

미국

미국에서의 AI 저작권 소송(6)

고려대학교 법학전문대학원/교수
이대희

미국 내 생성형 AI 관련 주요 저작권 소송으로서 다음 사건들을 중심으로 검토한다. 각 사건은 제기 법원, 사건 번호, 제소일을 기준으로 정리하였다.

1. Gracenote Media Services, LLC v. OpenAI Foundation

뉴욕 남부 연방지방법원, No. 1:26-cv-01947, 2026. 3. 10.

2. Koon v. Google, LLC

일리노이 북부 연방지방법원, No. 1:26-cv-02582, 2026. 3. 4.

3. Thaler v. Perlmutter

DC 연방항소법원, 130 F.4th 1039, 2025. 3. 18.; DC 연방지방법원, No. 1:22-cv-01564, 2022. 6. 2.

4. Businnessing LLC v. Runway AI, Inc.

뉴욕 남부 연방지방법원, No. 1:26-cv-01655, 2026. 2. 27.

5. David Vance Gardner v. Runway AI, Inc.

캘리포니아 중부 연방지방법원, No. 2:26-cv-01941, 2026. 2. 23.

6. Youngblood v. Meta Platforms, Inc.

캘리포니아 북부 연방지방법원, No. 3:26-cv-01100, 2026. 2. 4.

7. Nicole Chmura v. Snap Inc.

캘리포니아 중부 연방지방법원, No. 2:26-cv-01732, 2026. 2. 18.

8. Ace Cam, Inc. v. Runway AI Inc.

뉴욕 남부 연방지방법원, No. 1:26-cv-01362, 2026. 2. 18.

9. Kleiner v. Adobe Inc.

캘리포니아 북부 연방지방법원, No. 5:26-cv-01218, 2026. 2. 9.

10. Youngblood v. Nvidia Corporation

캘리포니아 북부 연방지방법원, No. 5:26-cv-00916, 2026. 1. 29.

11. Concord Music Group, Inc. v. Anthropic PBC

캘리포니아 북부 연방지방법원, No. 5:26-cv-00880, 2026. 1. 28.

12. Ted Entertainment, Inc. v. Snap Inc.

캘리포니아 중부 연방지방법원, No. 2:26-cv-00754, 2026. 1. 23.

1. Gracenote Media Services, LLC v. OpenAI Foundation

- 뉴욕 남부 연방지방법원 (1:26-cv-01947, 2026.3.10.)

(1) 사실관계(원고 주장)

1. 원고 Gracenote Media Services는 미디어 콘텐츠 전달자로서 Gracenote Programs Database 및 그 데이터 등 강력하고 독창성 있는 메타데이터(metadata) 제품(콘텐츠를 설명하는 정보를 검색, 추천, 분석이 가능하도록 구조화시킨 데이터베이스 제품, 이하 ‘데이터’로 총칭)을 개발해 왔는데, 이는 AI 학습을 위해서도 매우 가치있는 데이터이다. Gracenote의 데이터는 수백만 개에 달하는 서술 방식에 의한 설명, 비디오 설명, 식별자 등의 요소들을 포함하고, 이들 요소들은 Gracenote의 편집자들이 서술, 편집, 선정, 배열한 것이다. 피고 OpenAI는 AI 제품을 통하여 Gracenote의 데이터를 복제 및 배포하였다.

2. Gracenote Programs Database는 반세기 이상에 걸쳐 TV쇼, 영화 및 이와 관련된 주제들에 대한 수백만 개의 프로그램의 요소들을 서로 연관시키고 배열시킨 관계형 데이터베이스(relational database, 키를 통하여 연결해서 관리하는 데이터베이스)이다. Gracenote Programs Database의 상호 연결된 관계형 구조는 복잡한 디지털 엔터테인먼트 생태계에서 콘텐츠 식별, 관리 권리, 성과 추적, 광고 최적화를 위한 표준화된 진정한 출처를 제공한다.

3. Gracenote는 Gracenote Programs Database를 저작권 등록하였다.

4. OpenAI는 GPT-3, 4, 5를 학습시키기 위하여 Gracenote의 데이터를 사용·복제하고, 이를 이용하여 LLM을 개발하여 LLM에 기반한 제품 및 서비스를 개발·운영·개선하고 제3자에게 제공하고, LLM 및 이러한 제품·서비스 이용자들에게 침해 콘텐츠를 제공하고, 학습데이터 이용허락 시장에서 Gracenote의 역할을 약화·침식하였다.

(1) 위반 주장 사항(원고)

1. 저작권 직접침해: 피고들은 Gracenote Data 또는 그 일부를 포함하는 복제물로 구성된 학습 데이터 셋을 구축함으로써, Gracenote의 저작물에 대한 배타적 권리를 직접 침해하였다.

2. 대위침해: OpenAI, LLC 및 일부 운영 법인들이 Gracenote Data를 복제·저장·학습·배포¹⁾하여 저작권을 직접 침해하였고, OpenAI Foundation 등 다른 피고들²⁾은 이러한 행위를 지시하거나 통제할 수 있는 권한을 보유하고 이로부터 발생하는 경제적 이익을 향유하였으므로, 대위책임의 요건을 충족한다.

3. 기여침해: GPT 기반 제품의 출력과 관련된 직접 침해의 주체가 최종 사용자(end-user)라고 할지라도, 피고들은 ① Gracenote Data 복제물을 이용자에게 배포할 수 있는 LLM 모델을 공동 개발하고, ② 해당 데이터를 이용하여 LLM을 구축·학습시키고, ③ RAG(검색 증강 생성)에 의하여 Gracenote Data에

1) 원고는 왜 배포가 이루어지는지 정확하게 언급하지 않고 피고가 Gracenote Data의 복제물을 이용자에게 배포할 수 있는 LLM 모델을 공동 개발했다고만 주장하고 있다. 이는 사용자가 프롬프트를 입력하면 Gracenote Data와 동일하거나 유사한 내용이 출력되어 이용자에게 제공되고 이에 따라 복제물이나 파생물이 배포된다고 주장하는 것으로 보인다.

2) 이 소송에서 OpenAI Foundation(f/k/a OpenAI, Inc.), OpenAI Group PBC, OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo, LLC, OpenAI Global, LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC 등이 피고들인데, 직접침해가 인정되지 않는 경우에 대비하기 위하여 대위침해를 주장하고 있는 셈이다.

기반한 결과물을 생성하고³⁾, 모델을 미세조정하고, 파라미터(parameter)를 선택하고 가중치(weight)를 설정하는 방식 등에 의하여, AI 제품에서 실제로 어떤 콘텐츠가 결정될지 결정함으로써, 이용자의 직접침해(사용자가 프롬프트를 입력하여 Gracenote Data의 표현적 요소가 포함된 응답을 제공받는 경우, 복제나 2차적저작물 작성에 해당)에 중대하게 기여하고 이를 직접 지원하였다. 또한 피고들은 이용자에게 의한 복제나 2차적저작물 작성·배포를 인지하였거나 인지할 수 있었다.⁴⁾

2. Kogon v. Google, LLC

- 일리노이 북부 연방지방법원(1:26-cv-02582, 2026.3.4.)
- 원고 David “Davo Sounds” Woulard 등은 아티스트, 작곡가, 프로듀서로서 다수의 음악저작물 및 녹음저작물(sound recording)에 대하여 저작권을 가지고 있는 저작권자들이다(집단소송)

(1) 사실관계(원고 주장)

1. 피고 Google은 음악 생성과 유통을 포괄하는 AI 시스템을 개발하였는데, 수천만 개의 인터넷 음악 데이터를 이용하여 MuLan과 MusicLM과 같은 모델을 학습시키고, 이를 기반으로 Lyria, Lyria 2, Lyria 3와 같은 음악 생성 모델을 구축·제품화하였다. Lyria 3의 경우 보컬, 가사, 화음 및 백킹 보컬(backing vocal, 메인 보컬을 보조하면서 화음과 분위기를 만드는 보조 보컬 파트)을 포함한 음악을 생성할 수 있는데, Gemini 앱에 통합되어 대규모 사용자에게 제공되고 있다. 또한 Google은 ProducerAI라는 별도의 플랫폼을 통하여 이용자가 ‘전체 곡 하나에 해당하는 길이의 곡’을 생성, 공유 및 게시할 수 있도록 하였고, Dream Track과 같은 기능을 통하여 YouTube 내에서 AI 생성 음악을 활용할 수 있도록 하였다. 이와 함께 Google은 YouTube와 Content ID라는 음악 유통 및 권리관리 인프라를 직접 운영하면서, 생성된 음악이 기존 음악과 동일한 시장에서 유통되도록 하는 구조를 형성하였다. 따라서 Google이 개발한 것은 개별 모델에 그치지 않고, 음악 데이터를 학습하여 생성형 AI 음악을 생산하고 이를 플랫폼을 통해 배포·이용하게 하는 일련의 시스템이라 할 수 있다.

