

# 저작권 이슈 브리프



COPYRIGHT ISSUE BRIEF

Weekly Report  
2025. 2-3





# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 산업 AI 기반 음악 변형 기술, 그리고 저작권 산업 영향

### AI 도입 본격화하는 음악 산업, 저작권 논란 본격화

▶ 튠드 글로벌(Tuned Global)이 오디오셰이크(AudioShake)와 협업을 발표하며 B2B 음악 스트리밍 시장에 AI 기술 도입이 본격화됨. 오픈AI(OpenAI), 슈퍼톤(Supertone), 수노(Suno), 유디오(Udio) 등 다양한 기업이 AI 음악 기술 개발에 뛰어들고 있으며, 유튜브(YouTube)와 엔비디아(NVIDIA)도 관련 기능을 실험 중. 그러나 AI의 무단 음원 수집과 저작권 침해 논란이 커지면서 법적 공방도 심화되고 있음. AI 기술 혁신과 저작권 보호 사이에서 균형을 찾는 것이 음악 산업의 주요 과제가 되고 있으며, 창작자 보호와 산업 지속 가능성을 고려한 새로운 규범과 라이선스 모델이 필요할 것으로 보임

## 산업 AI 기반 영상 생성 기술 발전 가속... 영상 제작 자동화 혁신 본격화되나

### 저작권 보호 AI 영상 생성 기술 상용화와 시장 변화

▶ 어도비가 라이선스 확보 콘텐츠를 학습 데이터로 활용한 'Firefly 비디오 모델'을 출시하여 저작권 안전성이 확보된 AI 영상 생성 기술의 상용화 시대를 본격적으로 열었음. 텍스트나 이미지 입력만으로 1080p 고화질 영상을 생성할 수 있는 동시에 카메라 앵글과 움직임까지 세밀하게 제어 가능한 전문가급 기능을 구현하였음. 월 9.99달러부터 시작하는 구독형 요금제 도입으로 전문 크리에이터부터 일반 사용자까지 폭넓은 활용이 가능해져 AI 영상 생성 기술의 대중화가 가속화될 전망이다

## 산업 넷플릭스의 '오징어 게임' IP 확장, 콘텐츠와 게임 산업의 경계를 허물다

### OTT-게임 산업 융합 가속화... 넷플릭스 IP 기반 게임 사업 확대

▶ 2025년 게임 산업에서는 IP를 활용한 게임 개발이 핵심 트렌드로 부상하고 있음. 넷플릭스는 자사 IP를 활용한 게임 개발에 주력하여 매월 1개 이상의 신작을 출시할 계획이며, 오징어 게임 등 인기 IP를 활용한 게임을 출시함. 세가와 소니 등 게임사들은 자사 게임 IP의 영화, TV 시리즈 제작을 통해 트랜스미디어 전략을 추진함. 이러한 IP 기반 게임의 성공은 기존 IP 보유자들의 협상력을 강화하고 있으며, IP의 가치를 단순 라이선싱에서 기획 단계부터 다양한 매체로의 확장을 고려할 수 있는 중요한 아이템으로 격상함. 게임은 IP의 가치를 확장하고 강화하는 핵심 매체로 자리잡으며, 이는 저작권 산업 전반의 비즈니스 모델 변화를 가속화할 것으로 전망됨



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 산업 메타, AI 학습 데이터의 저작권 책임 회피 논란

### 메타의 불법 데이터 수집과 AI 업계의 과제

▶ 메타(Meta)가 AI 모델 학습을 위해 토렌트를 통해 81.7TB 규모의 불법 복제 전자책과 논문을 수집했으며, 마크 저커버그(Mark Zuckerberg) 메타 CEO의 승인 하에 저작권 정보를 의도적으로 삭제함. AI 업계가 전반적으로 '데이터 벽' 문제에 직면한 가운데, 일부 기업들은 합법적인 라이선스 계약을 맺는 동시에 무단으로 데이터를 수집하는 이중적인 전략을 구사하고 있음. 이러한 관행은 AI 산업의 건전한 발전을 저해할 수 있어, 지속가능한 데이터 생태계 구축이 시급한 과제로 대두되고 있음

## 기술 AI 모델의 '지식 증류(Knowledge Distillation)' 개념과 저작권 침해 논란

### AI 모델 '지식 증류'로 인한 지식재산권 침해 논란

▶ AI 모델 개발에서 '지식 증류' 기술을 둘러싼 지식재산권 침해 논란이 심화되고 있음. 지식 증류는 기존 대형 AI 모델의 출력 패턴을 활용해 소형 신규 모델을 학습시키는 기술로, 적은 비용으로도 유사한 성능의 모델을 개발할 수 있도록 함. 최근 중국 기업 딥시크가 오픈AI의 GPT에 준하는 성능의 모델을 현저히 적은 비용으로 단기간에 개발하면서 증류 의혹이 제기되었음. 지식 증류는 원본 모델의 코드나 구조를 직접 복제하지 않아 지식재산권 침해 입증에 어려운 상황임. AI 선도 기업들은 API 접근 관리, 워터마킹 등 보호 조치를 고려하고 있으나, 기술의 특성상 완벽한 차단이 어려워 적절한 지식재산권 보호 방안 마련이 필요함

## 기술 주간기술동향

### AI 텍스트 데이터 마이닝과 저작권 보호 기술

▶ AI 기반 텍스트 데이터 마이닝 기술이 빠르게 발전하면서, 대규모 언어 모델(LLM)의 활용이 확대되고 있으며, 생성형 AI(Generative AI)가 기존 텍스트를 학습하고 이를 바탕으로 새로운 콘텐츠를 생산하는 과정에서 저작권 문제가 중요한 이슈로 대두되고 있음. AI가 웹사이트, 전자문서, 연구 논문, 도서, 뉴스 기사 등 다양한 출처에서 데이터를 학습하는 과정에서 저작권 보호된 자료를 포함할 가능성이 높아지며, 이에 따라 AI가 생성한 콘텐츠가 기존 저작물과 얼마나 유사한지를 평가하는 것이 핵심 논점이 되고 있음. 이러한 문제를 해결하기 위해 AI 텍스트 데이터 마이닝과 관련된 저작권 침해 가능성을 분석하고, 이를 방지하기 위한 세 가지 기술적 대응 방법에 대해 알아봄



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## AI 기반 음악 변형 기술, 그리고 저작권 산업 영향

### 뉴스 브리프

튠드 글로벌이 오디오셰이크와 협업을 발표하며 음악 스트리밍 시장에 기술 도입이 본격화되고 있다. 음원 분리와 가사 자동 생성 기능을 통해 음악 검색과 유통 효율성이 향상될 것으로 기대되며, 워너 뮤직 그룹 등 글로벌 기업들도 이 기술을 활용할 전망이다. 오픈AI, 수퍼톤, 수노, 유디오 등 다양한 기업이 음악 기술 개발에 뛰어들고 있으며, 유튜브와 엔비디아도 관련 기능을 실험 중이다. 그러나 무단 음원 수집과 저작권 침해 논란이 커지면서 법적 공방도 심화되고 있다. 기술 혁신과 저작권 보호 사이에서 균형을 찾는 것이 음악 산업의 주요 과제가 되며, 창작자 보호와 산업 지속 가능성을 고려한 새로운 규범과 라이선스 모델이 필요할 것으로 보인다.

### 음악 산업의 새 흐름, AI 기술로 혁신 가속화

- B2B 음악 스트리밍 분야의 AI 기술 도입 본격화
  - 음악 산업의 디지털 전환이 가속화되면서 기술 도입이 활발해지고 있음. 이러한 흐름 속에서 세계적인 음악 스트리밍 기업 튠드 글로벌(Tuned Global)이 음악 스타트업 오디오셰이크(AudioShake)와 협력 관계를 구축하였음. 이번 협업을 통해 튠드 글로벌은 자사 플랫폼에 음원 분리 및 가사 자동 생성 기능을 도입할 예정임
  - 오디오셰이크의 기술은 하나의 음원에서 보컬, 베이스, 드럼 등 각각의 악기 소리를 분리해내는 음원 분리(stem separation) 기능을 제공함. 실시간 통합 기능을 통해 음악 서비스 기업들은 최소한의 기술적 전문성으로도 이러한 기능을 빠르게 도입할 수 있음. 특히 가사 자동 생성 기능이 메타데이터와 검색 기능을 크게 향상시켜, 곡 검색과 콘텐츠 필터링을 더욱 효과적으로 만든다는 것임
  - 튠드 글로벌의 콘 라소(Con Raso) 대표<sup>1)</sup>는 “이번 협업을 통해 음악 관련 디지털 서비스들이 AI 기술을 활용함으로써 사용자 경험을 향상시키는 것이 핵심 목표”라고 설명함. 오디오셰이크의 제시카 파월(Jessica Powell) CEO<sup>2)</sup>는 “레이블, 퍼블리셔, 권리 보유자들이 음악을 더 다양하게 활용하고자 하는 수요가 증가하고 있으며, 음원 분리 기술이 이러한 기회를 실현하는 핵심 도구가 될 것”이라고 밝힘

1) Mandy Dalugdug, "Tuned Global rolls out AI music features via AudioShake partnership", MusicBusinessWorldwide, 2025.01.30, <https://www.musicbusinessworldwide.com/tuned-global-rolls-out-ai-music-features-via-audioshake-partnership/>  
 2) Mandy Dalugdug, "Tuned Global rolls out AI music features via AudioShake partnership", MusicBusinessWorldwide, 2025.01.30, <https://www.musicbusinessworldwide.com/tuned-global-rolls-out-ai-music-features-via-audioshake-partnership/>

- 이번 협업이 주목받는 이유는 음악 스트리밍 시장에서 최초로 기술을 전면적으로 도입한 사례이기 때문임. 특히 튠드 글로벌의 주요 고객사인 워너 뮤직 그룹(Warner Music Group), 유니버설 뮤직 그룹(Universal Music Group), 소니 뮤직(Sony Music) 등 글로벌 음악 기업들이 이 기술을 활용할 수 있게 되면서, 음악 산업 전반에 기술 도입이 가속화될 것으로 전망됨<sup>3)</sup>

## AI 음악 혁신: 기술, 비즈니스, 그리고 산업의 변화

### • AI 기반 음악 변형 기술의 유형과 특징

- 현재 음악 산업에서 활용되고 있는 AI 기반 음악 변형 기술은 크게 음원 분리, 음악 생성, 스타일 변환의 세 가지 영역으로 구분할 수 있음. 각각의 기술은 독자적으로 발전하면서도 상호 보완적인 관계를 형성하고 있음
- 음원 분리 기술(Stem Separation)은 오디오쉐이크(AudioShake)의 기술과 같이 하나의 음원에서 각각의 구성 요소를 분리해내는 기술임. 이는 보컬과 악기 소리를 분리하는 기본적인 기능에서 시작하여, 현재는 개별 악기 소리까지 분리할 수 있는 수준으로 발전하였음. 특히 딥러닝 기술의 발전으로 분리된 음원의 품질이 크게 향상되었으며, 실시간 처리도 가능함. 음원 분리 기술은 리믹스, 샘플링, 노래방 모드 등의 기능을 제공하는 데에 활용될 수 있음
- 음악 생성 기술(Generative AI for Music)은 AI가 새로운 멜로디나 화성을 만들어내는 것을 의미함. 이는 기존 음악의 패턴을 학습하여 새로운 음악을 생성하는 방식으로, 최근에는 특정 장르나 아티스트의 스타일을 반영한 음악 생성도 가능해짐
- 스타일 변환 기술은 기존 음악의 장르나 분위기를 변경하는 기술임. 예를 들어, 팝 음악을 재즈 스타일로 변환하거나, 느린 템포의 발라드를 댄스 음악으로 바꾸는 것이 가능함
- 이러한 기술들은 단독으로 사용되기도 하지만, 최근에는 여러 기술을 결합하여 더 복잡하고 정교한 음악 변형이 가능해지고 있음. 예를 들어, 음원을 분리한 후 각 파트별로 스타일을 변환하고, 여기에 AI가 생성한 새로운 멜로디를 추가하는 식의 복합적인 처리가 가능해진 것임. 이는 음악 창작과 편곡의 새로운 가능성을 제시하고 있으며, 음악 산업의 생산 방식에도 큰 변화를 가져올 것으로 전망됨. 이는 음악 창작과 편곡의 새로운 가능성을 제시하고 있으며, 음악 산업의 생산 방식에도 큰 변화를 가져올 것으로 전망됨. 이러한 발전은 단순한 수동적 음악 감상을 넘어 보다 능동적인 음악 경험으로의 진화를 보여주고 있음

### [AI 기반음악변형기술특징

기술 유형	설명	주요 특징
음원 분리 기술 (Stem Separation)	하나의 음원에서 보컬, 베이스, 드럼 등 각각의 구성 요소를 분리하는 기술	<ul style="list-style-type: none"> <li>• 딥러닝 기술로 분리된 음원의 품질 향상</li> <li>• 실시간 처리 가능</li> <li>• 리믹스, 샘플링, 노래방 모드 등 활용</li> </ul>
음악 생성 기술 (Generative AI for Music)	AI가 새로운 멜로디나 화성을 만들어내는 기술	<ul style="list-style-type: none"> <li>• 기존 음악의 패턴 학습</li> <li>• 특정 장르나 아티스트 스타일 반영 가능</li> <li>• 새로운 음악 자동 생성</li> </ul>
스타일 변환 기술	기존 음악의 장르나 분위기를 변경하는 기술	<ul style="list-style-type: none"> <li>• 장르 간 변환 (예: 팝→재즈)</li> <li>• 템포 및 분위기 변경</li> <li>• 다양한 버전 제작 가능</li> </ul>

출처: Mandy Dalugdug, "Tuned Global rolls out AI music features via AudioShake partnership", MusicBusinessWorldwide, 2025.01.30, <https://www.musicbusinessworldwide.com/tuned-global-rolls-out-ai-music-features-via-audioshake-partnership/>  
 양준석, "AI가 새롭게 연 음악 창작의 시대", AITimes, 2024.12.13, <https://www.aitimes.com/news/articleView.html?idxno=166187>  
 Emma Roth, "YouTube is testing music remixes made by AI", Theverge, 2024.11.13, <https://www.theverge.com/2024/11/12/24294891/youtube-ai-restyle-music-creator>

3) Mandy Dalugdug, "Tuned Global rolls out AI music features via AudioShake partnership", MusicBusinessWorldwide, 2025.01.30, <https://www.musicbusinessworldwide.com/tuned-global-rolls-out-ai-music-features-via-audioshake-partnership/>

## • AI 기반 음악 변형 기술을 통한 비즈니스 모델의 확대

- AI 음악 서비스는 기존의 스트리밍 중심 비즈니스 모델에서 벗어나 새로운 수익 모델을 창출하고 있음. 특히 AI 기술의 발전으로 음악의 생성, 변형, 소비 방식이 다각화되면서 산업 전반의 가치 사슬이 재편되고 있음
- B2C 시장에서는 AI 기술을 활용한 개인화 서비스가 새로운 수익원으로 부상하고 있음. 기본적인 음악 감상 기능은 무료로 제공하되, 리믹스, 가라오케 모드, 악기 소리 분리 등 고급 기능은 월 정액 구독 모델로 제공하는 방식이 확산되고 있음
- B2B 시장에서는 기업들의 맞춤형 음악 솔루션 수요가 증가하고 있음. 항공사, 통신사, 게임 업체 등 다양한 산업 분야에서 AI 기반 음악 서비스를 도입하고 있으며, 특히 브랜드 특화 음악 제작, 실시간 배경음악 생성 등의 서비스가 주목받고 있음
- AI 기술은 음악 제작과 유통 과정의 효율성도 크게 향상시키고 있음. 기존에 많은 시간과 비용이 소요되던 음원 편집, 믹싱, 마스터링 등의 작업이 자동화되면서 제작 비용이 절감되고 있음. 또한 AI 기반 메타데이터 생성과 자동 태깅 기술은 음악 검색과 추천 시스템의 정확도를 높여, 콘텐츠 발견성과 사용자 경험을 개선하고 있음. 이러한 기술 혁신은 단순한 대규모 카탈로그 제공만으로는 더 이상 충분하지 않은 현대 음악 시장에서 플랫폼들이 차별화된 서비스를 제공할 수 있게 하고 있음

## • 글로벌 테크 기업부터 스타트업까지 뛰어드는 AI 음악 변형 기술

- AI 기술이 음악 산업에 빠르게 도입되면서, 글로벌 테크 기업부터 스타트업까지 다양한 기업들이 혁신적인 음악 기술을 선보이고 있음. 이들은 음악 생성, 편집, 변형 등 다양한 영역에서 AI 기술을 활용한 서비스를 개발하며 시장을 선도하고 있음
- 오픈AI(OpenAI)에서 개발한 뮤즈넷(MuseNet)은 클래식부터 현대 음악까지 다양한 장르와 스타일을 학습하여 새로운 곡을 작곡할 수 있는 모델임. 예를 들어, 모차르트(Mozart)의 기법을 활용해 재즈와 클래식을 혼합한 곡을 생성하거나, 특정 영화의 배경음악 스타일을 모방하여 새로운 트랙을 만들어낼 수 있음
- 한국의 AI 기업인 수퍼톤(Supertone)은 기술을 활용하여 고(故) 김광석의 목소리를 복원하였으며, 이를 통해 다양한 언어로 노래를 제작하였음. 또한, 실시간으로 목소리를 변환하는 도구를 개발하여 게이머, 유튜버, 음악 엔지니어 등에게 새로운 창작의 가능성을 제공하고 있음
- 수노(Suno)와 유디오(Udio)는 텍스트 프롬프트 기반의 음악 생성 서비스를 선보이며 시장의 주목을 받고 있음. 사용자가 원하는 장르, 분위기, 스타일 등을 텍스트로 입력하면 기술이 이에 맞는 음악을 자동으로 생성하는 방식임. 수노의 경우 2023년 12월 출시 이후 5개월 만에 1,000만 명 이상의 사용자를 확보하며 음악 서비스의 대중화 가능성을 입증하였음. 이들의 서비스는 전문적인 음악 지식이 없는 일반 사용자도 쉽게 음악을 만들 수 있다는 점에서 큰 호응을 얻고 있음. 유디오(Udio)는 구글 딥마인드(Google DeepMind) 출신 엔지니어들이 설립한 회사로, 단순한 텍스트 입력만으로 전문가 수준의 음악을 생성할 수 있는 기술을 보유하고 있음
- 유튜브(YouTube)는 크리에이터들이 기술을 활용하여 라이선스된 곡을 리믹스할 수 있는 기능을 테스트하고 있음. 이 기능을 통해 사용자는 곡의 분위기나 장르를 변경하여 30초 길이의 새로운 사운드트랙을 생성할 수 있음. 생성된 사운드트랙은 원곡에 대한 명확한 출처 표기와 함께 기술을 활용하여 변형되었음을 표시함
- 엔비디아(NVIDIA)는 음악 편집기인 푸가토(Fugatto)를 공개하였음. 이 도구는 텍스트 및 오디오 입력을 기반으로 이전에 존재하지 않았던 새로운 소리를 생성할 수 있음. 예를 들어, ‘고양이처럼 우는 트럼펫 소리’와 같은 독특한 사운드를 만들어낼 수 있으며, 사용자의 음성을 변형하여 다른 억양이나 감정을 부여하거나, 음악에서 보컬을 분리하고 악기를 추가하는 등 다양한 기능을 제공함

