 수도 있다는 판단을 내린 바 있습니다.

반면, 이번 OpenAI 판결의 경우, 독일 저작권법에는 미국 저작권법 제107조나 우리나라 저작권법 제35조의5와 같은 취지의 공정이용 규정이 존재하지 않는 관계로, 아래와 같이 TDM 면책 규정의 적용 범위를 중심으로 저작권 침해 여부를 판단하였습니다.

② TDM 면책 규정의 적용 범위

독일 저작권법 제44b조는 TDM 면책 규정을 두고 있는데, 제1항에서는 TDM은 특히 패턴, 추세 및 상관관계에 대한 정보를 획득하기 위해 하나 이상의 디지털 또는 디지털화된 저작물을 자동으로 분석하는 것이라고 정하고 있으며, 제2항에서는 **TDM을 위해 합법적으로 접근할 수 있는 저작물에 대해서는 복제를 예외적으로 허용**하고 있습니다.

OpenAI 판결은 기계학습을 위한 모든 기술적 복제가 일률적으로 TDM 면책 범위에 포함되는 것이 아니라, **학습 과정에서 저작물이 재현 가능한 수준으로 암기되는 경우에는 그 적용 범위에서 벗어나 저작권 침해를 구성한다고 보았습니다**.

그간 생성형 AI 사업자들은 'LLM이 특정 데이터를 저장, 복제하지 않고 확률적 합성으로 출력물을 만든다'고 항변해 왔는데, OpenAI 판결이 그대로 확정되는 경우에는 향후 생성형 AI의 학습단계 내지 출력단계에서의 저작물의 복제권 침해여부 판단에 있어서 중요한 기준이 될 것으로 보입니다.

③ 출력 단계에서 AI 사업자에 대한 책임 인정

앞서 살펴본 미국 판결들은 주로 학습단계에서의 저작권 침해 여부에 초점을 두고 있어, 언어모델 생성물의 출력단계에서의 구체적인 책임 귀속에 대해서는 아직 명확한 법리가 제시되지 않은 상황입니다. 반면 OpenAI 판결의 경우, 출력단계에서 사용자가 단순한 프롬프트만 입력했음에도 원 저작물과 거의 동일한 가사가 반복적으로 생성되는 경우, 이를 복제권 및 공중이용 제공권 침해로 인정하면서, 그 침해의 주체를 프롬프트를 입력한 사용자가 아니라, AI 사업자로 판단하였습니다. AI 사업자가 학습데이터를 선택하고 모델 아키텍처를 설계 운영한다는 점에서 암기 위험을 통제할 지위에 있다고 보았기 때문입니다. 이러한 OpenAI 판결은 향후 출력단계에서의 저작권 침해 책임의 귀속 주체를 판단함에 있어 중요한 기준을 제시하였다고 볼 수 있습니다.

4. 향후 전망 및 시사점

이번 OpenAI 판결은 유럽에서 생성형 AI와 저작권 침해를 직접적으로 다룬 중요한 판결 중 하나로, 기계학습 과정에서 발생하는 암기 현상이 복제권 침해에 해당할 수 있고, 이는 TDM 면책 규정의 적용 대상에 해당하지 않는다고 판단하였다는 점에서 의미가 있습니다.

또한 OpenAI 판결은 AI 사업자에게 데이터 선정 및 모델 설계에 관한 책임을 명시적으로 인정함으로써, 인공지능 출력물에 따른 저작권 침해 책임이 누구에게 귀속되어야 하는지에 관하여 중요한 판단기준을 제시하였다는 점에 있어서도 의의가 있습니다.