2. Google은 학습데이터 수집단계에서의 복제(YouTube 등에서 제공되는 수천만 개의 음악 영상의 데이터셋), 전처리 및 모델 입력 변환 과정에서의 복제 및 변환적 이용(추출된 오디오를 표준화된 형식으로 변환하는 등), 학습 및 개발 과정에서의 반복적 복제·접근(단일 이용이 아니라, 전처리, 중복 제거, 필터링, 모델 성능 평가(evaluation), 미세조정(fine tuning), 강화학습(reinforcement learning)⁵⁾ 등 다양한 단계에

3) 이것은 이미 학습된 LLM이 내부에 저장된 일반적 지식만으로 답을 생성하는 대신, 특정 과제에 맞는 신뢰 가능한 외부 자료를 함께 입력받아 그 자료를 근거로 응답을 생성하도록 만드는 방식인 ‘그라운드링(grounding)’을 의미한다. RAG은 질문에 대해 관련 정보를 외부에서 검색(retrieval)한 뒤 그 내용을 모델 입력에 포함시켜 답을 생성하도록 하는 구체적인 구현 방식을 의미하는 것으로서, 그라운드링을 실현하는 기술 중 하나라 할 수 있다.

4) 그 외에도 원고는 피고들에 대하여 부당이득(unjust enrichment)도 주장하였다.

5) 강화학습은 AI 에이전트(행동하고 보상을 받는 학습 주체)가 환경과 상호작용하면서 여러 행동을 시도하고, 그 결과로 주어지는 보상을 바탕으로 장기적인

서 동일한 데이터에 대한 반복적 접근과 복제, 이 과정에서 원고들 저작물을 지속적으로 샘플링·비교·재사용하고, 초기 모델 학습에 사용된 데이터는 이후 Lyria 2, Lyria 3 개발에도 계속 활용), 저장·보유를 통한 지속적 복제 상태 유지[학습 과정에서 생성된 오디오 데이터 및 그 파생 표현(토큰, 임베딩 등)을 시스템 내에 저장되어 이후 모델 개발과 제품 개선에 계속 사용], 결과물 생성 및 유통과 관련된 이용(Lyria 3 및 ProducerAI 등은 학습된 모델을 통해 새로운 음악을 생성하고 이를 사용자에게 제공·배포하는데, 이러한 결과물이 학습데이터에 포함된 저작물의 표현을 반영) 등의 형태로 원고들의 저작물을 이용하였다.

3. Google은 원본 파일에서 오디오를 추출하는 등의 과정에서 저작권 관리정보(CMI, 아티스트 성명, 음악 제목, ISRC 코드(International Standard Recording Code),⁶⁾ 음반사 및 출판사 정보, 저작권 고지 등을 구별하는 메타데이터로 내장)를 제거·변경하였다.

또한, Lyria 3의 가사 생성 및 보컬 정렬 기능을 학습시키는 과정에서 제목, 저작자, 출판사 등의 정보를 제거하였고, 시스템을 구축하기 위하여 사용한 저작물의 저작자 정보가 결여되어 있는 AI 결과물을 배포하였다.

4. Google은 암호, 유료장벽(paywall),⁷⁾ API 제한,⁸⁾ 스트리밍 DRM 등 접근통제로 보호되는 녹음 및 가사를 획득한 후 학습시키는 과정에서 이러한 접근통제 조치를 우회·회피하였다.

(2) 위반 주장 사항(원고)⁹⁾

1. 녹음저작물의 복제에 의한 저작권 직접침해: 녹음저작물은 데이터셋 및 파이프라인(생성형 음악시스템 개발, 학습, 미세조정, 평가, 운영)에서 복제되었는데, 오디오의 획득·추출, 내부 포맷으로 변환, 클립(clip) 단위로 분할, 모델이 처리할 수 있는 표현으로의 토큰화, 저장소 및 여러 차례에 걸친 학습과정에서의 저장, 모델 변환, 평가, 제품 개선 등의 과정에서 각각 복제가 이루어졌다. 또한 중복 제거(deduplication),¹⁰⁾ 품질 필터링, 안전 검토, 평가, 회귀 테스트(regression testing),¹¹⁾ 미세조정, 후속 모델에서의 반복적 개선 등 엔지니어링 및 제품 개발 과정에서도 복제가 이루어졌다.

보상을 최대화하는 행동 전략을 학습하는 머신러닝 방법이다. LLM 분야에서 사용되는 RLHF(Reinforcement Learning from Human Feedback)는 이러한 강화학습을 인간 피드백에 적용한 방식으로, 사람이 모델의 여러 응답 중 더 적절한 답변을 선택하거나 평가하면, 모델은 그 피드백을 바탕으로 인간이 선호하는 방향의 답변을 생성하도록 조정되는 학습방법이다.

6) 녹음저작물을 식별하기 위한 국제표준 녹음코드로서, 예컨대 US-S1Z-99-00001의 경우, US(국가 코드), S1Z(등록자 코드), 99(기준 연도), 00001(해당 녹음물의 고유 번호) 등을 의미한다.

7) 유료 구독 기반 접근통제로서, 구독료를 지급하여야 접근 가능한 기술적 보호조치를 의미한다.

8) API 접근제한으로서, 해당 API를 사용할 수 있는 주체, 방식, 사용량 등을 제한하는 인터페이스 API를 의미한다.

9) 미국에서 음악과 관련된 저작물은 음악저작물(musical work, 가사 및 악곡)과 녹음저작물(sound recording, 저작자는 음반제작자와 실연자이지만 대체로 음반제작자임)이고, 음악의 가사 및 악곡은 각각 독립적으로 보호된다. 이 소송에서 원고들은 싱어송 라이터이거나 음반회사로서 양 저작물 모두와 관계되므로, 양 저작물에 대한 저작권 침해와 가사 및 악곡에 대한 저작권 침해를 각각 별도로 주장하고 있다.

10) 수집된 음원, 가사, 클립 또는 파일 중 동일하거나 실질적으로 동일한 데이터가 복수로 존재하는 경우 이를 제거·통합하여 학습 데이터셋 내 중복 반영을 방지하는 처리 과정을 의미한다.

11) 회귀 테스트는 AI 모델이나 제품이 업데이트된 이후에도 기존의 기능, 성능 및 출력 품질이 저하되지 않았는지를 확인하기 위해 기존 테스트를 다시 수행하는 검증 절차를 의미한다.

2. 녹음저작물 배포에 따른 저작권 직접침해: 모델의 개발·운영 과정에서 외부 업체·계약자·서비스 제공자 등 제3자에게 오디오 녹음, 클립 또는 그 파생 표현을 제공·송신하거나 접근을 가능하게 하였다.

3. 음악저작물의 가사의 복제에 의한 저작권 직접침해

4. 음악저작물의 악곡의 복제에 의한 저작권 직접침해

5. 저작권 관리정보(CMI, Copyright Management Information)의 제거·변경: Google은 원고들의 음악저작물 및 녹음저작물에 메타데이터로 내장되어 있는 저작권 관리정보(CMI, 녹음저작물에는 아티스트 설명, 트랙 명칭, 앨범 정보, 크레딧, 저작권 표시 및 고유 식별자가 메타데이터로 내장, 음악저작물의 악곡과 가사는 제목, 저작자, 출판사 및 권리귀속에 관한 식별정보 포함)를 학습데이터 처리 과정에서 제거·변경하거나 읽을 수 없게 만들고, CMI가 제거·변경된 복제물을 외부 업체 등에 제공함으로써, DMCA § 1202(b)¹²⁾를 위반하였다.

6. 기술적 보호조치 무력화 및 무력화 도구의 거래

① Google과 Google의 데이터 벤더·중개업체는, DRM 시스템, 세션 키 기반 스트리밍 암호화, 인증 절차, 다운로드 방지 구조(architecture) 등 저작권자가 적용한 접근통제를 무력화하고, ② 이러한 무력화를 가능하게 하는 자동화 도구를 사용·조달하여 Google의 학습 파이프라인에 저장·처리함으로써, ③ Google은 무력화 기술을 거래하고 자신에게 대규모로 복호화된 오디오를 제공하거나 접근통제를 무력화하는 서비스를 제공한 데이터 벤더·중개업체를 조달하거나 이들과 조율·협력함으로써, 접근통제 및 권리통제 무력화를 위한 도구의 거래금지 규정(§1201(a)(2), §1201(b)(1))을 위반하였다.

