

# 저작권 이슈 트렌드



COPYRIGHT ISSUE TREND



한국저작권위원회  
KOREA COPYRIGHT COMMISSION

# CONTENTS

## 저작권 이슈 트렌드

Biweekly Report | 통권 제78호(2026. 3-2)

- 오디오 워터마킹 기술의 진화와 블록체인 기반 소유권 검증
- AI 코드 개발 시대, 오픈소스 라이선스 체계의 새로운 쟁점
- 애플 뮤직, 메타데이터를 기반으로 한 AI 투명성 태그 도입



# 오디오 워터마킹 기술의 진화와 블록체인 기반 소유권 검증

## 뉴스 브리프

디지털 음원 시장이 급성장하면서 오디오 콘텐츠의 불법 복제와 무단 사용이 심각한 문제로 떠올랐다. 스트리밍 플랫폼과 팟캐스트의 확산은 창작자의 권리 보호를 더욱 복잡하게 만들고 있다. 이러한 환경에서 오디오 워터마킹 기술이 저작권 침해를 추적하고 증명할 수 있는 핵심 수단으로 주목받는다. 오디오 워터마킹은 사람의 귀로는 감지할 수 없는 디지털 서명을 음원에 삽입해 소유권 정보를 보존한다. 최근에는 블록체인과 결합해 변조 불가능한 증거 체계를 구축하고 있다. 하지만 포맷 변환 과정에서 워터마크가 손실될 위험, 음질 저하 가능성, 산업 표준 부재 등의 기술적 한계도 여전히 존재한다. 본 보고서는 심리음향 원리를 활용한 워터마크 임베딩 메커니즘과 블록체인 기반 검증 시스템의 작동 방식을 분석하고, 이 기술이 저작권 보호 생태계에 어떤 변화를 가져올 수 있는지 전망한다.

## 뉴스 플러스

### I. 서론 : 디지털 오디오 시대의 저작권 보호 과제

#### • 오디오 불법 복제 확산과 포렌식 워터마킹 기술의 부상

디지털 음원 시장의 폭발적 성장은 창작자와 권리자에게 새로운 기회를 제공했지만, 동시에 불법 복제와 무단 배포라는 고질적 문제를 심화시켰다. 스트리밍 플랫폼을 통해 음원이 전 세계로 즉시 전송되고, 소셜미디어와 콘텐츠 공유 사이트에서 오디오 파일이 무분별하게 재가공되면서 원본 출처를 추적하기가 사실상 불가능해졌다. 특히 팟캐스트, 오디오북, AI 음성 합성 콘텐츠가 급증하면서 저작권 침해 양상은 더욱 다양하고 복잡한 형태로 진화하고 있다.



이러한 환경에서 포렌식 워터마킹 기술은 저작권 침해를 사후에 추적하고 법적 증거로 활용할 수 있는 핵심 수단으로 부상했다. 포렌식 워터마킹은 오디오 파일에 사람의 귀로 들을 수 없는 디지털 서명을 삽입해 유통 경로를 식별하고, 무단 사용이 발생했을 때 원본 소유권을 증명한다. 자동화된 음향 지문 인식 시스템은 웹상에서 침해 콘텐츠를 스캔하며, 일부 시스템은 95% 이상의 탐지 성공률을 기록하고 있다. 불법 복제와 변조를 원천적으로 차단하는 방어 체계로 자리 잡아가고 있다.

### • 워터마킹 기술 발전의 배경과 산업적 필요성

오디오 워터마킹 기술은 1990년대 후반 디지털 저작권 관리(Digital Rights Management, 이하 DRM) 체계의 일환으로 처음 도입되었으나, 초기에는 음질 저하와 낮은 강인성 문제로 확산에 한계가 있었다. 그러나 2000년대 중반 이후 심리음향\* 모델이 정교화되고 주파수 변환 기법이 발전하면서, 사람의 청각으로 감지할 수 없으면서도 압축과 변환 과정에서 살아남는 워터마크 설계가 가능해졌다. 이는 단순히 기술적 진보에 그치지 않고, 음원 유통 구조가 물리적 매체에서 스트리밍 중심으로 전환되면서 실시간 추적과 자동 검증이 필수가 된 산업 환경 변화와 맞물려 있다.

특히 최근에는 AI 생성 음성과 딥페이크 오디오의 확산이 새로운 위협 요인으로 떠오르면서, 워터마킹 기술의 중요성이 재조명되고 있다. AI 생성 음성 콘텐츠는 원본 창작자의 음성을 무단으로 모방하거나 변형할 수 있으며, 이 과정에서 복제권, 동일성유지권 등 저작권 문제가 발생할 수 있다. 워터마킹은 이러한 AI 산출물에도 원본 출처 정보를 각인해 출처 정보를 유지할 수 있는 기술적 기반을 제공한다. 동시에 블록체인의 결합을 통해 워터마크 자체의 무결성을 검증하고, 소유권 이전 이력을 투명하게 기록하는 시스템이 구축되면서 기술적 신뢰도가 강화되고 있다. 이제 워터마킹은 단순한 보안 기술을 넘어, 디지털 오디오 생태계 전반의 신뢰 구조를 뒷받침하는 인프라로 진화하고 있다.