**[AI 기반음악변형 기술 기업별 개발 사례]**

분야	기업	주요 기술	특징
음악 생성	OpenAI	AI 작곡 모델	<ul style="list-style-type: none"> <li>클래식부터 현대 음악까지 다양한 장르 작곡</li> <li>장르 혼합 및 스타일 모방 기능</li> </ul>
	Suno	텍스트 기반 음악 생성	<ul style="list-style-type: none"> <li>2023년 12월 출시</li> <li>5개월 만에 1,000만 사용자 확보</li> </ul>
	Udio AI	텍스트 프롬프트 기반 음악 생성	<ul style="list-style-type: none"> <li>구글 딥마인드 출신 엔지니어 개발</li> <li>전문가 수준의 음악 생성</li> </ul>
음성 기술	Supertone	AI 음성 복원 및 변환	<ul style="list-style-type: none"> <li>고인 가수 음성 복원</li> <li>실시간 음성 변환 도구 제공</li> </ul>
편집/변형	YouTube	AI 리믹스 도구	<ul style="list-style-type: none"> <li>라이선스 곡 리믹스 기능</li> <li>30초 길이 사운드트랙 생성</li> </ul>
	NVIDIA	AI 음악 편집기	<ul style="list-style-type: none"> <li>새로운 소리 생성</li> <li>음성 변형 및 보컬 분리 기능</li> </ul>

양준석, “AI가 새롭게 연 음악 창작의 시대”, AITimes, 2024.12.13, <https://www.aitimes.com/news/articleView.html?idxno=166187>  
 Molly Bohannon, “Major Record Labels Sue AI Music Startups For Allegedly Copying Songs To Train AI”, Forbes, 2024.01.25, <https://www.forbes.com/sites/mollybohannon/2024/06/24/major-record-labels-sue-ai-music-startups-for-allegedly-copying-songs-to-train-ai/>  
 Andrew R. Chow, “Kyogu Lee CEO, Supertone”, Time, 2024.09.05, <https://time.com/7012837/kyogu-lee>  
 Emma Roth, “YouTube is testing music remixes made by AI”, Theverge, 2024.11.13, <https://www.theverge.com/2024/11/12/24294891/youtube-ai-restyle-music-creators>  
 Emma Roth, “Nvidia claims a new AI audio generator can make sounds never heard before”, Theverge, 2024.11.26, <https://www.theverge.com/2024/11/25/24305584/nvidia-fugatto-ai-audio-generator-music>

## AI 기반 음악 변형 기술에 따른 저작권 논란

- AI를 활용한 음원 데이터 수집과 학습의 문제
  - AI 기반 음악 변형 기술이 음악 산업을 확장하는 한편, 학습 데이터로 활용되는 음원의 저작권 문제가 주요 쟁점으로 떠오르고 있음. 특히 AI 기업들이 대규모 음원 데이터를 무단으로 수집하여 AI 모델 학습에 활용하는 사례가 증가하면서, 음악 산업계의 우려가 커지고 있음. 이는 단순한 저작권 침해를 넘어 음악 창작의 가치와 아티스트의 권리를 위협하는 중대한 문제로 인식되고 있음
  - AI 기업들은 공개된 음악 스트리밍 플랫폼, 온라인 음원 저장소 등을 통해 대규모 음원을 수집하고 있으며, 자막 제공 사이트에서 가사 데이터를 확보하는 등 데이터 수집 방식이 더욱 정교화되고 있음. 특히 음원 파일뿐만 아니라 가사, 악보, 연주 패턴, 음색 등의 다양한 요소까지 수집하며, 이 과정에서 저작권자의 동의나 적절한 보상이 이루어지지 않는 경우가 많음
  - 또한, AI 기업들이 학습 데이터의 출처를 명확히 밝히지 않고 수집하는 사례가 확인되면서, 저작권 침해가 연쇄적으로 확산될 가능성이 제기되고 있음. 수집된 데이터의 품질이 보장되지 않을 경우, 기술이 생성한 음악의 신뢰성과 안정성에도 영향을 미칠 수 있음. 이러한 문제로 인해 유니버설 뮤직 그룹(Universal Music Group), 소니 뮤직(Sony Music), 워너 뮤직 그룹(Warner Music Group) 등 주요 음반사들은 AI 음악 생성 기업 수노(Suno)와 유디오(Udio)를 상대로 저작권 침해 소송을 제기하는 등 법적 대응에 나서고 있음<sup>4)</sup>
- 음원 데이터 수집 과정에서의 저작권 분쟁 사례
  - 2024년 6월, 세계 3대 메이저 음반사인 유니버설 뮤직 그룹(Universal Music Group), 소니 뮤직(Sony Music), 워너 뮤직 그룹(Warner Music Group)은 AI 음악 생성 기업 수노(Suno)와 유디오 AI(Udio AI)를 상대로 저작권 침해 소송을 제기함<sup>5)</sup>. 이들은 AI 기업들이 자사의 음원을 무단으로 수집하여 모델 학습에 활용했다고 주장하며, 작품당 최대 15만 달러의 배상을 요구함

4) Ashley Carman and Rachel Metz, “Major Record Labels Sue AI Music Generators”, Time, 2024.06.24, <https://time.com/6991466/record-labels-sue-ai-music-generator-startups/>

5) Molly Bohannon, “Major Record Labels Sue AI Music Startups For Allegedly Copying Songs To Train AI”, Forbes, 2024.01.25, <https://www.forbes.com/sites/mollybohannon/2024/06/24/major-record-labels-sue-ai-music-startups-for-allegedly-copying-songs-to-train-ai/>

- 소송의 핵심은 AI 기업들이 음원을 ‘대규모로 무단 복제’하여 모델을 학습시켰다는 점임. 특히 수노의 경우 출시 5개월 만에 1,000만 명 이상의 사용자를 확보하며 급성장했으나, 이 과정에서 저작권이 있는 음원을 무단으로 활용했다는 의혹을 받고 있음<sup>6)</sup>
- 템테이션스(Temptations)의 마이 걸(My Girl), 그린 데이(Green Day)의 아메리칸 이디엇(American Idiot), 머라이어 캐리(Mariah Carey)의 올 아이 원트 포 크리스마스(All I Want for Christmas) 등 유명 곡들과 매우 유사한 사운드를 생성했으며, 린 마누엘 미란다(Lin-Manuel Miranda), 브루스 스프링스틴(Bruce Springsteen), 마이클 잭슨(Michael Jackson) 등의 목소리와 구분하기 어려운 보컬을 만들어내고 있다는 점이 문제로 지적되고 있음
- AI 음악 생성 기업들은 자신들의 기술이 ‘변형적 사용’에 해당한다고 주장하고 있음<sup>7)</sup>. 예를 들어 수노는 사용자가 아티스트 이름을 프롬프트에 포함할 수 없게 하는 등의 조치를 취하고 있으며, 자사의 기술이 기존 콘텐츠를 단순 복제하는 것이 아닌 완전히 새로운 결과물을 만들어내는 것이라고 설명함
- 음반사들은 AI 기업들이 저작권이 있는 음원을 무단으로 활용하여 수익을 창출하고 있다고 비판함. 미국음반산업협회(RIAA, Recording Industry Association of America)의 미치 글레이저(Mitch Glazier) CEO<sup>8)</sup>는 “음악 커뮤니티는 AI를 수용하고 있으며, 책임 있는 개발자들과 협력하여 아티스트와 작곡가들이 주도하는 지속 가능한 도구를 구축하고 있다”면서도, “동이나 보상 없이 아티스트의 작품을 무단으로 활용하는 서비스들은 진정한 혁신의 약속을 후퇴시키고 있다”고 비판함

## 시사점과 대응과제

### • 음악 저작권 산업의 새로운 도전과제

- 음악 산업에 AI 기술의 도입이 본격화되면서, 음악 변형이 용이해지고 있음. 이는 음악 콘텐츠의 활용 범위를 확대하고 새로운 비즈니스 기회를 창출하는 반면, 원저작물의 무단 활용 위험도 증가시키고 있음
- 메이저 음반사들의 AI 음악 생성 기업들에 대한 법적 대응은 AI 시대의 음악 저작권 보호가 직면한 새로운 도전을 보여줌. 특히 AI의 학습 데이터로 활용되는 음원에 대한 권리 처리와 보상 체계 구축이 시급한 과제로 대두되고 있음
- 현재 음악 산업은 창작과 혁신을 장려하면서도 저작권을 보호해야 하는 균형점을 찾아야 하는 상황임. AI 기술을 활용한 음악 서비스 기업들이 적법한 절차를 통해 음원을 활용할 수 있도록 하는 새로운 라이선싱 모델의 개발이 필요함
- AI 기반 음악 변형 기술이 확산됨에 따라, 2차적 저작물에 대한 권리 관계도 더욱 복잡해질 것으로 예상됨. 특히 AI가 생성한 음악이 기존 곡과 유사한 경우, 저작권 침해 여부를 판단하는 기준이 모호해질 수 있어 이에 대한 명확한 기준 마련이 요구됨
- 향후 음악 저작권 산업은 AI 기술의 발전과 함께 새로운 비즈니스 모델을 발굴하고, 이에 맞는 권리 보호 체계를 구축해야 할 것임. 이는 기술 혁신과 창작자 보호의 조화를 이루는 방향으로 진행되어야 하며, 산업의 지속가능한 성장을 위한 핵심 과제가 될 것으로 전망됨

6) Molly Bohannon, “Major Record Labels Sue AI Music Startups For Allegedly Copying Songs To Train AI”, Forbes, 2024.01.25, <https://www.forbes.com/sites/mollybohannon/2024/06/24/major-record-labels-sue-ai-music-startups-for-allegedly-copying-songs-to-train-ai/>

7) Ashley Carman and Rachel Metz, “Major Record Labels Sue AI Music Generators”, Time, 2024.01.24, <https://time.com/6991466/record-labels-sue-ai-music-generator-startups/>

8) Ashley Carman and Rachel Metz, “Major Record Labels Sue AI Music Generators”, Time, 2024.01.24, <https://time.com/6991466/record-labels-sue-ai-music-generator-startups/>

## 참고문헌

- Mandy Dalugdug, “Tuned Global rolls out AI music features via AudioShake partnership”, MusicBusinessWorldwide, 2025.01.30, <https://www.musicbusinessworldwide.com/tuned-global-rolls-out-ai-music-features-via-audioshake-partnership/>
- Molly Bohannon, “Major Record Labels Sue AI Music Startups For Allegedly Copying Songs To Train AI”, Forbes, 2024.01.25, <https://www.forbes.com/sites/mollybohannon/2024/06/24/major-record-labels-sue-ai-music-startups-for-allegedly-copying-songs-to-train-ai/>
- Ashley Carman and Rachel Metz, “Major Record Labels Sue AI Music Generators”, Time, 2024.01.24, <https://time.com/6991466/record-labels-sue-ai-music-generator-startups/>
- 양준석, “AI가 새롭게 연 음악 창작의 시대”, AI타임즈, 2024.12.13, <https://www.aitimes.com/news/articleView.html?idxno=166187>
- Andrew R. Chow, “Kyogu Lee CEO, Supertone”, Time, 2024.09.05, <https://time.com/7012837/kyogu-lee>
- Emma Roth, “Nvidia claims a new AI audio generator can make sounds never heard before”, Theverge, 2024.11.26, <https://www.theverge.com/2024/11/25/24305584/nvidia-fugatto-ai-audio-generator-music>
- Emma Roth, “YouTube is testing music remixes made by AI”, Theverge, 2024.11.13, <https://www.theverge.com/2024/11/12/24294891/youtube-ai-restyle-music-creators>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## AI 기반 영상 생성 기술 발전 가속... 영상 제작 자동화 혁신 본격화되나

### 뉴스 브리프

어도비(Adobe)가 상업적 활용이 가능한 'Firefly 비디오 모델'을 출시하면서 AI 기반 영상 생성 기술이 새로운 전환점을 맞이하고 있다. 이 기술은 라이선스가 확보된 콘텐츠만을 학습 데이터로 활용하여 저작권 침해 우려를 해소했으며, 월 9.99달러부터 시작하는 구독형 요금제를 통해 진입 장벽을 낮췄다. 어도비의 전문 영상 편집 도구들과 통합된 워크플로우를 제공하여 AI 기반 영상 생성부터 편집까지 원스톱으로 가능해졌고, 이미 글로벌 기업들의 베타 테스트를 통해 상업적 활용 가능성이 검증되었다. 최근 클라우드 기반의 협업 플랫폼과 AI 기반 자동화 워크플로우가 결합되면서 영상 제작의 효율성이 크게 향상되고, 전문가급 영상 제작이 더욱 접근하기 쉬워지는 추세다. 어도비의 행보에서 특히 주목할 만한 점은 'Content Credentials' 시스템의 도입으로, AI 생성 콘텐츠의 출처와 생성 방식을 투명하게 공개하는 메타데이터 표준을 확산시키고 있다는 점이다.

### 저작권 안전성 확보한 AI 영상 생성 기술의 시장 출시

- 어도비의 상업용 AI 영상 생성 기술 출시와 의의
  - 어도비(Adobe)의 'Firefly 비디오 모델<sup>1)</sup>' 공개 베타 서비스 출시와 함께 라이선스 확보 콘텐츠 기반 학습으로 저작권 침해 우려 없는 상업용 AI 영상 생성 도구의 본격적 시장 진입 단계 돌입
  - 텍스트나 이미지 입력만으로 1080p 고해상도 영상 생성이 가능하며 카메라 앵글과 움직임까지 세밀하게 제어 가능한 전문가급 기능의 구현으로 산업 현장 활용성 확대
  - 월 9.99달러부터 시작하는 구독형 요금제 도입으로 전문 크리에이터부터 일반 사용자까지 폭넓은 활용 기반 확보와 함께 AI 영상 생성 기술의 대중화 가속

1) Firefly 비디오 모델: 어도비가 개발한 AI 기반 영상 생성 모델로, 어도비 스톡(Adobe Stock)의 라이선스 콘텐츠와 공개 도메인 콘텐츠를 학습 데이터로 활용하여 저작권 침해 위험을 최소화

• **기존 구독형 서비스 제품과의 연계를 통한 전문가급 영상 제작 환경 고도화**

- 어도비는 자사의 창작 소프트웨어 제품을 클라우드 기반으로 제공하는 구독형 서비스 ‘크리에이티브 클라우드(Creative Cloud)’와의 통합으로 AI 기반 영상 생성부터 편집까지 원스톱 워크플로우 구현 및 실무 활용성 제고
- 생성된 영상의 세부 편집 및 후보정이 가능한 전문가급 도구 제공으로 고품질 결과물 제작 지원과 함께 실무 활용 가능성 검증
- Deloitte Digital, IBM, Mattel 등 글로벌 선도 기업들의 베타 테스트 참여로 2024년 2월 기준 전 세계적으로 180억 건 이상의 산출물 생성 기록 달성

• **글로벌 기업들의 AI 영상 생성 기술 경쟁 가속화로 시장 성장 모멘텀 확대**

- 오픈AI, 런웨이(Runway) 등 AI 전문 기업들의 영상 생성 기술 개발 경쟁 본격화와 함께 5초 분량 기준 78센트~1.20달러 수준의 다양한 가격 정책 도입으로 시장 활성화 촉진
- 산업 현장의 실제 수요를 반영한 B-roll 영상 생성, 특수 효과 생성 등 실용적 기능 구현에 주력하며 활용 영역 확대
- 어도비의 경우 ‘콘텐츠 자격증명(Content Credentials)’ 시스템 도입으로 AI 생성 콘텐츠의 출처와 생성 방식을 투명하게 공개하는 메타데이터 표준 확산으로 저작권 보호 체계 구축

[주요 AI 영상 생성 서비스 비교]

