www.copyright.or.kr

저작권 이슈 브리프

COPYRIGHT ISSUE BRIEF

Weekly Report

2026. 4-2

한국저작권위원회
KOREA COPYRIGHT COMMISSION

저작권 이슈 브리프

SUMMARY | 산업/기업 | 기술

산업 구글, 최대 3분 길이의 음악을 생성할 수 있는 AI 모델 '리리아 3 프로' 출시

▶ 2026년 3월, 구글은 AI 음악 생성 모델 '리리아 3'의 확장 버전인 '리리아 3 프로'를 출시했다. 기존의 리리아 3는 최대 30초 길이의 음악만 생성할 수 있었으나, 프로 버전에서는 최대 3분까지 가능해졌다. 인트로, 코러스 등 특정 구간을 프롬프트로 지정해 제어할 수 있는 기능도 추가되었다. 한편 리리아 3 시리즈에는 기존 창작물을 보호하기 위한 장치도 마련되어 있다. 이용자가 특정 아티스트의 스타일을 모방한 음악을 생성하려 할 경우, 장르나 분위기 수준에서만 반영하도록 제한된다. 또한, 생성된 음원은 기존 음원과의 유사 여부를 점검하는 필터링 과정을 거치게 되며, 모든 음원에는 AI 생성물임을 식별할 수 있는 워터마크인 '신스ID'가 적용된다. 구글은 리리아 3 프로를 제미나이·구글 비즈 등 자사 서비스에 탑재하고, 외부 개발자용 API도 개방하여 활용 범위를 넓히고 있다.

산업 오픈AI, AI 영상 생성 서비스 '소라' 종료 발표

▶ 오픈AI는 2026년 3월 24일 AI 영상 생성 서비스 '소라' 종료 계획을 발표했으며, 웹사이트와 애플리케이션은 4월 26일, 개발자용 API는 9월 24일 각각 종료될 예정이다. 오픈AI는 이번 조치에 대해 핵심 서비스에 자원을 집중하기 위한 결정이라고 설명했다. 소라 2는 2025년 9월 공개 이후 저작권과 초상권 논란이 지속되어 왔으며, 이용자들이 저작권이 있는 캐릭터와 공인의 초상을 활용해 영상을 대량 생성하면서 창작자들의 반발이 이어진 바 있다. 한편, 디즈니는 자사와 창작자의 권리가 보호된다는 조건 아래 AI 기업과의 협력 가능성을 밝혀왔고, 이에 따라 2025년 12월 오픈AI에 10억 달러를 투자하고 캐릭터 라이선스 계약을 체결했다. 그러나 소라의 서비스 종료 발표로 계약을 체결한 지 약 3개월 만에 투자금이 집행되지 않은 채 해당 계약은 철회되었다.

산업 디저의 AI 생성 음원 탐지 기술 라이선싱과 권리 보호 체계

▶ 현재 프랑스 음원 스트리밍 서비스인 디저로 유입되는 일평균 약 6만 건의 AI 생성 음원 중 약 85%가 스트리밍 조작을 통해 저작권료를 편취하기 위해 등록된 것으로 밝혀진 가운데, 디저는 해당 문제에 대응하기 위해 AI 탐지 기술을 개발하였다. 디저는 AI 생성 음원을 자동으로 탐지하고 태그를 부여해 추천 알고리즘에서 제외함으로써 이용자에게 선택권을 제공하고, 인간 아티스트의 저작권 수익 보호를 도모하고 있다. 해당 AI 탐지 기술은 헝가리 실연자 권리 보호국 EJI와 프랑스 저작권협회 SACEM에 제공되었으며, 디저는 이를 포함한 종합 B2B 플랫폼을 개편해 발표하였다. 이는 스트리밍 플랫폼이 단순 유통 채널을 넘어 권리 보호 기술 공급자로 역할을 확장하며 새로운 수익 모델을 창출할 수 있음을 보여주는 사례이다.

저작권 이슈 브리프

산업 AI 에이전트의 콘텐츠 생산 참여 확대

▶ 워드프레스닷컴이 2026년 3월 MCP 표준을 기반으로 AI 에이전트가 웹사이트의 콘텐츠 생성부터 게시 및 관리 전반을 수행하는 기능을 도입했다. 전 세계 웹사이트의 43% 이상에 운영 기반을 제공하는 플랫폼 특성상, AI 에이전트가 보편적인 콘텐츠 생산 도구로 정착하는 변곡점이 될 것으로 평가된다. 다만 기존 사례에서 생성 콘텐츠의 투명성·품질 문제가 제기되어 왔으며, 신뢰성과 투명성 확보 여부에 따라 플랫폼 차원의 기능 보급이 제한될 가능성도 있다. 이번 사례는 인간의 역할이 창작자에서 관리자로 이동하는 흐름을 보여주는 동시에, 저작물의 작성 주체와 책임 판단 기준이 운영자·플랫폼 중심으로 재해석될 가능성을 제기한다. 결과적으로 기술 도입에 따른 진입 장벽 완화와 창작 공정 효율화라는 긍정적 가능성과 함께, 생산 주체의 투명한 공개와 인간의 실질적 개입 여부가 콘텐츠 관리 책임 및 저작권 성립의 핵심적인 판단 기준으로 작용할 것으로 예상된다.

산업 음악 핑거프린팅 기술을 활용한 뮤직스매치의 실시간 가사 저작권 탐지

▶ 2026년 3월, 뮤직스매치는 음악 핑거프린팅 기술을 활용해 저작권이 있는 가사의 사용 여부를 실시간 탐지하는 서비스 '센티널'을 공개했다. 센티널은 뮤직스매치가 보유한 세계 최대 규모의 가사 데이터베이스를 기반으로, 텍스트의 고유 패턴을 식별 정보로 변환해 원저작물과 대조한다. 이를 통해 가사의 일부 인용이나 순서 변형까지 밀리초 단위로 식별할 수 있으며, 탐지 결과를 창작물, 사용 허가 저작물, 보호 대상 저작물, 자유 이용 저작물의 네 유형으로 분류해 유형별 대응이 가능하다. 다만 향후 가사가 아닌 영역으로 확장할 경우 기술적 난이도가 높아질 수 있고, 오탐지로 인한 정상 콘텐츠 이용 제한 가능성도 과제로 남아 있다.

기술 주간 기술 동향

▶ 최근, AI 모델이 특정 데이터를 학습하지 않았음을 입증해야 하는 필요성이 더욱 증대되고 있다. 현재 AI 기업들은 학습 데이터 구성을 공개하지 않으며, 기존 기술은 데이터 포함 여부는 탐지할 수 있으나 포함되지 않았음을 증명하는 데는 한계가 있다. 이러한 문제를 해결하기 위해 개발된 'PRISM'은 AI 모델이 생성하는 토큰 확률 순위를 분석하여 포함되지 않았음을 검증하는 기술이다. PRISM은 검증 대상 모델과 참조 모델의 예측 패턴을 비교하여 순위 상관계수를 계산하며, 학습하지 않은 데이터는 낮은 상관계수를 나타낸다. 실험 결과 95% 이상의 정확도로 비학습 데이터를 식별했으며, 기존 기법 대비 20~30% 향상된 성능을 보였다. PRISM은 대규모 데이터셋 검증 시 계산 비용 증가하며, 모델 구조 변화에 따라 패턴이 변동될 수 있다는 한계가 있으나, AI 투명성 제고와 저작권 분쟁 해결을 위한 핵심 도구로 자리잡을 가능성을 보여준다.