\* 심리음향(psychoacoustics): 물리적인 소리 자극과 인간의 청각 시스템 간의 관계를 연구하는 학문. 소리를 어떻게 지각하고 심리적·생리적으로 어떤 반응을 보이는지 연구함

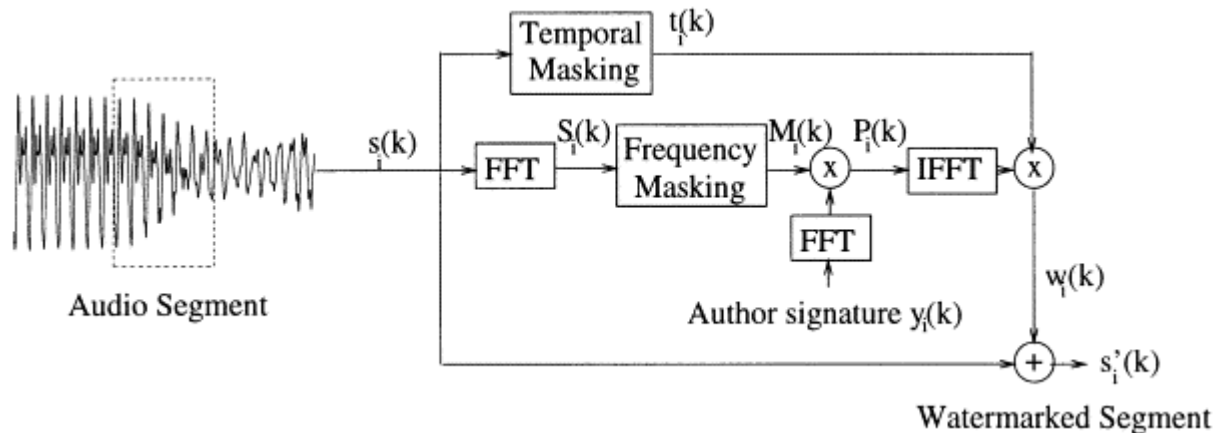
## II. 본론: 오디오 워터마킹의 기술적 메커니즘과 저작권 보호

### • 심리음향 마스킹을 활용한 워터마크 임베딩 원리

심리음향 오디오 워터마킹은 인간의 청각 시스템이 특정 주파수 대역의 소리를 인지하지 못하는 현상을 이용해 디지털 서명을 음원에 삽입하는 기술이다. 이 기술의 핵심은 마스킹\*으로, 큰 소리가 작은 소리를 가리는 청각적 특성을 활용한다. 예를 들어, 록 음악의 드럼 소리가 크게 울릴 때 배경의 미세한 기타 소리를 듣기 어려운 것처럼, 워터마크는 원본 오디오 신호에 가려져 사람의 귀로는 감지되지 않는 영역에 배치된다. 이를 통해 음질 저하 없이 저작권 정보를 음원 내부에 각인할 수 있다.

기술적으로는 주파수 마스크링(frequency masking)과 시간 마스크링(temporal masking) 두 가지 방식이 활용된다. 주파수 마스크링은 특정 주파수 대역에서 강한 신호가 인접한 약한 신호를 가리는 현상을 이용하며, 시간 마스크링은 큰 소리 직전이나 직후의 짧은 시간 동안 작은 소리가 인지되지 않는 특성을 활용한다. 워터마크는 이 두 마스크링 영역을 정밀하게 계산해 삽입되므로, 청취자는 원본과 워터마크된 파일을 구별할 수 없다. 국제음반산업협회(International Federation of the Phonographic Industry, 이하 IFPI)는 워터마크 시스템이 SNR\*\* 20dB 이상을 유지하고, 초당 20비트 이상의 데이터를 전송할 수 있어야 한다는 벤치마크를 제시하고 있다.

[그림] 주파수 마스크링과 시간 마스크링을 활용한 음향워터마크 삽입 개요



출처: Mitchell D Swanson, "Robust audio watermarking using perceptual masking", ScienceDirect, 1998.05.28., <https://www.sciencedirect.com/science/article/abs/pii/S0165168498000140>

심리음향 워터마킹의 산업적 의미는 음질과 보안 사이의 균형을 처음으로 실용 수준에서 달성했다는 점에 있다. 과거 DRM 시스템은 암호화 방식으로 접근 자체를 차단했지만, 한번 해독되면 무방비 상태가 되었다. 반면 심리음향 워터마킹은 콘텐츠를 자유롭게 유통하면서도 출처 추적이 가능하도록 설계되어, 스트리밍 시대의 개방형 유통 구조와 양립할 수 있다. 저작권 보호 측면에서는 워터마크가 음원 파일 자체에 내장되므로, 파일이 복사되거나 재배포되더라도 소유권 정보가 함께 이동한다는 점이 핵심이다. 이는 사후 추적과 법적 증거 확보를 가능하게 하며, 침해 발생 시 원본 권리자를 명확히 식별할 수 있는 기술적 토대를 제공한다.

\* 마스크링(masking): 특정 소리가 다른 소리에 의해 가려져서 들리지 않는 현상으로, 주파수 마스크링과 시간 마스크링으로 구분됨

\*\* 신호 대 잡음비(Signal-to-Noise Ratio, SNR): 유효한 신호의 전력과 배경 잡음 전력의 비율로 주로 데시벨(dB) 단위로 표기

### • 주파수 영역 변환과 비가청 데이터 삽입 메커니즘

오디오 워터마크를 삽입하는 과정은 시간 영역에서 주파수 영역으로의 변환을 통해 이루어진다. 원본 신호는 푸리에 변환 또는 이산 코사인 변환 등 수학적 주파수 변환 기법을 거쳐 개별 주파수 성분으로 분해되며, 이 과정에서 각 주파수 대역의 에너지 분포가 분석된다. 워터마크 데이터는 심리음향 모델이 계산한 마스킹 임계값 이하의 주파수 영역에 삽입되어, 청각적으로 감지되지 않으면서도 디지털 신호로는 명확히 존재하는 상태를 유지한다.

