

저작권 이슈 브리프



COPYRIGHT ISSUE BRIEF

Weekly Report
2026. 8-4



한국저작권위원회
KOREA COPYRIGHT COMMISSION

본 보고서는 EC21R&C(컨설팅사)에서 작성하였고, 국내외 저작권 기술·산업 동향을 조사한 자료로 한국저작권위원회 의견이 반영되어 있지 않습니다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

산업 제미나이의 워터마크 선택제 도입으로 본 AI 생성 콘텐츠 표시 체계의 변화

▶ 생성형 AI로 제작된 이미지와 영상, 음악이 다양한 분야에서 활용되면서, 생성물에 자동으로 붙는 워터마크가 전문적이고 창의적인 작업에서 오히려 제약으로 작용한다는 지적이 이용자 사이에서 확산되어 왔다. 구글은 제미나이 이용자를 대상으로 한 의견 수렴을 거쳐 기존의 적용 방식을 다시 검토했다. 그 결과 화면에 드러나는 가시적 워터마크는 이용자가 적용 여부를 선택하되, 파일 내부에 담기는 비가시 출처 정보는 그대로 유지하는 이원적 관리 방식이 도입되었다. 이 과정에서 이용 지역과 계정 유형에 따라 적용 범위가 달라지는 차등 구조도 함께 확인된다. 이러한 변화는 표시를 완화하면서도 출처 추적 기능은 유지하려는 균형점 모색으로 해석될 수 있으며, 서로 다른 AI 기업의 대응 방식과 대조되면서 향후 생성물 표시 정책 설계 전반에 참고될 가능성이 있다.

산업 AI 음악 생성 서비스의 권리 검사와 산출물 식별 및 저작권 확인 지점의 이동

▶ AI 음악 생성 기업 수노는 특정 아티스트나 저작권이 있는 곡을 지정하는 요청문을 허용하지 않고 업로드된 음원과 가사를 검사해 온 데 더해, 2026년 8월 6일 뮤직매치의 권리 확인 서비스 센티넬을 첫 고객으로 도입했다. 센티넬은 요청문에 저작권이 있는 작곡물과 가사, 음원이 포함됐는지를 판별하며, 수노는 그 결과와 귀속 가능성을 나타내는 점수를 콘텐츠 관리와 규정 준수 여부 검토, 제품 운영 판단에 활용한다고 설명했다. 아울러 AI 산출물이 다른 스트리밍 플랫폼에 올라가 수익화되는 등의 오용에 대응하기 위해 오디오 워터마킹과 핑거프린팅을 적용하고, 대량 유통을 막기 위한 다운로드 정책도 새로 두겠다고 밝혔다. 이는 소송과 판결을 통한 사후 처리를 대체하는 것이 아니라 생성과 산출 단계에 사전 확인 절차를 추가하는 방식이지만, 식별 도구의 구현 방식이 공개되지 않았고 탐지 결과가 이용 허락이나 저작권료 지급으로 이어지는 절차도 확인되지 않는다.

산업 메타의 자체 워터마크 도입 사례로 본 AI 생성 여부 검증 체계의 파편화

▶ AI 생성 콘텐츠가 확산되면서 비가시성 워터마크와 C2PA 콘텐츠 자격증명 등 출처 검증 체계가 주요 대응 수단으로 자리 잡고 있다. 그러나 메타는 C2PA 운영위원회에 참여하면서도 자체 워터마크인 '콘텐츠 실'을 별도로 도입해 기존 표준과의 호환성 문제가 제기되었다. 콘텐츠 실은 뮤즈 이미지로 생성하거나 편집한 이미지에만 적용되며, 이용자가 직접 확인하려면 별도의 전용 탐지 도구를 사용해야 하는 등 적용 범위와 이용 경로에 제한이 있다. 이처럼 사업자별 독자 기술이 늘어날 경우 같은 콘텐츠도 탐지 도구에 따라 검증 결과가 달라지고, 이용자가 여러 도구를 직접 확인해야 하는 불편이 커질 수 있다. 이에 따라 향후에는 서로 다른 식별 기술 간 연동성을 높이고, 단계별 책임을 구체화해 일관된 검증 체계를 마련할 필요가 있다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

산업 스트리밍 콘텐츠 불법 유통 대응을 위한 AI 기반 통합 탐지 및 추적 기술

▶ 실시간 스트리밍이 확대되고 불법 이용 방식도 계정 공유와 무단 재배포 등으로 다양해지면서, 개별 보호 기술과 분리된 이용 기록만으로는 침해 여부를 신속하게 판단하기 어려운 한계가 나타나고 있다. 이에 빅블루마블은 DRM, 동시 접속 보호, 포렌식 워터마킹 등에서 수집한 이용 기록을 AI로 연계 분석해 의심 행위를 탐지하고, 유출 가입자나 시청 세션을 추적한 뒤 접근 제한 등 후속 조치까지 이어지는 방식을 제시했다. 이는 실시간 콘텐츠 보호가 개별 기술을 따로 운영하는 방식에서 벗어나 여러 이용 기록을 함께 분석하고 대응까지 연계하는 방향으로 변화하고 있음을 보여준다. 다만 정상 이용이 의심 행위로 잘못 분류될 가능성과 AI가 법적 판단을 대신할 수 없다는 한계가 있는 만큼, 판단 근거를 확인하는 절차와 후속 조치 기준을 함께 마련할 필요가 있다.

산업 발매 전 AI 학습 방해 기술의 작동 방식과 음원 관리 플랫폼 연동 사례

▶ 음원 유통 기업 심포닉은 2026년 8월 5일 AI 권리 보호 플랫폼 아티셴드와 제휴하고, 발매 이전 음원을 보호하는 도구 뮤직셴드를 자사 고객에게 제공한다고 밝혔다. 뮤직셴드는 기계가 인식하는 음향 특징을 사람이 인지하기 어려운 수준으로 변조하고, 감지하기 어려운 잡음을 음원에 삽입해 생성형 AI가 유의미한 특징을 학습하기 어렵게 만드는 방식으로 설계됐다. 아티셴드의 도구는 심포닉의 유통 관리 플랫폼 심포닉MS에서 단일 인증을 통해 이용할 수 있으며, 학습 차단과 진위 탐지, 창작 증명 등으로 구성된다. 본 사례는 침해를 사후에 확인해 대응하는 대신 개입 시점을 발매 이전으로 앞당기려는 접근에 해당하나, 실제 방어 성능과 적용 범위는 아직 확인되지 않은 상태다.

기술 주간 기술 동향

▶ 딥러닝 모델 학습에 저작권 이미지가 무단으로 활용되면서, 비인가 학습을 막는 기술과 소유권을 입증하는 워터마킹 기술이 각각 발전해왔다. 그러나 두 기술은 모두 픽셀을 미세하게 조정하는 방식이어서, 함께 적용하면 서로 간섭해 워터마크 판독 오류가 늘어나는 문제가 있었다. 복원 가능한 학습 불가 데이터는 이미지와 모델 출력의 연관성을 끊는 방식으로 변형을 만들고, 변형이 있는 상태와 없는 상태를 각각 담당하는 두 개의 디코더를 두어 이 문제를 해소한다. 인가 사용자는 전달받은 변형 정보를 제거해 정상적으로 모델을 학습시킬 수 있으며, 유출된 데이터에 대해서도 워터마크 확인이 가능하다. 실험에서는 비인가 학습 차단과 낮은 판독 오류율이 함께 확인된다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

제미나이의 워터마크 선택제 도입으로 본 AI 생성 콘텐츠 표시 체계의 변화

가시적 워터마크에 대한 이용자 불편 확산과 정책 재검토 요구

• 워터마크가 전문적 창작 활용에서 유발해 온 제약

- 구글(Google)의 생성형 AI 서비스 제미나이(Gemini)는 이미지와 영상, 음악 생성물의 화면 우측 하단에 반짝임 형태의 워터마크(watermark)를 자동으로 삽입해 왔으며, 이 표시는 파일을 열어보지 않고도 AI 생성물임을 곧바로 알아볼 수 있게 하는 신호로 기능해 옴
- 이처럼 화면에 드러나는 가시적 워터마크는 일반적인 이용 환경에서는 간편한 식별 수단으로 작동했으나, 발표 자료나 마케팅 자산처럼 완결된 형태가 요구되는 작업에서는 워터마크가 화면 일부를 차지해 추가 편집을 요구하는 요인으로 지목되기도 함
- 이러한 가운데 AI 생성물의 워터마크를 제거하려는 이용자 시도는 지속적으로 확산되고 있으며, AI 생성물의 식별 기능과 창작 활용도를 함께 확보하는 표시 방식에 대한 검토 필요성이 제기됨