구글, 최대 3분 길이의 음악을 생성할 수 있는 AI 모델 '리리아 3 프로' 출시

기존 모델인 '리리아 3' 대비 '리리아 3 프로'의 주요 변화

- **최대 3분 길이의 음악 생성 및 구간별 제어 기능 추가**

- 구글(Google)은 2026년 2월 '리리아 3(Lyria 3)'를 공개했으며, 2026년 3월 확장 버전인 '리리아 3 프로(Lyria 3 Pro)'를 출시함
- 리리아 3는 구글이 개발한 AI 기반 음악 생성 모델로, 이용자가 텍스트 프롬프트를 통해 원하는 장르·분위기·악기 등을 지정하면 이를 반영한 음악을 자동으로 생성해 주는 도구임
- 기존 리리아 3는 최대 30초 길이의 음악만 생성할 수 있었으나, 프로 버전에서는 최대 3분 길이의 음악 생성이 가능해짐
- 길이가 늘어남에 따라 곡 구성을 보다 정교하게 설계할 수 있게 되었으며, 인트로(도입부)나 코러스(후렴) 등 특정 구간을 프롬프트로 지정하여 세밀하게 제어할 수 있는 기능이 추가됨

[표1] '리리아 3' vs '리리아 3 프로' 비교표

구분	리리아 3	리리아 3 프로
출시 시점	2026년 2월	2026년 3월
최대 생성 길이	30초	최대 3분
구간 제어	미지원	인트로·코러스 등 지정하여 제어 가능

출처: 참고문헌 종합하여 재구성

리리아 3 시리즈, 기존 창작물의 모방을 제한하도록 설계

- **특정 아티스트 스타일 모방 제한 설계 및 생성물 관리 방식**

- 리리아 3 시리즈는 이용자가 특정 아티스트의 스타일을 모방한 음악을 생성하려 할 경우에도, 해당 스타일을 그대로 재현하지 않고 장르나 분위기만 반영하도록 설계됨. 또한 생성된 음원은 기존 음원과의 유사도를 점검하는 필터링 절차를 거치도록 설계됨
- 구글은 생성된 음원이 유통되는 과정에서도 출처를 식별할 수 있도록 조치함. 구체적으로, 리리아 3 시리즈로 생성된 모든 음원에는 구글이 개발한 워터마크인 신스ID(SynthID)*가 적용되어, 해당 음원이 구글 AI에 의해 생성된 콘텐츠임을 확인할 수 있음

- 아울러 구글은 리리아 3 시리즈 학습에 유튜브 및 파트너사와의 계약, 관련 법률, 서비스 약관에 따라 사용 권한을 확보한 데이터만을 활용했다고 명시함[1)]

*신스ID(SynthID)는 구글이 개발한 AI 생성물 식별 기술로, 사진·음성·영상·텍스트 등 다양한 콘텐츠에 사람의 눈이나 귀로는 인지할 수 없는 워터마크를 삽입하여, 해당 콘텐츠가 AI로 생성되었는지를 확인할 수 있도록 하는 기능

구글, 리리아 3 프로를 자사 서비스 및 외부 플랫폼으로 확장

• 소비자·기업·개발자·전문가 대상 서비스 확장

- 구글은 리리아 3 프로를 음악 전문가용 창작 도구에 한정하지 않고, 소비자·기업·개발자 등 다양한 이용자층을 대상으로 자사 서비스 전반에 확장 적용함
- (소비자) 생성형 AI 서비스인 제미나이(Gemini)에서 브이로그·팟캐스트 등에 활용할 수 있는 맞춤형 배경 음악 제작을 지원함. 영상 제작 도구인 구글 비즈(Google Vids)에서도 콘텐츠 분위기에 맞는 배경 음악을 직접 생성할 수 있도록 함
- (기업) 기업용 AI 플랫폼인 버텍스 AI(Vertex AI)에서도 대규모 음원 생성을 지원함
- (개발자) 구글 AI 스튜디오(Google AI Studio) 및 제미나이 API*를 통해 외부 개발자가 별도의 음악 생성 모델 개발 없이 자체 앱이나 웹서비스에 해당 기능을 연동할 수 있음
- (전문가) 아티스트·프로듀서·작곡가를 위한 협업 창작 도구인 프로듀서AI(ProducerAI)에 탑재하여 전문적인 음악 제작 과정에 활용할 수 있도록 함

*API(Application Programming Interface): 외부 개발자가 특정 소프트웨어의 기능을 자신의 서비스에서 호출하여 사용할 수 있도록 제공하는 연결 통로로, 직접 기능을 개발하지 않고도 해당 기능을 활용할 수 있도록 지원함

[그림1] 제미나이 내 리리아 3 프로 이용 화면. 음악 생성 도구 선택 화면(좌), 프롬프트 입력 화면(우)

출처: 구글 제미나이(Gemini) 화면 캡처

참고문헌

• Myriam Hamed Torres, “Lyria 3 Pro: Create longer tracks in more Google products”, Google, 2026.03.25., https://blog.google/innovation-and-ai/technology/ai/lyria-3-pro/
• Ivan Mehta, “Google launches Lyria 3 Pro music generation model”, TechCrunch, 2026.03.25., https://techcrunch.com/2026/03/25/google-launches-lyria-3-pro-music-generation-model/

1) Myriam Hamed Torres, “Lyria 3 Pro: Create longer tracks in more Google products”, Google, 2026.03.25., https://blog.google/innovation-and-ai/technology/ai/lyria-3-pro/

오픈AI, AI 영상 생성 서비스 '소라' 종료 발표

소라의 서비스 종료 일정과 오픈AI의 방침

• 소라의 서비스 종료 일정 및 이용자 데이터 처리 방침

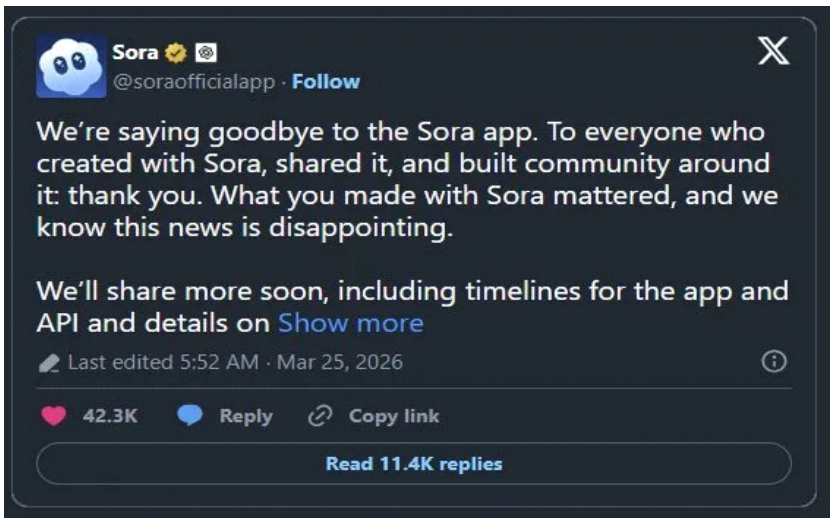
- 오픈AI는 2026년 3월 24일 자사의 AI 영상 생성 서비스인 소라를 종료한다고 발표함[1]. 소라 웹사이트와 애플리케이션은 2026년 4월 26일, 개발자용 API*는 9월 24일에 각각 종료될 예정임
- 이용자는 서비스 종료 전 소라 라이브러리에서 자신의 콘텐츠를 직접 다운로드할 수 있으며, 종료 후에도 제한된 기간 동안 콘텐츠를 추가로 다운로드 할 수 있음
- 추가로 다운로드 할 수 있는 기간이 종료되면 소라 이용과 관련된 데이터가 모두 영구적으로 삭제될 예정임