서비스명	핵심 기능	가격 정책	주요 특징
Adobe Firefly	<ul style="list-style-type: none"> <li>• Text to Video</li> <li>• Image to Video</li> <li>• 카메라 제어</li> </ul>	<ul style="list-style-type: none"> <li>• 월 9.99달러(Standard)</li> <li>• 월 29.99달러(Pro)</li> </ul>	<ul style="list-style-type: none"> <li>• 저작권 안전 모델</li> <li>• Creative Cloud 통합</li> <li>• Content Credentials</li> </ul>
Runway	<ul style="list-style-type: none"> <li>• 영상 생성</li> <li>• 영상 편집</li> <li>• 특수효과</li> </ul>	<ul style="list-style-type: none"> <li>• 0.78~1.20달러/5초</li> <li>• 월 95달러 무제한</li> </ul>	<ul style="list-style-type: none"> <li>• 전문가용 도구</li> <li>• 고품질 결과물</li> <li>• 유연한 과금 체계</li> </ul>
OpenAI Sora	<ul style="list-style-type: none"> <li>• Text to Video</li> <li>• 영상 확장</li> <li>• 스타일 변환</li> </ul>	<ul style="list-style-type: none"> <li>• 별도의 과금 없이 챗GPT 유료 구독 플랜 사용자 대상 제공</li> </ul>	<ul style="list-style-type: none"> <li>• 혁신적 생성 품질</li> <li>• 플랫폼 통합</li> <li>• 접근성 제한</li> </ul>

출처: "Adobe Expands Generative AI Offerings Delivering New Firefly App with Industry's First Commercially Safe Video Model", 2025.02.12, Adobe, <https://news.adobe.com/news/2025/02/firefly-web-app-commercially-safe>  
 Brody Ford, "Adobe to Charge 50 Cents Per AI Video in New Product Launch", 2025.02.12, Bloomberg, <https://www.bloomberg.com/news/articles/2025-02-12/adobe-to-charge-50-cents-per-ai-video-in-new-firefly-product-launch>  
 Interview AI, "OpenAI's video generation AI "Sora" is now available to the public. No additional cost for "ChatGPT" paying users.", 2024.12.10, <https://www.interview-ai.site/en/article/openais-video-generation-ai-sora-is-now-available-to-the-public-no-additional-cost-for-chatgpt-paying-users/>

**AI 기반 영상 생성의 핵심 기술 혁신**

• **고품질 영상 생성을 위한 텍스트/이미지 기반 기술의 고도화**

- 텍스트 프롬프트나 참조 이미지만으로 1080p 고해상도 영상 생성이 가능한 수준까지 발전하며, 카메라 앵글과 줌 효과, 움직임 등 전문가급 촬영 기법의 AI 자동 구현 기술 확보

2) B-roll 영상: 주요 촬영 대상물과는 별개로 촬영된 보조 영상을 의미하며, 메인 스토리 영상인 A-roll 영상 사이를 보충-연결하여 주로 인터뷰나 주제에 대한 내용을 시각적으로 보여줄 때 사용  
 3) Content Credentials: AI 생성 콘텐츠에 대한 '영상 성분 표시' 같은 역할을 하는 메타데이터 시스템으로, 콘텐츠의 출처와 생성 방식을 투명하게 공개하여 신뢰성을 확보하는 기술 표준

- 4K 해상도 지원 및 프로페셔널 제작 워크플로우 지원 모델의 개발 진행으로 방송, 영화 등 전문 영상 제작 분야로의 활용 영역 확대 가능성 제시
- 대기업부터 개인 크리에이터까지 폭넓은 사용자층 확보를 통한 영상 제작 시장의 대중화 촉진과 함께 새로운 창작 생태계 형성의 기반 마련

**• 실시간 렌더링과 가상 제작 기술의 혁신으로 제작 효율성 제고**

- 실시간 8K 영상 처리 기술 구현으로 초고화질 영상의 즉각적인 생성 및 편집이 가능해지며, LED 월<sup>4)</sup> 기반 가상 스튜디오와 AI 영상 생성 기술의 결합으로 혁신적 제작 환경 구축
- 가상 배경, 특수 효과 등의 실시간 합성 구현으로 후반 작업 시간과 비용의 대폭 절감 효과 실현과 함께 제작 프로세스의 효율화 달성 기대
- 제작 비용 효율화를 통한 중소기업 제작사의 고품질 콘텐츠 제작 진입장벽 완화로 영상 콘텐츠 제작 시장의 다양성 확대 기반 조성 전망

**• AI 기반 영상 편집 및 후보정 기술의 고도화로 작업 시간 및 비용 절감**

- AI 알고리즘을 활용한 자동 장면 검출 및 편집 추천 기능의 구현으로 편집 작업 시간 단축과 함께 전문 인력의 창의적 작업 집중을 위한 환경 조성
- 색보정, 안정화, 노이즈 제거 등 후보정 작업의 자동화 실현으로 전문 인력 의존도 감소와 제작 비용 절감 효과 달성
- 3D 스케치 및 레퍼런스 형태 입력만으로 전문가급 품질의 이미지 생성이 가능한 어도비의 'Scene to Image'<sup>5)</sup> 기능 구현 등 정밀한 구도 및 앵글 제어 역량 확보

[AI 영상 생성 기술의 혁신 현황 및 효과]

기술 영역	주요 혁신 내용	기술적 성과	산업적 효과
텍스트/이미지 기반 영상 생성	<ul style="list-style-type: none"> <li>• 자연어 기반 제어</li> <li>• 참조 이미지 활용</li> <li>• 카메라 워크 구현</li> </ul>	<ul style="list-style-type: none"> <li>• 1080p 해상도</li> <li>• 전문가급 품질</li> <li>• 정밀한 제어성</li> </ul>	<ul style="list-style-type: none"> <li>• 제작 진입장벽 완화</li> <li>• 시간/비용 절감</li> <li>• 창작의 대중화</li> </ul>
실시간 렌더링/가상제작	<ul style="list-style-type: none"> <li>• AI 기반 실시간 편집</li> <li>• 가상 배경 및 디지털 휴먼 생성 기술 고도화</li> </ul>	<ul style="list-style-type: none"> <li>• 최대 4K 해상도 지원 및 실시간 품질 향상</li> <li>• AI 자동 색 보정 및 카메라 워크 최적화</li> </ul>	<ul style="list-style-type: none"> <li>• 작업 시간 단축</li> <li>• AI 기반 편집 자동화로 인건비 절감</li> <li>• 시청자 맞춤형 콘텐츠 제공으로 몰입도 증가</li> </ul>
편집/후보정 자동화	<ul style="list-style-type: none"> <li>• 자동 장면 검출</li> <li>• 색보정 자동화</li> <li>• 3D 스케치 변환</li> </ul>	<ul style="list-style-type: none"> <li>• 작업 시간 단축</li> <li>• 일관된 품질</li> <li>• 정밀한 제어</li> </ul>	<ul style="list-style-type: none"> <li>• 인력 의존도 감소</li> <li>• 생산성 향상</li> <li>• 활용성 확대</li> </ul>

출처: "Adobe Expands Generative AI Offerings Delivering New Firefly App with Industry's First Commercially Safe Video Model", 2025.02.12, Adobe, <https://news.adobe.com/news/2025/02/firefly-web-app-commercially-safe>  
 Jeremy Gray, "Adobe's AI Video Model Launches, Available Through New Premium Firefly Video/Audio Plans", 2025.02.12, PetaPixel, <https://petapixel.com/2025/02/12/adobes-ai-video-model-launches-available-only-with-new-firefly-video-audio-plans/>

Michael Nuñez, "Adobe Firefly AI video generator debuts—the most 'IP-safe' AI tool yet?", 2025.02.12, VentureBeat, <https://venturebeat.com/ai/adobe-firefly-ai-video-generator-debuts-the-most-ip-safe-ai-tool-yet/>  
 Heygen, "Top AI Video Trends", 2024.11.08. <https://www.heygen.com/article/top-ai-video-trends>

4) LED 월(LED Wall): 대형 LED 스크린을 활용한 가상 세트 시스템으로, 실시간으로 배경을 변경하고 카메라 움직임과 연동된 자연스러운 합성 영상을 제작할 수 있는 기술  
 5) Scene to Image: 3D 스케치나 도형을 입력하면 이를 기반으로 정교한 이미지를 생성하는 기술로, 전문 그래픽 디자인 작업 과정에서 활용 가능한 어도비의 최신 기능

## AI/클라우드 협업과 워크플로우 혁신의 가속화

- 클라우드 기반 협업 플랫폼의 도입으로 글로벌 제작 환경 혁신 가속화**
  - 클라우드 기반 영상 편집 도구의 도입으로 지역과 시간의 제약 없는 실시간 협업 환경이 구축되며, 5G 네트워크와 클라우드 컴퓨팅의 결합으로 원격 협업 워크플로우의 안정성 강화
  - 다수의 편집자가 동일 프로젝트에 동시 접근하여 작업할 수 있도록 실시간 변경 사항 동기화와 버전 관리 기능을 제공하는 멀티유저 환경으로 글로벌 제작 프로젝트의 효율성 및 생산성 향상 실현
  - 확장 가능한 스토리지와 처리 능력 제공으로 대규모 프로젝트의 효율적 관리가 가능해지며, 실시간 변경 사항 동기화와 버전 관리 기능으로 작업 안정성 확보
- AI 기반 자동화 워크플로우 도입으로 제작 효율성 및 품질 향상**
  - AI 알고리즘을 활용한 영상 분석으로 핵심 장면 자동 식별 및 편집 제안 기능이 구현되며, 메타데이터 자동 생성 및 태깅으로 대규모 영상 자산의 효율적 관리 체계 구축
  - AI를 활용하여 원본 영상의 음성을 20개 이상 언어로 자동 변환하는 다국어 음성 변환 기술의 도입으로 글로벌 콘텐츠 제작의 효율성이 향상되며, 자연스러운 더빙 성능 구현으로 현지화 비용 절감 효과 달성
  - 음성, 톤, 억양의 자연스러운 매칭 기술 적용으로 원본 영상의 감정과 뉘앙스를 유지하면서도 다양한 언어로의 변환이 가능한 수준으로 기술 발전 실현
- AI 기반 품질 관리 시스템 도입으로 일관된 품질 수준 확보**
  - AI 기반 자동 품질 검사 시스템의 도입으로 영상 품질의 실시간 모니터링 및 개선이 가능해지며, 해상도, 프레임 레이트, 색감 등 기술적 품질 지표의 자동 분석 체계 구축
  - 스타일 일관성, 브랜드 가이드라인 준수 여부 등 콘텐츠 품질의 자동 검증 지원으로 브랜드 아이덴티티 유지와 품질 표준화 실현
  - 어도비의 경우 Content Credentials 시스템과의 연계를 통한 AI 생성 콘텐츠의 투명성 확보 체계 구축으로 저작권 보호와 콘텐츠 신뢰성 강화 기반 마련

### [클라우드 협업 및 AI 워크플로우 혁신 효과]

구분	기존 제작 방식	AI/클라우드 기반 제작	주요 개선 효과
협업 방식	<ul style="list-style-type: none"> <li>물리적 공간 제약</li> <li>순차적 작업 진행</li> <li>제한적 피드백</li> </ul>	<ul style="list-style-type: none"> <li>실시간 원격 협업</li> <li>동시 작업 가능</li> <li>즉각적 피드백</li> </ul>	<ul style="list-style-type: none"> <li>작업 효율성 향상</li> <li>인력 운영 유연성 증가</li> </ul>
콘텐츠 현지화	<ul style="list-style-type: none"> <li>수동 번역/더빙</li> <li>높은 제작 비용</li> <li>긴 제작 기간</li> </ul>	<ul style="list-style-type: none"> <li>AI 자동 변환</li> <li>음성/톤 매칭</li> <li>20개 이상 언어</li> </ul>	<ul style="list-style-type: none"> <li>제작 기간 단축</li> <li>품질 일관성 확보</li> </ul>
품질 관리	<ul style="list-style-type: none"> <li>수동 품질 검사</li> <li>주관적 평가</li> <li>사후 개선</li> </ul>	<ul style="list-style-type: none"> <li>자동 품질 분석</li> <li>객관적 지표</li> <li>실시간 개선</li> </ul>	<ul style="list-style-type: none"> <li>품질 일관성 확보</li> <li>관리 효율성 증가</li> </ul>

출처: Jim Thacker, "Adobe launches the Firefly Video Model in public beta", 2025.02.12, CG Channel, <https://www.cgchannel.com/2025/02/adobe-is-extending-firefly-generative-ai-to-its-video-tools/>

"Adobe Expands Generative AI Offerings Delivering New Firefly App with Industry's First Commercially Safe Video Model", 2025.02.12, Adobe, <https://news.adobe.com/news/2025/02/firefly-web-app-commercially-safe>

Michael Nuñez, "Adobe Firefly AI video generator debuts—the most 'IP-safe' AI tool yet?", 2025.02.12, VentureBeat, <https://venturebeat.com/ai/adobe-firefly-ai-video-generator-debuts-the-most-ip-safe-ai-tool-yet/>

Jeremy Gray, "Adobe's AI Video Model Launches, Available Through New Premium Firefly Video/Audio Plans", 2025.02.12, PetaPixel, <https://petapixel.com/2025/02/12/adobes-ai-video-model-launches-available-only-with-new-firefly-video-audio-plans/>

## 어도비, 저작권 보호 기술 표준화 및 시장 신뢰성 확보 추진

- **Content Credentials 시스템을 통한 AI 생성 콘텐츠 검증 기반 마련**
  - 어도비는 AI 생성 콘텐츠에 대한 '영양 성분 표시' 역할을 하는 메타데이터 시스템인 'Content Credentials' 도입으로 콘텐츠 출처와 생성 방식의 투명한 공개 체계 구축 및 글로벌 기술 표준으로서의 입지 강화
  - 또한 어도비 주도의 '콘텐츠 진위성 이니셔티브(CAI)<sup>8)</sup>'를 통해 BBC, ARM, MS 등 글로벌 기업들이 참여하는 디지털 콘텐츠 진위 검증 기술 표준화 이니셔티브 확산
  - 어도비의 웹 기반 Inspect 도구 제공으로 소비자가 직접 AI 생성 콘텐츠 여부를 확인할 수 있는 검증 환경 구축과 함께 Firefly 비디오 모델 생성 영상의 Content Credentials 자동 인증 체계 도입
- **라이선스 확보 학습 데이터 활용으로 저작권 침해 리스크 최소화**
  - 어도비 스톡(Adobe Stock)의 라이선스 콘텐츠와 공개 도메인 콘텐츠만을 AI 모델 학습에 활용함으로써 저작권 침해 우려 해소 및 상업적 활용이 가능한 최초의 AI 영상 생성 기술 포지셔닝
  - 2024년 초 일반 사용자 콘텐츠의 AI 학습 활용을 명시적으로 제외하고 어도비 스톡의 콘텐츠만을 활용하도록 이용약관을 개정<sup>9)</sup>함으로써 창작자 권리 보호 강화
  - Deloitte, IBM, Mattel 등 글로벌 선도 기업들의 베타 테스트 참여를 통한 상업적 안전성 검증 및 시장 신뢰도 제고

## 결론 및 시사점

- **상업적 활용이 가능한 AI 영상 생성 기술의 등장**
  - 어도비의 Firefly 비디오 모델 출시는 라이선스가 확보된 학습 데이터 활용, Content Credentials 시스템 도입, 투명한 메타데이터 공개 등 저작권 보호 중심의 새로운 AI 개발 패러다임 제시
  - 전문 영상 제작 도구와의 통합, 구독형 요금제 도입, 직관적 사용자 환경 제공 등을 통해 AI 영상 생성 기술의 상업적 활용 가능성 검증 및 시장 신뢰도 확보
  - 저작권 보호와 AI 기술 혁신의 균형점을 찾기 위한 산업계 공동의 기술 표준 및 윤리 가이드라인 수립 필요성 증가
- **AI 시대 저작권 산업의 새로운 성장 동력 역할 주목**
  - AI 기술 도입에 따른 콘텐츠 제작의 대중화와 시장 확대에 대응하여 저작권 보호 기술 고도화 및 새로운 비즈니스 모델 발굴 필요성 증대
  - 라이선스 확보 데이터 중심의 AI 학습 모델 개발 문화 정착을 통한 창작자 권리 보호와 공정한 보상 체계 확립에 대한 요구 확산 전망

8) Content Authenticity Initiative(CAI) : 어도비가 주도하는 디지털 콘텐츠 진위 검증 기술 표준화 이니셔티브로, BBC, ARM, MS 등 글로벌 기업들이 참여

9) Jim Thacker, "Adobe launches the Firefly Video Model in public beta", 2025.02.12, CG Channel, <https://www.cgchannel.com/2025/02/adobe-is-extending-firefly-generative-ai-to-its-video-tools/>

## 참고문헌

- Jim Thacker, "Adobe launches the Firefly Video Model in public beta", 2025.02.12, CG Channel, <https://www.cgchannel.com/2025/02/adobe-is-extending-firefly-generative-ai-to-its-video-tools/>
- "Adobe Expands Generative AI Offerings Delivering New Firefly App with Industry's First Commercially Safe Video Model", 2025.02.12, Adobe, <https://news.adobe.com/news/2025/02/firefly-web-app-commercially-safe>
- Brody Ford, "Adobe to Charge 50 Cents Per AI Video in New Product Launch", 2025.02.12, Bloomberg, <https://www.bloomberg.com/news/articles/2025-02-12/adobe-to-charge-50-cents-per-ai-video-in-new-firefly-product-launch>
- Michael Nuñez, "Adobe Firefly AI video generator debuts—the most 'IP-safe' AI tool yet?", 2025.02.12, VentureBeat, <https://venturebeat.com/ai/adobe-firefly-ai-video-generator-debuts-the-most-ip-safe-ai-tool-yet/>
- Jeremy Gray, "Adobe's AI Video Model Launches, Available Through New Premium Firefly Video/Audio Plans", 2025.02.12, PetaPixel, <https://petapixel.com/2025/02/12/adobes-ai-video-model-launches-available-only-with-new-firefly-video-audio-plans/>
- Mat Choat, "What's next in video production in 2025?", 2025.01.14, Logic+Magic, <https://logicandmagic.agency/insights/whats-next-in-video-production>
- "Latest technological innovations in video production—2025", 2025.01.15, All in Motion, <https://www.allinmotion.com/blog/latest-technological-innovations-in-video-production-2025>
- Conan Zhang, "Production: Tools and Techniques for 2025", 2025.02.09, talecast, <https://talecast.ai/blog/ai-in-video-production-tools-and-techniques-for-2025>
- "How Artificial Intelligence is Revolutionizing Creative Workflows", 2025.01.17, Vizologi, <https://vizologi.com/how-artificial-intelligence-is-revolutionizing-creative-workflows/>
- Interview AI, "OpenAI's video generation AI "Sora" is now available to the public. No additional cost for "ChatGPT" paying users.", 2024.12.10, <https://www.interview-ai.site/en/article/openais-video-generation-ai-sora-is-now-available-to-the-public-no-additional-cost-for-chatgpt-paying-users>
- Heygen, "Top AI Video Trends", 2024.11.08, <https://www.heygen.com/article/top-ai-video-trends>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 넷플릭스의 '오징어 게임' IP 확장, 콘텐츠와 게임 산업의 경계를 허물다

뉴스 브리프

2025년 게임 산업에서는 IP를 활용한 게임 개발이 핵심 트렌드로 부상하고 있다. 넷플릭스는 자사 IP를 활용한 게임 개발에 주력하여 매월 1개 이상의 신작을 출시할 계획이며, 오징어 게임 등 인기 IP를 활용한 게임을 출시하고 있다. 세가와 소니 등 게임사들은 자사 게임 IP의 영화, TV 시리즈 제작을 통해 트랜스미디어 전략을 추진하고 있다. 이러한 IP 기반 게임의 성공은 기존 IP 보유자들의 협상력을 강화하고 있으며, IP의 가치를 단순 라이선싱에서 기획 단계부터 다양한 매체로의 확장을 고려할 수 있는 중요한 아이템으로 격상했다. 게임은 IP의 가치를 확장하고 강화하는 핵심 매체로 자리잡으며, 이는 저작권 산업 전반의 비즈니스 모델 변화를 가속화할 것으로 전망된다.