삽입된 워터마크는 단순한 비트 패턴이 아니라, 소유권 정보와 유통 경로를 식별할 수 있는 고유 식별자 형태로 인코딩된다. 예를 들어, 특정 음원이 스트리밍 플랫폼 A를 통해 배포되었다면 플랫폼 코드와 배포 시점, 사용자 ID 등이 워터마크에 포함될 수 있다. 이 데이터는 오류 정정 코드와 함께 삽입되어, MP3 압축이나 포맷 변환 과정에서 일부 손실이 발생하더라도 복원 가능하도록 설계된다.

이 메커니즘이 저작권 보호에 기여하는 방식은 두 가지로 나뉜다. 첫째, 워터마크는 파일의 메타데이터가 아닌 오디오 신호 자체에 삽입되므로, 메타데이터를 삭제하거나 파일명을 변경하는 단순 조작으로는 제거할 수 없다. 둘째, 주파수 영역에 분산 삽입된 워터마크는 부분 복사나 샘플링 과정에서도 일부가 살아남아 원본 추적 단서를 제공한다.

### • 블록체인 기반 소유권 검증과 변조 방지 시스템

블록체인 기술은 오디오 워터마킹의 신뢰성을 한 단계 끌어올리는 보완 인프라로 작동한다. 블록체인은 분산 원장 구조를 통해 데이터를 여러 노드에 복제 저장하며, 한번 기록된 정보는 암호학적으로 연결되어 사후 변조가 사실상 불가능하다. 오디오 워터마킹과 결합될 때, 블록체인은 워터마크 자체의 생성 시점과 소유권 이력을 불변 기록으로 남겨 워터마크가 위조되지 않았음을 증명한다. 이는 워터마크 데이터를 해시 값으로 변환해 블록체인에 등록하고, 이후 검증 시 해시를 비교함으로써 원본 워터마크와의 일치 여부를 확인하는 방식으로 구현된다.

블록체인 기반 워터마킹은 생성, 유통, 검증의 전 과정에서 투명성과 무결성을 보장한다. 워터마크가 오디오 파일에 삽입되는 시점에 암호학적 해시가 생성되어 블록체인에 타임스탬프와 함께 기록되며, 파일이 유통되거나 소유권이 이전될 때마다 새로운 트랜잭션이 추가되어 전체 이력이 투명하게 추적된다. 침해가 의심되는 파일이 발견되면 내장된 워터마크를 추출해 블록체인에 등록된 원본 해시와 대조함으로써 출처와 정당성을 검증할 수 있다. 이러한 구조는 중앙 서버 방식의 취약점을 근본적으로 해결하며, 해킹이나 관리 주체 폐업에도 소유권 기록이 소실되지 않는다. 또한 스마트 계약을 활용하면 워터마크 검증과 동시에 라이선스 조건을 확인하고, 사용료나 로열티를 권리자에게 자동으로 지급하는 시스템 구축이 가능해 저작권 보호와 보상 체계를 통합적으로 운영할 수 있다.

## 결론 : 기술적 진화와 산업 생태계 변화

### • 오디오 워터마킹이 저작권 보호 체계에 미치는 영향과 과제

오디오 워터마킹 기술은 디지털 음원 유통 환경에서 저작권 보호의 패러다임을 사후 추적 중심으로 전환시키고 있다. 과거 접근 차단 방식의 DRM이 사용자 경험을 저해하며 한계를 드러낸 반면, 워터마킹은 콘텐츠의 자유로운 유통을 허용하면서도 침해 발생 시 출처를 정확히 추적할 수 있는 기술적 균형점을 제시한다. 심리음향 원리를 활용한 비가청 워터마크 삽입, 블록체인 기반 소유권 검증, 자동화된 침해 탐지 시스템의 결합은 음질 저하 없이 강력한 보호 체계를 구축할 수 있음을 입증했다. 이는 스트리밍 플랫폼이 급증하는 현 시점에서, 저작권 보호와 콘텐츠 접근성을 동시에 충족하는 실용적 해법으로 평가된다.

그러나 기술적 완성도와 산업 확산 사이에는 여전히 간극이 존재한다. 최신 오디오 처리 기술의 발전은 워터마크 제거 공격의 성공률을 높이고 있으며, 표준화되지 않은 워터마킹 방식은 플랫폼 간 상호 운용성을 저해한다. 또한 워터마크가 기술적 증거로 기능할 뿐 법적 침해 판단과는 별개라는 점에서, 공정 이용과 라이선스 관리를 통합하는 제도적 보완이 필수적이다. 향후 과제는 워터마킹 기술의 내구성을 높이는 동시에, 국제 표준 제정과 법적 효력 인정 범위를 명확히 하는 것이다. 이를 위해서는 기술 개발과 제도 정비가 병행되어야 하며, 권리자와 플랫폼, 이용자 간의 균형 잡힌 합의가 필요하다.

### 참고문헌

- Michael Sumner, "Blockchain Audio Watermarking for Piracy Protection", ScoreDetect, 2026.03.12., <https://www.scoredetect.com/blog/posts/blockchain-audio-watermarking-for-piracy-protection>
- Michael Sumner, "Psychoacoustic Audio Watermarking: How It Works", ScoreDetect, 2025.05.16., <https://www.scoredetect.com/blog/posts/psychoacoustic-audio-watermarking-how-it-works>
- Mitchell D Swanson, "Robust audio watermarking using perceptual masking", ScienceDirect, 1998.05.28., <https://www.sciencedirect.com/science/article/abs/pii/S0165168498000140>

### 기술용어

순번	용어	설명
1	심리음향 (psychoacoustics)	물리적인 소리 자극과 인간의 청각 시스템 간의 관계를 연구하는 학문. 소리를 어떻게 지각하고 심리적·생리적으로 어떤 반응을 보이는지 연구함
2	마스킹 (masking)	특정 소리가 다른 소리에 의해 가려져서 들리지 않는 현상으로, 주파수 마스킹과 시간 마스킹으로 구분됨
3	신호 대 잡음비 (Signal-to-Noise Ratio, SNR)	유효한 신호의 전력과 배경 잡음 전력의 비율로 주로 데시벨(dB) 단위로 표기