* API(Application Programming Interface): 서로 다른 프로그램이나 서비스 간에 정해진 방식으로 요청과 응답을 주고받을 수 있도록 하는 통로를 의미

• 오픈AI, 핵심 서비스에 집중하기 위해 소라 종료 결정

- 오픈AI 대변인은 핵심 서비스에 자원을 집중하기 위해 소라를 종료하며, 해당 연구팀은 로봇 개발 연구로 전환할 계획이라고 밝힘
- 오픈AI의 애플리케이션 총괄 담당자 피지 시모(Fidji Simo)는 2026년 3월 초, 직원들에게 부수적인 사업을 축소하고 코딩·엔터프라이즈 도구 등 핵심 사업에 집중하겠다는 방침을 밝힌 것으로 전해짐

[그림] 소라 서비스 종료 공지 게시물



출처: Dominic Patten, "Disney's $1B Investment In OpenAI DOA As Sam Altman Pulls Sora Plug: "The Deal Is Not Moving Forward"", Deadline, 2026.03.24., https://deadline.com/2026/03/sora-shut-down-disney-investment-1236764689/

1) Katelyn Chedraoui, "OpenAI's Once Viral Sora AI Video App Is Being Discontinued", CNET, 2026.03.24., https://www.cnet.com/tech/services-and-software/openais-once-viral-sora-ai-video-app-is-being-discontinued/

소라 2의 저작권·초상권 논란 및 디즈니와의 계약 체결·철회

• 소라 2 출시 이후의 저작권·초상권 논란

- 오픈AI는 2024년 2월 소라를 처음 공개했으며, 같은 해 12월 챗GPT 프로와 챗GPT 플러스 가입자에 한해 제한적으로 서비스를 제공했음
- 오픈AI는 2025년 9월 '소라 2'를 대중에 공개하면서, 저작권자가 별도로 사용 중단을 요청하지 않는 한 저작물이 AI 영상 생성에 활용되는 것을 허용하는 옵트아웃 정책을 적용함
- '소라 2' 출시 이후, 소라 이용자들이 스폰지밥, 마리오, 피카츄 등 저작권이 있는 IP를 활용해 영상을 대량으로 생성하면서 논란이 확산됨
- 저작권 침해 외에도 영화배우나 정치인 등 공인의 초상을 활용한 딥페이크 영상이 확산되자 할리우드 스튜디오와 창작자들 사이에서 강한 반발이 일었음

• 저작권·초상권 논란 속 오픈AI와 디즈니의 라이선스 계약 체결

- 디즈니는 2025년부터 AI 기업들을 상대로 저작권 보호 조치를 적극적으로 전개해 옴
- 2025년 6월 NBC유니버셜(NBCUniversal)과 함께 AI 이미지 생성 기업인 미드저니(Midjourney)를 상대로 소송을 제기했고, 같은 해 9월 AI 챗봇 플랫폼 캐릭터.AI(Character.AI)에 저작권 침해 경고장을 발송했음
- 다만 디즈니는 자사와 창작자의 권리가 보호된다는 조건 하에, AI 기업들과 공정한 협력에 열려 있다는 입장을 밝혔음
- 이러한 배경에서 디즈니는 2025년 12월 11일 오픈AI와 10억 달러(원화 약 1조 5,405억 원)[2] 규모의 투자 계약과 캐릭터 라이선스 계약을 체결하였음
- 디즈니는 해당 계약이 이러한 캐릭터 IP 보유 기업과 AI 기업 간 협력 모델의 사례가 되기를 기대한 것으로 알려짐
- 해당 계약에 따르면 소라 이용자가 마블, 스타워즈, 디즈니 애니메이션, 픽사 등 디즈니 소유 캐릭터 200여 개를 활용해 콘텐츠를 생성하는 것이 허용될 예정이었음
- 계약에는 캐릭터 라이선스만 포함되고 공인의 초상이나 음성 사용은 제외되었으며, 양사는 '이용자의 안전과 창작자의 권리를 보호하는 책임 있는 AI 사용'을 합의 조건으로 명시했음

• 디즈니, 소라 서비스 종료 통보에 따라 계약 철회

- 디즈니는 2026년 3월 24일 오픈AI와 소라 관련 회의를 진행했으나, 회의 종료 약 30분 뒤 소라 서비스가 종료된다는 사실을 통보받은 것으로 알려짐
- 이에 따라 2025년 12월 계약을 체결한 지 약 3개월 만에, 투자금이 집행되지 않은 상태에서 디즈니-오픈AI 간의 투자·라이선스 계약이 철회됨
- 디즈니 대변인은 오픈AI가 영상 생성 사업에서 철수한 결정을 존중한다며, 앞으로도 IP와 창작자의 권리를 보호하면서 AI 기업들과 협력해 나가겠다고 밝힘

2) 1달러=1,530.50원(2026.04.01, KEB 하나은행 매매기준율 적용, 이하 동일)

AI 기업들의 사업 방향 전환: 생성형 미디어에서 코딩 등 실무 제품으로

• AI 기업들의 사업 재편과 오픈AI의 향후 방향

- 2025년 하반기에 생성형 미디어가 주목받았던 것과 달리, 2026년에는 AI 기업들이 코딩 도구 등 실무 중심 제품에 집중하는 흐름이 나타나고 있음
- 미국 기술 매체 CNET은 오픈AI의 소라 서비스 종료에 대해, 주요 AI 기업들이 생성형 미디어 시장의 수익성에 확신을 갖지 못하고 있음을 보여준다고 분석함
- 다만 오픈AI는 영상 생성 사업에서 완전히 철수하는 것은 아니며, 향후 챗GPT 내 영상 생성 기능이 유지될 가능성이 있다고 밝힘

참고문헌

- Katelyn Chedraoui, "OpenAI's Once Viral Sora AI Video App Is Being Discontinued", CNET, 2026.03.24., https://www.cnet.com/tech/services-and-software/openais-once-viral-sora-ai-video-app-is-being-discontinued/
- Jim Vejvoda, "OpenAI Shuts Down Sora Generative Video App, Disney Pulls Out of Investment and Licensing Deal", IGN, 2026.03.25., https://www.ign.com/articles/openai-shuts-down-sora-generative-video-app-disney-pulls-out-of-investment-and-licensing-deal
- Todd Spangler, "OpenAI Just Spiked Bob Iger's Final Big Strategic Deal. For Disney, Maybe That's Lucky", Variety, 2026.03.25., https://variety.com/2026/digital/news/why-openai-disney-ended-sora-deal-bob-iger-1236698901/
- OpenAI, "What to know about the Sora discontinuation", 2026.03.28., https://help.openai.com/en/articles/20001152-what-to-know-about-the-sora-discontinuation