• 이용자 의견 수렴을 계기로 한 정책 재검토 확산

- 한편, 구글의 부사장 조시 우드워드(Josh Woodward)는 본인의 SNS를 통해 제미나이 이용 과정의 불편 사항을 묻는 조사를 실시함. 이에 수천 명의 이용자가 이미지 생성 모델인 나노 바나나(Nano Banana)의 워터마크 제거를 요구한 것으로 확인됨¹⁾
- 이용자 의견을 반영한 구글 제미나이의 기능 조정은 워터마크 선택 설정의 도입으로 이어졌으며, 이후 순차 적용을 거치면서 화면 표시 방식이 기본 적용에서 이용자 선택 적용으로 변경된 것을 확인함

가시 표시 선택제와 비가시 출처 정보의 이원 관리 구조 도입

• 워터마크 선택제의 적용 범위와 설정 방식

- 구글은 제미나이 웹과 모바일 환경, 그리고 구글의 AI 기반 영상 편집 도구인 플로우(Flow)의 설정 화면에 미디어 워터마크 항목을 신설했으며, 이용자가 이 항목에서 화면 표시의 적용 여부를 직접 지정하는 방식으로 구성됨
- 선택 설정은 이미지 생성 모델인 나노 바나나, 영상 생성 모델인 옴니(Omni) 및 음악 생성 모델인 리리아(Lyria)에 각각 적용되어 이미지와 영상, 음악이라는 세 가지 생성물 유형 전반을 포괄하며, 특정 유형에만 표시를 남기는 부분 적용은 제공되지 않음

1) Manisha Priyadarshini, "Google's Gemini app adds a toggle to disable AI watermarks, with some exceptions", DigitalTrends, 2026.08.14., <https://www.digitaltrends.com/computing/googles-gemini-app-adds-a-toggle-to-disable-ai-watermarks-with-some-exceptions/>

- 설정 결과는 이후 생성되는 모든 생성물에 일괄 반영되나 이미 제작된 생성물의 표시를 사후에 제거하는 기능은 포함되지 않으며, 구글 검색의 생성 기능에는 아직 적용되지 않아 추후 확대가 예정된 상태임
- 다만 이 설정은 모든 이용자에게 동일하게 제공되지 않으며, 한국, 인도 및 베트남에서는 구글의 유료 요금제인 AI 울트라(AI Ultra) 구독자만 이용할 수 있고 업무용 계정과 학교 계정에는 설정 자체가 제공되지 않는 것으로 확인됨

• 비가시적 출처 정보 유지를 통한 추적 가능성 확보

- 화면 표시를 끄더라도 구글의 비가시적 워터마크 기술인 신스ID(SynthID)와, C2PA*(Coalition for Content Provenance and Authenticity)가 정의한 콘텐츠 자격증명(Content Credentials)은 제미니에서 생성되거나 편집된 지원 대상 미디어에 그대로 유지됨
- 신스ID는 생성물 데이터 자체에 식별 정보를 새기고 콘텐츠 자격증명은 파일에 붙는 서명된 기록으로 생성 도구와 편집 이력을 남기는 방식으로, 두 기술은 서로 다른 방식으로 출처 확인 기능을 담당함
- 이에 따라 화면 표시가 없는 생성물이라도 제미니나 구글 검색을 통해 지원되는 출처 신호의 탐지 여부를 확인할 수 있으며, 가시적 워터마크의 부재가 곧 식별 불가로 이어지지 않는 검증 체계가 갖춰짐
- 이번 조치는 표시 자체를 없앤 것이 아니라 사람이 눈으로 읽는 층위와 기계가 판독하는 층위를 분리해, 전자는 이용자 선택에 맡기고 후자는 예외 없이 유지하는 이원적 관리 방식이 형성됨

* C2PA(Coalition for Content Provenance and Authenticity): 콘텐츠의 출처와 편집 이력을 증명하기 위해 결성된 산업 연합체

시사점: AI 생성물 표시 체계 설계의 균형점 모색

• 서로 다른 AI 기업의 표시 방식 대조와 이원적 대응 확산 가능성

- 텍스트 생성물 영역에서는 앤트로픽(Anthropic)이 유럽연합의 AI 투명성 규정에 대응하기 위해 자사의 생성형 AI 서비스 클로드(Claude)의 텍스트와 파일 생성물에 워터마크를 도입해, 화면 표시를 선택 사항으로 돌린 구글과 상반된 방향의 선택이 같은 시기에 나타남
- 가시 표시와 비가시 출처 정보를 분리해 운영하는 방식은 이용자 편의와 식별 기능을 함께 확보하는 구조로 평가될 수 있어, 생성물 표시 정책을 설계하는 다른 사업자에게도 참고될 여지가 있음

• 가시 표시 완화와 출처 검증 강화 병행이 시사하는 향후 과제

- 화면 표시가 선택 사항이 된 이후 이용자는 생성물의 겉모습만으로 AI 생성물 여부를 판단하기 어려워졌으며, 파일에 담긴 출처 정보를 별도 도구로 읽어내는 절차를 거쳐야 함
- 구글은 워터마크 선택제 발표와 함께 콘텐츠 자격증명을 검증하는 오픈소스 라이브러리 크레덴티오(Credentio)를 공개했으며, 개발자가 파일을 외부 서버로 보내지 않고 자체 애플리케이션 안에서 매니페스트(manifest)*와 디지털 서명을 분석해 출처 기록을 확인할 수 있게 함
- 다만 이러한 검증 기능은 개발자가 자체 서비스에 적용해야 작동하므로, 도구 보급이 기술 도입과 나란히 진행될 때 식별 기능이 유지될 수 있다는 과제가 남음

* 매니페스트(manifest): 콘텐츠 자격증명을 구성하는 데이터 단위로, 생성 도구와 편집 이력 등의 출처 정보가 담기며 디지털 서명으로 변경 여부를 확인할 수 있음

참고문헌

- Ivan Mehta, "Google will now allow users to remove visible watermark from its AI generations", TechCrunch, 2026.08.14., <https://techcrunch.com/2026/08/14/google-will-now-allow-users-to-remove-visible-watermark-from-its-ai-generations/>
- Manisha Priyadarshini, "Google's Gemini app adds a toggle to disable AI watermarks, with some exceptions", DigitalTrends, 2026.08.14., <https://www.digitaltrends.com/computing/googles-gemini-app-adds-a-toggle-to-disable-ai-watermarks-with-some-exceptions/>
- Sherif Hanna, "Introducing Credentio: Open Source C++ Library for C2PA Content Credentials from Google", Google Developers Blog, 2026.08.13., <https://developers.googleblog.com/introducing-credentio-open-source-c-library-for-c2pa-content-credentials-from-google/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

AI 음악 생성 서비스의 권리 검사와 산출물 식별 및 저작권 확인 지점의 이동

음악 분야 AI 산출물의 권리 확인 공백과 사후 대응의 한계

• 소송과 라이선스로 처리되어 온 학습 단계의 권리 문제

- 2026년 7월 31일 독일 뮌헨지방법원(Munich Regional Court)은 독일 음악저작권 관리단체 게마(Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte, GEMA)가 관리하는 음악 저작물을 AI 음악 생성 기업 수노(Suno)가 라이선스 없이 학습해 저작권을 침해했다고 판단했으며¹⁾, 이는 AI 음악 학습과 관련해 저작권 침해 책임을 인정한 유럽의 주요 판결 중 하나로 확인됨
- 수노는 해당 판결이 자사 기술의 작동 방식을 잘못 파악했다며 항소를 검토 중이라고 밝혔고, 유니버설 뮤직 그룹(Universal Music Group) 및 소니 뮤직 엔터테인먼트(Sony Music Entertainment)와의 소송에서는 AI 학습이 공정이용(Fair Use)*에 해당한다는 입장을 유지함¹⁾
- 반면 워너 뮤직 그룹(Warner Music Group)은 2025년 11월 수노와 합의하고 라이선스 제휴를 체결해, 수노를 둘러싼 저작권 문제가 한편에서는 소송으로 이어지고, 다른 한편에서는 합의와 라이선스 방식으로 처리되는 서로 다른 양상이 나타나는 등 쟁점이 계속되고 있음