# AI 코드 개발 시대, 오픈소스 라이선스 체계의 새로운 쟁점

## 뉴스 브리프

2026년 3월, 파이썬 오픈소스 라이브러리 chardet의 유지보수자가 AI 코딩 도구 클로드 코드를 활용하여 코드를 전면 재작성한 뒤, 재작성된 코드의 라이선스를 카피레프트 라이선스인 LGPL에서 허용적 라이선스인 MIT로 변경하는 사건이 발생하였다. 이에 대해 chardet의 원저작자는 이의를 제기하였다. 핵심 쟁점은 AI가 재작성한 코드가 원본의 파생 저작물에 해당하는지 여부이다. 최근 AI가 학습 데이터를 재조합하여 원본과 구조적 유사성이 낮은 산출물을 생성해내는 사례가 증가하는 가운데, 이번 chardet 관련 분쟁은 기존의 오픈소스 라이선스 체계가 직면한 어려움을 조명하고 있다.

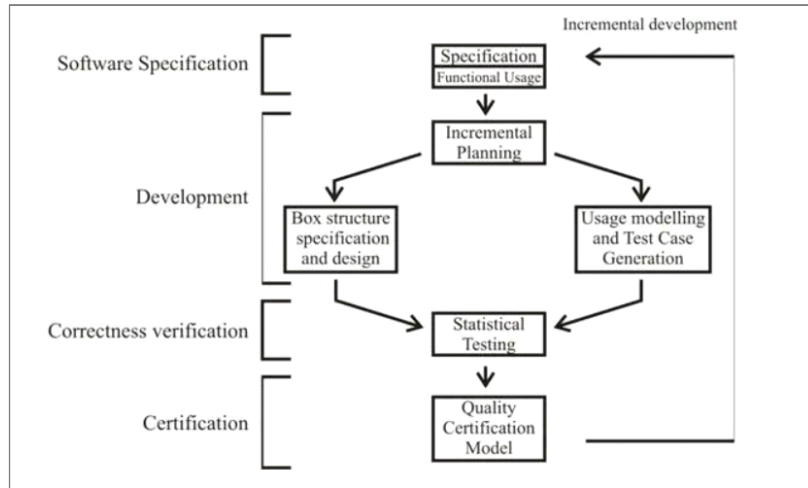
## 뉴스 플러스

### I. 서론 : AI 코딩 환경의 변화와 오픈소스 라이선스의 새로운 쟁점

#### • AI 코딩의 부상과 ‘클린룸’ 방식의 위기

소프트웨어 개발 분야에서 ‘클린룸(Clean Room)’은 원본 코드의 기능을 재현한 신규 코드를 개발하면서도, 원본 코드의 저작권을 침해하지 않을 수 있는 방식으로 활용되어 왔다. ‘클린룸’ 방식은 기능을 재현하고자 원본 코드를 분석하는 팀과, 신규 코드를 개발하는 팀을 물리적으로 분리한다. 분석팀은 원본 코드에 대한 기능적 분석 결과를 개발팀에 제공하고, 개발팀은 원본 코드를 확인하지 않은 채 분석팀으로부터 제공받은 분석 결과만을 바탕으로 새로운 코드를 개발한다. 이러한 과정을 통해 최종 산출물이 원본으로부터 직접적으로 파생된 저작물이 아님을 입증하는 구조이다. 1982년 미국의 컴퓨터 회사 컴팩(Compaq)이 ‘클린룸’ 방식으로 아이비엠 바이오스(IBM BIOS)의 클론을 제작한 사례가 대표적인 선례로 평가된다.

[그림] '클린룸' 도식화



출처: Archie, "Cleanroom Software Engineering – Short Notes", My Thinking, 2011.12.17., <https://archieanugrah.wordpress.com/2011/12/17/cleanroom-software-engineering-short-notes/>

‘클린룸’ 방식은 통상 수 주에서 수 개월의 기간이 소요되고, 다수의 인력을 필요로 하였다. 그러나 대규모 언어 모델(Large Language Model, 이하 LLM)을 기반으로 하는 AI 코딩 도구는 원본 코드의 기능을 수 시간 내에 분석·재구현할 수 있는 역량을 갖추게 되었으며, 이 과정에서 원본 코드와 구조적 유사성이 극히 낮은 산출물을 생성할 수 있다. 문제는 AI 코딩 도구가 학습 과정에서 원본 코드를 이미 학습하였을 가능성이 높다는 점이다. 이로 인해 AI 코딩 환경에서는 ‘클린룸’ 방식의 핵심인 ‘원본 코드를 보았는가 혹은 보지 않았는가’의 기준이 적용되기 어렵다는 지적이 제기되고 있다.

### • chardet 라이선스 분쟁

이러한 논의가 구체적인 사건으로 표면화된 것이 2026년 3월 발생한 파이썬 오픈소스 라이브러리\* chardet의 라이선스 변경 관련 분쟁이다. chardet의 유지보수자 댄 블렌차드(Dan Blanchard)는 AI 코딩 도구인 클로드 코드(Claude Code)를 활용하여 chardet의 코드를 재작성한 뒤, 기존 LGPL가 아닌 MIT 라이선스를 적용한 chardet 7.0.0을 공개하였다. 블렌차드는 chardet 7.0.0 개발 과정에서 ‘클린룸’과 유사한 방식을 활용하였다고 설명하였다.