디저의 AI 생성 음원 탐지 기술 라이선싱과 권리 보호 체계

AI 생성 음원의 대규모 유입과 탐지 기술 개발 배경

• AI 생성 음원 유입 현황 및 부정 행위 실태

- 프랑스의 음원 스트리밍 서비스 디저(Deezer)는 2026년 1월 플랫폼에 일평균 약 6만 건의 완전 AI 생성 음원이 유입되고 있다고 밝혔으며, 이는 전체 일일 신규 등록 음원의 약 39%에 해당하는 규모임
- 이 중 최대 85%의 AI 생성 음원 스트리밍이 조회수를 인위적으로 늘려 저작권료 편취를 시도한 것으로 판명되어 수익 배분 풀에서 제외되었으며, 이는 AI 생성 음원이 저작권료 시스템을 악용하는 사기 수단으로 활용되고 있음을 보여줌
- 참고로, 디저는 180개국 이상에서 서비스를 제공하고 프랑스, 독일, 영국, 브라질, 미국에 600명 이상의 직원을 보유한 세계 최대 독립 음악 플랫폼 중 하나로서, 음악이 인간의 창작물이며 권리자가 보호받아야 한다는 신념을 바탕으로 업계 최고 수준의 AI 생성 음원 탐지 기술을 구축함

• 디저의 AI 생성 음원 자동 인식 및 라벨링 시스템 구축

- 디저는 스트리밍 플랫폼 중 유일하게 AI 생성 음원을 자동으로 탐지하고 명확하게 태그를 부여하는 시스템을 운영하고 있으며, 이를 추천 알고리즘에서 배제하여 이용자에게 선택권을 제공함
- 해당 기술은 AI와 시스템을 활용한 비정상적인 스트리밍을 차단하고, 인간 아티스트의 저작권 수익이 부당하게 분배되는 것을 방지하는 데 기여함
- 디저의 해당 기술 적용은 스트리밍 플랫폼 기술을 저작권 관리 영역으로 확장한 사례로 볼 수 있음

저작권 단체로의 AI 탐지 기술 라이선싱 확대와 권리 보호 체계

• 헝가리 실연자 권리 보호국 EJI와의 계약 체결

- 헝가리 실연자 권리 보호국(Eloadomuveszi Jogvedo Iroda, 이하 EJI)는 디저의 AI 생성 음원 탐지 기술을 라이선스한 최초의 집중관리단체*가 되었으며, 이를 통해 공개된 음원에서 생성형 AI 개입 여부를 탐지할 역량을 확보함
- EJI는 아티스트를 기계와의 경쟁으로부터 보호하는 솔루션을 적극적으로 모색하고 있으며, AI 생성 음원을 필터링하는 것이 필수적이지만 AI 학습 자체가 아티스트의 동의와 보상을 전제로 이루어져야 한다고 강조함
- EJI는 생성형 AI의 도움으로 제작된 음원에 대해서는 저작권료를 지급하지 않는 정책을 시행하고 있으며, 이는 기술적 탐지와 권리 보호 정책이 연계된 사례로 평가됨

- 이번 계약은 집중관리단체가 AI 시대에 대응하기 위해 스트리밍 플랫폼의 탐지 기술을 도입한 최초 사례로서, 저작권 관리 체계의 기술적 고도화 가능성을 시사함

* 집중관리단체(Collective Management Organization): 저작권자를 대신해 저작물 이용 허락, 사용료 징수·분배 등을 관리하는 기관

• 프랑스 저작권협회 등 음악 산업 전반으로의 기술 확산

- 디저는 2025년 이미 프랑스 저작권협회 사쎔(Societe des Auteurs, Compositeurs et Editeurs de Musique, SACEM)에 이 기술을 라이선스한 바 있으며, 저작권 단체들이 AI 생성 음원을 식별하고 관리할 수 있는 기술적 기반을 제공함
- 디저는 음악 산업 전반에 AI 탐지 기술을 제공함으로써 투명성 확보와 인간 창작자 권리 보호 운동에 동참하고 있으며, 이는 스트리밍 서비스가 단순 유통 채널을 넘어 권리 보호 기술 공급자로 역할을 확장하는 사례로 평가됨

• 디저의 B2B 플랫폼 전략과 기술 라이선싱 사업화

- 디저는 2026년 3월 파트너십 플랫폼을 전면 개편하여 발표하였으며, AI 생성 음원 탐지 기술을 포함한 5개 사업 부문으로 구성된 종합 B2B 플랫폼을 구축함
- 디저는 미국의 라디오 플랫폼 소노스(Sonos)와의 파트너십을 통해 음악을 제공하고, 오랑주(Orange), 프낙 다르티(FNAC Darty) 등 통신사 및 유통사와 협력하며, 던킨도너츠(Dunkin' Donuts), 컨버스(Converse)의 매장에 음악 솔루션을 제공하는 등 다각화된 B2B 서비스를 운영함
- AI 탐지 기술의 라이선싱 사업화는 스트리밍 플랫폼이 콘텐츠 유통뿐 아니라 권리 보호 기술 제공을 통해 새로운 수익 모델을 창출할 수 있음을 입증한 사례로 평가됨

시사점: 음악산업의 AI 생성 음원 관리를 위한 과제

• 집중관리단체의 AI 탐지 역량 확보 필요성

- 헝가리 EJI가 디저의 AI 탐지 기술을 도입한 사례는 집중관리단체가 AI 시대에 대응하기 위해 기술 역량을 확보해야 할 필요성을 보여주는 상징적 사건임
- 디저 사례에서 보듯이 AI 생성 음원이 일평균 6만 건 이상 유입되고 그중 85%가 부정 행위로 판명되는 상황에서, 저작권 단체가 자체적으로 AI 음원을 식별할 수 있는 기술적 기반을 갖추는 것이 권리 보호의 전제 조건으로 부상하고 있음
- 다만 AI 생성 음원 필터링만으로는 충분하지 않으며, AI 학습 단계에서부터 아티스트의 사전 동의와 보상 체계가 확립되어야 한다는 주장이 제기되고 있으며, 이는 기술적 탐지와 법제도적 규율이 병행되어야 함을 시사함

• 플랫폼 차원의 AI 생성 음원 투명성 확보와 산업 표준화 과제

- 디저의 AI 탐지 기술 상용화는 음악 스트리밍 시장에서 AI 생성 음원의 표시 및 구분이 단순한 윤리적 선택이 아니라 산업 표준으로 자리 잡을 가능성을 제시함
- 디저가 AI 생성 음원을 탐지하여 추천 알고리즘에서 배제하고 이용자에게 선택권을 제공하는 방식은, 플랫폼이 시장 투명성과 창작자 권리 보호를 동시에 추구할 수 있는 모델로 평가됨
- 디저의 B2B 플랫폼 강화 전략은 스트리밍 서비스가 단순히 음원 유통 채널에 머무르지 않고, 권리 보호 기술 제공자로서 새로운 수익 모델을 창출할 수 있음을 보여줌

참고문헌

- James Hanley, "Deezer licenses AI music detection technology to Hungarian rights organization EJI", MusicBusinessWorldwide, 2026.03.27., https://www.musicbusinessworldwide.com/deezer-licenses-ai-music-detection-technology-to-hungarian-rights-organization-eji/
- Deezer, "Deezer announces revamped partnership platform Deezer for Business to drive continued growth", 2026.03.19., https://newsroom-deezer.com/2026/03/deezer-announces-revamped-partnership-platform-deezer-for-business-to-drive-continued-growth/