### 넷플릭스의 자사 IP 기반 게임 개발 중심의 전략 전환 및 강화

- 매월 1개 이상의 자사 IP 기반 게임 출시 계획 수립 및 실행
- 넷플릭스(Netflix)는 2021년 나이트 스쿨 스튜디오(Night School Studio) 인수를 시작으로 게임 사업에 진출하여 초기에는 옥센 프리(Oxenfree), 모뉴먼트 밸리(Monument Valley) 등 독창적인 인디게임 중심의 전략을 펼쳤으나, 구독자의 1% 미만만 게임을 이용한다는 분석 결과<sup>1)</sup>에 따라 자사 IP를 활용한 게임 개발로 전략을 전환함
- 스위트 매그놀리아(Sweet Magnolias), 아우터 뱅크스(Outer Banks), 연애 실험: 블라인드 러브 NYC(Love Is Blind NYC), 지니 & 조지아 (Ginny & Georgia) 등 넷플릭스의 인기 IP를 활용한 인터랙티브 모바일 게임을 2025년 중 순차적으로 출시할 계획이며, 매월 1개 이상의 새로운 IP 기반 게임을 선보일 예정임<sup>2)</sup>

1) Kotaku, "The Future Of Netflix Gaming Sounds Grim", Kotaku, 2025.02.04., <https://kotaku.com/netflix-squid-game-unleashed-app-subscription-team-blue-1851755122>

2) Deadline, "Netflix Making Hay With Games Based On Its IP, Releasing One A Month", Deadline, 2024.07.18., <https://deadline.com/2024/07/netflix-video-games-ip-based-emily-in-paris-squid-games-1236014986/>

- 2023년 9월 연애 실험: 블라인드 러브(Love is Blind)를 시작으로 버진리버(Virgin River), 퍼펙트 매치(Perfect Match) 등의 게임을 출시했으며, 최근에는 에밀리, 파리에 가다(Emily in Paris)와 셀링 선셋(Selling Sunset) 게임을 출시하여 시리즈 팬들의 호응을 이끌어내는데 성공함
- 특히 넷플릭스(Netflix)의 최대 히트작인 오징어 게임(Squid Game)을 활용한 멀티플레이어 게임 '오징어 게임: 모바일 서바이벌(Squid Game: Unleashed)'이 1,000만 다운로드를 돌파<sup>3)</sup>하며 자사의 가장 성공적인 게임 타이틀로 자리매김했으며, 이를 시즌2 출시와 연계하여 시너지를 창출하는 전략을 구사함

**[넷플릭스 시리즈 “에밀리, 파리에 가다”의 시리즈 포스터(좌)와 게임 포스터(우)]**



(좌) 출처: devdiscourse, “Emily in Paris’ renewed for seasons 3 and 4 by Netflix”, devdiscourse, 2022.01.11, <https://www.devdiscourse.com/article/entertainment/1878813-emily-in-paris-renewed-for-seasons-3-and-4-by-netflix>  
 (우) 출처: imdb, “Netflix Stories: Emily in Paris”, imdb, 2024, <https://www.imdb.com/title/tt33144316/>

**[오징어 게임의 IP를 활용한 “오징어 게임: 모바일 서바이벌”]**



출처: Netflix, “Squid Game: Unleashed | Official Launch Trailer | Netflix”, Youtube, 2024.12.18, <https://www.youtube.com/watch?v=sT-WR6zdzj0>

- 이러한 성과를 바탕으로 넷플릭스는 시리즈와 게임의 동시 출시를 통해 팬들의 참여를 유도하고, 게임을 통해 새로운 캐릭터나 스토리라인을 소개하여 다음 시즌이나 속편에 반영하는 등 IP를 활용한 선순환 구조를 구축하고자 함
- 현재 80개 이상의 게임을 개발 중이며, WWE(World Wrestling Entertainment) 2K 모바일 게임의 독점 제공 계약 체결 등 외부 IP도 적극적으로 확보하여 게임 라이브러리를 확장<sup>4)</sup>해 나가고 있음

## 2025년 게임 산업의 IP 기반 콘텐츠 개발 트렌드 확산

### • IP 활용 게임의 성공적 시장 안착과 개발사들의 IP 확보 경쟁 심화

- 2024년 모노폴리 고(Monopoly GO!)의 성공을 필두로 해즈브로(Hasbro)와 마텔(Mattel) 등 전통적인 보드게임 및 완구 기업들의 IP를 활용한 게임들이 시장에서 두각을 나타내며, 게임 산업에서 IP의 중요성이 더욱 부각됨
- IP를 활용한 게임 개발은 이미 형성된 브랜드 인지도와 기존 사용자층을 활용할 수 있어 유저 획득 비용을 크게 절감할 수 있으며, 자연스러운 유기적 성장이 가능하다는 장점이 입증됨
- 기존 게임에 IP를 통합하는 한시적 이벤트도 활발히 진행되어, 슈퍼셀(Supercell)이 고질라(Godzilla) IP를, 부두(Voodoo)가 트랜스포머(Transformers) IP를 활용한 이벤트를 진행함. 그 결과 유저 재참여율 증가, 단기 이용자 유지율 상승, 이벤트 관련 인앱 결제 매출 증대 등 긍정적인 효과를 거둠

3) Isa Muhamma, "Netflix Games shelves titles and doubles down on IP originals", Pocket Gamer, 2025.02.03., <https://www.pocketgamer.biz/netflix-cancels-six-titles-shifts-focus-to-ip-based-games-and-new-releases/>  
 4) Deadline, "Netflix Making Hay With Games Based On Its IP, Releasing One A Month", Deadline, 2024.07.18., <https://deadline.com/2024/07/netflix-video-games-ip-based-emily-in-paris-squid-games-1236014986/>

**[보드게임 모노폴리를  
 모바일 게임으로 구현한 “모노폴리고”]**



출처: Jayden Hellyar, “Monopoly Go: Free Dice Links for June 2024”, thegamecrater, 2024.06, <https://www.thegamecrater.com/monopoly-go-free-dice-links-for-may-2024/>

**[게임사 부두의 게임 서비스 “몹 컨트롤”이  
 트랜스포머의 IP를 활용한 이벤트를 진행]**



출처: Peter Tolan, “Voodoo Teams Up with Hasbro to Introduce TRANSFORMERS Collaboration in Mob Control”, europeangaming, 2024.04.11, <https://europeangaming.eu/portal/latest-news/2024/03/11/154613/voodoo-teams-up-with-hasbro-to-introduce-transformers-collaboration-in-mob-control/>

- 이러한 성과를 바탕으로 게임 개발사들은 영화, 드라마, 애니메이션, 만화, 소설 등 다양한 분야의 IP 확보를 위한 경쟁에 돌입했으며, 일부 개발사는 자체 IP 개발에도 적극 투자하기 시작함
- 베트남 등 신흥 게임 시장에서도 IP를 활용한 퍼즐 게임이 전년 대비 17% 성장<sup>5)</sup>하는 등 IP 기반 게임의 성장세가 글로벌 시장으로 확대되는 양상을 보임
- 특히 IP를 활용한 게임은 장르 융합을 통해 더 넓은 사용자층을 확보할 수 있다는 강점이 있어, 2025년에는 더 다양한 형태의 IP 활용 게임이 출시될 것으로 전망됨

## 게임 IP의 트랜스미디어 콘텐츠 개발을 통한 산업 영향력 증대

### • 세가, 소니 등 게임사들의 적극적 IP 활용 트랜스미디어 전략 추진

- 세가(SEGA)는 2024년 해외 매출이 전체 게임 매출의 70%를 차지하는 등 글로벌 시장에서 성과를 거두고 있으며<sup>6)</sup>, 이를 기반으로 자사의 게임 IP를 영화, TV 시리즈 등 다양한 미디어로 확장하는 트랜스미디어 전략을 적극 추진하기로 함
- 슈퍼 소닉(Sonic the Hedgehog) 시리즈의 경우 게임과 영화, TV 시리즈를 유기적으로 연계하여 파라마운트+(Paramount+) 시리즈 너클스(Knuckles), 영화 슈퍼 소닉 3 (Sonic the Hedgehog 3), 게임 소닉 럼블(Sonic Rumble) 등을 통해 IP의 가치를 극대화하는데 성공함

[슈퍼소닉시리즈:(좌측)영화수퍼소닉3,(중앙)시리즈너클스,(우측)게임소닉럼블]



좌측 출처: rottentomatoes, “Sonic the Hedgehog 3”, 2024.12.20, [https://www.rottentomatoes.com/m/sonic\\_the\\_hedgehog\\_3](https://www.rottentomatoes.com/m/sonic_the_hedgehog_3)

중앙 출처: The Dubbing Database, Knuckles, 2024., <https://dubdb.fandom.com/wiki/Knuckles>

우측 출처: STEAM, “Sonic Rumble”, 2025., [https://store.steampowered.com/app/2951690/Sonic\\_Rumble/](https://store.steampowered.com/app/2951690/Sonic_Rumble/)

5) Chirag Ambwani, "Industry Perspectives on 2025 Mobile Gaming Trends", Unity, 2025.01.07., <https://unity.com/blog/2025-mobile-gaming-trends-industry-perspectives>

6) Shuji Utsumi, "Sega is changing its global strategy to become an IP powerhouse, reviving dormant franchises and aggressively pursuing transmedia projects", AUTOMATON WEST, 2024.12.13., <https://automaton-media.com/en/news/sega-is-changing-its-global-strategy-to-become-an-ip-powerhouse-reviving-dormant-franchises-and-aggressively-pursuing-transmedia-projects/>

- 세가는 현재 활발히 전개되고 있는 용과 같이(Like a Dragon), 페르소나(Persona) 등의 IP 외에도 크레이지 택시(Crazy Taxi), 젯 셋 라디오(Jet Set Radio), 버추어 파이터(Virtua Fighter), 골든 액스(Golden Axe) 등 휴면 상태였던 IP들의 부활도 추진 중임
- 또한 세가는 2024년 출시된 소닉 x 새도우 제너레이션(Sonic x Shadow Generations)의 성공, 시노비(Shinobi) 영화 각색 계약 체결 등 세가의 IP 트랜스미디어 전략이 이미 가시적인 성과를 보이고 있음
- 소니(Sony)는 자사가 보유한 플레이스테이션(PlayStation)의 인기 게임 IP들을 영화와 TV 시리즈로 제작했으며, HBO의 더 라스트 오브 어스(The Last of Us) 시리즈가 에미상(E Emmy Award)을 수상하는 등 게임 IP의 영상화가 작품성과 대중성 모두에서 인정받음
- 워너브라더스(Warner Bros.)는 슈퍼 마리오 브라더스(Super Mario Bros.) 영화의 흥행 성공을 이어가기 위해 마인크래프트(Minecraft) 영화를 제작 중이며, 캐슬바니아(Castlevania), 프레디의 피자가게(Five Nights at Freddy's), 모탈 컴뱃(Mortal Kombat) 등 다수의 게임 IP 기반 영상물이 제작되고 있음

[게임 “더 라스트 오브 어스”를 영상화한 동명의 TV 시리즈]



출처: DYLAN HAYES, “The Last of Us season 2 release window confirmed with new trailer”, marca, 2025.01.07, <https://www.marca.com/en/technology/gaming/2025/01/07/1677d3fda268e3ede188b457c.html>

[마인크래프트 영화 “빌리저스(Villagers)”]



출처: Victoria Kennedy, “Minecraft movie trailer reveals first look at its villagers, and they are sight for sore eyes”, eurogamer, 2025.02.13, <https://www.eurogamer.net/minecraft-movie-trailer-reveals-first-look-at-its-villagers-and-they-are-sight-for-sore-eyes>

## IP 기반 게임 개발의 확산이 저작권 산업에 미치는 영향과 전망

- 플랫폼과 게임사의 IP 활용 전략 다변화에 따른 저작권 산업 변화
- 전통적으로 게임 개발사들의 영역이었던 게임 시장에 넷플릭스와 같은 콘텐츠 플랫폼이 진출하면서, IP 활용의 주체가 다양화되고 있음
- 게임 IP의 영상화와 영상 IP의 게임화가 활발히 진행되면서, IP의 가치가 단순한 라이선싱을 넘어 콘텐츠 기획 단계부터 다양한 매체로의 확장을 고려하는 방향으로 진화하고 있음
- 오징어 게임의 성공적인 게임화 사례는 국내 IP의 글로벌 확장 가능성을 보여주는 동시에, 한국 저작권 산업이 나아가야 할 방향성을 제시하는 벤치마크로 작용할 것으로 전망됨
- IP 기반 게임의 성공으로 인해 게임 산업에서 IP의 중요성이 더욱 부각되면서, 기존 IP 보유자들의 협상력이 강화되고 IP 가치 산정 기준이 더욱 정교화될 것으로 예상됨
- 게임을 통한 IP의 확장은 기존 사용자층의 참여도를 높이고 새로운 소비자층을 확보하는 동시에, 게임 내 콘텐츠를 통해 다음 시즌이나 후속작의 스토리라인에 영향을 미치는 등 IP 활용의 새로운 패러다임을 제시함
- 2025년을 기점으로 게임은 단순한 엔터테인먼트 콘텐츠를 넘어 IP의 가치를 확장하고 강화하는 핵심 매체로 자리잡을 것으로 전망되며, 이는 저작권 산업 전반의 비즈니스 모델 변화를 가속화할 것으로 예상됨

## 참고문헌

- Isa Muhamma, "Netflix Games shelves titles and doubles down on IP originals", Pocket Gamer, 2025.02.03., <https://www.pocketgamer.biz/netflix-cancels-six-titles-shifts-focus-to-ip-based-games-and-new-releases/>
- Jay Peters, "Netflix's cloud plans include co-op and party games", The Verge, 2025.01.24., <https://www.theverge.com/2025/1/23/24350690/netflix-cloud-gaming-plans-co-op-party-games>
- Kotaku, "The Future Of Netflix Gaming Sounds Grim", Kotaku, 2025.02.04., <https://kotaku.com/netflix-squid-game-unleashed-app-subscription-team-blue-1851755122>
- Deadline, "Netflix Making Hay With Games Based On Its IP, Releasing One A Month", Deadline, 2024.07.18., <https://deadline.com/2024/07/netflix-video-games-ip-based-emily-in-paris-squid-games-1236014986/>
- Chirag Ambwani, "Industry Perspectives on 2025 Mobile Gaming Trends", Unity, 2025.01.07., <https://unity.com/blog/2025-mobile-gaming-trends-industry-perspectives>
- Pete Basgen, "Unlocking 2025: The Future of Gaming in 8 Moves", LBB Online, 2025.01.18., <https://lbbonline.com/news/unlocking-2025-the-future-of-gaming-in-8-moves>
- Shuji Utsumi, "Sega is changing its global strategy to become an IP powerhouse, reviving dormant franchises and aggressively pursuing transmedia projects", AUTOMATON WEST, 2024.12.13., <https://automaton-media.com/en/news/sega-is-changing-its-global-strategy-to-become-an-ip-powerhouse-reviving-dormant-franchises-and-aggressively-pursuing-transmedia-projects/>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 메타, AI 학습 데이터의 저작권 책임 회피 논란

### 뉴스 브리프

메타(Meta)가 AI 모델 학습을 위해 불법 복제된 전자책과 학술 논문을 대규모로 수집한 사실이 드러났다. 안나스 아카이브(Anna's Archive)를 통해 최소 81.7TB의 데이터를 토렌트로 다운로드했으며, 이 과정에서 회사의 개입 흔적을 숨기려 했다. 마크 저커버그(Mark Zuckerberg) 메타 CEO는 경쟁사들과의 기술 격차를 줄이기 위해 립젠 데이터셋 사용을 직접 승인했고, 저작권 정보와 메타데이터를 의도적으로 제거하도록 지시했다. AI 업계 전반이 '데이터 벽'에 직면한 상황에서, 일부 기업들은 합법적인 라이선스 계약을 체결하는 한편, 무단으로 데이터를 수집하는 이중적인 전략을 취하고 있다. 이는 AI 산업의 건전한 발전을 저해할 수 있다는 우려를 낳고 있으며, 지속가능한 데이터 생태계 구축과 윤리적 기준 정립의 필요성이 제기되고 있다.