[표] LGPL vs MIT 라이선스 유형 비교

구분	LGPL	MIT
라이선스 유형	카피레프트(copyleft)	허용적(permissive)
핵심 요구 조건	라이브러리 수정 시 동일 라이선스 적용 및 공개 의무	저작권 표시 유지
파생 저작물에 대한 규제	라이브러리 수정분에 한해 LGPL 유지	없음

출처: 자유소프트웨어재단(FSF) 라이선스 문서, 오픈소스 이니셔티브(OSI) 라이선스 분류 기준을 바탕으로 재구성



이에 대해 2011년 이후 개발 활동을 중단하였던 chardet의 창작자 마크 필그림(Mark Pilgrim)은 “chardet 7.0.0의 라이선스를 임의로 변경하는 것은 LGPL의 라이선스 요구 조건을 명백히 위반하는 행위”라며 이의를 제기<sup>1)</sup>하였다. 블랜차드는 chardet 7.0.0이 기존 chardet으로부터 독립적인 코드라고 주장하였으나, 필그림은 블랜차드가 chardet의 유지보수자로서 원본 코드에 장기간 노출되었으며, chardet 7.0.0 제작에 사용된 AI 모델 역시 원본 코드를 학습하였을 가능성이 높아 ‘클린룸’ 방식에 위배된다고 지적하였다.

\* 오픈소스 라이브러리(open source library): 무료 또는 일부 조건 하에 사용할 수 있는 소스 코드를 모아둔 패키지

## II. 본론: chardet 사건의 전개와 주요 쟁점

### • chardet의 역사

chardet은 미국 소프트웨어 기업 모질라(Mozilla)가 2006년 공개한 문자 인코딩 감지 코드를 파이썬 환경에서 구동할 수 있도록 재개발된 프로젝트이다. 필그림은 당초 chardet을 LGPL 라이선스로 공개하였다. LGPL은 카피레프트(copyleft)\* 라이선스의 한 종류로, 라이브러리를 특정 소프트웨어 내에서 자유롭게 사용할 수 있도록 허용하되, 라이브러리 자체를 수정할 경우 수정된 버전에도 동일한 LGPL 라이선스를 적용하여 공개하도록 요구한다. chardet은 2026년 3월 기준 연간 약 1억 3,000만 회의 다운로드 횟수를 기록하는 등 파이썬 생태계의 핵심 인프라로 평가되고 있다.

필그림은 2011년 소프트웨어 개발 활동을 전면 중단하였는데, 이후 chardet의 유지보수 업무는 블랜차드에게로 이관되었다. 블랜차드는 2012년 7월 chardet 1.1 버전 이후 모든 릴리스를 담당해 온 것으로 확인된다. 결과적으로 chardet은 파이썬 생태계에서 가장 유명한 라이브러리 중 하나임에도 불구하고, 약 10년 이상 1인 유지보수 체제로 운영되어 온 것이다.

\* 카피레프트(copyleft): 모든 프로그램·정보·미디어는 소수에게 독점되어서는 안 되며, 자유롭게 사용되고 수정될 수 있어야 한다는 개념

### • chardet 코드 재작성 및 라이선스 변경 과정

블랜차드는 2014년부터 chardet을 파이썬 표준 라이브러리에 포함시키는 것을 주요 목표로 설정하였다. 그러나 파이썬 표준 라이브러리는 MIT, BSD 등 일부 라이선스만을 수용하고 있어, LGPL 라이선스가 적용된 chardet은 구조적으로 편입이 불가능한 상태였다. 블랜차드는 라이선스 외에도 chardet의 처리 속도와 정확도를 개선해야 한다고 강조하였으며, 본업과 병행하는 무보수 유지보수 환경에서 이를 독자적으로 수행할 여력이 제한적이었다고 설명하였다.

1) Simon Willison, “Can coding agents relicense open source through a “clean room” implementation of code?”, Simon Willison’s Weblog, 2026.03.05., <https://simonwillison.net/2026/Mar/5/chardet/>



2026년 2월, 블렌차드는 미국 AI 기업 앤트로픽(Anthropic)의 AI 코딩 도구인 클로드 코드를 활용하여 chardet의 코드를 전면 재작성하는 과정에 착수하였다. 그는 먼저 Claude의 브레인스토밍 기능을 통해 chardet의 주요 기능을 분석한 설계도를 작성하였다. 이후 별도의 세션에 chardet의 원본 코드에 접근하지 않도록 설계된 프로젝트를 구축하여 chardet 7.0.0 개발을 시작하였다. 블렌차드는 약 5일간의 작업을 거쳐 chardet 7.0.0을 완성하였고, 신규 모델의 성능이 원본 대비 약 48배 향상되었다고 보고하였다. 아울러, 블렌차드는 코드 표절 검증 도구인 제이플래그(JPlag)을 활용하여 chardet 7.0.0과 기존 버전 간의 최대 유사도가 1.29%에 불과하다는 분석 결과를 제시하며, chardet 7.0.0을 MIT 라이선스로 공개하였다.

### • 필그림의 이의 제기와 ‘파생 저작물’ 관련 논쟁

chardet 7.0.0 공개 이후, 필그림은 라이선스 변경 조치가 LGPL의 요구 조건을 위반한다고 주장하였다. 그의 핵심 논거는 두 가지로 요약된다. 첫째, LGPL에 따르면 라이선스가 적용된 코드를 수정하여 배포할 경우 동일한 LGPL 라이선스를 유지하여야 한다는 점이다. 둘째, 유지보수자가 원본 코드에 장기간 노출되어 있었으므로 이번 사례는 ‘클린룸’ 구현에 해당하지 않는다는 점이다. 필그림은 chardet 7.0.0의 라이선스를 LGPL로 복구할 것을 요청한 뒤, 추가적인 논쟁에는 참여하지 않았다.