AI 에이전트의 콘텐츠 생산 참여 확대

웹 호스팅 플랫폼의 AI 에이전트 도입 및 작동 방식

• 워드프레스닷컴의 AI 에이전트 기반 콘텐츠 생성 기능 도입

- 웹 호스팅 플랫폼 워드프레스닷컴(WordPress.com)은 2026년 3월 AI 에이전트(AI agent)*가 웹사이트 콘텐츠를 작성·편집·게시하고, 댓글 관리·메타데이터 수정·태그 및 카테고리 구성까지 수행할 수 있는 기능을 도입함
- 해당 기능은 모델 컨텍스트 프로토콜(model context protocol, MCP)**을 기반으로 작동하며, 사용자가 자연어 명령어로 지시하면 클로드(Claude)·커서(Cursor) 등 MCP를 지원하는 AI 모델과 연동할 수 있음
- AI 생성 콘텐츠는 원칙적으로 사용자의 최종 승인을 거치도록 설계되어 있으나, 사용자 설정에 따라 AI가 콘텐츠를 직접 게시하는 방식도 선택할 수 있음
- 전 세계 웹사이트의 43% 이상이 워드프레스닷컴을 통해 운영된다는 점을 고려할 때, 이번 기능 도입은 AI 에이전트가 보편적인 콘텐츠 생산 도구로 정착하는 변곡점이 될 것으로 평가됨

* AI 에이전트(AI agent): 사용자의 자연어 지시를 기반으로 목표를 이해하고, 외부 시스템과 연동하여 콘텐츠 생성·수정·게시 등 일련의 작업을 자율적으로 수행하는 인공지능 기반 실행 주체

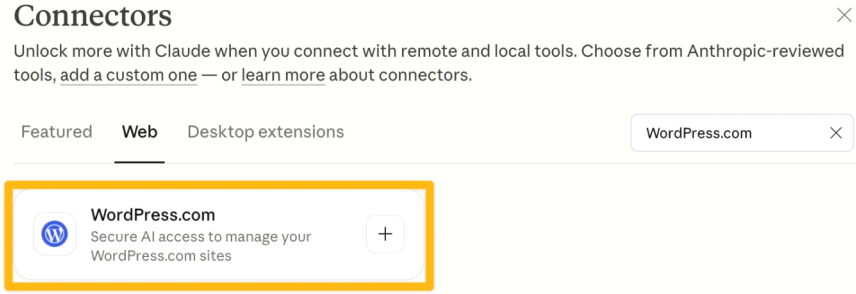
** 모델 컨텍스트 프로토콜(model context protocol, MCP): AI 에이전트가 외부 플랫폼의 데이터와 기능에 접근할 수 있도록 연결 방식을 표준화한 프로토콜

• AI 에이전트의 권한과 작동 방식

- 워드프레스닷컴의 AI 에이전트 기능은 사이트 콘텐츠, 설정 및 분석 데이터에 대한 읽기 권한을 시작으로, 게시물 생성·편집·게시로 이어지는 쓰기 권한까지 단계적으로 확장되는 구조로 설계됨
- AI 에이전트는 콘텐츠를 생성하기 전 사이트의 테마와 디자인을 먼저 참조하도록 설계되어, 기존 사이트의 색상·폰트·레이아웃 패턴에 부합하는 결과물을 도출하도록 작동함
- 모든 변경 사항은 사이트 활동 로그(activity log)*에 기록되어 추적이 가능하나, 독자에게 AI 생성 콘텐츠임을 별도로 알리는 표시 기능은 포함되어 있지 않음

* 활동 로그(activity log): 워드프레스닷컴이 제공하는 사이트 변경 이력 추적 기능으로, AI 에이전트가 수행한 작업 내역도 동일하게 기록됨

[그림] 워드프레스닷컴의 클로드 MCP 커넥터 설정 화면



출처: WordPress.com Support, "Edit your site with Claude", 2026.03.18., https://wordpress.com/support/model-context-protocol-mcp-settings/connect-claude/

선행 사례의 한계와 콘텐츠 생산 주도권 변화

• 기존의 AI 콘텐츠 실험 사례와 한계

- 워드프레스닷컴의 기능 도입에 앞서 AI 기술기업 앤트로픽(Anthropic)은 2025년 6월 자사 AI 모델인 클로드가 직접 게시물을 작성하는 블로그 '클로드 익스플레인스(Claude Explains)'를 공개하였으나, 약 1주 만에 서비스를 종료한 바 있음
- 해당 블로그는 클로드가 생성한 초안을 사람이 검수·보완하는 방식으로 운영되었으나, AI 생성 여부와 인간의 편집 범위가 명확히 표시되지 않았다는 점에서 투명성 부족에 대한 비판을 받았음
- 블룸버그(Bloomberg) 및 G/O 미디어(G/O Media) 등 주요 미디어사도 AI 작성 기사의 사실관계 오류와 품질 저하 문제로 인해 편집진의 반발과 독자의 비판에 직면했던 선례가 존재함
- 이러한 사례들은 향후 AI 생성 콘텐츠가 다양한 플랫폼으로 확산되는 단계에서도 콘텐츠의 신뢰성과 투명성 확보가 기술 안착을 결정짓는 리스크가 될 수 있음을 시사함

• 콘텐츠 생산 주도권의 이동과 저작권 체계의 변화

- AI는 기존에는 인간의 콘텐츠 창작을 보조하는 도구로 활용되었으나, 앞으로는 초안 생성부터 게시·댓글 관리까지 전 과정을 수행하고 인간은 최종 승인 역할만 담당하는 형태로 확대될 가능성이 나타남
- 이러한 변화는 콘텐츠 생산 과정에서 인간의 역할이 창작자에서 관리자로 이동하는 방향을 의미하며, AI 에이전트에게 창작의 주도권 중 일부가 부여되는 흐름으로도 볼 수 있음
- 창작 주도권이 AI 에이전트에게 일부 부여됨에 따라 저작물 작성 주체와 책임 판단 기준이 운영자 및 플랫폼 중심으로 재해석될 가능성이 제기되며, 특히 인간의 개입 수준에 따라 저작권 인정 범위와 표시 기준이 달라질 수 있음
- 이러한 흐름은 현행 저작권 체계의 근간인 '인간 중심 창작 구조'에 대한 새로운 해석의 필요성을 제기하며, 향후 생산 주체의 투명한 공개와 인간의 실질적 개입 정도가 콘텐츠 관리 책임 및 저작권 인정 여부를 결정하는 기준으로 다뤄질 가능성을 시사함

AI 에이전트의 콘텐츠 생산 참여에 따른 기대와 우려

• 진입 장벽의 완화와 창작 공정의 효율화

- 웹 호스팅 플랫폼에서 AI 에이전트의 콘텐츠 생성·게시·관리 기능이 제공되면서, 인간의 직접적인 콘텐츠 생산 없이도 사이트를 유지·운영하는 것이 기술적으로 가능한 환경이 형성되고 있음
- 기술적·경제적 여건이 부족한 개인이나 소규모 운영자의 진입 장벽을 낮추고, 반복 작업을 AI에 위임함으로써 전문가가 핵심 창작에 집중할 수 있는 구조가 형성됨
- 실제로 AI 에이전트가 생성한 콘텐츠가 독자에게 AI 모델의 작성 방식이나 표현 특성에 대한 직접적인 참고 자료로 활용되는 등 AI 생성 콘텐츠의 활용에 관한 논의가 확장되는 양상을 보임