* 공정이용(Fair Use): 저작권자의 허락 없이 저작물을 사용하더라도, 일정한 조건을 충족하면 저작권 침해로 보지 않는 제도

• AI 산출물의 유통 이후의 생성 출처 및 권리 확인 공백

- 수노를 둘러싼 저작권 분쟁에서는 이용자가 AI 산출물을 다른 스트리밍 플랫폼에 올려 수익을 얻는 방식이 주요 쟁점 중 하나로 지목됐으며, 이는 산출물이 AI 생성 서비스 밖으로 유통된 이후에는 생성 출처를 추적 및 확인하기 어려워지는 문제로 해석됨
- 이탈리아 음악 데이터 기업 뮤직매치(Musixmatch)의 공동대표 리오 카라에프(Rio Caraeff)는 어떤 곡이 만들어지고 저작권이 있는 자료가 입력되는지를 플랫폼이 파악할 수 있게 하는 기반이 필요하다고 평가하며²⁾, 생성 과정에서 이를 확인할 수 있는 인프라의 필요성을 강조함
- 이러한 가운데 AI 생성 콘텐츠에 표시를 요구하는 유럽연합 인공지능법(EU AI Act)*의 관련 규정이 2026년 8월 2일부터 일반 적용되기 시작해, AI 생성 사실의 표시가 제도 차원에서도 요구되는 상황임

* 유럽연합 인공지능법(AI Act): 유럽연합이 AI의 개발과 이용에 적용하는 규정으로, AI로 만든 콘텐츠에 대한 표시 요건을 포함함

1) Tim Ingham, "Suno becomes first customer of Musixmatch's Sentinel, which screens AI prompts and outputs for copyrighted material", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/suno-first-customer-of-musixmatch-s-sentinel-which-screens-ai-prompts-and-outputs-for-copyrighted-material/>

2) Andre Paine, "Suno shares principles for 'responsible AI' as it adopts Musixmatch's copyright detection service", Music Week, 2026.08.06., <https://www.musicweek.com/digital/read/suno-shares-principles-for-responsible-ai-as-it-adopts-musixmatch-s-copyright-detection-service/094675>

입력 검사와 산출물 식별을 함께 두는 구조의 형성

• 요청문과 업로드 자료를 검사하는 입력 단계

- 수노는 자사 서비스에서 특정 아티스트나 저작권이 있는 곡을 지정하는 요청문을 허용하지 않으며, 요청문에 아티스트 이름이 포함되면 해당 이름을 제거하고 음악적 특성을 서술하는 방향으로 바꿔 처리한다고 설명함³⁾
- 또한 업로드된 음원 파일과 가사에 타인의 저작물이 무단 사용됐는지 확인하기 위해 오디오 매직 (Audible Magic), ACR클라우드(ACRCloud) 등 음원 인식 기업과 뮤직매치의 기술을 활용한다고 밝힘³⁾
- 이와 함께 커뮤니티 가이드라인을 개정해 실제 녹음으로 오인될 수 있는 음원과 실존 인물의 음성 및 유사성(likeness)을 허락 없이 사용하는 행위를 명시적으로 금지함⁴⁾

• 외부 권리 정보를 활용한 탐지 계층의 결합

- 뮤직매치의 권리 확인 서비스 센티넬(Sentinel)은 생성형 미디어와 이용자 생성 콘텐츠(User-Generated Content, UGC)* 플랫폼을 대상으로 3월 27일 출시된 서비스로, 2026년 8월 6일 수노를 첫 고객으로 확보함
- 제품 설명상 센티넬은 AI 서비스 이용자의 요청문과 그에 따른 산출물을 함께 살펴 생성 과정의 어느 단계에서든 저작권이 있는 가사를 찾아내며, 가사 대조로 이용자 생성 콘텐츠를 권리자의 저작물과 연결해 저작권료 귀속에 활용할 수 있음³⁾
- 수노는 센티넬을 자사 서비스에 결합해 사용자의 요청문에 저작권이 있는 작곡물, 가사 및 음원이 포함됐는지를 판별하고, 그 결과와 함께 저작물과의 일치 및 귀속 가능성을 나타내는 점수를 받아 콘텐츠 관리와 규정 준수 여부 검토, 제품 운영 관련 판단에 활용한다고 설명함

* 이용자 생성 콘텐츠(User-Generated Content, UGC): 기업이나 플랫폼이 아닌 이용자가 직접 제작해 게시한 콘텐츠. 영상, 이미지, 글, 음악 등이 이에 해당함

• 생성 출처를 표시하고 외부 유통을 제한하는 출력 단계

- 수노는 AI 산출물이 다른 스트리밍 플랫폼에 올라가 수익화되는 등의 오용에 대응하기 위해 오디오 워터마킹(audio watermarking)*과 핑거프린팅(fingerprinting)**을 적용하겠다고 밝혔으며, 이들 식별 도구는 변조에는 견디면서도 이용자의 청취 경험에는 영향을 주지 않도록 설계됐다고 설명함⁴⁾
- 다만 식별 도구를 어떤 기술로 구현할지는 공개하지 않았으며, 워터마킹의 경우 구글(Google)의 비가시적 워터마크 기술인 신스ID(SynthID)와 같은 기존 기술을 사용할지 여부도 밝히지 않음
- 이와 함께 스트리밍 플랫폼으로의 대량 유통을 막기 위한 다운로드 정책을 새로 두겠다고 밝혔으나, 구체적인 제한 조건 등 세부 사항은 공개되지 않음

* 오디오 워터마킹(audio watermarking): 음원에 사람이 듣기 어려운 신호를 넣어 이후에도 그 음원의 출처를 확인할 수 있게 하는 기술

** 핑거프린팅(fingerprinting): 음원의 음악적 특징을 추출해 대조용 정보로 만든 뒤, 같거나 비슷한 음원을 찾아내는 기술

[표 1] 수노의 단계별 권리 확인 및 식별 수단

단계	적용 수단	확인 및 식별 대상	성격
생성 요청문	아티스트명 제거, 센티넬	기존 저작물의 포함 여부	사전 차단 및 탐지
업로드 음원 및 가사	오디블 매직, ACR클라우드, 뮤직매치	이용자가 올린 자료	무단 이용 검사
생성 산출물	센티넬, 워터마킹, 핑거프린팅	저작물 포함 여부, 생성 출처	탐지 및 출처 표시
다운로드 및 외부 유통	다운로드 정책, 워터마킹, 핑거프린팅	대량 유통, 외부 플랫폼 오용	유통 제한 및 오용 대응

출처: 참고문헌 종합하여 재구성

3) Tim Ingham, "Suno becomes first customer of Musixmatch's Sentinel, which screens AI prompts and outputs for copyrighted material", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/suno-first-customer-of-musixmatchs-sentinel-which-screens-ai-prompts-and-outputs-for-copyrighted-material/>

4) Ivan Mehta, "Amid legal battles, Suno says it will start watermarking songs", TechCrunch, 2026.08.06., <https://techcrunch.com/2026/08/06/amid-legal-battles-suno-says-it-will-start-watermarking-songs/>

시사점: 생성 서비스의 1차 권리 확인 역할 확대와 남은 과제

• 생성 단계의 1차 권리 확인이 추가되는 의미

- 수노의 이번 조치는 소송과 판결을 통한 사후 처리를 대체하는 것이 아니라, AI를 통해 생성, 산출 단계에 사전 확인 절차를 추가하는 방식으로 이해할 수 있음
- 뮤직매치와 같이 음악 권리 정보를 축적한 사업자가 판별 기능을 공급하고 AI 기업이 이를 자사 서비스에 결합하는 분업 형태가 나타나면서, 권리 확인 기능을 외부 사업자로부터 도입해 서비스에 적용하는 방식도 가능해지고 있음
- 이러한 방식의 적용 가능성은 음악에만 국한될 것으로 보기 어려우며, 이용자가 자료를 입력해 결과물을 만들어내는 다른 AI 생성 서비스에서도 유사한 확인 절차가 요구될 여지가 있음

• 자율 조치의 검증과 권리 처리 연계에 남은 과제

- 다만 수노가 AI 산출물 식별 도구의 구체적인 구현 방식을 공개하지 않은 상태에서는 공개된 정보만으로 제3자가 식별 방식과 검출 가능성을 검증하기 어려워, 자율 조치의 실효성을 외부에서 평가하는 데 한계가 있음
- 또한 식별 도구를 통해 확인되는 정보를 실제로 어떻게 공개할지는 아티스트와 플랫폼이 결정할 문제라는 수노의 입장을 고려하면, AI 산출물에 식별 정보를 넣는 것과 그 정보를 이용자에게 표시하는 것은 별개의 단계로 남음
- 마찬가지로 센티넬이 권리자 저작물과의 매칭 및 저작권료 귀속을 지원하더라도, 탐지 결과가 실제 이용 허락이나 저작권료 지급으로 이어지는 운영 절차는 공개 자료에서 확인되지 않으므로 확인 단계와 권리 처리 단계를 잇는 설계가 과제로 남음