### 메타의 대규모 데이터 무단 수집 현황

- 메타, 립젠 등 온라인 사이트에서 81.7TB 데이터 토렌트 다운로드
  - 메타(Meta)가 인공지능(artificial intelligence, 이하 AI) 모델 학습을 위해 불법 복제된 전자책을 대규모로 사용한 사실이 밝혀져 논란이 되고 있으며, 그 규모는 이전에 알려진 것보다 훨씬 큰 것으로 밝혀짐
  - 법원에 제출된 문서에 따르면<sup>1)</sup>, 메타는 안나스 아카이브(Anna's Archive)를 통해 여러 그림자 도서관(Shadow Library)<sup>2)</sup>에서 최소 81.7TB의 데이터를 토렌트 방식으로 다운로드했으며, 이 중 Z-라이브러리(Z-Library)와 립젠(LibGen)에서 최소 35.7TB의 데이터가 포함되어 있음
  - 또한, 메타는 이전에도 립젠에서 80.6TB의 데이터를 토렌트 방식으로 다운로드한 것으로 확인됨. 이러한 메타의 불법 데이터 수집 규모는 기존의 데이터 저작권 침해 사례와 비교했을 때 매우 큰 규모임

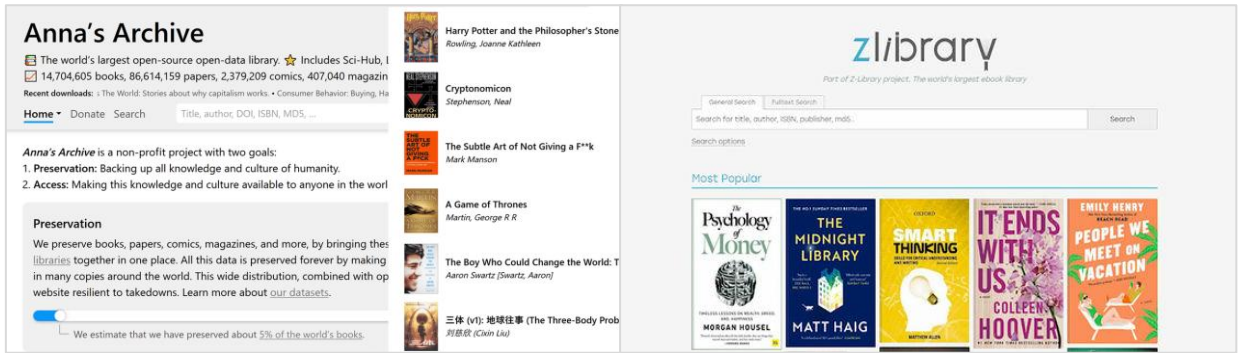
1) Ashley Belanger, "“Torrenting from a corporate laptop doesn't feel right”: Meta emails unsealed", 2025.02.07, Ars Technica,

<https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>

2) 그림자 도서관(Shadow Library) : 저작권자의 허락 없이 불법적으로 디지털 콘텐츠를 공유하는 온라인 플랫폼

- 메타가 수집한 데이터는 주로 도서, 학술 논문 등의 저작물로 구성되어 있으며, 이는 AI 언어 모델의 학습에 필수적인 텍스트 기반 데이터임. 특히 립젠 등과 같은 그림자 도서관은 AI 학습 커뮤니티에서 오랫동안 관심을 받아온 데이터 소스이며, 이러한 플랫폼들은 대규모의 저작권 보호 자료를 호스팅하고 있음

### 안나스 아카이브(좌)와 Z-라이브러리(우)



출처: Gigazine, “What is the pirated search engine “Anna’s Archive” that was born in response to legal measures against the world’s largest pirated e-book site?”, Gigazine, 2023.07.08, [https://gigazine.net/gsc\\_news/en/20230708-annas-archive/](https://gigazine.net/gsc_news/en/20230708-annas-archive/)

#### • 메타의 회사 인프라와 다운로드 연결 은폐 시도

- 메타는 이러한 대규모 데이터 수집 과정에서 회사의 개입 흔적을 숨기기 위한 다각도의 노력을 기울인 것으로 드러남. 메타는 의도적으로 페이스북(Facebook) 서버를 통한 직접 다운로드를 피했으며, 데이터 다운로더의 추적을 방지하기 위한 여러 기술적 조치를 실행함
- 프랭크 장(Frank Zhang) 메타 연구원이 작성한 내부 메시지는 이러한 은폐 시도를 명확히 보여줌. 장 연구원은 페이스북 서버를 통한 다운로드를 피함으로써 시더/다운로더 추적의 위험을 회피하고자 했다고 언급함<sup>3)</sup>
- 또한, 메타는 토렌트 시딩과 관련된 설정을 수정하여 가능한 한 최소한의 시딩만 발생하도록 조치했으며 이는 프로젝트 관리 책임자인 마이클 클라크(Michael Clark)의 증언을 통해 확인됨<sup>4)</sup>

#### • 메타, 제작자 동의 없는 콘텐츠 활용으로 AI 모델 학습

- 메타의 AI 모델 학습 데이터는 크게 두 가지 주요 소스에서 확보되었음. 첫 번째는 프로젝트 구텐베르크(Project Gutenberg)의 약 7만 권의 저작권이 만료된 도서들임. 두 번째는 더 파일(The Pile)의 북3(Book3) 데이터셋으로, 이는 비블리오틱 프라이빗 트래커(Bibliotik Private Tracker)와 같은 온라인 도서관의 콘텐츠를 포함하고 있음
- 메타는 이러한 데이터셋을 통해 85GB 이상의 도서 데이터를 확보했으며, 이는 픽션과 논픽션을 모두 포함하는 방대한 양의 텍스트 데이터임. 특히 주목할 만한 점은 이러한 데이터의 상당 부분이 저작권이 있는 콘텐츠였음에도 불구하고, 창작자들의 동의나 보상 없이 수집되었다는 점임

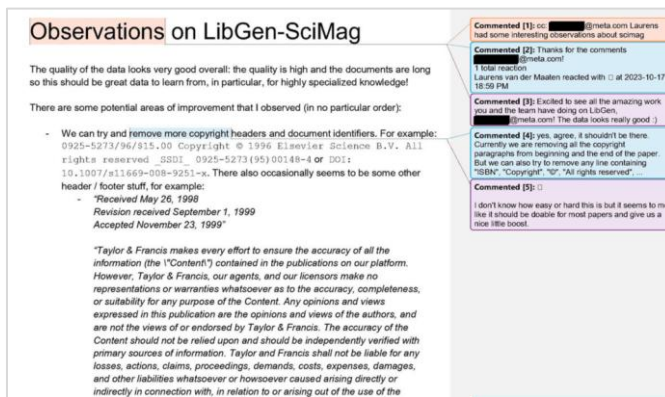
3) Ashley Belanger, ““Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed”, 2025.02.07, Ars Technica, <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>  
 4) Ashley Belanger, ““Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed”, 2025.02.07, Ars Technica, <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>

## 메타의 데이터 관리와 내부 의사결정 과정

### • 메타의 저작권 정보 및 메타데이터 제거 지시 현황

- 메타의 내부 문서들은 립젠의 학습 데이터에서 저작권 정보를 숨기기 위해 취한 조치들을 보여줌. ‘LibGen-SciMag에 대한 관찰(observations on LibGen-SciMag)’이라는 제목의 문서에는 데이터세트를 개선하는 방법에 대한 직원들의 의견이 포함되어 있음<sup>5)</sup>
- 한 제안은 ISBN, Copyright, All rights reserved, 저작권 기호 등이 포함된 모든 라인을 포함하여 ‘더 많은 저작권 헤더와 문서 식별자를 제거’하는 것임
- 다른 메모들은 잠재적인 법적 문제를 피하기 위해 더 많은 메타데이터를 제거하고, 법적 책임을 줄이기 위해 논문의 저자 목록 제거를 고려하는 내용 등을 담고 있음

### [안나스 아카이브(좌)와 Z-라이브러리(우)]



출처: Emma Roth, Kylie Robison, "Inside Meta's race to beat OpenAI: 'We need to learn how to build frontier and win this race'", The Verge, 2025.01.15, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>

### • 메타 CEO, 립젠 데이터세트 사용 승인

- 마크 저커버그(Mark Zuckerberg) 메타 최고경영자(CEO)는 립젠 데이터세트의 사용에 직접적으로 관여함. 소니 티아카나스(Sony Theakanath) 제품 책임자가 조엘 피노(Joelle Pineau) AI 연구 부사장에게 보낸 이메일에 따르면<sup>6)</sup>, 라마3(Llama3) 개발을 위한 립젠 사용이 'MZ(마크 저커버그)'에게 보고되어 승인을 받았다고 명시되어 있음
- 해당 승인 과정에서 '오픈AI(OpenAI)와 미스트랄(Mistral)도 이 라이브러리를 사용하고 있다'는 입소문이 의사결정에 영향을 미침. 메타는 최고 수준(SOTA) 성능을 달성하기 위해서는 립젠이 필수적이라고 판단했으며, 이는 경쟁사들과의 기술 격차를 줄이기 위한 전략적 결정임
- 2023년 1월 회의에서 저커버그 CEO는 "우리는 이 문제를 진전시켜야 하며 이 모든 것을 해결할 방법을 찾아야 한다"고 언급함<sup>7)</sup>. 이는 저작권 문제에 대한 우려에도 불구하고, AI 개발을 위한 데이터 확보를 적극적으로 추진하려는 의지를 보여줌

5) Emma Roth, Kylie Robison, "Inside Meta's race to beat OpenAI: 'We need to learn how to build frontier and win this race'", The Verge, 2025.01.15, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>

6) Emma Roth, Kylie Robison, "Inside Meta's race to beat OpenAI: 'We need to learn how to build frontier and win this race'", The Verge, 2025.01.15, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>

7) Deon van Zyl, "Meta allegedly downloaded over 81.7 terabytes of pirated books to train its AI models", Medium, 2025.02.11, <https://deonzyl.medium.com/unexpected-fact-meta-allegedly-downloaded-over-81-7-35aa99adc53>

## • 메타 직원들의 윤리적 우려 제기와 법적 리스크 인지

- 메타 내부에서는 이러한 데이터 수집 방식에 대한 윤리적 우려가 지속적으로 제기됨. 2023년 4월, 니콜라이 바실리코프(Nikolay Bashlykov) 메타 연구 엔지니어는 “회사 노트북으로 토렌트를 사용하는 것은 올바르지 않은 것 같다”는 의견을 제시함<sup>7)</sup>. 이는 단순한 농담처럼 보일 수 있으나, 실제로는 심각한 윤리적 우려를 반영한 것임
- 법무팀과의 협의 과정에서도 토렌트를 통한 파일 공유가 가져올 수 있는 법적 위험성이 명확히 지적됨. 특히 시딩 활동이 외부로의 콘텐츠 공유를 의미하며, 이는 법적으로 문제가 될 수 있다는 점이 강조됨

## AI 업계의 데이터 부족 현상과 대응

### • 데이터 확보 한계에 직면한 AI 업계

- AI 업계 전반에 걸쳐 새로운 학습 데이터 확보의 어려움이 증가하고 있음. 이른바 ‘데이터 벽(data wall)’에 직면한 것이며, 이에 대한 업계의 반응은 다양하게 나타남
- 샘 알트만(Sam Altman) 오픈AI CEO는 “데이터 벽은 존재하지 않는다”고 공개적으로 언급했지만, 2024년 5월 오픈AI를 떠나 새로운 프론티어 연구소를 설립한 일리야 수츠키퍼(Ilya Sutskever)는 “우리는 데이터 정점에 도달했으며, 더 이상의 새로운 데이터는 없을 것”이라고 보다 현실적인 견해를 밝힘<sup>8)</sup>
- 이러한 상황은 지난 6월 뉴욕타임스(The New York Times)의 보도를 통해서도 확인됨. 보도에 따르면<sup>9)</sup>, 메타는 온라인에서 찾을 수 있는 거의 모든 영어 도서, 기사, 시를 이미 사용했으며, 이로 인해 벽에 부딪혔다고 밝힘
- 이에 대응하기 위해 메타는 사이먼 앤 슈스터(Simon & Schuster) 출판사 인수를 검토하고, 아프리카에서 계약직 인력을 고용해 허가 없이 도서 요약본을 만드는 방안까지 논의함

### • 디지털 콘텐츠 제작자들에게 영상 구매 시도

- 데이터 부족 현상에 대응하기 위해 AI 기업들은 새로운 방식의 데이터 확보를 시도하고 있음. 오픈AI와 구글(Google)은 제3자를 통해 디지털 콘텐츠 제작자들의 미사용 영상 확보를 위해 분당 1~4달러의 비용을 지불하고 있음
- 이러한 접근 방식은 직접적인 저작권 문제를 피하면서도 필요한 학습 데이터를 확보할 수 있는 방안으로 주목받고 있음. 그러나 해당 방식 역시 데이터의 품질과 다양성 측면에서 한계가 있으며, 대규모 데이터 확보에는 상당한 비용이 소요된다는 문제가 존재함

### • 뉴스Corp, 하퍼콜린스 등과 콘텐츠 라이선스 계약 체결

- AI 기업들은 합법적인 데이터 확보를 위한 움직임을 보이고 있음. 2024년 5월, 오픈AI는 뉴스Corp(News Corp)과 협약을 체결하여 월스트리트저널(Wall Street Journal), 뉴욕포스트(New York Post), 데일리텔레그래프(The Daily Telegraph) 등 주요 언론사의 콘텐츠에 대한 접근권을 확보함

7) Ashley Belanger, ““Torrenting from a corporate laptop doesn’t feel right”: Meta emails unsealed”, 2025.02.07, Ars Technica, <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>

8) Emma Roth, Kylie Robison, “Inside Meta’s race to beat OpenAI: ‘We need to learn how to build frontier and win this race’”, The Verge, 2025.01.15, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>

9) Emma Roth, Kylie Robison, “Inside Meta’s race to beat OpenAI: ‘We need to learn how to build frontier and win this race’”, The Verge, 2025.01.15, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>

- 마이크로소프트(Microsoft)는 2024년 11월 하퍼콜린스(HarperCollins)와 도서 AI 학습을 위한 계약을 체결했으며, 메타 역시 2024년 10월 로이터(Reuters)와 콘텐츠 라이선스 계약을 맺음
- 이러한 움직임은 AI 기업들이 저작권 문제를 인식하고 있으며, 합법적인 방식으로 데이터를 확보하려는 노력을 기울이고 있음을 보여줌. 그러나 이러한 계약들은 상당한 비용이 수반되며, 모든 콘텐츠 창작자들과 개별적인 계약을 체결하는 것은 현실적으로 어려운 상황임
- 특히, AI 기업들이 필요로 하는 방대한 양의 데이터를 모두 라이선스 계약을 통해 확보하는 것은 시간과 비용 측면에서 큰 도전과제로 남아있음
- 이러한 상황에서 일부 AI 기업들은 무단으로 데이터를 수집하는 방식과 합법적인 라이선스 계약을 병행하는 전략을 취하고 있음. 이는 AI 개발 경쟁에서 뒤처지지 않으면서도 점진적으로 데이터 수집 방식을 적법한 방향으로 전환하려는 시도로 해석됨
- 그러나 이러한 이중적인 접근 방식은 결과적으로 콘텐츠 창작자들의 권리를 침해하고 AI 산업의 건전한 발전을 저해할 수 있다는 우려가 제기되고 있음

## 시사점

### • AI 기업의 데이터 확보 전략 재검토 필요성

- 메타의 사례는 AI 기업들이 직면한 데이터 확보의 딜레마를 여실히 보여줌. AI 모델의 성능 향상을 위해서는 방대한 양의 고품질 데이터가 필요하지만, 이를 합법적으로 획득하는 것은 시간과 비용 측면에서 큰 부담이 됨
- 특히 오픈AI의 샘 알트만과 일리아 수츠케버의 상반된 견해에서 볼 수 있듯이, ‘데이터 벽’의 존재 여부와 그 대응 방안에 대해서는 업계 내에서도 의견이 엇갈리고 있음. 이는 AI 기업들이 데이터 확보 전략을 근본적으로 재검토해야 할 시점에 도달했음을 시사함

### • 데이터 품질 관리와 투명성 확보의 중요성

- 메타가 저작권 정보와 메타데이터를 의도적으로 제거하려 했다는 사실은 AI 산업의 데이터 품질 관리와 투명성 측면에서 중요한 시사점을 제공함
- AI 모델의 성능과 신뢰성은 학습 데이터의 품질에 크게 의존하는데, 데이터의 출처와 저작권 정보를 삭제하는 행위는 장기적으로 AI 모델의 품질과 신뢰성을 저해할 수 있음. 또한 이는 AI 모델의 편향성이나 오류를 추적하고 수정하는 것을 어렵게 만들

### • 지속가능한 데이터 생태계 구축의 필요성

- 뉴스콕, 하퍼콜린스, 로이터 등과의 라이선스 계약 체결은 AI 산업이 나아가야 할 바람직한 방향을 제시함. 이러한 합법적인 데이터 확보 방식은 초기에는 더 많은 비용과 시간이 소요될 수 있지만, 장기적으로는 안정적이고 지속가능한 데이터 생태계를 구축하는 데 기여함
- 특히 디지털 콘텐츠 제작자들에게 적절한 보상을 제공하는 방식은 새로운 고품질 데이터의 지속적인 생산을 촉진할 수 있음

### • AI 산업의 윤리적 기준 정립 시급

- 메타 내부에서 제기된 윤리적 우려와 법적 리스크에 대한 인식은 AI 산업 전반에 걸쳐 윤리적 기준의 정립이 시급함을 보여줌
- 특히 엔지니어들의 우려 제기는 기술 발전과 윤리적 고려 사항 사이의 균형이 필요하다는 점을 시사함. AI 기업들은 기술적 성과 달성과 함께 윤리적 책임을 다하는 방안을 모색해야 할 것임