이에 대해 블렌차드는 자신이 원본 코드에 오랜 기간 노출되어 있었음을 인정하면서도, “‘클린룸’ 방법론은 최종 산출물이 파생 저작물이 아님을 보장하기 위한 ‘수단’ 중 하나이며 ‘목적’ 자체가 아니다”라고 반론<sup>2)</sup>하였다. 또한, JPlag 유사도 분석 결과를 근거로 chardet 7.0.0이 구조적으로 기존 버전과 독립적임을 입증할 수 있다고 주장하였다.

한편, 일부 전문가들은 블렌차드의 주장에 대한 반론을 제기하였다. 특히 클로드 AI가 재작성 과정에서 chardet 원본 버전에 포함된 일부 메타데이터 파일을 참조한 사실이 확인되었는데, 이는 chardet 7.0.0의 독립성 주장을 약화시키는 요소로 평가되었다. 또한, 블렌차드가 코드를 직접 작성하지는 않았으나 설계-검토-반복 개선의 전 과정에 깊이 관여하였다는 점에서, 원본 코드에 대한 그의 사전 지식이 결과물에 간접적으로 반영되었을 가능성도 제기되었다.

### • 오픈소스 업계 관계자들의 반응

이번 chardet 라이선스 분쟁은 오픈소스 업계에서 광범위한 논쟁을 촉발하였다. 각 진영의 입장은 뚜렷하게 구분된다. 카피레프트 옹호 진영에서는 미국의 자유소프트웨어재단(Free Software Foundation, 이하 FSF) 사무총장 조이 쿠이만(Zoë Kooyman)이 입장을 밝혔다. 쿠이만은 LLM이 코드 재작성 과정에서 원본 코드를 학습 데이터로 흡수한 이상, 신규 코드가 독립적이라는 평가를 부여할 수 없다고

2) Simon Willison, “Can coding agents relicense open source through a “clean room” implementation of code?”, Simon Willison’s Weblog, 2026.03.05., <https://simonwillison.net/2026/Mar/5/chardet/>

지적하였다. 아울러 머신러닝을 통한 저작권 우회 시도가 급증하는 현 시점에서, 원본 코드의 저작권을 보호하기 위한 규제가 더욱 엄격하게 적용되어야 한다는 입장을 표명하였다.

반면, 오스트리아의 오픈소스 개발자 아민 로나처(Armin Ronacher)는 AI를 통해 코드를 재작성하는 방식을 “카피레프트 라이선스의 제약에서 벗어날 수 있는 경로”라고 평가<sup>3)</sup>하였다. 그는 기존 코드를 폐기하고 처음부터 새로 작성한 결과물은 동일한 기능을 수행하더라도 새로운 저작물이라는 견해를 제시하였다. 다만 일각에서는 이러한 논리가 정당화될 경우, 현존하는 모든 카피레프트 라이선스 프로젝트가 AI를 경유하여 독점적인 라이선스로 전환될 수 있다는 우려를 제기하였다.

### III. 결론 : AI 시대 소프트웨어 저작권 보호의 과제

#### • 카피레프트 라이선스의 선순환 구조 약화 가능성

chardet 분쟁은 AI를 활용한 코드 생성 기술이 카피레프트 라이선스 체계를 약화시키고 있음을 보여주는 사례이다. 카피레프트 라이선스는 파생 저작물에 원본과 동일한 라이선스 조건을 부과함으로써 소프트웨어를 합법적으로 공유하는 한편, 소프트웨어의 기능적인 개선을 촉진할 수 있는 선순환적 성격의 메커니즘이다. 그러나 카피레프트 라이선스는 ‘파생 저작물’의 판별이 가능하다는 전제 위에 성립한다. AI 코딩 에이전트가 원본 코드와 구조적 유사성이 극히 낮은 산출물을 단기간에 생성할 수 있게 된 현재, ‘파생 저작물’의 경계는 모호해지고 있다.

이번 chardet 사건에서 확인된 바와 같이, AI를 활용한 코드 재작성과 라이선스 변경이 기술적으로 용이해질 경우, 카피레프트 라이선스의 핵심인 ‘공유와 환원’의 선순환 구조가 왜곡될 가능성이 있다. 이러한 문제는 소프트웨어 영역에 국한되지 않을 것으로 보인다. 최근 AI가 텍스트·이미지·음악 등 다양한 저작물을 학습하고 산출물을 생성할 수 있는 환경에서, 보다 정밀한 저작권 보호 체계를 마련하는 것이 시급한 과제로 대두되고 있다.

3) Luka Kladaric, “License Laundering and the Death of Clean Room”, Shiftmag, 2026.03.10., <https://shiftmag.dev/license-laundering-and-the-death-of-clean-room-8528/>



## 참고문헌

- Thomas Claburn, “Chardet dispute shows how AI will kill software licensing, argues Bruce Perens”, The Register, 2026.03.06., [https://www.theregister.com/2026/03/06/ai\\_kills\\_software\\_licensing/](https://www.theregister.com/2026/03/06/ai_kills_software_licensing/)
- CSDN, “Claude 5天重写老库引全网争议, 维护者擅自更换开源协议, 退网15年原作者突然现身: 不准改”, 36KR, 2026.03.12., <https://www.36kr.com/p/3719844907070852>
- Luka Kladaric, “License Laundering and the Death of Clean Room”, Shiftmag, 2026.03.10., <https://shiftmag.dev/license-laundering-and-the-death-of-clean-room-8528/>
- Simon Willison, “Can coding agents relicense open source through a “clean room” implementation of code?”, Simon Willison’s Weblog, 2026.03.05., <https://simonwillison.net/2026/Mar/5/chardet/>
- Kyle Orland, “AI can rewrite open source code—but can it rewrite the license, too?”, Ars Technica, 2026.03.11., <https://arstechnica.com/ai/2026/03/ai-can-rewrite-open-source-code-but-can-it-rewrite-the-license-too/>
- Archie, “Cleanroom Software Engineering – Short Notes”, My Thinking, 2011.12.17., <https://archieanugrah.wordpress.com/2011/12/17/cleanroom-software-engineering-short-notes/>