참고문헌

- Ivan Mehta, "Amid legal battles, Suno says it will start watermarking songs", TechCrunch, 2026.08.06., <https://techcrunch.com/2026/08/06/amid-legal-battles-suno-says-it-will-start-watermarking-songs/>
- Andre Paine, "Suno shares principles for 'responsible AI' as it adopts Musixmatch's copyright detection service", Music Week, 2026.08.06., <https://www.musicweek.com/digital/read/suno-shares-principles-for-responsible-ai-as-it-adopts-musixmatch-s-copyright-detection-service/094675>
- Tim Ingham, "Suno becomes first customer of Musixmatch's Sentinel, which screens AI prompts and outputs for copyrighted material", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/suno-first-customer-of-musixmatchs-sentinel-which-screens-ai-prompts-and-outputs-for-copyrighted-material/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

메타의 자체 워터마크 도입 사례로 본 AI 생성 여부 검증 체계의 파편화

공통 표준의 정착과 메타의 자체 워터마크 도입

• AI 생성 여부 검증을 위한 공통 표준의 정착

- AI 생성 이미지가 온라인에서 대량으로 유통되면서, 콘텐츠 생성 단계에 이미지에 삽입되는 비가시성 워터마크(invisible watermark)가 AI 생성 여부를 검증하는 주요 수단으로 자리 잡고 있음
- 이러한 흐름 속에서 구글(Google LLC)과 메타(Meta Platforms, Inc.) 등 주요 사업자가 참여하는 산업 연합체 C2PA*의 '콘텐츠 자격증명(Content Credentials)'은 콘텐츠의 출처와 편집 이력을 증명하는 공통 표준으로 통용되고 있으며, 공개 검증 포털을 통해 개방형으로 운영됨
- 아울러 구글이 개발한 워터마크 신스ID(SynthID)는 자사 서비스의 탐지 기능에 그치지 않고 경쟁사인 오픈AI(OpenAI)에도 채택되는 등, 사업자 간 공통 기술의 활용이 확대되며 워터마크 방식도 점차 수렴되는 모습이 관찰됨

* C2PA(Coalition for Content Provenance and Authenticity): 콘텐츠의 출처와 편집 이력을 증명하기 위해 결성된 산업 연합체

• 공통 표준 참여와 별개로 추진된 메타의 자체 워터마크 도입

- 이처럼 공통 표준이 확산되는 가운데, 메타의 콘텐츠 정책 심의 기구인 감독위원회(Oversight Board)는 2026년 3월, 오해를 유발할 수 있는 AI 생성 콘텐츠에 대응하기 위해 메타에 자체 탐지 도구를 적극적으로 활용할 것을 권고함¹⁾
- 메타는 이 권고에 따라 2026년 7월 이미지 생성 모델 '뮤즈 이미지(Muse Image)'를 출시하고, 해당 모델로 생성하거나 편집한 이미지에 자동 삽입되는 비가시성 워터마크 '콘텐츠 실(Content Seal)'과 그 삽입 여부를 확인할 수 있는 전용 탐지 도구를 함께 공개함²⁾
- 다만 메타는 C2PA 운영위원회에 참여하고 있음에도 콘텐츠 실은 C2PA와 연동되지 않는 독자 기술로 설계됨에 따라, 공개 직후부터 기존 표준과의 호환성이 주요 쟁점으로 제기됨

1) Jess Weatherbed, "Meta made its own AI detection system. It should have just used Google's", The Verge, 2026.07.22., <https://www.theverge.com/tech/968680/meta-ai-detection-labeling-content-seal-watermarks-synthid>

2) Karissa Bell, "Meta built an AI detection tool to ID images and video created with its new models", Engadget, 2026.07.07., <https://www.engadget.com/2210223/meta-built-an-ai-detection-tool-to-id-images-and-video-created-with-its-new-models/>

자체 워터마크 운영에 따라 파편화된 사업자별 검증 체계

• 표준 간 비호환으로 발생하는 검증의 공백

- 미국 IT 매체 더 버지(The Verge)가 뮤즈 이미지로 생성한 이미지를 구글의 AI 챗봇 제미니(Gemini)와 C2PA 공개 검증 포털에서 직접 테스트한 결과, 두 도구 모두 이를 AI 생성물로 탐지하지 못하는 것으로 드러남³⁾
- 또한 하나의 파일에 서로 다른 식별 신호를 동시에 삽입할 경우 상호 간섭이 발생하는지에 대한 질문에 메타가 공식 설명을 내놓지 않아, 복수의 식별 신호를 병용할 수 있는지 확인되지 않은 상태임
- 아울러 외부 플랫폼과의 연동에서도 메타는 AI 라벨을 운영하는 틱톡(TikTok), 링크드인(LinkedIn) 등과 협력하겠다는 의사만 밝힌 상태로, 콘텐츠가 메타 외부로 유통될 경우 식별 신호가 외부 플랫폼의 AI 라벨 표시로 이어지지 않을 수 있다는 우려가 제기됨

[표 1] 주요 AI 생성 여부 검증 수단 비교

구분	콘텐츠 실	신스ID	C2PA 콘텐츠 자격증명
운영 주체	메타 자체 기술	구글 개발, 타 사업자 채택	산업 연합 표준
적용 범위	뮤즈 이미지 생성 및 편집 이미지	구글, 오픈AI 생성 콘텐츠	표준 채택 기업의 콘텐츠
탐지 경로	전용 탐지 도구	제미니 내 통합	공개 검증 포털
확인 횟수 제한	일일 확인 횟수 제한 적용	이용 한도 적용	제한 없음

출처: Jess Weatherbed, "Meta made its own AI detection system. It should have just used Google's", The Verge, 2026.07.22., <https://www.theverge.com/tech/968680/meta-ai-detection-labeling-content-seal-watermarks-synthid>

• 콘텐츠 실의 제한된 적용 범위와 확인 경로

- 콘텐츠 실은 기존 표준과 호환되지 않는 데 이어, 자체 적용 범위 역시 뮤즈 이미지로 생성하거나 편집한 이미지에 한정됨. 이에 따라 메타가 2023년부터 제공해 온 기존 모델이 생성한 이미지와 영상 형식의 AI 생성 콘텐츠는 아직 적용 대상에 포함되지 않음
- 적용 대상뿐 아니라 이용자의 확인 경로도 제한적임. 콘텐츠 실의 식별 신호는 메타의 소셜 미디어 플랫폼에서 AI 생성 표시에 활용되지만, 이용자가 개별 이미지의 AI 생성 여부를 직접 확인하려면 해당 플랫폼 내에서가 아니라 별도의 전용 탐지 도구를 이용해야 함
- 전용 탐지 도구에는 일일 확인 횟수 제한도 적용되어 이용자가 다수의 콘텐츠를 대량으로 확인하기는 어려움. 이는 확인 횟수에 제한을 두지 않는 C2PA 공개 검증 포털과 대비되는 지점임

• 자사 제품 중심의 설계 및 운영과 정립되지 않은 검증 방향

- 콘텐츠 실은 메타의 자체 기술 사양과 제품에 맞춰 구축되어 뮤즈 이미지로 생성하거나 편집한 이미지에만 적용됨. 메타는 이러한 설계가 다른 사업자와 마찬가지로 자사 제품에 특화하여 기술을 설계한 결과라고 설명함
- 이러한 자사 제품 중심의 접근은 운영 방식에서도 나타남. 메타는 소셜 미디어의 AI 생성 표시에 콘텐츠 실과 함께 이와 별도인 메타데이터(metadata)도 활용한다고 밝혔으나, 해당 메타데이터의 구체적인 내용과 활용 방법은 공개하지 않음

³⁾ Jess Weatherbed, "Meta made its own AI detection system. It should have just used Google's", The Verge, 2026.07.22., <https://www.theverge.com/tech/968680/meta-ai-detection-labeling-content-seal-watermarks-synthid>

- 메타 산하 소셜 미디어 플랫폼 인스타그램(Instagram) 총괄 책임자 애덤 모세리(Adam Mosseri)는 AI 생성 콘텐츠를 걸러내는 데에는 반대하면서도 AI 생성 표시의 필요성에는 동의하는 등, 메타 내부에서도 대응 방향이 하나로 정리되지 않은 상태임³⁾