7. 허위 CMI 형성: AI 생성 음악에 사용자나 Google을 창작자, 소유자로 표시하여 실제 권리자를 왜곡하는 허위의 CMI를 형성하고, 이에 따라 이용자들은 해당 결과물을 적절한 콘텐츠로 오인하게 되고, 결과적으로 저작권 침해를 촉진하거나 은폐하는 효과가 발생하게 함으로써, 허위 CMI의 제공 등을 금지하는 규정(§1202(a))을 위반하였다.

8. 기여침해: Google은 ①이용자들의 직접침해를 ②인식하고 있었고, ③직접침해에 실질적으로 기여하였다.

① 이용자들의 직접침해

Google의 생성형 음악시스템을 통하여 저작물에 포함된 표현을 복제·재구성하거나, 실질적으로 유사한 결과물을 생성하고, 이를 다운로드하여 복제물을 만들고, YouTube 등 플랫폼에 업로드·공유함으로써 배포하며, 변형 기능을 통하여 2차적저작물을 생성하였다.

② Google의 인식

Google은 출력 필터가 완벽하지 않음을 스스로 인정하고, 권리자 신고 시스템과 워터마킹을 운영하며, Content ID를 통하여 저작권 상태와 권리자를 충분히 인지하고 있었다. 그럼에도 이러한 시스템이 침해 결과

12) 미국 저작권법 제1202조(b)는 저작권자 또는 법률의 허락 없이 CMI를 제거·변경하거나, CMI가 무단 제거·변경된 사실을 알면서 이를 포함한 저작물 등을 배포·수입·공개실연하는 행위를 금지한다. 단, 그 행위가 저작권 침해를 유발·가능·용이하게 하거나 은폐할 것임을 알았거나 알 만한 합리적 근거가 있는 경우에 적용된다.

물을 생성·유통하는 데 사용될 수 있음을 알거나, 적어도 고의적으로 외면하였다.

③ 실질적 기여

Google은 침해 결과물을 생성하는 모델과 이를 실행하는 서버 및 인터페이스를 제공하고, 결과물을 다운로드·공유할 수 있는 기능과 YouTube 연계 유통 구조를 구축하였다. 또한 편집 및 이어쓰기 도구를 통해 추가적인 2차적저작물 생성을 가능하게 하고, 학습 과정에서 CMI를 제거하여 결과물의 출처와 권리관계를 은폐함으로써 침해를 촉진하였다.

9. 대위침해: Google은 ① 이용자들의 직접침해를 ② 감독·통제할 권한과 능력(모델, 서버 및 인터페이스를 통제하고, 프롬프트 및 출력에 관한 정책을 설정·집행하며, 필터링·유사성 검사·사용 제한을 관리하고, 이용자의 활동을 모니터링·차단·계정 정지할 수 있고, 결과물의 저장·다운로드·게시·공유 경로를 전반적으로 통제)을 가지고 있고, ③침해로부터 직접적인 경제적 이익을 획득(유료 기능과 구독 서비스, 플랫폼 이용 확대를 통해 수익을 창출하고, 생성된 결과물이 기존 음악을 대체할수록 제품의 상업적 매력이 증가하며, 사용자 상호작용 및 생성 결과물을 모델 개선에 활용하여 추가적인 경제적 가치를 창출)함으로써, 이용자들의 직접 침해를 통제할 수 있는 지위에 있고, 이에 따라 대위침해 책임이 있다.

10. 일리노이주법에 따른 퍼블리시티권 침해: 일리노이주는 개인의 이름, 음성 등 정체성(identity)을 상업적으로 이용하는 것을 통제할 권리(퍼블리시티권)를 인정하고 있는데(765 ILCS 1075/10), 특정 개인의 음성을 모방한 AI 음성(무단 디지털 복제물)의 생성 및 유통 등을 규제하고 있다(765 ILCS 1075/30(a), 30(b), 30(d)). Google은 원고들의 녹음에서 보컬을 복제·수집하여 그 음성의 고유한 특성을 추출·모델링하고, 이를 기반으로 Lyria 3 등 생성형 음악시스템에서 실제 아티스트의 목소리와 유사하게 들리는 보컬을 생성하여 상업적 서비스로 제공하고, 특정 아티스트를 참조하는 프롬프트를 허용하여 유사한 스타일과 음성을 구현하도록 하였다. Google은 원고들로부터 이러한 이용에 관한 서면 동의를 받지 않았고 이에 의하여 일리노이주법에 따른 퍼블리시티권을 침해하였다.¹³⁾

3. THALER v. PERLMUTTER

- DC 연방 항소법원(130 F.4th 1039, 2025.3.18.); DC 연방 지방법원(1:22-cv-01564, 2022.6.2.)

(1) 사실관계: 소송 제기 및 경과

1. 원고의 저작권 등록 신청

원고(항소인) Stephen Thaler는 인공지능 시스템인 ‘Creativity Machine’이 생성한 ‘A Recent Entrance to Paradise’라는 결과물(수풀이 우거진 보라색 톤의 풍경 속에 기찻길이 놓여 있고, 그 길이 어

13) 또한 원고들은 연방상표법 위반, 일리노이주의 생체정보보호법(Biometric Information Privacy Act) 및 기만적 거래행위 방지법(Uniform Deceptive Trade Practices Act) 위반, 일리노이주 법(코몬로)에 따른 부당이득 획득, 일리노이주 소비자 기망 및 기만적 영업금지법(Consumer Fraud and Deceptive Business Practices Act) 위반 등을 주장하고 있다.

두운 터널 입구로 연결되는 이미지)에 대하여 저작권 등록을 신청하였다(2018.11.3.). 원고는 신청서에 저작자를 Creativity Machine이라고 표시하고, 해당 결과물이 기계에서 작동되는 컴퓨터 알고리즘이 자율적으로 생성되었다고 설명하고, 업무상저작물로서 Creativity Machine의 소유자인 Thaler 자신이 저작권자임을 주장하였다.

2. 저작권청의 등록거절

저작권청(USCO)은 등록신청된 결과물이 인간에 의하여 창작된 것이 아니라고 하여 등록을 거절하였다(2019.8.12.). 원고는 재심을 청구하였고 저작권청은 다시 등록을 거절하였고(2020.3.30.), 원고는 재심을 또다시 청구하였으나 저작권청은 최종적으로 등록을 거절하였다(2022.2.14.).

3. 원고의 소송 제기

원고는 저작권법에 의하면 인간 저작자만을 전제로 하지 않으므로 AI가 생성한 저작물도 보호 대상이 될 수 있고, AI 생성물에 대한 보호를 배제하는 판례가 존재하지 않고, Thaler 자신이 재산권에 관한 코몬로 원칙에 따라 작품에 대한 권리를 가지고, 업무상저작물 원칙에 따라 자신이 권리를 가지고, 이러한 해석이 과학과 기술의 진보를 촉진하려는 저작권법 및 헌법의 목적에도 부합한다고 주장하면서, 저작권청이 해당 결과물을 저작권 등록하도록 명령하는 구제명령을 청구하는 소송을 제기하였다.

4. 지방법원 및 연방 항소법원 판단

지방법원은 저작권 등록거부는 적법하다고 하여 원고 패소판결을 내렸고(2023.8.18.), 원고는 이에 불복하여 DC 연방 항소법원에 항소하였다. 항소법원은 지방법원의 판결을 인용하여 원심을 유지하는 판결을 하였다(2025.3.18.).

5. 연방대법원 상고 허가 청구 및 판결 확정

원고는 항소심 판결에 불복하여 연방대법원의 심리를 구하기 위하여 상고를 허가하도록 하는 청구를 하였는데(2025.10.9.), 연방대법원은 원고의 이러한 청구를 기각하였고(2026.3.2.), 이에 따라 항소심 판결은 최종 확정되었다(이하에서는 항소심 판결을 살펴보기로 한다).

(2) 항소심 판결

1. 연방 저작권법에 의하면, 저작물은 인간에 의하여 창작되어야 하는데, 원고 Thaler는 저작권 등록을 신청하면서 인간이 아닌 Creativity Machine을 저작자로 기재하였고, 따라서 저작권청은 원고의 저작권 등록 신청을 적절하게 거절하였다.