## 기술용어

순번	용어	설명
1	오픈소스 라이브러리 (open source library)	무료 또는 일부 조건 하에 사용할 수 있는 소스 코드를 모아둔 패키지
2	카피레프트 (copyleft)	모든 프로그램·정보·미디어는 소수에게 독점되어서는 안 되며, 자유롭게 사용되고 수정될 수 있어야 한다는 개념



# 애플 뮤직, 메타데이터를 기반으로 한 AI 투명성 태그 도입

## 뉴스 브리프

애플 뮤직은 2026년 3월 AI 투명성 태그 정책을 발표하고, 아트웍·트랙·작곡·뮤직비디오 네 가지 유형으로 AI 활용 지점을 세분화하여 메타데이터에 표시할 수 있는 체계를 마련했다. 그러나 이 정책은 공급자의 자발적 신고를 전제로 하며, 메타데이터 시스템은 파일의 기술적 무결성만 검증할 뿐 입력된 정보의 진위는 확인하지 않는다. 본 보고서는 애플 뮤직 메타데이터 시스템의 기술적 구조와 AI 투명성 태그의 통합 방식을 분석하고, 자율 신고 방식의 구조적 한계 속에서 실효성 있는 투명성 체계 구축을 위한 과제를 살펴본다.

## 뉴스 플러스

### I. 서론: AI 생성 음악의 확산과 투명성 요구

#### • 애플 뮤직의 AI 투명성 태그 도입

AI 생성 음악의 급증으로 음악 유통 구조 전반에서 AI 투명성 요구가 확대되고 있다. 글로벌 음악 스트리밍 플랫폼인 디저(Deezer)가 2026년 1월 공개한 자료에 따르면, 플랫폼에는 하루 평균 약 6만 곡 이상의 AI 생성 음악이 업로드되고 있으며, 이는 전체 신규 등록 곡의 약 39%를 차지하는 수준이다.<sup>1)</sup> 문제는 이러한 음악이 AI 활용 여부를 명시하지 않은 채 유통되고 있다는 점이다. 이에 따라 AI 생성 여부를 명확히 표시해야 한다는 요구가 음악 산업 전반에서 제기되고 있다.

이러한 흐름 속에서 애플 뮤직은 2026년 3월 4일 AI 투명성 태그(Transparency Tags)를 발표했다. 음원 공급자가 AI 사용 여부를 메타데이터(metadata)\*로 제출하도록 하는 방식으로, 애플 뮤직은 “콘텐츠에 적절한 태그를 지정하는 것이 음악 산업이 AI 관련 정책을 개발하는 데 필요한 데이터와 도구를 확보하는 첫 단계”라고 설명했다.<sup>1)</sup> 애플 뮤직은 AI 투명성 태그를 통해 공급자의 자발적 신고를 기반으로 AI 생성 음악을 식별하고 관리하는 체계를 구축하고자 한다.

\* 메타데이터(metadata): 데이터를 위한 데이터로서 음악 관련해서는 곡명, 아티스트, 장르 등을 기록할 수 있음

1) Murray Stassen, "Apple Music launches AI transparency tags — but only if labels and distributors declare them", Music Business Worldwide, 2026.03.04., <https://www.musicbusinessworldwide.com/apple-music-launches-ai-transparency-tags-but-only-if-labels-and-distributors-choose-to-declare-them/>

## II. 본론: 애플 뮤직 메타데이터 시스템과 투명성 체계의 한계

### • 애플 뮤직의 메타데이터 형식과 구성 항목

애플 뮤직의 메타데이터는 XML\* 형식으로 구조화되어 있으며, 공급자는 정해진 규격에 따라 파일을 작성하여 제출한다. 메타데이터 파일은 반드시 UTF-8 인코딩(UTF-8 Encoding)\*\*으로 작성되며, 파일의 최상단에는 버전과 인코딩 방식을 밝히는 선언부가 포함된다. 선언부 아래에는 애플 뮤직이 정한 네임스페이스(namespace)\*\*\*가 표시된다. 이러한 네임스페이스는 XML 태그가 준수해야 할 스키마(schema)\*\*\*\*를 선언하며, 애플 뮤직은 이를 통해 파일이 규격에 맞게 작성되었는지 검증한다.

메타데이터 파일에는 여러 필수 항목이 포함되며, 각 항목은 특정한 역할을 수행한다. 언어(language) 항목은 메타데이터가 어느 언어로 작성되었는지 지정하고, 데이터 파일(data file) 항목은 실제 음원 파일이 무엇인지 가리킨다. 또한 음원 파일이 전송 과정에서 손상되지 않았는지 확인할 수 있게 해주는 체크섬(checksum)이라는 항목도 존재한다. 애플은 공급자가 제출한 메타데이터의 체크섬 값과 실제로 수신한 파일의 체크섬 값을 비교하여, 두 값이 일치하지 않으면 해당 파일을 거부한다. 이외에도 애플의 기술 사양서는 메타데이터를 작성할 때 지켜야 할 여러 규칙을 명시하고 있으며, 이러한 규칙을 위반한 파일은 처리가 지연되거나 아예 받아들여지지 않을 수 있다. 이와 같이 애플 뮤직의 메타데이터 시스템은 구조화된 형식과 검증 절차를 통해 데이터의 일관성과 신뢰성을 확보하고 있다.