• 창작물 희소성 약화와 신뢰성 저하 우려

- AI 생성 콘텐츠의 무분별한 확산은 인간 창작물의 고유한 가치와 차별성을 약화시키고 콘텐츠의 진정성에 대한 독자의 신뢰 저하로도 이어질 수 있다는 시각도 존재함

- 선행 사례에서 나타난 AI 생성 콘텐츠의 품질 관리 문제는 플랫폼 기반 확산 단계에서도 해소되지 않은 채 더 넓은 범위로 이어질 가능성이 있음
- 이와 함께 일부 언론사에서 AI 도입이 생산성 향상뿐 아니라 고용 축소의 수단으로 활용하는 움직임이 나타나는 등 콘텐츠 생산 영역 내 인간 창작자의 역할 변화에 대한 논의가 업계 전반으로 확산되는 양상임

참고문헌

- Sarah Perez, "WordPress.com now lets AI agents write and publish posts, and more", TechCrunch, 2026.03.20., https://techcrunch.com/2026/03/20/wordpress-com-now-lets-ai-agents-write-and-publish-posts-and-more/
- Kyle Wiggers, "Anthropic's AI is writing its own blog — with human oversight", TechCrunch, 2025.06.03., https://techcrunch.com/2025/06/03/anthropics-ai-is-writing-its-own-blog-with-human-oversight/
- Kyle Wiggers, "Anthropic's AI-generated blog dies an early death", TechCrunch, 2025.06.09., https://techcrunch.com/2025/06/09/anthropics-ai-generated-blog-dies-an-early-death/
- WordPress.com Support, "Edit your site with Claude", 2026.03.18., https://wordpress.com/support/model-context-protocol-mcp-settings/connect-claude/

음악 핑거프린팅 기술을 활용한 뮤직스매치의 실시간 가사 저작권 탐지

AI·UGC 확산에 따른 음악 가사 저작권 침해 문제

• 콘텐츠 생성 속도 증가에 따른 무단 이용 사례 확산

- 생성형 AI 도구와 이용자 제작 콘텐츠(User-Generated Content, UGC)* 플랫폼의 확산으로 디지털 콘텐츠의 제작 및 업로드가 급증하면서, 저작권이 있는 음악이나 가사가 무단으로 포함되어 유통되는 사례가 늘고 있음
- 생성형 AI 도구가 텍스트, 이미지, 음원 등을 자동으로 생성하는 과정에서는 학습 데이터에 포함된 저작물의 일부가 이용자도 인지하지 못한 사이 결과물에 반영될 수 있음
- 이로 인해 저작권 침해 콘텐츠가 플랫폼에 게시되더라도 즉시 걸러내기 어렵고, 권리자 또한 자신의 저작물이 무단으로 이용되고 있다는 사실을 제때 파악하기 어려움

* 이용자 제작 콘텐츠(User-Generated Content, UGC): 일반 이용자가 직접 제작하여 온라인 플랫폼에 업로드하는 텍스트, 이미지, 영상, 음악 등의 디지털 콘텐츠

• 사후 삭제 방식의 대응 한계와 실시간 탐지의 필요성

- 현재 대부분의 플랫폼은 권리자가 침해 사실을 발견한 뒤 삭제를 요청하는 사후 통지 및 삭제(notice and takedown) 방식에 의존하고 있으나, 콘텐츠의 생성 속도가 빨라지면서 대응에 한계가 나타남
- 특히 저작물 전체가 아닌 일부만 인용되거나 변형 또는 편집된 형태로 활용된 경우에는, 기존의 파일 대조 방식이나 해시값* 비교 방식만으로는 침해 여부를 정확히 판별하기 어려움
- 이러한 한계를 배경으로, 음악 데이터 기업을 중심으로 콘텐츠가 플랫폼에 게시되는 시점에 저작권 침해 여부를 실시간으로 탐지하는 기술이 개발되고 있음

* 해시값(hash value): 디지털 파일의 고유한 식별 코드로, 파일 내용이 동일하면 같은 값을 생성하지만 일부만 변경되어도 전혀 다른 값이 산출되는 특성을 가짐

센티널의 기술 구조와 플랫폼 적용 방식

• 핑거프린팅 기반의 가사 유사도 분석 방식

- 글로벌 음악 데이터 및 기술 기업 뮤직스매치(Musixmatch)는 2026년 3월 음악 핑거프린팅* 기술을 활용해 저작권이 있는 가사의 사용 여부를 실시간으로 탐지하는 서비스인 센티널(Sentinel)을 공개함
- 센티널은 20만 개 이상의 음악 출판사가 활용하는 세계 최대 규모의 가사 데이터베이스를 기반으로, 가사가 부분적으로 인용된 경우까지 수 밀리초 내에 식별할 수 있음[1)]

1) MusixMatch, “Musixmatch launches real-time music copyright detection service”, PR Newswire, 2026.03.27., https://www.prnewswire.com/news-releases/musixmatch-launches-real-time-music-copyright-detection-service-302727233.html

- 이 기술은 파일을 직접 대조하는 방식이 아니라, 텍스트의 고유한 패턴을 분석하여 식별 정보로 변환한 뒤 데이터베이스에 등록된 원저작물의 식별 정보와 비교함
- 기존의 해시값 비교나 파일명 대조 방식은 파일이 조금이라도 달라지면 동일한 저작물로 인식하지 못하는 반면, 핑거프린팅 방식은 가사의 일부만 인용되거나 순서가 바뀌어도 원저작물과의 유사성을 판별할 수 있음

  * 음악 핑거프린팅(music fingerprinting): 음원이나 가사의 고유한 특징을 수치화한 식별 정보를 생성하고, 이를 데이터베이스에 등록된 원저작물의 식별 정보와 대조하여 일치 여부를 판별하는 기술

• API 연동을 통한 플랫폼 적용과 콘텐츠 유형 분류

- 센티널은 API*를 통해 기존 플랫폼에 연동할 수 있어, 플랫폼이 자체 시스템을 별도로 개편하지 않고도 해당 기능을 비교적 용이하게 추가할 수 있음
- 탐지 결과는 독자적 창작물(original), 사용 허가 저작물(licensed), 보호 대상 저작물(copyrighted), 자유 이용 저작물(public domain)**의 네 가지로 분류되어, 플랫폼이 각 콘텐츠의 권리 상태에 맞는 대응 방침을 정하는 데 활용될 수 있음

  * API(Application Programming Interface): 서로 다른 소프트웨어 간에 데이터를 주고받을 수 있도록 정의된 통신 규약