저작권법은 저작권이 저작물의 저작자에게 원시적으로 귀속된다고 규정하면서(§201(a)) 저작자는 정의하고 있지 않다. 그러나 저작권법에 대한 전통적인 해석에 의하면, 저작자는 인간만을 지칭하고, 저작권법의 여러 규정에 의하더라도 저작자는 인간이어야 한다는 것을 알 수 있고, 저작권청도 1976년 저작권법이 제정되기 전부터 저작자는 인간을 의미하는 것으로 일관되게 해석해 왔다. 따라서 본 항소법원도 의회가

1976년 저작권법을 제정할 당시 저작권청이 저작자에 대하여 오랫동안 해석해 왔던 입장을 채택한 것으로 추론한다.

① ①저작권은 의회가 창설한 재산권이고,¹⁴⁾ 저작권법은 저작자가 저작권을 소유한다고 규정하고 있으므로(§201(a)), 재산을 소유할 수 없는 존재는 저작자가 될 수 없고, ②저작권법은, 저작권의 존속기간에 관한 문언(저작자의 생존기간, 저작자 사망 후 70년, 저작자 사망에 관한 기록, 저작자의 사망 시점을 알 수 없는 경우, 업무상저작물에 대한 존속기간은 공표 후 95년 또는 창작 후 120년¹⁵⁾ 등)(§§302(c), (d), (e))을 통하여, 저작자의 생존기간이나 인간의 생존기간에 근접하는 기간을 저작권 존속기간의 기준으로 하고 있고, ③기계는 저작권이 상속되는 저작자의 배우자나 생존한 손·자녀(§203(a)(2))를 가지지 않고, ④기계는 저작권 양도를 위한 서명(§204(a), 법률에 의하지 않은 저작권 양도는 저작권자의 서명이 이루어진 서면에 의하지 않으면 무효)을 할 수 없고, ⑤기계는 국적이나 주소(미공표 저작물의 저작자는 국적과 주소에 관계없이 보호, §104(a))를 가지지 않으며, ⑥기계는 공동저작물(joint work)의 요건으로서 별도의 독립적인 부분을 하나의 전체로 결합시키고자 하는 ‘의사 또는 의도(intention)’(§101, 공동저작물은 각 저작자가 창작적으로 기여한 부분이 분리할 수 없거나 상호의존적인 하나의 전체로 결합될 의사를 가지고 작성한 저작물로 정의)라는 정신(mind)을 가지거나 어떠한 것을 의도하지 않으며, ⑦직·간접적으로 사용되어 특정한 결과를 발생시키기 위한 일련의 진술 또는 명령이라는 컴퓨터프로그램의 정의(§101), 기계(machine)는 장치(device)나 과정(process)과 동일한 의미이고 저작자를 보조하는 수단이지 저작자를 의미하지 않는다(§§102(a), 108(c)(2), 109(b)(1)(B)(i), 116(d)(1), 117(a)(1),(c), 401(a), 1001(2),(3)).¹⁶⁾ 이러한 문언을 종합하면 ‘인간성(humanity)’은 저작자성(authorship)의 필수조건으로 이해하는 것이 타당하다.

② 저작권청의 오랜 규칙도 저작권법이 인간 저작자를 전제로 한다는 것을 뒷받침하고 있다. 저작권청은 1973년 저작권 편람(Copyright Compendium)을 통하여 인간 저작자성 요건을 공식적으로 채택하였고(First Edition § 2.8.3(l)(a)(1)(b)), 컴퓨터프로그램에 대한 대응 연구를 위한 CONTU 보고서도 기계는 저작자가 될 수 없다고 결론을 내렸는데(CONTU, Final Report at 44 (1978)), 의회는 1980년 개정시 저작자에 관한 규정을 그대로 유지함으로써 이러한 결론을 존중하였다. 따라서 저작자를 인간 저작자로 해석하여야 한다는 것은 1976년 저작권법 제정 당시에 이미 확립된 것이었다.

③ (A) 원고는 저작자의 자연적 의미가 인간에 한정되지 않는다고 주장하는데, 저작자를 기계로 대체하는 경우 기계의 자녀·배우자·주소·국적·생명·사망 등 저작권법이 규정하고 있는 문언이 문제되고, 기계가 저작자이면서 동시에 저작자가 사용하는 도구라는 상충된 의미를 가지게 된다(예컨대 §117(d)(1)).¹⁷⁾ 또한

14) “의회는, 저작물과 발견(discoveries)에 대하여 저작자와 발명자에게 제한된 기간동안 배타적 권리를 보장함으로써, 과학과 유용한 기술의 발전을 촉진할 권한을 가진다.” 미국 연방헌법, Article I, Section 8, Clause 8.

15) 법원은 인간 저작자를 전제하는 규정의 하나로서 업무상저작물을 언급하고 있는데, 이는 법인이 업무상저작물의 저작자가 되므로 AI도 저작자가 될 수 있다는 주장을 차단하기 위한 것이다. 곧 법인도 저작자가 예외적으로 될 수 있도록 하면서 그 경우에도 존속기간을 인간과 달리 별도로 존속기간을 규정하는 것에 불과하고, 따라서 업무상저작물의 존속기간에 관한 규정도 여전히 인간을 중심으로 설계되어 있다는 것이다.

16) 이들 규정들은 기계 등을 언급하고 있는데, 예컨대 102조 (a)항은 저작물의 고정과 관련하여 “기계나 장치”를 이용하여 인식·복제·기타 전달될 수 있는 표현을 위한 매개체...”라고 언급하고 있다.

원고는 저작권법이 고용주(법인)를 업무상저작물의 저작자로 간주하므로(considered)(§101 업무상저작물 정의)¹⁸⁾ AI도 저작자가 될 수 있다고 주장하고 있지만, 간주한다는 것은 고용주가 실제 저작자로 인정된다는 것이 아니라 인간 저작자가 작성한 저작물에 대한 권리를 법적으로 고용자에게 이전해 주는 장치에 불과한 것이고, 업무상저작물의 경우에는 반드시 인간 저작자가 먼저 존재하여야 하는데, AI의 경우에는 이전해 줄 저작권이 애초에 존재하지 않으므로, 원고의 주장은 타당하지 않다.

(B) 원고는 인간 저작자성 요건으로 인하여 AI가 생성한 저작물에 대한 저작권 보호가 부당하게 제한된다고 주장하고 있다. 저작권이라는 독점적 권리를 부여하는 주된 목적은 저작자의 노동에 의하여 공중이 이익을 얻게 된다는 공익이 발생한다는 것이고, 이러한 공익목적을 위하여 사진, 녹음, 녹화, 컴퓨터프로그램 등의 기술이 저작권에 의하여 보호되어 왔지만, 이러한 보호는 법률 용어의 의미 변경이 아니라 의회의 입법에 이루어진 것이다. 또한 원고가 주장하는 바와 같이, 인간 저작자성 요건으로 인하여 ‘AI를 이용하여 작성된 저작물(works made with AI)’에 대한 보호가 저해되는 것도 아닌데, 이와 관련하여 첫째, 인간 저작자성 요건은 인공지능 활용 저작물의 보호를 배제하는 것이 아니라, 단지 저작자 지위를 인간에게 한정하는 기능을 수행할 뿐이고,¹⁹⁾ 둘째, 원고는 인간 저작자성 요건에 의하여 AI 관련 창작 유인을 약화시켜 저작물 생산을 감소시킬 것이라고 주장하고 있으나, 기계는 경제적 인센티브에 반응하지 않으므로 저작자 인정 여부가 창작량에 영향을 미치지 않으며, 인간은 여전히 AI를 활용한 저작물에 대해 저작권을 취득할 수 있어 창작 유인은 유지될 수 있고, 셋째, 의회가 1976년 이후 AI의 저작자성을 허용하는 저작권법 개정을 하지 않았다는 것은 인간 저작자성에 대한 기존의 사법적 해석을 묵시적으로 승인한 것이고, *Kelley v. Chicago Park District*(635 F.3d 290, 7th Cir. 2011)(‘야생화가 자라는 정원’에 대하여 인간에 의한 고정된 표현으로 보기 어렵다고 하여 저작물성 부인) 및 *Naruto v. Slater*(888 F.3d 418, 9th Cir. 2018)(침팬지가 카메라 셔터를 눌러 자신을 찍은 사진에 대하여 동물은 저작자가 될 수 없다고 하여 저작물성 부인) 등의 판례도 저작자는 인간에 한정된다는 입장을 확인하고 있으며, 넷째, 인간 저작자성 요건에 의하여 창작이 저해되더라도 이는 정책적 문제로서 의회가 해결해야 할 사항이고, 법원의 역할은 법을 있는 그대로 적용하는데 한정되므로, 인공지능과 저작권의 관계 역시 의회와 저작권청이 다루어야 한다.