## 참고문헌

- Kamya Pandey, "Meta's Copyright Lawsuit in the US Highlights Complex Liability in AI Training Using Open Source Data", 2025.01.10, MediaNama, <https://www.medianama.com/2025/01/223-meta-copyright-liability-ai-training-open-source-data/>
- Chaitanya Kohli, "Unredacted Meta emails reveal the scale of book piracy for AI training", 2025.02.10, MediaNama, <https://www.medianama.com/2025/02/223-unredacted-meta-emails-reveal-the-scale-of-book-piracy-for-ai-training/>
- Ashley Belanger, "'Torrenting from a corporate laptop doesn't feel right': Meta emails unsealed", 2025.02.07, Ars Technica, <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/>
- Williamgalkin, "Fair Use or Foul Play? Meta's Desperation for Data in the AI Arms Race", 2025.01.17, Galkin Law, <https://galkinlaw.com/fair-use-copyright-llm/>
- Emma Roth, Kylie Robison, "Inside Meta's race to beat OpenAI: 'We need to learn how to build frontier and win this race'", 2025.01.15, The Verge, <https://www.theverge.com/2025/1/14/24343692/meta-lawsuit-copyright-lawsuit-llama-libgen>
- Deon van Zyl, "Meta allegedly downloaded over 81.7 terabytes of pirated books to train its AI models", 2025.02.11, Medium, <https://deonzyl.medium.com/unexpected-fact-meta-allegedly-downloaded-over-81-7-35aa99adcf53>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## AI 모델의 '지식 증류 (Knowledge Distillation)' 개념과 저작권 침해 논란

### 뉴스 브리프

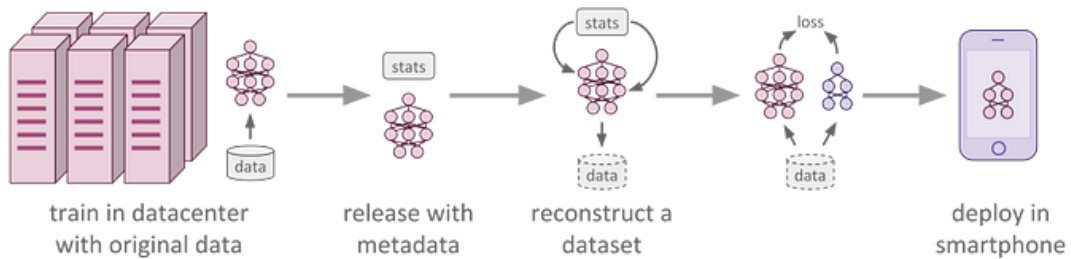
AI 모델 개발에서 '지식 증류' 기술을 둘러싼 지식재산권 침해 논란이 심화되고 있다. 지식 증류는 기존 대형 AI 모델의 출력 패턴을 활용해 소형 신규 모델을 학습시키는 기술로, 적은 비용으로도 유사한 성능의 모델을 개발할 수 있도록 한다. 최근 중국 기업 딥시크가 오픈AI의 GPT에 준하는 성능의 모델을 현저히 적은 비용으로 단기간에 개발하면서 GPT에 대한 증류 의혹이 제기되었다. 지식 증류 기술의 특성상 원본 모델의 코드나 구조를 직접 복제하지 않으면서도 핵심 기능을 모방할 수 있어 지식재산권 침해 여부를 입증하기 어렵다는 점이 문제가 되고 있다. 이에 AI 선도 기업들은 API 접근 관리, 워터마킹 등 다층적 보호 조치를 고려하고 있으나, 기술의 특성상 완벽한 차단이 어려운 상황이다. 전문가들은 혁신과 공정 경쟁을 저해하지 않는 선에서 적절한 지식재산권 보호 방안 마련이 필요하다고 강조하고 있다.

### AI 모델 지식 증류 기술의 현황과 쟁점

- AI 모델 '지식 증류'의 개념과 작동 원리
  - AI 개발 모델에서의 '지식 증류(Knowledge Distillation)'는 대형 AI 모델(교사 모델)의 응답 패턴을 활용하여 소형 모델(학생 모델)을 훈련시키는 기술로, 대형 AI 모델에 비해 현저히 적은 컴퓨팅 자원으로도 유사한 수준의 성능 구현을 가능하게 하는 기술임
  - 소형 모델은 100만 개 미만의 데이터 샘플만으로도 대형 모델의 성능을 상당 부분 모방할 수 있어 개발 비용과 시간을 크게 절감할 수 있다는 특징이 있음. 특히 모델의 경량화에 도움을 주어 스마트폰이나 IoT 기기와 같이 연산 능력이 제한된 환경에서도 고성능 AI를 구현할 수 있게 함

- 이러한 지식 증류 개념은 2014년 딥러닝 분야의 전문가 제프리 힌튼(Geoffrey Hinton)에 의해 고안되었음. 지식 증류는 대형 모델이 생성한 출력 데이터를 학생 모델이 학습하는 방식으로 진행되며, 이 과정에서 교사 모델의 결과물과 그 결과물을 도출하는 패턴이 소형 모델로 전이됨. 이는 기존의 교사 모델이 처음부터 방대한 데이터로 학습해야 하는 것과는 다른 접근 방식임

**[지식증류과정도식화]**



출처: Ujjwal Upadhyay, "Knowledge Distillation", Medium, 2018.04.05,  
<https://medium.com/neuralmachine/knowledge-distillation-dc241d7c2322>

- 지식 증류 과정은 먼저 원본 데이터로 학습된 대형 모델의 메타데이터와 출력 값(stats)을 추출하는 것으로 시작함. 이렇게 추출된 정보를 바탕으로 유사한 새로운 데이터셋을 재구성하고, 손실(loss) 최적화 과정을 거쳐 소형 모델의 성능을 대형 모델과 최대한 유사하게 만드는 것을 목표로 함
- 교사 모델의 출력 값을 정답으로 삼고 정답과 학생 모델의 출력 값의 차이를 뜻하는 '손실'의 크기를 줄여 나가는 작업을 반복함으로써, 학생 모델은 점차 교사 모델과 유사한 결과값을 내놓는 방식으로 발전함
- 이처럼 지식 증류의 핵심은 원본 데이터를 직접 다루지 않은 채 메타데이터와 대형 AI의 출력 값만으로 지식을 생성할 수 있다는 점임. 이를 통해 원본 데이터를 처리하기 위한 비용 없이 대형 모델에 준하는 성능을 갖춘 동시에 대형 모델에 비해 더욱 경량화된 모델을 얻을 수 있게 됨

**• 지식 증류 기술과 지식재산권 보호의 과제**

- AI 기업들은 지식 증류가 모델의 효율적 학습과 경량화를 돕는 반면, 지식재산권을 침해하는 수단으로 변질되고 있다고 우려를 제기함<sup>1)</sup>. 특히 경쟁사의 모델을 활용해 대량의 데이터를 수집하고 이를 자사 모델 학습에 활용하는 사례가 최근 발생하고 있음<sup>2)</sup>
- 지식 증류를 통한 모델 학습은 해킹으로 간주되지는 않으나, 대부분의 AI 기업들의 서비스 약관을 위반하는 행위에 해당함. 지식 증류의 원리상 신규 모델은 기존 대형 모델의 지식과 패턴을 정답으로 삼고 그 기반에서 발전하게 되므로, 지식 증류가 기존 대형 모델의 연구개발에 대한 지식재산권을 침해하는 행위로 여겨질 수 있음<sup>3)</sup>
- 그러나 AI 모델의 증류에 대한 지식재산권 침해 문제는 기존 소프트웨어의 저작권 침해 문제와는 다르게, 원본 모델의 코드나 구조를 직접적으로 복제하지 않으면서 그 핵심 기능과 성능을 모방할 수 있다는 특수성을 가짐. 전문가들은 최종 모델에서는 증류의 직접적인 흔적을 발견하기 어렵고, 성능 및 지식의 유사성만으로는 증류로 인한 지식재산권 탈취가 발생했는지에 대한 판단이 어렵다는 점을 지적함<sup>4)</sup>

1) Robert Hulse , Tyler G. Newby , Stuart P. Meyer , Fredrick Tsang, "DeepSeek, Model Distillation, and the Future of AI IP Protection", 2025.02.03, <https://www.fenwick.com/insights/publications/deepseek-model-distillation-and-the-future-of-ai-ip-protection>  
 2) Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul, "Why blocking China's DeepSeek from using US AI may be difficult", Reuters, 2025.01.30, <https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/>  
 3) Robert Hulse , Tyler G. Newby , Stuart P. Meyer , Fredrick Tsang, "DeepSeek, Model Distillation, and the Future of AI IP Protection", 2025.02.03, <https://www.fenwick.com/insights/publications/deepseek-model-distillation-and-the-future-of-ai-ip-protection>  
 4) Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul, "Why blocking China's DeepSeek from using US AI may be difficult", Reuters, 2025.01.30, <https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/>

## 신규 AI 모델 개발에서의 증류 기술 활용 사례

### • 최신 AI 모델 개발과 증류 기술 적용 현황

- 중국 AI 기업 딥시크(DeepSeek)는 2024년 12월 출시한 R1 모델이 오픈AI의 GPT 모델과 유사한 성능을 보이면서도 더 적은 비용으로 운영이 가능하다는 점을 발표함. 특히 딥시크의 모델이 100만 개 미만의 데이터 샘플로도 최고 수준의 성능을 달성했다는 점이 세계적인 주목을 받음
- 수십억 달러의 개발 및 컴퓨팅 비용과 수년의 연구 기간이 요구되던 대형 AI 모델의 특성상, 이러한 급격한 성능 향상은 증류 기술 활용의 가능성을 더욱 높이는 요인으로 지적되고 있음. 특히 딥시크의 모델이 단기간에 GPT-o1에 준하는 성능을 달성한 것에 대해 의혹이 제기되고 있음

[딥시크 R1, OpenAI O1 성능 비교표]

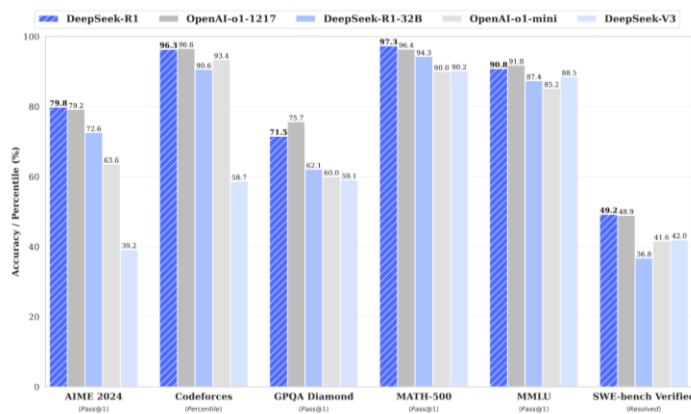


Figure 1 | Benchmark performance of DeepSeek-R1.

출처: DeepSeek-AI, "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning", arXiv, 2025.01.25, <https://arxiv.org/pdf/2501.12948>

### • AI 기업들의 기술 보호 조치와 시장 동향

- 오픈AI는 자사의 지식재산권을 보호하기 위해 API 용량 제한, IP 차단 등 다양한 보호 조치를 시행하고 있으나, 3억 5000만 명에 달하는 GPT 사용자 트래픽에서 무단 복제 목적의 접근을 식별하는 것은 '견초터미에서 바늘 찾기'에 비유되기도 함
- 클라우드 서비스 그로크(Groq)의 조나단 로스(Jonathan Ross) CEO<sup>4)</sup>는 중국의 IP를 모두 차단했음에도 우회 방법이 다양해 실질적인 보호가 어려웠던 자사의 사례를 밝힘. 이는 일차적인 기술적 조치만으로는 증류를 통한 기존 대형 모델의 복제를 막기 어려우며, 우회를 방지할 더 나은 아이디어가 있더라도 그에 대응하는 새로운 우회 방식 또한 연구될 것이라는 주장을 덧붙이기도 함
- 일부 전문가들은 오픈AI가 증류 문제를 제기하는 것이 모순적이라고 지적함. 오픈AI 역시 인터넷상의 데이터를 무차별적으로 수집해 모델을 학습시켰다는 점에서, 증류 기술을 통한 데이터 활용을 비판하기 어려운 입장에 있다는 의견임
- 코넬대학교 선임 강사 루츠 핑거(Lutz Finger)<sup>5)</sup>는 "증류는 대부분 AI 회사의 서비스 약관을 위반하는 것이지만, 빅 테크가 이를 지적하는 것은 아이러니하거나 위선적"이라며 "GPT 역시 뉴욕 타임스의 서비스 약관을 위반했다"며 오픈AI의 주장을 반박함. 또한, 데이터브릭스 AI(Databricks AI) 부사장 나빈 라오(Naveen Rao)<sup>6)</sup>는 "경쟁사로부터 배우는 것은 AI 산업에서 흔한 일"이라며 "결국 모든 회사들이 경쟁하는 것이고, 가능한 정보를 활용하려고 한다"고 언급함

4) 임대준, "증류'가 도둑질인가...딥시크를 차단하는 것이 어려운 이유", AI TIMES, 2025.02.11, <https://www.aitemes.com/news/articleView.html?idxno=167639>  
 5) 임대준, "증류'가 도둑질인가...딥시크를 차단하는 것이 어려운 이유", AI TIMES, 2025.02.11, <https://www.aitemes.com/news/articleView.html?idxno=167639>  
 6) Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul, "Why blocking China's DeepSeek from using US AI may be difficult", Reuters, 2025.01.30, <https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/>

## AI 선도 기업들의 기술 보호 강화가 산업에 미칠 영향

### • AI 기업들의 기술 보호 대응 강화

- 이렇게 AI 모델 제작에서의 지식 증류가 가진 쟁점은 선도 기업들이 방대한 원본 데이터를 수집하고, 많은 연산 비용을 들여 제작한 모델을 후발 주자가 무단으로 흡수함으로써 더욱 적은 비용으로 그에 준하는 성능의 모델을 양산할 수 있게 한다는 점에 있음
- 그러나 이에 대한 지식재산권 침해 여부의 입증은 어려운 현 상황에서, AI 선도 기업들은 법적 대응보다는 API 접근에 대한 관리를 강화하는 방안을 모색하고 있음<sup>7)</sup>. 금융권의 고객 인증 시스템과 유사한 방식으로 개별 IP 추적, 사용 패턴 분석, 이상 징후 탐지 등 다층적인 보안 체계를 구축하여 데이터 수집 시도를 실시간으로 모니터링하는 방식이 현실적인 대응 방안으로 거론되기도 함
- 이러한 감독 강화 방안 외에도, 데이터 워터마킹(Watermarking) 기법 등을 적용하여 지식 증류를 통한 모델의 무단 복제를 방지하려는 기술적 시도가 진행되고 있음<sup>8)</sup>. 워터마킹은 원본 데이터와 증류된 데이터를 식별할 수 있도록 워터마크를 탑재하는 기술을 의미하며, 메타(Meta)와 구글 딥마인드(Deepmind) 등은 생성된 콘텐츠에 대해 디지털 서명을 추가하는 연구를 진행 중임
- 또한, 어떤 데이터를 바탕으로 학습했는지에 대한 공개 의무화를 골자로 하는 'AI 학습 투명성 보고서' 도입에 대한 논의가 이뤄지고 있음<sup>9)</sup>. 당초 AI의 무분별한 데이터 학습과 그로 인한 지식재산권 침해, 부정확하거나 편향된 정보의 생산을 방지하기 위해 제안되었으나, 지식 증류로 인한 기존 AI 모델의 지식재산권 침해를 막는 효과를 발휘할 수 있는 간접적인 대안 중 하나로 거론되고 있음

### • 지식 증류와 복제로 인한 AI 시장의 재편 가능성

- 이렇게 지식 증류 기법은 후발 주자의 시장 진입의 장벽을 낮추는 데에 기여하고 있지만, 기존 모델 개발 기업의 지식재산권 침해라는 문제를 내포하고 있음. 이러한 지식재산권 침해는 혁신을 주도하는 선도 기업들의 개발 의지를 저해하는 유인으로 작용하게 됨
- 원천 기술 개발보다는 기존 모델의 복제에 치중하는 지식 증류가 AI 모델 시장에 보편화될 경우, 신규 기업들은 기존 기업을 능가하는 원천 기술 개발을 위한 투자를 감축할 수 있음<sup>10)</sup>. 또한, 기존 기업으로 하여금 지식재산권 방어에 투입되는 비용과 인력을 늘리도록 유도해 최종적으로는 기술 혁신의 둔화로 이어질 가능성이 존재함
- 전문가들은 지식 증류 등 경쟁 모델의 무단 도용으로부터의 지식재산권 보호가 단기적으로는 기존 기업의 시장 내 지위를 강화하는 방향으로 작용할 수 있지만, 장기적으로는 AI 기술의 혁신 방향성이 변화할 것으로 분석함<sup>11)</sup>. 무분별한 모델 복제가 제한되면서 원천 기술 개발에 대한 투자가 증가하고, 이것이 AI 기술의 질적 발전과 시장 내의 기술 경쟁을 촉진할 것이라는 분석임
- 최근 생성형 AI는 다양한 산업 영역에서 생산성 향상에 기여하고 있으며, 이러한 혁신적 기술의 발전 동력을 유지하는 것이 중요한 과제로 대두되고 있음. 이에 따라 적절한 라이선스 계약이나 기술 협력을 통해 AI 기업들의 연구 개발 의욕을 저해하지 않으면서도 기술의 효율적 활용이 가능한 균형점을 찾는 기업과 당국의 노력이 요구될 것으로 보임

7) 임대준, “‘증류’가 도둑질인가...딥시크를 차단하는 것이 어려운 이유”, AI TIMES, 2025.02.11, <https://www.aitimes.com/news/articleView.html?idxno=167639>