  ** 자유 이용 저작물(public domain): 저작권 보호 기간이 만료되었거나 권리자가 권리를 포기하여 누구나 자유롭게 이용할 수 있는 상태의 저작물

[표1] 기존 탐지 방식과 센티널의 핑거프린팅 탐지 방식 비교

비교 항목	파일 대조 / 해시값 방식	센티널 핑거프린팅 탐지 방식
분석 대상	파일명, 해시값 등 파일 자체 정보	어구 배열, 문장 구성 등 텍스트 패턴
부분 인용 탐지	불가	밀리초 단위로 식별 가능
콘텐츠 분류	일치 또는 불일치만 판별	창작물, 허락 콘텐츠, 보호 대상, 자유 이용의 4개 분류
데이터 기반	개별 파일 정보	20만 이상 출판사, 250개 이상 언어의 가사 DB

출처: James Hanley, “Musixmatch launches ‘Sentinel’ service to detect when copyrighted music and lyrics are used in AI and user-generated content”, Music Business Worldwide, 2026.03.26., https://www.musicbusinessworldwide.com/musixmatch-launches-sentinel-service-to-detect-when-copyrighted-music-and-lyrics-are-used-in-ai-and-user-generated-content/

• 3대 음악 출판사 이용 허락 계약과 탐지 데이터 확보

- 뮤직스매치는 2025년 10월 소니 뮤직 퍼블리싱(Sony Music Publishing), 유니버설 뮤직 퍼블리싱 그룹(Universal Music Publishing Group), 워너 채플 뮤직(Warner Chappell Music) 3대 음악 출판사와 AI 이용 허락 계약을 체결함
- 이 계약을 통해 확보한 1,500만 곡 이상의 저작물 데이터는 센티널을 포함한 분석 도구 개발에 활용되는 것으로 알려짐
- 특히 이용 허락 계약을 통해 권리자의 동의 아래 탐지 데이터를 확보하는 방식은, 권리자 허가 없이 저작물을 수집해 AI에 활용하는 사례가 늘고 있는 상황에서 합법적으로 충분한 양과 질의 원본 데이터를 확보할 수 있다는 점에서 의미가 있음

뮤직스매치의 향후 전망과 잔존 과제

• 탐지 범위 확장 계획과 오탐지·법적 판단 연계 등 기술적 과제

- 뮤직스매치는 현재 가사 중심의 탐지 기능을 시작으로, 향후 가사를 넘어 보다 넓은 범위의 저작권 식별 및 권리 관리로 센티널의 적용 범위를 확장할 계획임
- 다만 음원이나 멜로디의 유사도 판별은 가사 대조보다 기술적 난이도가 높으며, 오탐지로 인해 정상 콘텐츠가 제한될 가능성도 남아 있음
- 또한 이러한 기술적 탐지 수단은 저작권 침해 여부를 법적으로 확정하는 것이 아니라 침해 가능성이 있는 콘텐츠를 사전에 걸러내는 역할을 하므로, 실효성을 갖추려면 법적 판단 체계 및 업계 합의와 연계될 필요가 있음

참고문헌

• MusixMatch, “Musixmatch launches real-time music copyright detection service”, PR Newswire, 2026.03.27., https://www.prnewswire.com/news-releases/musixmatch-launches-real-time-music-copyright-detection-service-302727233.html
• Stuart Dredge, “Musixmatch unveils Sentinel music copyright-detection service”, Music Ally, 2026.03.26., https://musically.com/2026/03/26/musixmatch-unveils-sentinel-music-copyright-detection-service/
• James Hanley, “Musixmatch launches ‘Sentinel’ service to detect when copyrighted music and lyrics are used in AI and user-generated content”, Music Business Worldwide, 2026.03.26., https://www.musicbusinessworldwide.com/musixmatch-launches-sentinel-service-to-detect-when-copyrighted-music-and-lyrics-are-used-in-ai-and-user-generated-content/

주간 기술 동향

AI 학습 데이터 제외 여부 검증 기술, PRISM

• 생성형 AI 저작권 분쟁에서, '학습하지 않았음'에 대한 입증 중요성 증가

음악 저작권 관리 회사 BMG(Bertelsmann Music Group)는 2025년 12월 인공지능 기업 앤트로픽(Anthropic)을 상대로 저작권 침해 소송을 제기했다. BMG는 앤트로픽이 저스틴 비버, 브루노 마스, 아리아나 그란데, 롤링 스톤스 등 유명 아티스트의 가사를 무단으로 수집해 자사의 챗봇 클로드(Claude) 학습에 사용했다고 주장했다. 소송에서 BMG는 앤트로픽이 웹사이트 자동 스크래핑 도구와 불법 온라인 저장소를 통해 보호받는 저작물을 복제했으며, 클로드 사용자들이 해당 저작물을 재생성할 수 있도록 방조했다고 밝혔다.

이번 소송은 생성형 AI 기업들이 대규모 언어 모델을 학습시키기 위해 인터넷상의 방대한 텍스트 데이터를 수집하는 과정에서 저작권 보호 콘텐츠가 무분별하게 포함될 수 있다는 우려를 반영한다. AI 모델은 적게는 수천만 개에서, 많게는 수십억 개의 문서를 학습 데이터로 활용하는데, 이 과정에서 저작권자의 허가 없이 보호받는 창작물이 사용되는 경우가 빈번하다. 특히 앤트로픽은 회사 설립 초기부터 이러한 관행을 지속해왔으며, BMG가 2025년 12월 발송한 중단 요구 서한에도 응답하지 않았다고 알려졌다.

현재 AI 기업들은 학습 데이터셋의 구성을 영업 기밀로 간주하며 외부 공개나 검증을 거부하는 경향이 있다. 이 때문에 권리자는 자신의 콘텐츠가 AI 학습에 사용되었는지 확인할 방법이 없으며, 법적 분쟁이 발생해도 침해 사실을 입증하기 어려운 상황이다. 기존의 멤버십 추론 공격 기술은 특정 데이터가 학습에 포함되었는지는 탐지할 수 있으나, 반대로 특정 데이터가 학습에 포함되지 않았음을 증명하는 데는 한계가 있다.

이러한 배경에서 AI 모델이 특정 데이터셋을 학습하지 않았음을 객관적으로 검증할 수 있는 기술의 필요성이 대두되고 있다. 저작권자와 규제 당국은 AI 기업에게 학습 데이터의 투명성을 요구하고 있으며, 이는 단순히 데이터 목록을 공개하는 수준을 넘어 특정 콘텐츠의 비포함 여부를 과학적으로 입증할 수 있는 검증 체계를 필요로 한다. 본 보고서에서는 순위 상관관계 분석을 활용한 비멤버십 검증 기술을 소개하며, 이 기술이 저작권 분쟁 해결과 AI 투명성 확보에 어떻게 기여할 수 있는지 분석한다. 특히 이 기술이 법적 분쟁에서 증거 자료로 활용될 수 있는 가능성과 함께, AI 산업 전반의 신뢰성 제고에 미칠 영향을 살펴본다.