④ 저작권법 자체가 인간 저작자성을 요구하고 있으므로 저작권법만으로 이 사건을 해결하기에 충분한 근거를 제공한다.

2. ① 원고는 업무상저작물 규정을 근거로, Creativity Machine은 자신의 피용자이므로, 자신이 저작자

17) 미국 저작권법 제117조는 컴퓨터프로그램 저작권에 관한 제한으로서 기계(machine)의 소유자가 기계의 유지관리나 수리를 위한 목적으로 컴퓨터프로그램을 복제할 수 있는 예외를 허용하고 있는데, (d)항은 기계의 ‘유지관리’ 및 ‘수리’를 “해당 기계가 그 원래의 사양 ... 에 따라 작동하도록 하기 위하여 그 기계를 정비하고 ... 사양...에 따라 작동하는 상태로 회복시키는 것으로 정의하고 있다.

18) “업무상저작물은 저작물이 종업원의 고용 범위에서 종업원에 의하여 작성되는 경우 ... 해당 저작물은 업무상저작물로 간주된다....”

19) 법원은 저작권청이 ‘인간’을 저작자로 표시한 결과물(Zarya of the Dawn)에 대해서도 등록을 거절하였고, AI가 인간 저작자의 저작물에 어느 정도 기여하여야 저작물로 인정되는지 여부에 대하여 의견이 일치하지 않지만, 이 사건에서는 ‘기계’를 저작자로 표시하고 AI만이 결과물을 생성한 것에 관한 것이므로, Zarya 사건과 무관하다고 밝히고 있다.

로 간주되어야 한다고 주장하고 있으나, 저작권법은 인간 저작자성을 전제로 하므로 업무상 저작물의 경우에도 최초 창작은 인간에 의해 이루어져야 한다. ㉓ 원고는 자신이 Creativity Machine을 제작·사용하였으므로 저작자라고 주장하고 있는데, 원고는 저작권청에서 이러한 주장을 포기하였고 따라서 본 법원은 이에 대하여 판단하지 않는다.

4. **Businessing LLC v. Runway AI, Inc.**

- 뉴욕 남부 연방지방법원 (1:26-cv-01655, 2026.2.27.)
- 시청각 콘텐츠를 YouTube에 업로드하는 콘텐츠 제작자인 원고 및 집단 구성원들이, AI 모델 학습을 목적으로 YouTube의 기술적 보호조치를 우회하여 대규모 영상 데이터를 수집한 피고 Runway AI를 상대로 제기한 집단소송이다.

(1) 사실관계 1: 피고가 주장하는 YouTube의 기술적 보호조치²⁰⁾

1. YouTube는 스트리밍(streaming) 중심 전달 방식으로 영상 콘텐츠에 대한 접근을 통제하고 있다. 이용자는 YouTube 플랫폼에서 영상을 시청할 수 있을 뿐, 그 기초가 되는 영상 파일 자체를 직접 취득하는 것은 아니다. 따라서 콘텐츠를 공개적으로 볼 수 있게 하였더라도, 이는 통제된 서비스 환경 내 시청을 허용한 것일 뿐 파일 수준의 접근까지 허용한 것은 아니다.

2. YouTube는 다운로드 기능을 제한하여 파일 수준 접근을 차단한다. 일반 이용자에게는 쉽게 이용 가능한 파일 다운로드 옵션이 제공되지 않고, Premium 이용자의 다운로드 기능도 원본 파일의 이전이나 소유가 아니라 YouTube 앱 내부의 제한적 오프라인 스트리밍에 불과하다. 따라서 Premium 이용자도 시청 각 파일을 다른 기기로 전송하거나 YouTube 서비스 밖에서 독립적으로 사용할 수 없다.

3. YouTube는 API 사용 제한 및 접근 경로 통제에 의하여 허용되지 않는 파일 추출을 방지한다. 명시적으로 허용된 기능이나 라이선스된 API를 통한 접근만 허용되며, 약관에 의하면 스크래핑, 무단 다운로드, 대량 추출, 데이터 마이닝 및 사전 서면 허가 없는 자동적인 접근은 금지된다. 또한 YouTube는 API를 주기적으로 업데이트하고 업데이트되지 않은 API나 다운로드 도구를 통한 파일 추출을 허용하지 않는다.

4. YouTube는 자동화된 접근, 스크래핑 및 대량 다운로드를 탐지·차단하기 위한 기술적 절차와 도구를 사용한다. 다운로드 활동을 모니터링하고, 일정 시간 내 과도한 다운로드 시도를 하는 IP 주소를 차단할 수 있다. 이러한 조치는 일반적 스트리밍 이용과 자동화된 대량 추출을 구별하여, AI 학습에 필요한 규모의 영상 파일 수집을 방지하는 기능을 한다.

20) YouTube 업로드 콘텐츠를 학습데이터로 이용한 이 사건은 Ted Entertainment, Inc. v. Meta Platforms, Inc. 케이스(4:25-cv-10931, 2025.12.23.)와 유사한데, Ted Entertainment 케이스와 YouTube가 적용한 기술적 보호조치에 대하여 보다 상세한 것은 [이대회, 미국에서의 AI 저작권 소송\(3\)\(저작권동향. 2026.3.10.\)](#) 참조. 또한 YouTube가 적용한 기술적 보호조치에 관하여 자세한 것은 이대회, 접근통제와 권리통제 구별의 필요성 -YOUT v. RIAA 케이스-(저작권동향 2025.11.3.) 참조.

5. YouTube의 이용약관은 기술적 조치와 결합하여 접근통제 구조를 강화하고 있다. 약관은 스크래핑, 무단 다운로드, 대량 추출, 데이터 마이닝을 금지하고, 이용자가 YouTube 서비스와 독립적으로 콘텐츠를 사용할 권리를 갖지 않는다고 하고 있다. 이러한 제한은 스트리밍 전달 방식, 다운로드 제한, API 통제, 자동화 차단 시스템과 함께 허용되지 않는 접근을 통제하는 구조를 형성한다.

6. YouTube는 제3자의 AI 학습 이용에 관하여 옵트인(opt-in) 방식의 허가 구조를 운영한다. 곧 창작자(저작권자, YouTube에 콘텐츠를 업로드하는 주체)가 제3자 학습 설정을 활성화하고 특정 제3자 AI 기업 또는 모든 제3자 기업을 명시적으로 선택한 경우에만 해당 영상의 제3자 AI 학습 이용이 허용된다. 기본값은 비활성화되어 있으므로, 창작자가 적극적으로 동의하지 않는 한 제3자 AI 기업은 YouTube 영상을 학습에 사용할 수 없다.

(2) 사실관계 II: 피고가 주장하는 기술적 보호조치의 회피

1. 피고는 YouTube 플레이어를 통한 스트리밍 방식이 아니라, yt-dlp와 같은 오픈소스 YouTube 다운로드 프로그램 및 자동화 도구를 사용하여 영상 파일을 확보하였다. yt-dlp는 YouTube 등 온라인 플랫폼의 영상과 음성을 다운로드하고, 개별 영상뿐 아니라 전체 재생목록도 내려받으며, 분리된 오디오 스트림과 비디오 스트림을 자동으로 병합할 수 있는 도구로 사용되었다.

2. 피고는 YouTube 영상 URL 또는 식별자를 다운로드 도구에 입력하여 개별 영상 또는 재생목록을 다운로드하였다. 이 방식은 YouTube 플랫폼 안에서 영상을 재생하는 것이 아니라, YouTube가 일반 이용자에게 제공하지 않는 경로를 통하여 원본 시청각 파일을 피고의 시스템으로 가져오는 방식이다.

3. 피고는 영상과 음성이 분리된 형태로 제공되는 경우에도 이를 병합하여 완전한 시청각 파일 패키지를 만들었다. 따라서 단순한 영상 시청이 아니라, 생성형 AI 모델 학습에 사용할 수 있는 파일 수준의 복제물을 생성·저장하였다.

4. 피고는 YouTube의 모니터링과 차단을 피하기 위하여 가상머신 및 IP 주소 갱신 방식을 사용하였다. YouTube는 다운로드 활동을 모니터링하고, 일정 시간 내 과도한 다운로드 시도가 있는 IP 주소를 차단할 수 있는데, 피고는 IP 주소를 갱신함으로써 동일한 출처에서 발생하는 대량 다운로드로 탐지·차단되는 것을 회피하였다.