시사점: 공통 표준으로의 수렴과 검증 책임 논의

• 검증 체계 파편화에 따른 판단 공백

- 사업자별 식별 기술이 호환되지 않는 경우 탐지 도구마다 인식할 수 있는 식별 신호가 달라져, 특정 도구의 미탐지 결과만으로는 AI 생성 여부를 판단하기 어려움. 실제로 메타의 탐지 도구는 '미탐지는 메타 AI로 처리되지 않았을 가능성이 높음'을 의미할 뿐이라고 안내함
- 이러한 한계는 이용자가 미탐지 결과를 실제 촬영물이라는 근거로 받아들일 경우 오인으로 이어질 수 있음. 사업자별 자체 워터마크가 늘어날수록 이용자는 여러 도구로 콘텐츠를 확인해야 하며, 어떤 도구로도 출처를 확인할 수 없는 영역이 확대될 가능성도 있음
- 이번 사례는 개별 사업자가 자체 식별 기술을 제공하는 것만으로는 이러한 공백을 해소하기 어렵다는 점과, 서로 다른 표준의 연동 및 공통 식별 기준 마련을 함께 추진할 필요가 있다는 점을 보여줌

• 이용자 점점 확인 기능과 표준 연동, 단계별 책임 구체화

- 메타는 이용자가 AI 생성 콘텐츠를 접하는 지점에서 바로 탐지 결과를 확인할 수 있도록 하는 방안을 검토 중이라고 밝혔으며, 업계와의 협력 의사와 콘텐츠 실 관련 추가 발표 계획도 함께 언급함
- 각 사업자의 식별 신호를 타 사업자의 탐지 도구에서도 인식할 수 있도록 표준을 연동하는 방안과, 탐지 결과를 사업자 간에 공유해 일관되게 표시하는 방안이 주요 과제로 논의될 전망이다
- 개별 사업자의 기술 공개만으로는 일관된 검증 체계를 구축하기 어려운 만큼, 저작권 보호와 콘텐츠 신뢰 확보를 위해 생성, 유통, 검증 단계별 책임을 구체화하는 작업도 필요함

참고문헌

- Jess Weatherbed, "Meta made its own AI detection system. It should have just used Google's", The Verge, 2026.07.22., <https://www.theverge.com/tech/968680/meta-ai-detection-labeling-content-seal-watermarks-synthid>
- Karissa Bell, "Meta built an AI detection tool to ID images and video created with its new models", Engadget, 2026.07.07., <https://www.engadget.com/2210223/meta-built-an-ai-detection-tool-to-id-images-and-video-created-with-its-new-models/>

저작권 이슈 브리프

SUMMARY

산업/기업

기술

스트리밍 콘텐츠 불법 유통 대응을 위한 AI 기반 통합 탐지 및 추적 기술

실시간 스트리밍 확산에 따른 기존 불법 유통 탐지 방식의 한계

• 실시간 콘텐츠 유통 확대와 불법 이용 방식의 다양화

- 대형 스포츠 경기와 음악 행사 등 시청 수요가 높은 콘텐츠가 클라우드를 통해 실시간으로 제공되는 사례가 늘면서, 방송 및 미디어 업계에서 콘텐츠 보호의 중요성이 커지고 있음
- 이와 함께 불법 이용 방식이 단순한 파일 복제를 넘어 계정 공유와 무단 재배포 등으로 다양해지면서, 콘텐츠 보호 과정에서 확인해야 할 검증 대상도 늘어나고 있음

• 보호 기술별로 분리 관리되는 이용 기록

- 그간 스트리밍 콘텐츠 보호에는 디지털 권한 관리(Digital Rights Management, 이하 DRM)*, 동시 접속 보호, 포렌식 워터마킹(forensic watermarking)** 등 여러 기술이 각각 적용되어 왔으며, 각 기술에서 수집되는 이용 기록도 서로 연계되지 않은 채 별도로 관리되는 경우가 많았음
- 그러나 하나의 불법 이용 행위와 관련된 이용 기록이 여러 시스템에 나뉘어 저장되면, 어느 한 시스템의 기록만으로는 정상 이용과 부정 이용을 구분하기 어려울 수 있음
- 이 경우 운영자가 여러 시스템의 이용 기록을 직접 대조해 침해 여부를 확인해야 하며, 이러한 방식은 중계 도중에도 신속한 대응이 필요한 실시간 콘텐츠의 경우 조치가 늦어지는 요인이 될 수 있음

* 디지털 권한 관리(Digital Rights Management): 디지털 콘텐츠나 문서의 무단 복사 및 배포를 막고 저작권이나 보안 권한을 통제하는 기술

** 포렌식 워터마킹(forensic watermarking): 콘텐츠에 육안으로 식별되지 않는 정보를 삽입해 유출 발생 시 해당 콘텐츠의 출처를 추적할 수 있도록 하는 기술

여러 보호 기술의 이용 기록을 연계한 탐지 및 추적 체계

• 여러 보호 기술에서 수집한 이용 기록을 AI로 함께 분석

- 오스트리아 미디어 기술 기업 빅 블루 마블(Big Blue Marble)*은 2026년 9월 국제 방송 및 미디어 기술 전시회인 IBC2026**에서 자사 보호 솔루션인 '콘텐츠 보호 스위트(Content Protection Suite)'에 AI 기반 부정 이용 탐지 기능을 추가해 공개할 예정이라고 밝힘¹⁾
- 빅 블루 마블에 따르면 이 기능은 DRM, 동시 접속 보호, 포렌식 워터마킹 등 여러 보호 기술에서 수집한 이용 기록을 함께 분석해, 개별 시스템만으로는 파악하기 어려운 불법 유통 가능성이 있는 패턴과 비정상적 계정 공유를 탐지하도록 설계됨

1) Big Blue Marble, "Big Blue Marble to unveil AI-assisted piracy detection and broadcast-grade streaming innovations at IBC2026", 2026.08.06., <https://www.bigbluemarble.com/en/resources/post/bbm-ibc2026/>

- 특히 AI가 여러 시스템에 나뉘어 있는 이용 기록을 서로 연결하고 의심 행위로 판단한 근거도 함께 제시해, 운영자가 이를 직접 대조하던 기존 방식에 비해 조사에 걸리는 시간을 줄일 수 있다는 것이 회사 측 설명임

* 빅 블루 마블(Big Blue Marble): 오스트리아에 기반을 둔 미디어 기술 기업으로, 방송 및 스트리밍 플랫폼 구축과 콘텐츠 보호 서비스를 제공하며 포렌식 워터마킹, 불법 재유통 탐지 등 영상 콘텐츠 보안 기술을 개발함

** IBC2026(International Broadcasting Convention 2026): 매년 네덜란드 암스테르담에서 열리는 국제 방송 및 미디어 기술 전시회로, 9월 11일부터 14일까지 개최될 예정임

[표 1] 스트리밍 콘텐츠 보호 기술별 역할과 탐지 범위 및 특징

구분	DRM	동시 접속 보호	포렌식 워터마킹	AI 통합 분석
주요 역할	콘텐츠 접근 및 복제 제어	계정당 세션 수 제한	콘텐츠에 식별 정보 삽입	여러 보호 기술의 이용 기록 종합 분석
탐지 및 추적 대상	비인가 접근	계정 공유	유출 출처 및 경로	여러 유형이 결합된 불법 이용
특징	콘텐츠 접근 단계에서 통제	계정 단위 이용 통제	유출 발생 후 출처 추적	서로 다른 시스템의 이용 기록을 연계해 분석

출처: Big Blue Marble, "Big Blue Marble to unveil AI-assisted piracy detection and broadcast-grade streaming innovations at IBC2026", 2026.08.06., <https://www.bigbluemarble.com/en/resources/post/bbm-ibc2026/>

• 포렌식 워터마킹을 활용한 유출 콘텐츠 출처 추적

- 빅 블루 마블에 따르면, AI 기반 부정 이용 탐지 기능과 함께 공개되는 AI 기반 포렌식 워터마킹 처리 절차는 불법 유통이 의심되는 스트림의 녹화본을 분석해, 해당 콘텐츠가 어느 가입자나 시청 세션에서 유출되었는지 추적할 수 있도록 설계됨
- 포렌식 워터마킹은 콘텐츠 자체에 삽입된 식별 정보를 이용하므로, 파일명이나 유통 경로가 바뀐 이후에도 처음 유출된 지점을 찾는 단서가 될 수 있음

• 탐지, 조사, 대응을 연계한 통합 처리 절차

- 빅 블루 마블은 다중 DRM(multi-DRM), 동시 접속 보호, 포렌식 워터마킹에 AI 기반 부정 이용 탐지 기능을 결합해, 의심 행위 탐지부터 조사와 후속 조치까지가 하나의 처리 절차 안에서 이어지도록 함
- 회사 측은 침해 가능성을 확인한 뒤 별도의 대응 절차를 거치던 기존 방식과 달리, 이를 통해 탐지 결과를 콘텐츠 접근 제한이나 전송 관리 등 후속 조치에 더 빠르게 반영할 수 있다고 설명함