[사례] 학습되지 않았음을 입증하는 'PRISM'

• 기술 개발 배경 및 현재 검증 기술의 한계

- 현재 AI 기업들은 자신들의 모델을 학습시킬 때 어떤 데이터를 사용했는지 공개하지 않으며, 이를 핵심 영업 비밀로 취급하고 있음
- 기존의 '멤버십 추론 공격'이라는 기술은 특정 데이터가 AI 학습에 사용되었는지를 알아내는 데는 효과적이지만, 반대로 '이 데이터는 학습에 사용되지 않았다'는 것을 증명하는 데는 한계가 있음
- 일반적인 저작권 침해 소송에서 권리자들은 AI 기업에게 "당신들이 우리 콘텐츠를 학습하지 않았다는 것을 증명하라"고 요구하지만, 이를 객관적으로 입증할 수 있는 표준화된 방법이 없었음
- 이 때문에 개인정보 보호나 저작권 보호를 위해 특정 데이터가 AI 학습에서 제외되었는지 확인할 수 있는 신뢰할 만한 검증 기술이 필요한 상황임

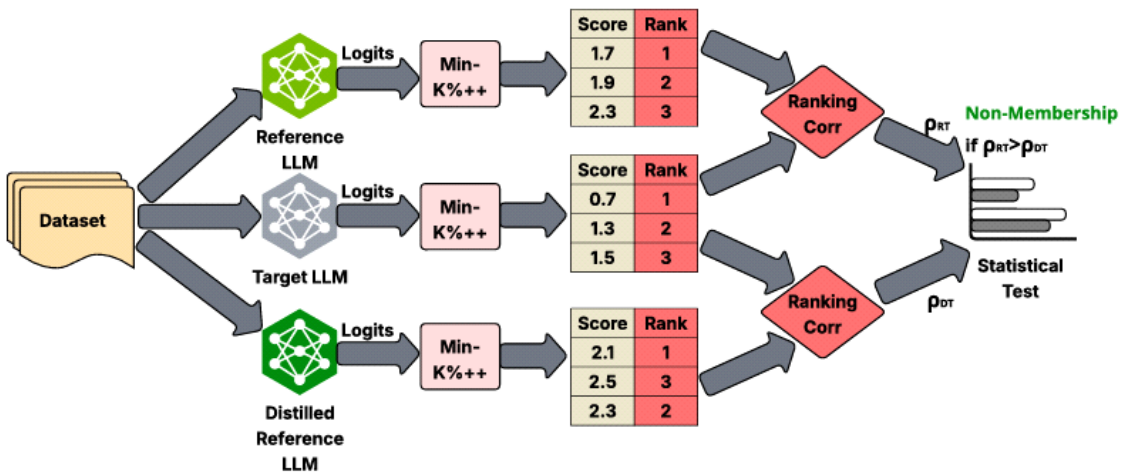
* 멤버십 추론 공격(Membership Inference Attack): 특정 데이터가 머신러닝 모델의 학습 데이터셋에 포함되어 있는지 확인하는 기술

• 핵심 기술 원리: 순위 상관관계 분석

- 확률 기반 순위 추론을 통한 멤버십 제외 데이터 검증(Probabilistic Rank-based Inference for Scrubbed Membership, 이하 PRISM)은 AI 모델이 문장을 생성할 때 각 단어 토큰에 부여하는 확률 순위를 분석하여, 해당 텍스트가 학습 데이터에 포함되었는지 판별하는 방식임
- AI는 이미 학습했던 텍스트를 다시 만나면 일관되게 같은 단어 순서를 예측하지만, 처음 보는 텍스트에 대해서는 매번 다른 예측을 내놓는 경향이 있음
- 이러한 차이를 수치화하기 위해 스피어만 순위 상관계수*라는 통계 기법을 사용하며, 학습한 데이터는 높은 점수를, 학습하지 않은 데이터는 낮은 점수를 받게 됨

* 스피어만 순위 상관계수(Spearman rank correlation): 두 변수 간의 순위 기반 단조 관계를 측정하는 비모수적 통계 기법으로 −1에서 1 사이의 값을 가짐. 0은 두 변수가 무관함을 나타내며, 1에 가까울수록 양의 상관관계, −1에 가까울수록 음의 상관관계를 가짐

[그림 1] PRISM의 작동 프로세스 개념도



출처: Pranav Shetty 외 3인, "Detecting Non-Membership in LLM Training Data via Rank Correlations", arXiv, 2026.03.24., https://arxiv.org/pdf/2603.22707

• PRISM의 검증 방식 및 작동 메커니즘

- PRISM은 검증하려는 데이터를 두 개의 AI 모델에 입력함. 하나는 검증 대상 모델이고, 다른 하나는 비교를 위한 참조 모델임. 두 모델이 각 단어를 예측할 때 부여하는 확률값 중 가장 낮은 값들의 평균을 계산하는데, 이를 'Min-K%++ 점수'라고 부름
- 두 모델에서 나온 점수들의 순위를 비교하여 상관관계를 계산함. 만약 두 모델의 예측 패턴이 크게 다르다면, 해당 데이터는 검증 대상 모델이 학습하지 않은 것으로 판정함
- 판정의 신뢰도를 높이기 위해 '부트스트래핑'이라는 통계 기법을 사용하여 95% 신뢰 구간을 설정함. 이는 100번 검증했을 때 95번은 정확한 결과를 보장한다는 의미임
- 참조 모델은 검증 대상 모델과 비슷한 구조를 가진 별도의 모델을 사용하거나, 검증 대상 모델을 간소화한 버전을 만들어 사용할 수 있음

• 실험 설계 및 성능 평가

- 연구팀은 의학 논문 데이터베이스, 학술 논문 저장소, 일반 웹 문서, 위키백과 등 8개의 서로 다른 종류의 데이터셋으로 PRISM을 실험함
- 실험 결과, AI가 학습하지 않은 데이터를 정확하게 식별하는 비율이 평균 95% 이상이었으며, 특히 전문 분야 데이터(의학, 학술)에서는 98% 이상의 정확도를 보임
- 기존의 다른 검증 기법들과 비교했을 때, 학습하지 않은 데이터를 찾아내는 정확도가 20~30% 더 높았음. 기존 기법들은 학습한 데이터를 찾는 데는 효과적이었지만, 학습하지 않은 데이터를 확인하는 데는 상대적으로 부정확했음

결론 및 시사점

• 기술적 한계 및 개선 과제

- PRISM은 대규모 데이터셋을 검증할 때 상당한 계산 비용이 발생할 수 있으며, 이는 실시간으로 학습 데이터 포함 여부를 확인해야 하는 상황에서 제약 요인으로 작용함. 특히 수백만 건 이상의 문서를 검증해야 할 경우 현실적인 시간과 비용 문제가 발생할 수 있음
- AI 모델의 구조나 학습 방식이 변화하면 토큰 예측 패턴도 달라질 수 있어, PRISM의 검증 정확도에 영향을 미칠 가능성이 존재함. 예를 들어 새로운 학습 알고리즘이나 데이터 증강 기법이 적용되면 순위 상관관계 패턴이 예상과 다르게 나타날 수 있음
- 악의적인 공격자가 의도적으로 학습 데이터를 조작하거나 PRISM의 검증을 회피하기 위한 적대적 샘플을 생성할 경우, 이에 대한 강건성이 충분히 검증되지 않았음. 향후 이러한 공격 시나리오에 대한 추가 연구가 필요함

참고문헌

- Pranav Shetty 외 3인, "Detecting Non-Membership in LLM Training Data via Rank Correlations", arXiv, 2026.03.24., https://arxiv.org/pdf/2603.22707
- Camila Curcio , "BMG Sues Anthropic Over Alleged Use of Copyrighted Lyrics in AI Training", Law Commentary, 2026.03.19., https://www.lawcommentary.com/articles/bmg-sues-anthropic-over-alleged-use-of-copyrighted-lyrics-in-ai-training