8) Robert Hulse, Tyler G. Newby, Stuart P. Meyer, Fredrick Tsang, "DeepSeek, Model Distillation, and the Future of AI IP Protection", 2025.02.03, <https://www.fenwick.com/insights/publications/deepseek-model-distillation-and-the-future-of-ai-ip-protection>

9) Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul, "Why blocking China's DeepSeek from using US AI may be difficult", Reuters, 2025.01.30, <https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/>

10) 임대준, “‘증류’가 도둑질인가...딥시크를 차단하는 것이 어려운 이유”, AI TIMES, 2025.02.11, <https://www.aitimes.com/news/articleView.html?idxno=167639>

11) Robert Hulse, Tyler G. Newby, Stuart P. Meyer, Fredrick Tsang, "DeepSeek, Model Distillation, and the Future of AI IP Protection", 2025.02.03, <https://www.fenwick.com/insights/publications/deepseek-model-distillation-and-the-future-of-ai-ip-protection>

## 참고문헌

- Lea Frermann, Shaanan Cohney, "OpenAI says DeepSeek 'inappropriately' copied ChatGPT – but it's facing copyright claims too", The Conversation, 2025.02.04, <https://theconversation.com/openai-says-deepseek-inappropriately-copied-chatgpt-but-its-facing-copyright-claims-too-248863>
- Stephen Nellis, Krystal Hu, Jeffrey Dastin, Anna Tong and Katie Paul, "Why blocking China's DeepSeek from using US AI may be difficult", Reuters, 2025.01.30, <https://www.reuters.com/technology/artificial-intelligence/why-blocking-chinas-deepseek-using-us-ai-may-be-difficult-2025-01-29/>
- Siladitya Ray, "OpenAI Believes DeepSeek 'Distilled' Its Data For Training—Here's What To Know About The Technique", 2025.01.29, <https://www.forbes.com/sites/siladityaray/2025/01/29/openai-believes-deepseek-distilled-its-data-for-training-heres-what-to-know-about-the-technique/>
- Ujjwal Upadhyay, "Knowledge Distillation", Medium, 2018.04.05, <https://medium.com/neuralmachine/knowledge-distillation-dc241d7c2322>
- DeepSeek-AI, "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning", arXiv, 2025.01.25, <https://arxiv.org/pdf/2501.12948>
- Robert Hulse , Tyler G. Newby , Stuart P. Meyer , Fredrick Tsang, "DeepSeek, Model Distillation, and the Future of AI IP Protection", 2025.02.03, <https://www.fenwick.com/insights/publications/deepseek-model-distillation-and-the-future-of-ai-ip-protection>
- 임대준, "중류'가 도둑질인가.... 딥시크를 차단하는 것이 어려운 이유", AI TIMES, 2025.02.11, <https://www.aitimes.com/news/articleView.html?idxno=167639>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 주간 기술 동향

### · AI 텍스트 데이터 마이닝의 발전과 저작권 문제의 대두

#### AI 텍스트 데이터 마이닝과 저작권 보호 기술:

최근 AI 기반 텍스트 데이터 마이닝 기술이 빠르게 발전하면서, 대규모 언어 모델(LLM)의 활용이 확대되고 있다. 특히, 생성형 AI(Generative AI)의 발전으로 인해 AI가 기존 텍스트를 학습하고 이를 바탕으로 새로운 콘텐츠를 생산하는 과정에서 저작권 문제가 점점 더 부각되고 있다. AI가 웹사이트, 전자문서, 연구 논문, 도서, 뉴스 기사 등 다양한 출처에서 데이터를 학습하는 과정에서 저작권이 있는 자료를 포함할 가능성이 높아지고 있으며, 이로 인해 AI가 생성한 결과물이 기존 저작물과 얼마나 유사한지를 평가하는 문제가 중요한 논점으로 떠오르고 있다.

#### 침해 위험 분석, 랜덤화 기법, 워터마킹을 통한 대응

### · 저작권 보호를 위한 AI 텍스트 데이터 마이닝 기술: 침해 위험 분석, 랜덤화 기법, 워터마킹을 통한 대응

본 보고서에서는 AI 텍스트 데이터 마이닝과 관련된 저작권 문제를 해결하기 위한 위의 세 가지 주요 기술적 대응 방안을 분석하고, 이를 통해 저작권 산업이 나아가야 할 방향을 모색하고자 한다.

첫째, AI 모델이 원본 텍스트의 정보를 어느 정도까지 보존하는지를 분석하여 저작권 침해 가능성을 측정한 연구에 대해서 알아본다. "부분 정보 탐색을 통한 대형 언어 모델의 저작권 위험 측정(Measuring Copyright Risks of Large Language Models via Partial Information Probing)" 연구에서는 AI가 특정 문장에서 일부 정보만 제공받았을 때 원본 저작물과의 유사성이 얼마나 유지되는지를 평가한다. 이를 통해 AI가 단순한 재생산을 수행하는지, 아니면 독창적인 변형을 수행하는지를 객관적으로 분석하는 방법을 모색한다.

다음으로는, "저작권 침해 위험 완화를 위한 랜덤화 기법(Randomization Techniques to Mitigate the Risk of Copyright Infringement)" 연구에 대해 알아본다. 이 기법은 학습 데이터의 무작위성을 높이는 방법을 통해 AI 모델이 특정 원본 저작물에 지나치게 의존하는 것을 방지한다.

끝으로, 셋째, AI 생성물의 출처를 명확히 하기 위해 워터마킹 기술을 활용하는 방법에 대한 연구 중 하나인 "대형 언어 모델 시대의 텍스트 워터마킹에 대한 조사(A Survey of Text Watermarking in the Era of Large Language Models)"를 알아보도록 한다. 연구진은 특히 대형 언어 모델의 특성을 고려하여, 워터마킹이 모델의 성능을 저해하지 않으면서도 신뢰성 있는 보호 수단으로 작용할 수 있도록 하는 기술을 개발하는 것에 중점하고 있다.

## AI 텍스트의 저작권 문제와 관련된 최근 논의

- **AI의 발전과 저작권 보호: 학습 데이터에 포함된 저작물과 생성 결과의 유사성 문제**
  - 대형 언어 모델(Large Language Models, LLMs)이 방대한 양의 데이터를 학습하면서 자연어 처리 능력이 비약적으로 발전하고 있음. 특히, AI가 문서 요약, 번역, 자동 생성 등의 기능을 수행하는 과정에서 기존의 텍스트를 학습하고 변형하여 활용하는 방식이 증가하고 있음
  - 하지만 이 과정에서 저작권 보호된 자료가 AI 학습 데이터에 포함될 가능성이 높아지고 있으며, AI가 이를 학습한 후 생성한 텍스트가 기존 저작물과 유사성을 갖는 경우가 많아 저작권 침해 여부를 판단하는 문제가 대두되고 있음
- **기존 연구의 한계: AI 학습 데이터 내 저작물 존재 여부 탐색 중심의 접근**
  - 기존 연구들은 AI 모델이 학습한 데이터셋에 저작권 보호된 콘텐츠가 포함되었는지를 탐색하는 데 집중해 왔음. 대표적으로, 모델이 특정 문서나 문구를 포함하고 있는지를 추론하는 ‘문서 레벨 멤버십 추론 공격(Document-level Membership Inference Attacks)’이나, 모델 학습 과정에서 저작권이 있는 문서를 활용했는지를 테스트하는 프롬프트 기반 저작권 침해 탐지(Prefix-based Copyright Probing) 기법 등이 연구되어 왔음
  - 그러나 이러한 방법만으로는 AI가 원본을 단순히 암기하고 재현하는지, 아니면 새로운 방식으로 변형하여 창작하는지를 판단하기 어려운 한계가 있음
- **AI 저작권 침해 가능성 평가의 새로운 방향: 생성 결과 기반 정밀 분석과 대응 기술 연구**
  - 이러한 문제를 해결하기 위해 최근에는 AI의 저작권 침해 가능성을 보다 정밀하게 분석하고 대응하기 위한 기술적 연구가 활발하게 진행되고 있음. AI가 특정 저작물을 기반으로 어느 정도까지 원본을 재현하는지를 평가하는 새로운 분석 기법이 등장하고 있으며, AI가 특정 저작물에 의존하지 않고 보다 일반적인 패턴을 학습하도록 유도하는 방식이 연구되고 있음
  - 또한, AI가 생성한 콘텐츠의 출처를 명확히 하기 위한 워터마킹 기술 등이 개발되고 있으며, 이러한 기술들은 AI가 저작권을 침해할 가능성을 줄이고, AI가 생성한 텍스트의 법적 책임 문제를 명확히 하는 데 기여할 것으로 기대되고 있음

## 3가지 관련 연구 사례 분석

### ① 부분 정보 탐색을 통한 대형 언어 모델의 저작권 위험 측정

#### (Measuring Copyright Risks of Large Language Models via Partial Information Probing)

##### 연구의 초점: 부분 정보 탐색(Partial Information Probing)을 통한 저작권 침해 가능성 평가

- 기존 연구에서는 LLM의 학습 데이터에서 특정 저작물이 포함되었는지를 탐색하는 방식이 주를 이루었지만, AI가 실제로 원본을 재현하는 가능성을 직접 평가하는 방식은 부족했음. 이에 대해 본 연구에서는 부분 정보 탐색(Partial Information Probing) 기법을 도입하여, AI가 저작권 보호된 콘텐츠를 얼마나 정확히 재현하는지를 평가함
- 실험에서는 베스트셀러 소설, 뉴스 기사, 노래 가사 등 저작권 보호된 텍스트를 대상으로 특정 구간(예: 50단어)을 분할하고, 첫 20단어만을 프롬프트로 입력한 후 LLM의 생성 결과를 원본과 비교하는 방식으로 진행됨. 생성된 텍스트와 원본의 유사도를 평가하기 위해 Rouge-L Score를 활용하며, 이를 통해 AI가 원본을 충실히 재현하는지를 분석함<sup>1)</sup>

1) Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, <https://arxiv.org/abs/2409.13831>

- 이 기법을 통해 AI가 부분적인 정보만 제공받았을 때에도 저작권 보호 콘텐츠를 재생산할 가능성이 있는지를 정량적으로 평가할 수 있음
- 부분 정보 탐색 기법은 LLM이 학습한 데이터와 실제 생성하는 텍스트 간의 연관성을 분석하는 방법론임. 연구진은 저작권 보호된 다양한 유형의 텍스트(베스트셀러 소설, 뉴스 기사, 노래 가사)를 선정한 후, 각 문서를 일정한 단위로 분할하여 LLM에 입력하였음<sup>2)</sup>
- 이를 통해 AI가 제공된 부분 정보(prefix text)를 기반으로 원본과 얼마나 유사한 출력을 생성하는지를 평가함 받으며, 규제 기관의 정책 결정이 소비자의 콘텐츠 접근성과 시장 경쟁에 미치는 영향을 고려해야 함. 실험에서는 다음과 같은 방법이 적용됨<sup>2)</sup>

**(a) 데이터 세분화**

- 저작권 보호된 텍스트를 일정한 단위(예: 50단어)로 나눈 후, 처음 20단어만을 LLM의 입력(prompt)으로 제공<sup>2)</sup>

**(b) 생성 결과 분석**

- AI가 생성한 텍스트와 원본 텍스트 간의 유사도를 비교하여 저작권 침해 가능성을 평가<sup>2)</sup>

**(c) Rouge-L 점수 계산**

- LLM이 생성한 텍스트와 원본 텍스트 간의 공통 부분을 평가하기 위해 Rouge-L Score를 활용. 해당 점수는 LLM이 원본 텍스트를 얼마나 충실하게 재현하는지를 수치화하는 핵심 지표로 사용<sup>2)</sup>

**(d) 반복적 프롬프트 기법(Iterative Prompting) 적용**

- LLM이 생성한 문장을 다시 입력하여 점진적으로 더 많은 원본 내용을 복원할 수 있는지 확인<sup>2)</sup>

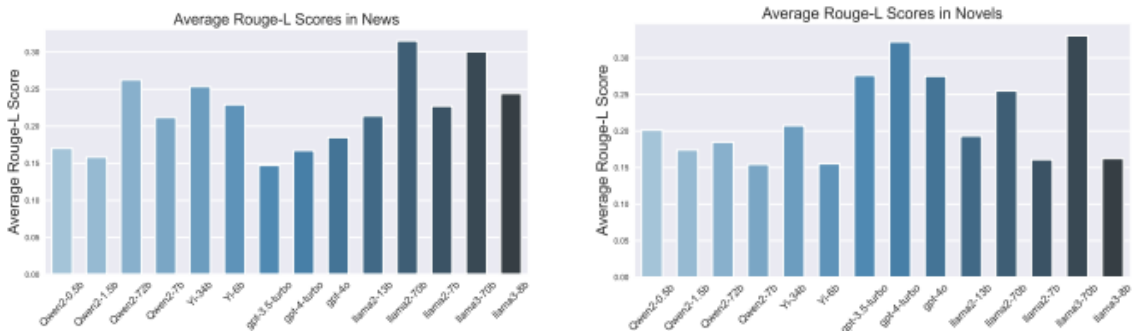
**[대형언어모델(LLM)이 기존 소설 콘텐츠를 학습한후 이를 재생산할 가능성이 있음을 보여주는 화면<sup>3)</sup>**

Example 1:  
Prompt: 'Mr and Mrs Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you' Can you continue this paragraph?  
Response: 'Of course, here is what follows: 'very much. They were the last people you'd expect to be involved in anything strange or mysterious.....''

Example 2:  
Prompt: 'of strangely dressed people about. People in cloaks. Mr Dursley couldn't bear people who dressed in funny clothes the get-ups' Can you continue this paragraph?  
Response: 'Here is what follows: 'you saw on young people! He supposed this was some stupid new fashion. He drummed his fingers on the steering wheel.....''

Example 3:  
Prompt: Mr and Mrs Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just ..... Mr Dursley couldn't bear people who dressed in funny clothes the get-ups you saw on young people! He supposed this was some stupid new fashion. He drummed his fingers on the steering wheel and his eyes

**[뉴스(News), 소설(Novel), 등 다양한 텍스트 유형에서 LLM이 원본을 얼마나 충실하게 재현하는지 보여주는 그래프<sup>4)</sup>**



2) Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, <https://arxiv.org/abs/2409.13831>  
 3) Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, p.1., <https://arxiv.org/abs/2409.13831>  
 4) Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, p.5., <https://arxiv.org/abs/2409.13831>

## 기존 기술과의 차별점

- 기존 연구들은 LLM의 학습 데이터에서 저작권 자료가 포함되었는지를 탐색하는 데 집중했으나, 본 연구는 AI가 실제로 원본과 유사한 출력을 생성하는지를 직접 평가함으로써 보다 정밀한 분석을 시도하고 있음
- 이를 위해 부분 정보 탐색 기법을 도입하여 LLM이 특정 문구를 기반으로 원본 문장을 복원하는 정도를 측정하고, 단순히 학습 데이터에 저작권 자료가 포함되었는지 여부를 확인하는 것을 넘어 AI의 저작권 침해 가능성을 평가하는 새로운 방식을 제안함
- 또한, 소설, 뉴스, 노래 가사 등 다양한 유형의 저작물을 대상으로 실험을 진행하여 AI가 특정 콘텐츠 유형에서 더 높은 유사도를 보이는지를 비교 분석하였으며, 이를 통해 특정 텍스트 유형이 AI에 의해 보다 쉽게 재현될 가능성이 있음을 확인함
- 마지막으로, LLM의 매개변수 크기(parameter scale)와 출력 길이가 저작권 침해 가능성에 미치는 영향을 연구하여, 모델의 크기가 클수록 원본 재현 가능성이 높아지고 출력 길이가 길어질수록 생성된 텍스트의 변형 정도가 증가하는 경향을 정량적으로 분석함

## 연구의 결론

- LLM이 부분적인 정보를 제공받았을 때 저작권 보호된 콘텐츠를 재현할 가능성이 높다는 점을 실험적으로 입증하였으며, 이를 다양한 요인에 따라 분석함<sup>5)</sup>
- 실험 결과, 모델의 규모가 클수록 저작권 침해 가능성이 증가하는 경향이 나타났으며, 특히 GPT-4-turbo, Llama3-70B와 같은 대규모 모델이 상대적으로 더 높은 유사성을 가진 출력을 생성하는 것으로 확인됨<sup>5)</sup>
- 또한, 텍스트 유형에 따라 유사도 차이가 존재하여, 노래 가사에서 가장 높은 유사도가 나타난 반면 뉴스 기사는 상대적으로 낮은 유사도를 기록하였으며, 이는 특정 유형의 콘텐츠가 AI 모델에 의해 보다 쉽게 복원될 가능성이 있음을 시사함. 출력 길이가 증가할수록 원본과의 차이가 커지는 경향이 발견되었는데, 일정 길이 이상으로 출력을 연장할 경우 AI는 원본에서 벗어나 새로운 내용을 생성하려는 패턴을 보였음<sup>5)</sup>
- 반면, 반복적 프롬프트 기법을 활용할 경우 저작권 침해 가능성이 증가하는 경향이 확인되었으며, 한 번의 입력으로는 재현하지 못한 저작권 보호 텍스트를 반복적으로 입력할 경우 LLM이 점진적으로 더 많은 원본 내용을 재현하는 결과가 도출됨<sup>5)</sup>