(3) 원고 위반 주장: 기술적 보호조치 규정 위반²¹⁾

YouTube의 이용약관과 관련 접근통제는 이용자들이 YouTube 플랫폼을 통해 영상을 시청하는 것과 그

21) YouTube 영상 콘텐츠를 학습데이터로 사용한 것이므로 '저작물' 이용과 관계되지만, 원고들은 저작권 침해는 주장하지 않고 기술적 보호조치 규정 위반만을 주장하고 있다. 원고들은 YouTube 업로드 영상 콘텐츠에 대하여 저작권 등록을 하지 않았고, 이에 따라 민사소송을 제기하는 조건인 저작권 등록을 하지 않았기 때문인 것으로 보인다.

기초가 되는 영상 파일 자체에 접근하는 것을 구별하고, 후자에 대해서는 허용되지 않는 접근을 방지하는 구조로 작동한다. 따라서 YouTube의 이용약관 및 접근통제는 저작권법이 기술적 보호조치 무력화 금지에 대하여 규정하고 있는 ‘효과적인 기술적 보호조치(effective technological measure)’²²⁾에 해당한다.

피고는 자동화 도구를 사용하여 YouTube의 접근장벽을 우회하고, 일반 대중에게 제공된 적 없는 기초 영상 파일을 추출하여 자신의 AI 모델 및 서비스 학습용 말뭉치 구축에 이용하였다. 피고의 이러한 행위는 YouTube의 이용약관 및 접근통제로 구성된 기술적 보호조치를 자동화 도구를 이용하여 무력화한 것이고, 이는 기술적 보호조치의 무력화 금지에 관한 규정(§1201) 위반에 해당한다.

5. David Vance Gardner v. Runway AI, Inc.

- 캘리포니아 중부 연방지방법원 (2:26-cv-01941, 2026.2.23.)
- 이 소송은 YouTube에 업로드된 시청각저작물의 소유자인 원고 David Vance Gardner가 시청각 콘텐츠를 생성하는 생성형 AI를 전문으로 하는 Runway AI, Inc.를 상대로 제기한 집단소송이다. YouTube가 적용한 기술적 보호조치나 피고의 기술적 보호조치 우회 등의 사실관계는, 원고가 Gardner라는 것과 기술적 보호조치(TPM) 규정 위반 외에 캘리포니아 주법상의 부정경쟁과 부당이득을 주장한 것 외에는, 앞서 고찰한 *Businessing LLC v. Runway AI, Inc.* 케이스(S.D.N.Y. 1:26-cv-01655, 2026.2.27.)²³⁾와 동일할 정도로 유사하다.

6. Youngblood v. Meta Platforms, Inc.

- 캘리포니아 북부 연방지방법원 (3:26-cv-01100, 2026.2.4.)
- 이 소송은 원고 Devin Youngblood, Nicole Chmura, Chris Rice가 자신들 및 자신들과 유사한 지위에 있는 YouTube 콘텐츠 창작자들을 대표하여 Meta Platforms를 상대로 제기한 집단소송이다.
- YouTube가 적용한 기술적 보호조치나 피고의 기술적 보호조치 우회 등의 사실관계와 위반주장은, 원고가 Youngblood이고 피고가 Meta Platforms라는 것을 제외하면, 앞서 고찰한 *Businessing LLC v. Runway AI, Inc.* 케이스(S.D.N.Y. 1:26-cv-01655, 2026.2.27.)²⁴⁾와 사실상 동일하다.

7. Nicole Chmura v. Snap Inc.

22) 미국 저작권법은 ‘효과적인 기술적 보호조치(effective technological measure)’를 정의하지 않고, 대신에 “어떠한 조치가 저작물에 대한 접근하기 위하여 ... 정보 입력 ...을 요구하는 경우, 그 조치는 ‘저작물에 대한 접근을 효과적으로 통제’하고 “어떠한 조치가 저작권자의 권리행사를 방지... 하는 경우, 그 조치는 ‘저작권자의 권리를 효과적으로 보호’한다”고 정의하고 있을 뿐이다. §1201(a)(3)(B), (b)(2)(B).

23) 이 글의 “4. *Businessing LLC v. Runway AI, Inc.*”(9페이지) 참고

24) 이 글의 “4. *Businessing LLC v. Runway AI, Inc.*”(9페이지) 참고

- 캘리포니아 중부 연방지방법원 (2:26-cv-01732, 2026.2.18.)
- 이 소송은 YouTube 콘텐츠 창작자인 원고 Nicole Chmura가 자신 및 이와 유사한 지위에 있는 모든 자들을 대표하여, 기술적 보호조치 무력화 금지에 관한 규정(§1201) 위반을 이유로, 생성형 AI 기능과 동영상 생성 AI 기술을 개발·상업화하려는 기술기업인 Snap Inc.를 상대로 제기한 집단소송이다.
- YouTube가 적용한 기술적 보호조치나 피고의 기술적 보호조치 우회 등의 사실관계와 위반 주장사항은 Businessing LLC v. Runway AI, Inc. 케이스(S.D.N.Y. 1:26-cv-01655, 2026.2.27.)²⁵⁾와 거의 동일하다.

8. Ace Cam, Inc. v. Runway AI Inc.

- 뉴욕 남부 연방지방법원 (1:26-cv-01362, 2026.2.18.)
- 이 소송은 YouTube에 시청각 콘텐츠를 제작하여 업로드한 원고 Ace Cam(Random Golf Clup이라는 명칭으로 영업)이 기술적 보호조치 무력화 금지규정(§1201)의 위반을 근거로, 텍스트·이미지 입력을 통해 동영상을 생성하는 생성형 AI 시스템을 개발·제공하는 기업인 Runway AI를 상대로, 자신 및 이와 유사한 지위에 있는 사람들을 대표하여 제기한 집단소송이다.
- YouTube가 적용한 기술적 보호조치나 피고의 기술적 보호조치 우회 등 사실관계와 위반 주장사항은 Businessing LLC v. Runway AI, Inc. 케이스(S.D.N.Y. 1:26-cv-01655, 2026.2.27.)²⁶⁾와 거의 동일하다.

9. Kleiner v. Adobe Inc.

- 캘리포니아 북부 연방지방법원 (5:26-cv-01218, 2026.2.9.)
- 이 소송은, 서적의 저작자인 원고 Arthur Kleiner가 자신 및 이와 유사한 지위에 있는 사람들을 대표하여, 모바일 기기에서의 문서 지원 작업에 최적화된 AI 소프트웨어인 SlimLM이라는 소형 언어 모델(SLM)을 학습시키기 위하여, 서적 및 저술물을 복제·저장·사용하여 저작권을 침해하였다는 것을 이유로, 피고 Adobe Inc.에 대하여 제기한 집단소송이다.

(1) 사실관계(원고 주장)

1. 원고의 저작물: 원고들의 서적은 약 196,640권의 도서로 구성되고 온라인 그림자도서관인 Bibliotik에서 파생된 Books3 데이터셋에 포함되어 있다. Books3 데이터셋은 AI 모델 학습을 위하여 EleutherAI

25) 이 글의 “4. Businessing LLC v. Runway AI, Inc.”(9페이지) 참고

26) 이 글의 “4. Businessing LLC v. Runway AI, Inc.”(9페이지) 참고

가 선별·구성한 The Pile 데이터셋에 포함되어 있으며, The Pile 데이터셋은 The-Eye로 알려진 웹사이트나 Hugging Face 웹사이트에서 다운로드할 수 있었다. Books3 데이터셋 전체는 Together AI가 만든 또 다른 데이터셋인 RedPajama에도 포함되어 있었다. 따라서 적어도 2023년 10월 이전에는 Hugging Face에서 The Pile 또는 RedPajama 데이터셋을 다운로드하면 Books3 데이터셋의 사본도 함께 다운로드하게 된다. 또한 SlimPajama는 RedPajama 데이터셋을 정제·중복 제거한 데이터셋이고 동시에 Common Crawl 데이터셋에 부분적으로 기반하거나 중복되는 데이터셋으로 알려져 있고, Common Crawl은 저작권으로 보호되는 자료를 포함하고 있다. 따라서 SlimPajama는 RedPajama 및 Common Crawl을 통해 저작권 있는 자료를 포함하게 된다.

원고 서적 → Books3 → The Pile / RedPajama → 다운로드 시 Books3 복제본 취득
RedPajama → 정제·중복 제거 → SlimPajama → Adobe의 SlimLM 학습에 사용

2. Adobe의 저작물 이용: Adobe는 자신의 SlimLM SLM을 SlimPajama-627B로 사전학습시켰고, RedPajama와 Common Crawl을 통하여 저작물을 SlimPajama 데이터셋에서 서적들을 다운로드받았다.