시사점: 실시간 콘텐츠 보호 체계의 변화와 과제

• 탐지와 후속 조치를 연계하는 보호 방식의 확대 가능성

- 빅블루마블 사례는 개별 보호 기술을 따로 운영하던 방식에서 벗어나, 각 기술에서 수집한 이용 기록을 연계해 탐지 결과를 후속 조치까지 이어가는 콘텐츠 보호 방향을 보여줌
- 특히 대형 실시간 이벤트는 짧은 시간 안에 많은 시청자가 몰리고 중계 중에도 불법 재배포가 이어질 수 있어, 탐지 이후 실제 대응까지 걸리는 시간을 줄이는 방식의 활용이 확대될 가능성이 있음

• AI 탐지의 한계와 후속 조치 기준 마련의 필요성

- 다만 AI 기반 부정 이용 탐지 기능은 여러 이용 기록을 토대로 불법 유통 가능성이 높은 행위를 찾아내는 수단이며, 실제 침해 여부와 책임 소재에 대한 법적 판단을 대신할 수는 없음
- 또한 이용 기록을 분석하는 과정에서 정상적인 시청 행위가 의심 행위로 잘못 분류될 가능성도 있는 만큼, 계정 제한이나 이용 차단 등의 조치에 앞서 AI가 제시한 판단 근거를 확인하는 절차가 필요함
- 따라서 실시간 콘텐츠 보호의 실효성을 높이기 위해서는 탐지 기술을 고도화하는 동시에, 잘못된 탐지로 인한 이용자 피해를 줄일 수 있도록 후속 조치의 기준과 확인 절차를 함께 마련할 필요가 있음

참고문헌

- Big Blue Marble, "Big Blue Marble to unveil AI-assisted piracy detection and broadcast-grade streaming innovations at IBC2026", 2026.08.06., <https://www.bigbluemarble.com/en/resources/post/bbm-ibc2026/>
- Julian Clover, "Big Blue Marble to show AI piracy tools at IBC", Broadband TV News, 2026.08.06., <https://www.broadbandtvnews.com/2026/08/06/big-blue-marble-to-show-ai-piracy-tools-at-ibc/>
- IBC, "Big Blue Marble to unveil AI-assisted piracy detection at IBC2026", 2026.08.10., <https://www.ibc.org/piracy-security/news/big-blue-marble-to-unveil-ai-assisted-piracy-detection-at-ibc/22794>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

발매 전 AI 학습 방해 기술의 작동 방식과 음원 관리 플랫폼 연동 사례

발매 이전 단계에 보호 수단을 배치하려는 배경

• 무단 AI 학습에 대한 우려와 사후 대응 시점의 한계

- AI 기업의 모델 학습 데이터 확보 경로를 둘러싼 우려가 최근 수년간 커졌으며, 음악 분야에서도 학습 데이터의 출처와 수집 방식을 둘러싼 문제 제기가 이어짐
- 음악 전문매체 뮤직 비즈니스 월드와이드(Music Business Worldwide)는 AI 음악 생성 서비스 수노(Suno)가 AI 모델 구축을 위해 음악 관련 플랫폼인 유튜브 뮤직(YouTube Music), 디저(Deezer), 지니어스(Genius) 등의 데이터를 스크래핑(scraping)*한 것이 유출된 소스 코드(source code)**에 나타났다고 보도함¹⁾
- 이는 학습 데이터의 수집 경로가 외부 유출 또는 보도를 통해 사후에 드러날 수 있음을 보여주는 사례로, 창작자가 자신의 음원이 AI 학습에 활용됐는지를 사전에 파악하기 어렵다는 문제의식과 연결됨

* 스크래핑(scraping): 웹사이트나 서비스에 게시된 데이터를 자동화된 방식으로 수집하는 행위

** 소스 코드(source code): 컴퓨터 프로그램이나 소프트웨어가 어떻게 작동하는지를 사람이 읽고 수정할 수 있는 형태로 작성한 명령어와 코드의 집합

• 음원 공개 이전을 기술적 개입 지점으로 설정하는 접근

- 음원 유통 기업 심포닉(Symphonic)의 최고경영자 조지 브레아(Jorge Brea)는 AI를 둘러싼 논의가 창작물이 이미 복제되거나 오용된 뒤에 시작해서는 안 된다고 언급함²⁾
- 기업의 설명에 따르면 음원, 보컬(vocal) 녹음 및 카탈로그(catalog)* 자산은 창작자가 이를 인지하기 전에 복제되거나 클로닝(cloning)** 또는 스크래핑될 수 있어, 사후 대응만으로는 적절한 보호 시점을 놓칠 수 있음²⁾
- 이에 심포닉은 창작물 보호가 오용이 발생한 뒤가 아니라 그 이전에 이루어져야 한다는 문제의식을 제시하며, 제휴 도구인 뮤직실드(MusicShield)를 음원 발매 이전 핵심 보호 수단으로 소개함²⁾

* 카탈로그(catalog): 음원 유통사 레이블·권리자 등이 관리하거나 보유하고 있는 여러 음원과 관련 권리를 묶어 이르는 말

** 클로닝(cloning): 특정인의 음성 특성을 학습해 그와 유사한 음성을 인공적으로 생성하는 기술

1) Rafaela Fornitani, "Symphonic partners with ArtyShield to give artists tools that protect their music and voices from unauthorized AI use", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/symphonic-partners-with-artyshield-to-give-artists-tools-that-protect-their-music-and-voices-from-unauthorized-ai-use/>

2) Randi Zimmerman, "Symphonic Partners with ArtyShield to Help Artists Protect Their Music from Unauthorized AI Exploitation", Symphonic Blog, 2026.08.05., <https://blog.symphonic.com/2026/08/05/protect-music-from-unauthorized-ai-use/>

[사례] 뮤직실드의 작동 방식과 관리 플랫폼 연동 구조

• 사람이 듣는 음질은 유지하면서 기계가 인지하는 음향 특징을 변조하는 방식

- 심포닉은 2026년 8월 5일 AI 대상 저작권 보호 플랫폼 아티실드(ArtShield)와의 제휴를 발표하고, 뮤직실드를 자사 고객에게 제공한다고 밝혔다³⁾
- 뮤직실드는 기계가 인지하는 음향 특징(machine-perceived acoustic features)*으로 표현되는 요소를 사람이 인지하기 어려운 수준으로 변조하면서도, 청취 경험은 유지되도록 설계됨
- 또한 사람이 감지하기 어려운 잡음을 음원에 삽입해 생성형 AI가 해당 음원에서 유의미한 특징을 학습하기 어렵게 만들며, 보호 처리된 음원을 학습 데이터로 활용할 경우 AI 산출물의 품질이 저하되도록 설계된 방식을 적용함

* 기계가 인지하는 음향 특징(machine-perceived acoustic features): AI가 음원을 분석할 때 참조하는 음향상의 특성을 가리키는 표현으로, 아티실드가 사용하는 명칭이며 세부 항목은 공개되지 않음

• 관리 플랫폼 연동과 보호 도구의 구성

- 아티실드의 도구는 심포닉의 유통 관리 플랫폼 심포닉MS(SymphonicMS)에서 단일 인증(Single Sign-On, SSO)*을 통해 이용할 수 있으며, 기본 기능은 무료로 제공되나 확장된 보호 기능은 유료 등급으로 제공됨
- 제공되는 도구는 뮤직실드와 보이스실드(VoiceShield), 베리튤(VeriTune), 베리보이스(VeriVoice) 등 5개 중 4개가 이용 가능하며, 인간 창작 증명 도구인 아티실드 서티파이(ArtShield Certify)는 준비 단계로 설명됨
- 심포닉은 이와 별도로 AI 기술 기업 수렐 AI(Sureel AI)와는 AI 이용 추적 및 옵트아웃(opt-out)** 분야에서, 음악 라이선싱 플랫폼 소스오디오(SourceAudio)와는 옵트인(opt-in)*** 방식의 AI 학습 데이터 라이선싱 분야에서 각각 협력하고 있음⁴⁾