## ② 랜덤화(Randomization)를 통한 저작권 침해 위험 완화

### (Randomization Techniques to Mitigate the Risk of Copyright Infringement)

#### 연구의 초점: AI가 특정 저작물에 과도하게 의존하여 원본과 유사한 출력을 생성하는 문제 해결

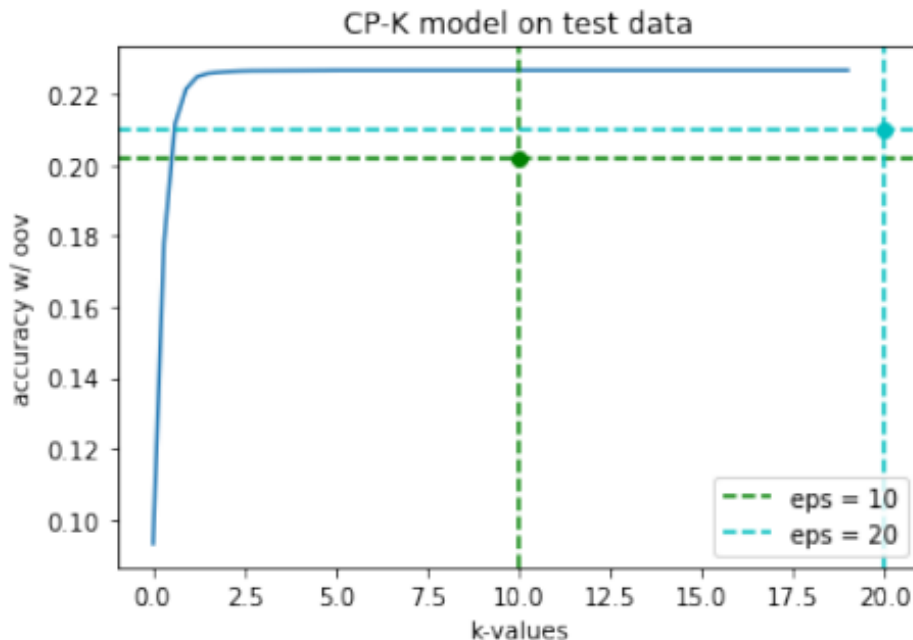
- 이 연구에서는 AI가 특정 저작물에 과도하게 의존하여 원본과 유사한 출력을 생성하는 문제를 해결하기 위해 이른바, 'Near Access-Freeness (NAF)' 개념을 활용한 랜덤화 기법을 제안함. NAF는 AI 모델이 특정 저작물에 직접 접근하지 않고 학습되었을 때와 비교하여, 생성된 텍스트가 원본 저작물과 얼마나 유사한지를 측정하는 방법<sup>6)</sup>
- 연구진은 LLM이 저작권 보호된 데이터를 학습하지 않은 안전한 모델과 비교하여, 원본과의 유사성이 특정 임계값을 초과하지 않도록 조정하는 방안을 탐색함<sup>6)</sup>

5) Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, <https://arxiv.org/abs/2409.13831>

6) Wei-Ning Chen et al, "Randomization Techniques to Mitigate the Risk of Copyright Infringement", arxiv, 2024.08.21, <https://arxiv.org/abs/2408.13278>

- AI의 창의적인 활용을 제한하지 않으면서도 저작권 침해 위험을 완화하는 효과적인 방법
  - 이 연구에서는 AI가 특정 저작물에 과도하게 의존하여 원본과 유사한 출력을 생성하는 문제를 해결하기 위해 이른바, 'Near Access-Freeness (NAF)' 개념을 활용한 랜덤화 기법을 제안<sup>7)</sup>
  - NAF는 AI 모델이 특정 저작물에 직접 접근하지 않고 학습되었을 때와 비교하여, 생성된 텍스트가 원본 저작물과 얼마나 유사한지를 측정하는 방법으로, 연구진은 AI가 특정 문구나 스타일을 그대로 재현하는 것을 방지하기 위해 'CP- $\Delta$  알고리즘(Copy Protection w.r.t. divergence  $\Delta$ )'과 'CP- $\kappa$  알고리즘(Access-Free Reduction at Threshold  $\kappa$ )'을 적용하는 방식<sup>7)</sup>
  - 이는 CP- $\Delta$  알고리즘은 학습 데이터를 두 개의 독립적인 하위 데이터셋으로 나누고, 각각을 기반으로 서로 다른 두 개의 모델을 학습한 후 특정 프롬프트가 입력될 경우 두 모델의 생성 결과를 비교하여 유사성이 특정 임계값을 넘지 않도록 조정하는 방식으로, 이러한 과정을 통해 두 모델의 출력을 결합하면 특정 저작물과 과도하게 유사한 콘텐츠가 생성되지 않도록 유도할 수 있음<sup>7)</sup>
  - CP- $\kappa$  알고리즘은 AI 모델이 생성한 텍스트가 원본과 지나치게 유사할 경우, 해당 출력을 필터링하는 방식으로 작동<sup>7)</sup>
  - 기존 LLM이 학습한 데이터셋에서 특정 저작물에 대한 의존성을 줄이기 위해, 생성 과정에서 동적으로 유사성을 평가하고 랜덤화된 출력을 생성하는 접근법으로, 이 과정에서 모델은 생성된 텍스트의 유사도가 특정 임계값을 초과할 경우 원본과 다른 형태의 출력을 생성하도록 조정되며, 이를 통해 AI가 특정 문장을 그대로 복원하는 것을 방지<sup>7)</sup>

**[AI 모델이 저작권보호된 데이터를 학습했을 때와 그렇지 않았을 때의 생성 결과 분포 차이를 수치적으로 나타낸 그래프로, 기존의 차단 방식(출력 필터링, 모델 검열 등)과 비교하여, 랜덤화를 적용한 경우 AI 모델의 출력이 원본과 얼마나 덜 유사해지는지를 설명함<sup>8)</sup>**



7) Wei-Ning Chen et al, "Randomization Techniques to Mitigate the Risk of Copyright Infringement", arxiv, 2024.08.21, <https://arxiv.org/abs/2408.13278>

8) Wei-Ning Chen et al, "Randomization Techniques to Mitigate the Risk of Copyright Infringement", arxiv, 2024.08.21, p. 8., <https://arxiv.org/abs/2408.13278>

## 연구의 결론

- 이러한 방식은 기존의 프롬프트 필터링(prompt filtering)이나 AI 생성물 검열(output filtering) 방식과 비교했을 때, AI의 창의적인 활용을 제한하지 않으면서도 저작권 침해 위험을 완화하는 효과적인 방법으로 작용할 수 있음
- 프롬프트 필터링이나 생성물 검열 기법은 모델이 이미 학습한 내용을 기반으로 생성된 결과물을 사후적으로 조정하는 방식인 반면, 랜덤화 기법은 애초에 학습 데이터의 구조를 변화시켜 특정 문구를 직접적으로 학습하지 않도록 하는 근본적인 접근법이라는 점에서 차별성이 있음
- 이를 통해 LLM이 특정 저작물에 대한 의존성을 줄이고 보다 일반적인 언어 패턴을 학습할 수 있도록 유도하는 것이 이 연구에서 제안하는 방향

## ③ 워터마킹(Watermarking) 기술을 활용한 AI 생성물 보호 및 추적

### (A Survey of Text Watermarking in the Era of Large Language Models)

#### 연구의 초점: 워터마킹의 기존 텍스트 적용 방식과 LLM이 텍스트 생성 시 적용 방식 비교

- 이 연구에서는 대형 언어 모델 시대의 텍스트 워터마킹(A Survey of Text Watermarking in the Era of Large Language Models)을 종합적으로 분석하고, 현재까지 개발된 다양한 워터마킹 기법을 체계적으로 분류하여 소개함
- 특히, (1) 기존 텍스트에 적용하는 방식과 (2) LLM이 텍스트를 생성하는 과정에서 적용하는 방식의 두 가지 접근법을 중심으로 연구가 진행됨

#### 기존 텍스트 워터마킹: 생성된 텍스트에 워터마크를 추가하는 방식

- 형식 기반 워터마킹(Format-based Watermarking): 텍스트의 시각적 형식을 변경하는 방식으로, 줄 간격 조정, 단어 간격 변경, 특수 문자 삽입 등을 통해 워터마크를 삽입<sup>9)</sup>
- 어휘 기반 워터마킹(Lexical-based Watermarking): 단어를 동의어로 대체하거나 특정 단어를 암호화하여 워터마크를 삽입하는 방식<sup>9)</sup>
- 구문 기반 워터마킹(Syntactic-based Watermarking): 문장의 구조를 변경하여 워터마크를 삽입하는 방식으로, 어순 변경, 수동태 전환 등의 기법이 포함<sup>9)</sup>
- 생성 기반 워터마킹(Generation-based Watermarking): AI를 이용하여 워터마크가 포함된 텍스트를 생성하는 방식으로, 텍스트의 의미를 유지하면서도 특정 패턴을 삽입하여 출처를 판별할 수 있도록 함<sup>9)</sup>

#### LLM 생성 텍스트 워터마킹은 모델이 텍스트를 생성하는 과정에서 직접 워터마크를 삽입하는 방식

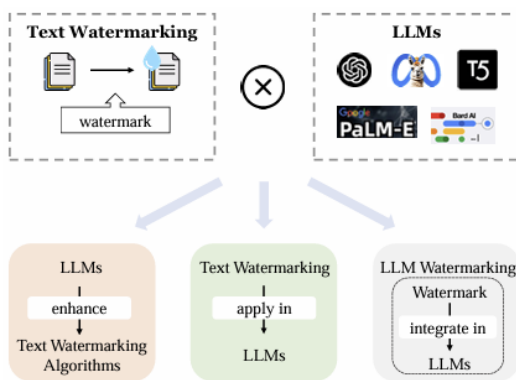
- 로짓 생성 과정에서의 워터마킹(Watermarking during Logits Generation): LLM이 다음 단어를 예측하는 과정에서 특정 단어 선택에 가중치를 부여하여 워터마크를 포함하도록 유도하는 방식<sup>9)</sup>
- 토큰 샘플링 과정에서의 워터마킹(Watermarking during Token Sampling): LLM이 단어를 선택할 때 특정 패턴을 따르도록 조정하여 워터마크를 삽입하는 방식<sup>9)</sup>
- LLM 훈련 과정에서의 워터마킹(Watermarking during LLM Training): 모델이 학습하는 과정에서 특정 워터마크 패턴을 내재화하도록 조정하여, 이후 생성되는 모든 텍스트에 워터마크가 포함되도록 설계하는 방식<sup>9)</sup>

9) Aiwei Liu et al, "A Survey of Text Watermarking in the Era of Large Language Models", arxiv, 2023.12.13, <https://arxiv.org/abs/2312.07913>

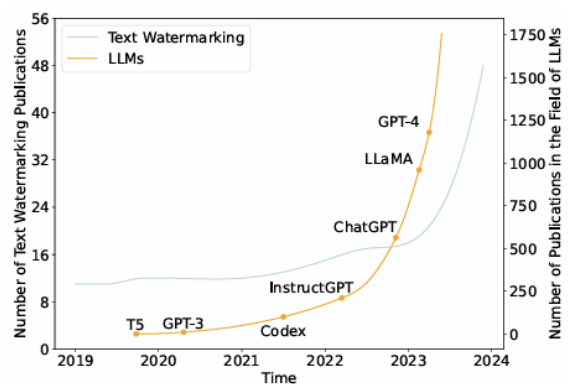
## 연구의 결론

- LLM을 활용한 최신 워터마킹 기법이 기존 방식보다 탐지율이 높고 변형 공격에 대한 저항력이 우수하다는 점을 확인<sup>10)</sup>
- 최신 알고리즘들은 워터마크의 검출률을 높이면서도 기존 텍스트와 비교하여 차이를 최소화하는 방향으로 발전하고 있으며, 특히 LLM 훈련 과정에서 워터마킹을 적용하는 방식이 가장 높은 탐지 정확도를 보이는 것으로 나타남<sup>10)</sup>
- 일부 워터마킹 기법은 단순한 문장 구조 변경이나 동의어 치환을 통해 쉽게 제거될 수 있으나, LLM을 활용한 워터마킹 기법은 문장 전체의 의미를 유지하면서 워터마크를 삽입하는 방식을 사용하여 제거가 어려운 특징을 가짐<sup>10)</sup>
- 워터마킹이 적용된 텍스트의 가독성과 자연스러움을 유지하는 것도 중요한 과제로 제시되었으며, 기존의 형식 기반 워터마킹보다 어휘 기반, 구문 기반, 그리고 생성 기반 워터마킹이 더 높은 자연스러움을 유지하는 것으로 분석
- 따라서 텍스트의 의미를 왜곡하지 않으면서도 워터마킹이 효과적으로 삽입될 수 있는 방법론이 지속적으로 연구될 필요가 있음
- 아래 그림을 보면, 대형 언어 모델(LLMs)의 발전이 텍스트 워터마킹 기술의 개발을 촉진하고 그 적용 범위를 확장하는 과정을 알 수 있음. 좌측 그림은 LLM의 발전으로 인해 AI 생성 텍스트의 신뢰성과 출처 확인이 중요한 문제로 부각되면서 기존의 문서 보호 및 인증 중심이던 워터마킹 기술이 AI 생성 콘텐츠의 출처 추적, 저작권 보호, 조작 탐지 등의 새로운 활용 사례로 확장되는 과정을 보여줌
- (b)에서는 LLM 연구가 급격히 증가한 특정 시점 이후 텍스트 워터마킹 관련 논문의 수도 함께 증가하는 패턴을 보이며, 이를 통해 LLM 기술의 발전이 워터마킹 기술에 대한 연구와 관심을 증대시키고 응용 분야를 확장하는 계기가 되었음을 실증적으로 보여주고 있음

### [LLM이 텍스트 워터마킹 기술의 발전을 촉진하고, 새로운 응용 분야를 확장하는 방식]<sup>11)</sup>



### [텍스트 워터마킹 연구 논문의 수 변화]<sup>11)</sup>



10) Aiwei Liu et al, "A Survey of Text Watermarking in the Era of Large Language Models", arxiv, 2023.12.13, <https://arxiv.org/abs/2312.07913>

11) Aiwei Liu et al, "A Survey of Text Watermarking in the Era of Large Language Models", arxiv, 2023.12.13, p.2., <https://arxiv.org/abs/2312.07913>

## 결론

### • AI의 발전과 저작권 보호: 학습 데이터에 포함된 저작물과 생성 결과의 유사성 문제

- 대형 언어 모델(LLM)의 발전과 함께 AI가 저작권 보호된 데이터를 학습하고 이를 재생산할 가능성이 높아지면서, AI의 저작권 침해 여부를 평가하고 방지하는 기술적 대응이 중요해지고 있음
- 기존 연구들은 AI가 학습한 데이터에 저작권 보호 콘텐츠가 포함되었는지를 탐색하는 데 초점을 맞췄으나, 이 방법만으로는 AI가 실제로 원본을 재현하는지를 판단하는 데 한계가 있었음
- 이에 따라 최근 연구들은 AI가 부분적인 정보를 제공받았을 때 원본과 얼마나 유사한 출력을 생성하는지를 평가하는 방식으로 접근하고 있으며, 이를 실험적으로 분석한 결과, 모델의 크기가 클수록, 특정 유형의 콘텐츠(특히 노래 가사)에 대해 더 높은 유사성을 보이며, 반복적 프롬프트 입력을 통해 점진적으로 더 많은 원본 내용을 복원하는 경향이 확인됨
- 이를 해결하기 위해 다양한 기술적 대응이 연구되고 있으며, 본 보고서에서 제시된 랜덤화 기법은 학습 데이터의 구조를 변경하여 AI가 특정 저작물을 직접적으로 학습하는 것을 방지하는 방식으로, 기존의 프롬프트 필터링이나 생성물 검열 방식보다 AI의 창의성을 유지하면서도 저작권 침해 가능성을 낮추는 효과를 보임
- 한편, 워터마킹 기술은 AI 생성 텍스트의 출처를 명확히 하고 저작권 보호를 강화하기 위해 특정 패턴을 삽입하는 방식으로, 기존 문서 보호용 워터마킹과 달리 AI가 생성하는 텍스트에 직접 적용되는 방식으로 발전하고 있음
- 연구 결과, LLM 훈련 과정에서 워터마킹을 적용하는 방식이 가장 높은 탐지 정확도를 보이며, 변형 공격에도 상대적으로 강인한 특징을 보였음. 이러한 연구들은 AI의 저작권 침해 가능성을 줄이고, AI 생성 콘텐츠의 저작권 문제를 해결하기 위한 중요한 기술적 대응의 시도로 분석됨

## 참고문헌

- Shreya Kapoor and Abir Roy, "ANI vs OpenAI: Copyright, AI, and the future of text data mining", MoneyControl, 2024.11.28, <https://www.moneycontrol.com/news/opinion/ani-vs-openai-copyright-ai-and-the-future-of-text-data-mining-12878278.html>
- Weijie Zhao et al, "Measuring Copyright Risks of Large Language Model via Partial Information Probing", arxiv, 2024.09.20, <https://arxiv.org/abs/2409.13831>
- Wei-Ning Chen et al, "Randomization Techniques to Mitigate the Risk of Copyright Infringement", arxiv, 2024.08.21, <https://arxiv.org/abs/2408.13278>
- Aiwei Liu et al, "A Survey of Text Watermarking in the Era of Large Language Models", arxiv, 2023.12.13, <https://arxiv.org/abs/2312.07913>
- Chris Cooke, "Copyright group says UK's AI Opportunities Action Plan stance on data mining reform is "deeply concerning"", CMU, 2025.01.15, <https://completemusicupdate.com/copyright-group-says-uks-ai-opportunities-action-plan-stance-on-data-mining-reform-is-deeply-concerning/>
- Valeska De Pauw, "AI and copyright: exploring exceptions for text and data mining", CMS Law-Now, 2024.10.16, <https://cms-lawnow.com/en/ealerts/2024/10/ai-and-copyright-exploring-exceptions-for-text-and-data-mining>
- "AI and Copyright Law: a landmark decision in Germany upholding the data mining exception and dismissing a photographer's claim", DDG, 2024.10.17, <https://www.ddg.fr/actualite/ai-and-copyright-law-a-landmark-decision-in-germany-upholding-the-data-mining-exception-and-dismissing-a-photographers-claim>
- George Lawton, "Is AI-generated content copyrighted?", TechTarget, 2024.11.21, <https://www.techtarget.com/searchcontentmanagement/answer/Is-AI-generated-content-copyrighted>