(2) 공정이용의 부인 주장: Adobe의 저작물 이용은 공정이용에 해당하지 않음

1. Adobe는 원고 및 집단 구성원들의 저작물을 적법하게 구매하거나 이용허락받을 수 있었지만, RedPajama 및 Common Crawl에서 유래한 SlimPajama 데이터셋을 통해 이를 다운로드·복제·저장·사용하였다. Bartz v. Anthropic PBC 케이스²⁷⁾에 의하면, 적법하게 구매하거나 접근할 수 있었던 원본 복제물을 불법 사이트에서 다운로드하는 행위는 이후의 공정이용(Bartz 케이스는 학습데이터를 해적 사이트에서 다운로드하는 것과 학습행위를 분리하여 분석하고 후자가 공정이용에 해당한다고 판시)에 합리적으로 필요한 것이 아니어서 공정이용이 될 수 없다. 또한 원고 저작물은 SlimPajama 데이터셋의 일부로서 저작물 전체가 다운로드·사용·스크래핑되었고, 즉시 폐기되지 않은 채 SlimLM 학습 및 장래 상업적 목적을 위해 보관 되었으므로, 단순히 일시적이거나 간접적인 이용으로 보기 어렵다. 나아가 SLM은 학습 말뭉치에 기반하여 앞선 문맥을 바탕으로 다음에 이어질 토큰(next token)을 확률적으로 예측하는 방식으로 작동하고, 주요 AI 모델들이 학습데이터를 상당 부분 역류(regurgitation)²⁸⁾할 수 있으므로, Adobe의 이용은 변형적 이용이 아니다.

2. ① Adobe는 엄청난 양의 가치있는 저작물을 교육·비평적 목적이 아닌 상업적인 목적으로 이용하였고, 자유시장에서 이용하지 않고 무단 복제물을 이용함으로써, 변형적 이용이 아니고, ② 원고의 저작물은 매우 창의적인 것이고, 원본 복제물을 구할 수 없었던 것이 아니라 구매 또는 이용허락받을 수 있었지만

27) 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025).

28) '역류'는 모델이 학습데이터에 포함된 표현을 새롭게 생성하는 것이 아니라, 거의 그대로 반복·재현하여 생성하는 것을 의미한다.

Adobe는 그렇게 하지 않았고,²⁹⁾ ③ 원고의 저작물 전체가 데이터셋에 편입되어 계속 복제·저장되었고, 썸네일(thumbnail) 이미지와 같이 저작물의 실질적 표현이 제거·약화된 변형본으로 복제된 것이 아니라,³⁰⁾ 내용이 실질적으로 추출 가능한 형태로 복제·저장되었고, ④ 저작물이 학습 라이브러리를 구축하는데 사용되었고, 서적에 대한 수요를 대체하는 불법 출처에서 획득되어 원고의 저작물 시장에 대한 피해(서적의 판매 및 이용허락, 서적과 무료 버전과의 경쟁, 학습데이터로서 원고 저작물에 대한 이용허락 시장은 이론적이 아니라 실제로 존재)가 명백하다. 따라서 공정이용의 4가지 요소 모두 공정이용이 부정되는 방향으로 작용한다.

(3) 침해 주장: 저작권 직접침해

Adobe는 ❶ 원고의 저작물이 포함된 Books3 데이터셋은 RedPajama 데이터셋에 포함되어 있었고, SlimPajama는 RedPajama를 정제·중복 제거하여 생성된 것이고 Common Crawl도 일부 포함하는 데이터셋인데, SlimPajama를 다운로드하거나 학습 파이프라인에 투입하거나 그 밖의 방식으로 취득하고, ❷ 내부 서버에 SlimPajama의 여러 복제물을 생성·저장하고, ❸ SlimLM SLM을 개발하고 학습데이터에 편입시키기 위하여 SlimPajama 데이터셋을 복제하였다. ① ‘원고의 저작물이 포함된 데이터셋’을 복사, 저장, 처리, 복제, 사용하고, ② ‘원고의 저작물로 학습된 SlimLM 모델’³¹⁾을 복사, 저장, 처리, 복제함으로써, 원고의 배타적 권리를 직접 침해하였다.

10. Youngblood v. Nvidia Corporation

- 캘리포니아 북부 연방지방법원 (5:26-cv-00916. 2026.1.29.)
- 이 소송은 YouTube 콘텐츠 제작자인 원고 Devin Youngblood가 YouTube 영상 파일을 대규모로 수집하여 AI 모델 학습에 사용하는 과정에서 YouTube가 적용한 기술적 보호조치(TPM)를 우회·무력화하여 접근 통제 TPM 무력화 금지규정(§1201(a))을 위반하였다고 주장하면서 NVIDIA를 상대로 제기한 집단소송이었다. 이 소송은 2026.2.3. 자발적으로 취하(voluntary dismissal without prejudice)³²⁾되었다.³³⁾

29) 원고의 이러한 주장은 공정이용의 제2요소와 직접적으로 관계되지 않고 제2요소와 제4요소가 만나는 지점에서 고려되는 사항이다. 원고가 이렇게 주장한 것은 저작물을 구할 수 있는데도 구하지 않고 이용한 경우에 공정이용에 해당하지 않는다는 주장을 강화하기 위한 것으로 보인다.

30) 미국이나 한국에서 썸네일 이미지 제공은 공정이용으로 판시되었는데, 썸네일은 미리보기용으로 제공되면 원래 이미지보다 크기가 축소되고 (확대하는 경우) 해상도가 떨어지는 것으로서 실질 내용이 약화된 변형된 복제물이다.

31) Adobe가 원고 저작물이 포함된 SlimPajama 데이터셋을 학습에 사용함으로써 원고 저작물의 보호되는 표현이 SlimLM 모델의 가중치 또는 파라미터에 반영되었다거나 AI 모델이 저작물의 표현을 학습·내재화한 결과물이고, 따라서 그러한 모델 자체를 복제하는 등의 행위도 저작권을 침해한다는 주장이다. 다만 이러한 주장에서는 원고의 저작물이 모델(유형의 매개체, tangible medium)에 고정되었는지 여부가 관건이 될 것이다.

32) 이 소송은 제소 가능성을 남긴 형태(without prejudice)로 취하한 것이므로, 원고가 청구 자체를 확정적으로 포기한 것이 아니라, 동일 또는 유사한 청구를 다시 제기할 가능성을 남긴 채 사건이 종료되었다.

33) 원고 Devin Youngblood 등은 2026.2.4. Meta Platforms를 상대로 이 소송과 거의 동일한 사실관계에 바탕하여 소송을 제기하였다(Youngblood v. Meta Platforms, Inc., 3:26-cv-0110). 이 글의 “6. Youngblood v. Meta Platforms, Inc.”(11페이지) 참고

11. Concord Music Group, Inc. v. Anthropic PBC

- 캘리포니아 북부 연방지방법원 (5:26-cv-00880, 2026.1.28.)
- 이 소송은 Concord Music 등 음악출판사들이 BitTorrent 방식으로 그림자도서관을 통하여 출판사들의 음악저작물, 악보, 가사를 대량으로 불법 취득하고, 이와 같이 취득한 자료를 Claude 모델의 학습 및 출력에 사용하였다고 주장하면서 Anthropic(법인), Dario Amodei(창립자 및 CEO), Benjamin Mann(창립자 및 기술직 구성원)에 대하여 제기한 소송이다(Concord II). 이들 출판사들은 이미 2023.10.18. Anthropic을 상대로 소송을 제기하였었는데(Concord I),³⁴⁾ Concord I은 Anthropic의 Claude 모델이 음악 가사를 학습데이터로 복제하고, 이용자 프롬프트에 따라 해당 가사를 출력함으로써 저작권을 침해하였다고 제기한 소송이었다. 양자는 모두 모두 Anthropic의 Claude 모델이 음악저작물을 학습 및 출력 과정에서 이용하였다고 주장하는 것은 동일하지만, Concord I의 핵심 쟁점은 생성형 AI 모델의 학습 과정에서의 복제와 출력 과정에서의 저작물 재현이 저작권 침해에 해당하는지 여부인데, Concord II는 AI 모델 학습 여부와 별개로, 토렌팅에 의한 불법 복제·배포 자체에 대하여 직접침해까지 확장하여 주장하고 있다. 또한 Concord I은 약 500곡의 노래 가사 학습·출력 문제를 중심으로 하지만, Concord II는 소송은 20,000개 이상의 음악저작물, 악보 및 가사를 토렌트로 불법 취득·배포하고 이를 학습·출력에 사용했다는 점을 문제 삼는다는 차이가 있다.