* 단일 인증(Single Sign-On, SSO): 한 번의 로그인으로 연결된 여러 서비스를 추가 인증 없이 이용할 수 있게 하는 방식

** 옵트아웃(opt-out): 권리자가 자신의 저작물이 AI 학습에 이용되는 것을 거부할 수 있도록 하는 방식

*** 옵트인(opt-in): 권리자가 사전에 동의한 저작물만 AI 학습에 이용할 수 있도록 하는 방식

[표 1] 아티실드 도구별 보호 및 검증 대상과 주요 기능

도구명	보호 및 검증 대상	주요 기능	제공 상태
뮤직실드 (MusicShield)	음원	무단 AI 학습 이용 방해	제공 중
보이스실드 (VoiceShield)	보컬 실연	무단 음성 클로닝 또는 합성 이용 방지	제공 중
베리튤 (VeriTune)	음원	AI 생성 가능성 분석	제공 중
베리보이스 (VeriVoice)	음성 녹음	AI 생성, 클로닝, 합성 조작 여부 분석	제공 중
아티실드 서티파이 (ArtShield Certify)	창작물	인간 창작 여부 증명	준비 중

출처: 참고문헌 종합하여 재구성

3) Randi Zimmerman, "Symphonic Partners with ArtyShield to Help Artists Protect Their Music from Unauthorized AI Exploitation", Symphonic Blog, 2026.08.05., <https://blog.symphonic.com/2026/08/05/protect-music-from-unauthorized-ai-use/>

4) Rafaela Fomitani, "Symphonic partners with ArtyShield to give artists tools that protect their music and voices from unauthorized AI use", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/symphonic-partners-with-artysield-to-give-artists-tools-that-protect-their-music-and-voices-from-unauthorized-ai-use/>

시사점: 사례의 기술적 의미와 확인 과제

• 보호 개입 시점이 발매 이전으로 앞당겨진 사례로서의 의미

- 본 사례는 침해 사실을 사후에 확인한 뒤 대응하는 대신, 음원 공개 전 파일 자체에 보호 처리를 적용해 개입 시점을 발매 이전으로 앞당기려는 접근에 해당함
- 이는 음원 파일 자체를 대상으로 하는 기술 계층을 발매 절차와 연결한 방식이어서, 유통 이후의 탐지에 의존해 온 접근과는 개입 지점이 다른 것으로 분석됨
- 또한 기업이 유통 관리 플랫폼을 통해 음원을 전달하는 기능과 창작자가 보호 도구를 이용할 수 있는 접점을 함께 제공하는 형태여서, 유통 단계가 맡는 역할이 넓어진 것으로 볼 수 있음

• 효과와 적용 범위에 관해 확인이 필요한 영역

- 다만 실제 AI 모델을 대상으로 한 방어 성능이나 정량적 효과 지표는 명확히 제시되지 않아, 뮤직실드의 실제 보호 수준을 판단하기 위해서는 추가 검증이 필요한 상태임
- 또한 이미 공개된 음원에도 동일한 보호 처리를 적용할 수 있는지와 무료, 유료 등급별 보호 범위가 어떻게 달라지는지는 현재까지 파악하기 어려움
- 각 도구가 하나의 권리 관리 절차로 연동되는 방식과 이러한 처리가 법적 판단에서 갖는 의미 역시 제시되지 않았으며, 권리 귀속과 침해 여부는 법적 해석에 따라 결정될 것으로 전망됨

참고문헌

- Joe Sparrow, "Symphonic partners with Artyshield to offer AI protection platform MusicShield", Music Ally, 2026.08.06., <https://musically.com/2026/08/06/symphonic-partners-with-artyshield-to-offer-ai-protection-platform-musicshield/>
- Rafaela Fornitani, "Symphonic partners with ArtyShield to give artists tools that protect their music and voices from unauthorized AI use", Music Business Worldwide, 2026.08.06., <https://www.musicbusinessworldwide.com/symphonic-partners-with-artyshield-to-give-artists-tools-that-protect-their-music-and-voices-from-unauthorized-ai-use/>
- Randi Zimmerman, "Symphonic Partners with ArtyShield to Help Artists Protect Their Music from Unauthorized AI Exploitation", Symphonic Blog, 2026.08.05., <https://blog.symphonic.com/2026/08/05/protect-music-from-unauthorized-ai-use/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

주간 기술 동향

복원 가능한 학습 불가 데이터와 이중 워터마크 결합 기술 RUE

• 비인가 학습 방지와 소유권 검증을 동시에 만족하는 저작권 보호 기술

딥러닝 기술의 발전은 대규모 이미지 데이터셋의 확보 없이는 불가능했다. 그러나 데이터 규모가 커질수록 저작권으로 보호받는 이미지가 소유자의 동의 없이 상업적 모델 학습에 활용되는 사례도 함께 늘고 있다. 문제는 이렇게 수집된 이미지가 대개 외부에 공개되지 않은 채 학습에만 사용되기 때문에, 소유자가 자신의 이미지가 무단으로 쓰였는지조차 확인하기 어렵다는 점이다. 결국 데이터 저작권 보호는 소유권을 증명하는 문제를 넘어, 이미지가 담고 있는 의미 정보 자체를 지켜야 하는 문제로 확장되고 있다.

이러한 배경에서 두 갈래의 대응 기술이 각각 발전해왔다. 하나는 이미지에 사람이 육안으로 인지할 수 없는 미세한 변형을 추가해 AI 모델이 유의미한 특징을 학습하지 못하도록 만드는 학습 불가 데이터(Unlearnable Examples) 기술로, 비인가 학습을 사전에 차단하는 데 초점을 둔다. 다른 하나는 이미지에 식별 정보를 삽입한 뒤 이를 추출해 소유권을 입증하는 디지털 워터마킹 기술로, 이미 유출된 이미지에 대한 사후 검증 수단으로 활용된다.

문제는 두 기술을 단순히 결합할 경우 서로 간섭을 일으킨다는 점이다. 워터마크와 학습 불가 변형이 모두 이미지의 픽셀 값을 미세하게 조정하는 방식으로 작동하기 때문에, 둘을 함께 적용하면 워터마크를 읽어내는 과정에서 오류가 크게 늘어난다. 학습 방지 효과를 높이기 위해 변형의 강도를 키울수록 워터마크 추출 성능은 더욱 떨어지는 상충 관계도 나타난다. 이는 두 기술을 병행하는 것만으로는 소유권 검증과 학습 방지라는 목표를 동시에 달성할 수 없음을 보여준다.

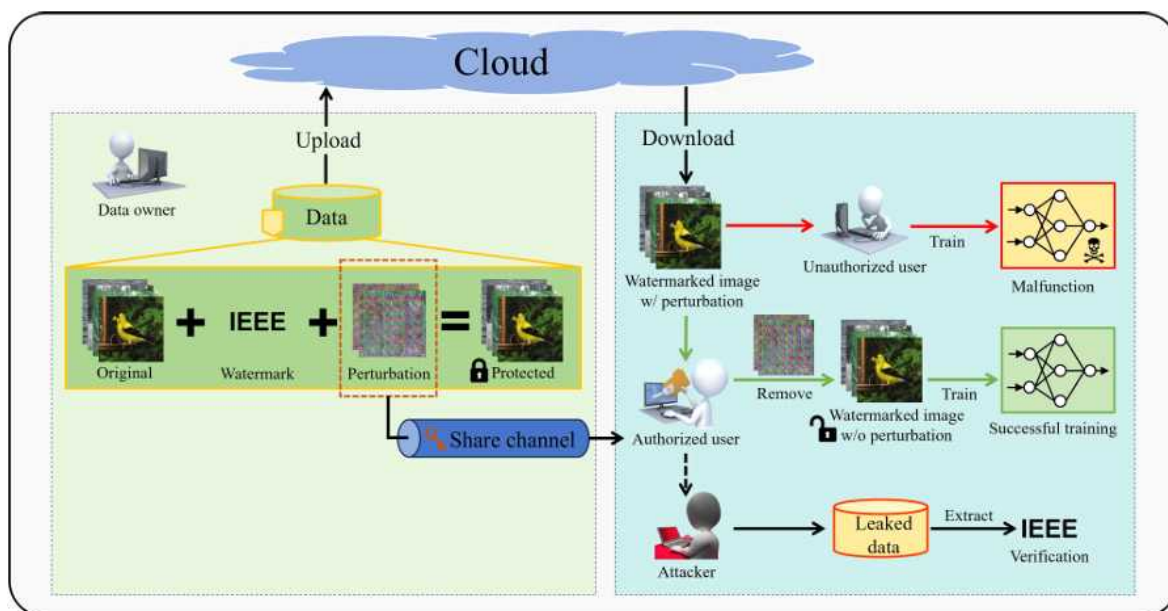
본 보고서에서는 이러한 한계를 극복하기 위해 등장한 복원 가능한 학습 불가 데이터(Reversible Unlearnable Examples, 이하 RUE) 기술을 다룬다. 이 기술은 계층 단위로 설계된 학습 불가 변형과 두 개의 독립 디코더를 사용하는 이중 워터마크 추출 구조를 결합하여 간섭 문제를 해소한다. 인가받은 사용자는 별도 경로로 전달받은 변형 정보를 이용해 원본에 가까운 이미지를 복원한 뒤 정상적으로 모델을 학습시킬 수 있으며, 비인가 사용자에게는 학습이 차단된 상태로 남는다. 변형이 적용된 상태와 제거된 상태 모두에서 워터마크 추출이 가능하다는 점도 이 기술의 특징이다.