일반적으로 AI 사업자는 사용자와의 관계에서 프롬프트 입력과 관련한 법적 책임을 사용자에게 귀속시키는 약관 규정을 두는 경우가 많으나, 위와 같은 판결 취지에 따르면 단순한 약관 규정만으로는 AI 사업자가 관련 책임에서 면책되기 어려울 것으로 보이는 바, OpenAI 판결을 계기로 AI 사업자가 출력물에 대한 저작권 침해의 책임을 어느 범위까지 부담하는지, 그리고 사용자와의 관계에서 책임 분담 기준을 어떻게 설정할 것인지에 관한 논의도 보다 활발해질 것으로 예상됩니다.

법무법인(유) 세종 콘텐츠엔터테인먼트 팀은 AI와 관련된 국내외 최신 판결 및 제도 개선 동향 등을 주시하면서 관련 법리를 심도 있게 연구해오고 있습니다. 또한 AI 저작권 이슈 검토, AI 활용에 관한 가이드라인 작성 등 자문 업무뿐만 아니라, 유관 기관들과 기업들에 대한 세미나와 강연도 활발하게 진행해오고 있는바, AI와 저작권 이슈에 있어 가장 선도적이고 효과적인 솔루션을 제시할 수 있는 역량을 갖추고 있습니다.

[1] LG München I, Urt. v. 11.11.2025 – 42 O 14139/24 (*GEMA v. OpenAI*)
[2] Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte(독일 음악저작권 관리 단체)
[3] 독일 저작권법(Urheberrechtsgesetz, 이하 'UrhG')
[4] 독일 저작권법 제19a조의 '공중이용제공권'은 저작물 등을 공중의 이용에 제공하는 것을 의미하며, 국내 저작권법상 '전송권'에 대응하는 개념입니다.
[5] Article 2(Reproduction right) Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part: 제2조(복제권) 회원국은 다음 각 호의 전부 또는 일부에 대하여, 그 형태나 방식의 여하를 불문하고, 직접적·간접적이거나 일시적·영구적인 복제를 허용하거나 금지할 수 있는 배타적 권리를 보장하여야 한다.
[6] 독일 저작권법(UrhG) 제44b조(텍스트 및 데이터 마이닝)
(1) 텍스트 및 데이터 마이닝은 특히 패턴, 추세 및 상관관계에 대한 정보를 획득하기 위해 하나 이상의 디지털 또는 디지털화된 저작물을 자동으로 분석하는 것이다.
(2) 텍스트 및 데이터 마이닝을 위해 합법적으로 접근할 수 있는 저작물의 복제가 허용된다. 복제품은 텍스트 및 데이터 마이닝에 더 이상 필요하지 않으면 삭제된다.
(3) 제2항 제1문에 따른 사용은 권리보유자가 유보하지 않은 경우에만 허용된다. 온라인으로 접근할 수 있는 저작물에 대한 사용 유보는 기계판독이 가능한 형태인 경우에만 유효하다.
[7] 법원은 OpenAI 측이 언어모델 안에 기억된 가사들에 대하여 공중의 접근을 허용하고 있고, 해당 가사들이 산출물로 다시 재현되고 있다는 점을 지적하고 있습니다.
[8] 한국저작권위원회, 저작권 동향 제16호, '뮌헨지방법원: GEMA vs. OpenAI 사건에서 생성형 인공지능의 저작권 침해 시사'(2025. 11.)
[9] 아직 판결은 확정되지 않은 상태이며, OpenAI는 제1심 판결에 대해 항소할 계획이라고 밝혔습니다.
[10] *Bartz v. Anthropic PBC*, 2025 WL 1741691 (N.D. Cal. June 23, 2025).
[11] *Kadrey v. Meta Platforms, Inc.*, 2025 WL 1752484 (N.D. Cal. June 25, 2025).
[12] *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc.*, 765 F. Supp. 3d 382 (D. Del. 2025).

관련구성원

김우균
변호사

02-316-4083
wgkim@shinkim.com

황지원
변호사

02-316-1799
jwhwang@shinkim.com

김소리
변호사

02-316-1907
srkim@shinkim.com

Copyright SHIN & KIM LLC. All rights reserved.