(1) 사실관계 주장³⁵⁾

1. 대표적인 온라인 침해 운영주체인 LibGen과 PiLiMi³⁶⁾은 악보와 가사를 수록한 송북(songbook), 악보집(sheet music), 기타 노래가사집 등 모든 장르의 서적을 포함하고 있는데, 피고들은 이들 주체로부터 해적판 도서들을 P2P 파일공유 프로토콜인 BitTorrent 방식으로 .epub, .pdf 및 .txt 파일 형식으로 다운로드받았다.

2. BitTorrent와 같은 P2P 프로토콜은 중앙 저장소를 통해 파일을 업로드하고 다운로드하는 방식이 아니라, 파일을 여러 조각(piece)으로 나누고 그 조각들을 많은 이용자들(peers) 사이에 동시에 즉각적으로 유포한다. 이용자가 파일의 한 조각을 다운로드하면 그 이용자는 즉시 다른 이용자들에게 해당 파일을 배포하게 되는데, 이에 의하여 해당 파일을 다운로드하는 모든 이용자가 동시에 다른 이용자들에게 그 파일을 배포³⁷⁾하는 역할을 하게 된다. BitTorrent상의 파일은 수천 명의 다른 이용자와 동시에 공유되어, 1:1 방

34) 음악출판사들은 2023.10.18. Anthropic를 상대로 테네시 중부 연방지방법원에 제기(이후 캘리포니아 북부 연방지방법원으로 이송, Concord Music Group, Inc. v. Anthropic PBC, 5:24-cv-03811)(Concord I)하였는데, Concord I에 관해서는 [이대회, 미국에서의 AI 저작권 소송\(5\)\(저작권동향 2026\) 참조](#).

35) 여기서는 Torrent 방식에 의한 복제·배포와 Concord I과의 차이점을 중심으로 살펴보기로 한다.

36) 미국 정부가 LibGen과 같은 불법 복제물 라이브러리를 폐쇄하자 이를 새도 버전(shadow version)으로 복제되어 만들어진 것이 Z-Library이다. 이후 FBI가 Z-Library도 폐쇄하자 그 새도 라이브러리 자체를 복제하거나 '미러링'하여 만들어진 라이브러리가 PiLiMi이다.

37) BitTorrent 네트워크에서 파일(또는 파일조각)을 다른 이용자에게 공유하는 행위는 한국법상으로 전송이 되지만 미국에서는 이러한 행위가 배포에 해당

식이 아니라 ‘수천:수천’ 방식으로 이루어진다. 따라서 이용자들이 BitTorrent 네트워크에 참여하면 파일 전송 속도가 느려지는 것이 아니라, 다운로드가 더 빨라지고 훨씬 더 많아진다. 또한 일단 침해 파일이 BitTorrent상에 배포되면 온라인 침해의 바이러스적 성격으로 인해 무한한 이용자 집단 전체로 대규모의 급속한 확산이 가능해지기 때문 이를 통제할 수 없고, 실제로 하나의 파일이 불과 몇 시간 만에 수천 번 배포될 수도 있다.

3. 피고들은 2021년 LibGen에서 최소한 500만 권, 2022년 PiLiMi에서 최소한 추가로 200만 권의 해적판 도서 복제물을 다운로드받았는데, 여기에는 출판사들이 소유한 음악저작물의 악보와 노래 가사를 포함하는 수백 권의 도서가 포함되어 있고, 개별 저작물과 함께 관련 저작권 관리정보를 포함하고 있다.

(2) Concord I 소송에서의 소장변경 실패와 본 소송의 제기

이 소송이 Concord I과 다른 점은 BitTorrent를 이용한 저작물의 다운로드(복제)·배포에 관한 것인데, 원고들은 이러한 복제·배포를 Bartz 케이스 판결³⁸⁾ 등을 통하여 알게 되었다. 원고들은 Concord I 소송에서 BitTorrent에 관하여 주장하기 위하여 소장을 수정하려고 하였으나 Anthropic이 반대하여 수정하지 못하였고, 이에 따라 이 소송을 새로 제기하게 되었다.

(3) 위반 주장 사항

1. 저작권 직접침해(모든 피고): 음악출판사들의 음악 악곡을 토렌트 방식으로 복제·배포하여 저작권을 직접 침해하였다.

2. 학습 및 결과물 생성에 의한 저작권 직접침해(Anthropic): ①Claude AI 모델의 학습 및 ②모델에 의한 결과물 생성에 의하여 음악저작물을 복제, 배포, 전시 및 2차적저작물 작성하여 저작권을 직접 침해하였다.

3. 기여침해(Anthropic): Anthropic 모델의 이용권자 및 사용자들은 음악저작물을 복제, 배포, 전시, 2차적저작물을 작성하여 저작권을 직접 침해하고 있다. Anthropic은 이러한 직접침해를 인식하고 있고, 침해를 가능하게 하고 조장하고 실질적으로 기여하고(AI 모델의 홍보, 웹사이트 및 API를 통한 접근 제공, 모델에 의하여 침해임을 알고 콘텐츠를 호스팅하고 전송 등), 따라서 기여침해 책임이 있다.

4. 대위침해(Anthropic): Anthropic 모델의 이용권자 및 사용자들은 저작권을 직접 침해하는데, Anthropic은 이러한 침해에 의하여 경제적 이익을 얻고, 모델의 입력과 출력을 통제하고 침해를 모니터링할 수 있는 등 침해행위를 감독하고 통제할 법적 권리와 능력을 가지고 있고, 따라서 대위침해 책임이 있다.

5. 저작권 관리정보 규정 위반(Anthropic): Anthropic은 저작권 관리정보(CMI, 저작자 명칭 및 기타 식

한다. 이 케이스에서 원고들이 BitTorrent를 통한 다운로드와 관련하여 복제권 침해 외에 배포권 침해를 주장하는 이유가 여기에 있다.

38) Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025).

별 정보)를 제거·변경하고, 제거·변경되었다는 것을 알면서 복제물을 배포함으로써, CMI 제거·변경 금지에 관한 규정(§1202(b)(1))과 제거·변경을 알면서 CMI를 저작물의 배포 등을 금지하는 규정(§1202(b)(3))을 위반하였다.

12. Ted Entertainment, Inc. v. Snap Inc.

- 캘리포니아 중부 연방지방법원 (2:26-cv-00754, 2026.1.23.)
- 원고 Ted Entertainment는 유튜브(YouTube)에 시청각 콘텐츠를 업로드하는 콘텐츠 창작자인데, 이미 2025.12.23. Meta Platforms, Inc.과 ByteDance, Inc.를 상대로 캘리포니아 북부 연방지방법원에 집단 소송을 제기하였었다(4:25-cv-10931; 5:25-cv-10933). 이 소송은 피고가 Snap이라는 것, 피고의 AI 모델·서비스의 명칭과 기능, 데이터셋 등을 제외하면, 이미 제기된 이들 소송과 사실관계나 위반주장 사항이 거의 동일하다.³⁹⁾
- Snap은 이용자가 사진, 짧은 영상, 메시지를 친구들과 주고받고, 일정 시간이 지나면 사라지는 방식의 커뮤니케이션 기능을 제공하는 소셜미디어·메시징 앱이다.

³⁹⁾ 이들 소송에 대하여 자세한 것은 이대희, [미국에서의 저작권 AI 소송\(3\)\(저작권동향 2026.3.10.\)](#) 참조.

참고 자료

- <https://www.courtlistener.com/docket/72384605/gracernote-media-services-llc-v-openai-foundation/>
- <https://www.courtlistener.com/docket/72377338/kogon-v-google-llc/>
- <https://www.courtlistener.com/docket/63356475/thaler-v-perlmutter/>
- <https://www.courtlistener.com/docket/72344359/businessing-llc-v-runway-ai-inc/>
- <https://www.courtlistener.com/docket/72318854/david-vance-gardner-v-runway-ai-inc/>
- <https://www.courtlistener.com/docket/72234164/youngblood-v-meta-platforms-inc/>
- <https://www.courtlistener.com/docket/72295615/chmura-v-snap-inc/>
- <https://www.courtlistener.com/docket/72291526/ace-cam-inc-v-runway-ai-inc/>
- <https://www.courtlistener.com/docket/72252618/kleiner-v-adobe-inc/>
- <https://www.courtlistener.com/docket/72206727/youngblood-v-nvidia-corporation/>
- <https://www.courtlistener.com/docket/72199828/concord-music-group-inc-v-anthropic-pbc/>
- <https://www.courtlistener.com/docket/72185144/ted-entertainment-inc-v-snap-inc/>