[사례] 복원 가능한 학습 불가 데이터 기반 저작권 보호 기술, RUE

• RUE 기술 등장 배경 및 기존 방식의 한계

- 딥러닝(deep learning) 환경에서 이미지 저작권 보호는 소유권 검증과 의미 정보 보호라는 두 목표로 구분되며, 각각 디지털 워터마킹과 학습 불가 데이터 기술이 독립적으로 발전해왔음
- 그러나 워터마크와 학습 불가 변형은 모두 이미지 픽셀을 미세하게 조정하는 방식이어서, 두 기술을 단순 결합하면 변형이 워터마크 신호에 영향을 주어 추출 오류율이 상승할 수 있는 부작용이 발생함
- 기존의 결합 시도들은 워터마크를 읽기 전에 변형을 제거하거나 가시적 표식을 활용하는 방식에 의존해, 변형이 남아 있는 상태에서는 권리 확인이 어렵거나 학습 방지 효과가 충분하지 않은 한계가 있음
- 복원 가능한 학습 불가 데이터(Reversible Unlearnable Examples, 이하 RUE)는 데이터 소유자, 인가 사용자, 비인가 사용자, 공격자라는 네 주체가 존재하는 실제 유통 환경을 전제로, 변형의 존재 여부와 무관하게 소유권 검증이 이루어질 수 있는 구조를 설계 목표로 삼음

[그림 1] 데이터 유통 과정에서의 RUE 작동 구조



출처: Binze Wang 외 4인, "Reversible Unlearnable Examples: Towards the Copyright Protection in Deep Learning Era", arXiv, 2026.08.06., <https://arxiv.org/pdf/2608.06211>

• 워터마크 임베딩 및 추출 구조

- 워터마크 처리는 정보를 심는 인코더(encoder)와 읽어내는 디코더(decoder)로 구성됨. 인코더는 원본 이미지와 삽입할 식별 정보를 함께 입력받아 워터마크가 삽입된 이미지를 만들고, 디코더는 해당 이미지에서 식별 정보를 다시 복원함
- 이때 인코더는 식별 정보를 이미지 특징과 결합할 수 있는 형태로 확장한 뒤 원본에서 추출한 시각적 특징과 합치는 방식으로 작동하며, 중요한 특징에 더 큰 비중을 두는 처리 단계를 거쳐 정보를 이미지 전반에 분산시킴
- 다만 정보를 삽입하는 과정에서 화질이 저하될 수 있어, 이를 억제하기 위해 판별기를 함께 학습시키는 구조를 채택함. 판별기는 입력된 이미지가 원본인지 워터마크가 삽입된 이미지인지 맞도록 훈련되며, 인코더는 이 판별기를 속이는 방향으로 학습되어 시각적으로 구별하기 어려운 결과물을 만들게 됨

- 한편 워터마크 추출은 삽입된 정보를 이루는 0과 1의 각 자리를 정확히 맞추는 문제로 다루어지며, 원래 심은 정보와 읽어낸 정보 사이의 차이가 작아지는 방향으로 학습이 진행됨
- 학습 과정에는 JPEG 압축을 모사한 처리 단계가 포함되어, 실제 유통 환경에서 흔히 발생하는 압축을 거친 뒤에도 워터마크가 유지되도록 강건성을 확보함

• SI가 이미지 내용을 배우지 못하게 만드는 원리

- 기존의 학습 불가 데이터 기술은 대체로 모델이 이미지를 잘못된 정답으로 분류하도록 유도하는 방식을 사용해왔으며, 이 방식은 특정 모델의 판단 기준에 맞춰 변형을 설계하기 때문에 다른 구조의 모델에는 효과가 약해지는 경향이 있음
- 이러한 한계를 넘기 위해 RUE는 이미지와 모델의 출력 사이의 연관성 자체를 끊는 방식을 제안함. 두 변수 간 정보 의존도를 나타내는 상호정보량(mutual information)을 낮추면, 모델은 이미지에서 의미 있는 단서를 얻기 어려워짐
- 즉 변형의 목표는 모델의 답을 특정 오답으로 바꾸는 것이 아니라, 입력 데이터에 상관없이 모델의 출력이 원본 내용과 무관해지도록 만드는 것임
- 이는 서로 다른 모델들이 같은 이미지에서 대체로 비슷한 시각적 특징에 주목한다는 점에 착안한 설계로, 그 공통된 특징을 가리면 여러 모델에 두루 통하는 효과를 기대할 수 있음
- 실제로 변형이 적용된 이미지를 학습한 모델들은 원래 사물의 형태나 질감이 아니라 삽입된 변형 영역에 주로 반응하는 것으로 확인되며, 변형을 만들 때 사용하지 않은 다른 구조의 모델에서도 유사한 현상이 나타남

• 이중 워터마크 추출 구조 및 복원 가능성

- 디코더는 원래 변형이 없는 이미지를 기준으로 학습되기 때문에, 변형이 더해진 이미지에서는 정보를 정확히 읽어내기 어려워짐. RUE는 이를 보완하기 위해 동일한 구조의 디코더를 하나 더 두고, 변형이 남아 있는 이미지에서 워터마크를 읽는 역할을 별도로 맡김
- 이렇게 두 디코더가 각각 변형이 없는 상태와 있는 상태를 담당하며, 이미지에 변형이 남아 있는지 외부에서 판단할 수 없는 상황을 전제하므로 두 디코더를 동시에 가동해 둘 중 하나라도 원래 심은 정보와 일치하면 소유권이 확인된 것으로 처리함
- 한편 변형은 이미지 한 장마다 따로 만들지 않고 같은 분류에 속한 이미지들이 하나의 변형을 공통으로 사용하도록 평균을 내어 정리되며, 이는 전달해야 할 정보량을 줄이는 효과가 있음
- 정리된 변형 정보는 안전한 경로를 통해 인가 사용자에게 전달되며, 인가 사용자가 이미지에서 해당 변형을 빼내면 워터마크만 남은 상태로 되돌아가 정상적인 모델 학습이 가능함

• 성능 검증 결과

- 실험은 사물 사진을 분류별로 정리해 놓은 공개 이미지 모음인 이미지넷(ImageNet)과 사이파10(CIFAR10), 품종별 반려동물 사진을 모은 옥스퍼드 데이터셋(Oxford-IIIT Pet Dataset)을 대상으로 진행되며, 하나의 모델로 만든 변형을 서로 다른 다섯 종류의 신경망에 적용해 학습 방지 효과와 워터마크 추출 성능을 함께 측정함
- 변형이 적용된 데이터로 학습한 모델들의 정확도는 무작위로 답을 고르는 수준까지 떨어지고, 변형을 제거한 뒤 학습한 경우에는 원본 데이터로 학습했을 때와 근접한 정확도를 회복하는 것으로 확인됨
- 워터마크 추출 오류율은 고해상도 데이터셋에서 0.4% 미만, 저해상도 데이터셋에서 2% 내외로 나타나며, 두 경우 모두 하나의 디코더만 사용한 방식보다 낮아 이중 구조가 간섭 완화에 기여함

결론 및 시사점

• 기술적 한계 및 개선 과제

- RUE는 비인가 학습 방지와 소유권 검증이라는 두 목표를 하나의 절차로 묶어 처리하며, 서로 간섭하던 두 기술을 함께 적용할 수 있는 방식을 제시함
- 특정 모델의 판단 기준에 의존하지 않는 변형 생성 방식을 택함으로써, 비인가 사용자가 어떤 구조의 모델을 사용하는지 알 수 없는 실제 환경에서도 보호 효과를 기대할 수 있음
- 변형이 남아 있는 상태와 제거된 상태 모두에서 워터마크 확인이 가능하므로, 인가 사용자를 거쳐 유출된 데이터에 대해서도 권리 주장의 근거를 확보할 수 있음
- 다만 변형 정보를 인가 사용자에게 전달하는 경로의 안전성에 보호 효과가 좌우되며, 향후 저작권 관리가 필요한 이미지 데이터셋의 거래 및 배포 환경에서 활용 가능성을 검토할 수 있음

참고문헌

- Binze Wang 외 4인, "Reversible Unlearnable Examples: Towards the Copyright Protection in Deep Learning Era", arXiv, 2026.08.06., <https://arxiv.org/pdf/2608.06211>