\* XML: 태그를 붙여 각 정보가 무엇을 의미하는지 함께 설명하는 마크업 언어로, 서로 다른 시스템 간에도 데이터를 일정한 형식으로 주고받기 위해 사용

\*\* UTF-8 인코딩(UTF-8 Encoding): 한글, 영어 등의 언어 문자를 컴퓨터가 읽을 수 있는 형태로 변환하는 방식

\*\*\* 네임스페이스(namespace): XML 문서에서 요소-속성 이름 충돌을 방지하기 위해 URI 기반으로 구분하는 식별 체계

\*\*\*\* 스키마(schema): XML 문서의 구조, 데이터 타입, 제약 조건 등을 정의하여 문서의 유효성을 검증하는 규칙 집합

### • AI 투명성 태그 도입에 따른 메타데이터 구조 확장

애플 뮤직이 AI 활용 여부를 표시하기 위해 도입한 AI 투명성 태그는 ▲아트워크(artwork), ▲트랙(track), ▲작곡(composition), ▲뮤직비디오(music video)로 나뉜다. 아트워크 태그는 앨범 단위에서 적용되며, 앨범 커버의 상당 부분이 AI로 생성된 경우 사용된다. 트랙 태그는 트랙 단위에서 적용되며, 사운드 레코딩의 상당 부분이 AI로 생성된 경우에 해당한다. 작곡 태그는 트랙에 구현된 음악 작곡의 상당 부분이 AI로 생성된 경우에 사용되며, 뮤직비디오 태그는 뮤직비디오 제작의 상당 부분이 AI로 생성된 경우 적용된다. 이러한 구분은 AI 활용 지점을 세분화하여 표시하려는 설계로 볼 수 있다.

애플 뮤직은 기존의 메타데이터 구조 내에 AI 투명성 태그를 위한 전용 항목을 새로 추가했다. 공급자는 이 항목에 아트워크, 트랙, 작곡, 뮤직비디오 중 어느 부분에 AI가 사용되었는지 선택하여 입력할 수 있다. AI 투명성 태그를 입력하지 않으면 AI가 사용되지 않은 것으로 간주되며, 애플 뮤직은 이를 "생략 시 없음으로 간주한다"고 기술 사양서에 명시했다. 이 방식은 기존 제출 구조를 크게 변경하지 않으면서 AI 정보를 추가할 수 있다는 점에서 효율적이다.

[그림] AI 투명성 태그 사용 예시 - 작곡, 뮤직비디오, 트랙 제작에서 AI가 사용되었음을 표시

```
<ai_transparencies>  
  <ai_transparency>Composition</ai_transparency>  
  <ai_transparency>Music Video</ai_transparency>  
  <ai_transparency>Track</ai_transparency>  
</ai_transparencies>
```

출처: Apple, "Apple Music Specification 5.3.25", 2026.03.20 접속 기준., <https://help.apple.com/itc/musicspec/en.lproj/static.html#itc57c967017>

### III. 결론: 실효성 있는 AI 투명성 태그 체계의 구축 방향

#### • AI 투명성 태그 자율 신고 방식의 한계와 과제

애플 뮤직의 AI 투명성 태그 도입은 음악 스트리밍 산업이 AI 시대의 투명성 문제를 인식하고 대응하기 시작했다는 점에서 의미를 가진다. 메타데이터라는 기존 인프라를 활용하여 AI 사용 정보를 통합하는 방식은 기술적으로 구현 가능하며, 공급자에게도 큰 부담을 주지 않는다. 또한 네 가지 태그 유형을 통해 AI 활용 지점을 세분화한 것은 단순한 AI 사용 여부를 넘어 창작 과정에서 AI가 개입한 단계까지 드러내려는 시도라는 점에서 의미가 있다.

그러나 애플 뮤직의 AI 투명성 태그 정책은 강제성이 없는 자발적 신고 방식에 기반하며, 참여 여부와 정보의 정확성은 공급자의 판단에 맡겨져 있다. 실제로 애플 뮤직의 메타데이터 시스템은 파일의 기술적 무결성만 검증할 뿐, 입력된 정보의 진위 여부는 확인하지 않는다. 또한 애플의 기술 사양서는 음악 작곡의 '상당 부분(material portion)'이 AI로 생성된 경우 태그를 사용하도록 규정하지만, '상당 부분'에 대한 정량적 기준은 제시하지 않는다. 명확한 기준이 없는 상황에서는 공급자마다 다른 해석이 적용될 수 있으며, 동일한 수준의 AI 활용에도 곡마다 태그 적용 여부가 달라지는 불일치가 발생할 가능성이 있다.

따라서 향후 AI 음악 투명성 체계가 실효성을 갖추기 위해서는 몇 가지 과제가 해결되어야 한다. 우선 산업 차원의 메타데이터 표준을 마련해 AI 활용 정보를 일관된 방식으로 기록할 필요가 있다. 또한 태그 정보의 신뢰성을 확보하기 위한 검증 메커니즘을 도입하고, 장기적으로는 AI 투명성 태그의 의무화 여부도 검토할 필요가 있다. 이러한 제도적 보완이 이루어질 때 AI 음악 투명성 체계는 실질적인 기능을 수행할 수 있을 것이다.



## 참고문헌

- Murray Stassen, "Apple Music launches AI transparency tags — but only if labels and distributors declare them", Music Business Worldwide, 2026.03.04., <https://www.musicbusinessworldwide.com/apple-music-launches-ai-transparency-tags-but-only-if-labels-and-distributors-choose-to-declare-them/>
- Apple, "Apple Music Specification 5.3.25", 2026.03.20. 접속 기준, <https://help.apple.com/itc/musicspec/en.lproj/static.html#itc57c967017>

## 기술용어

순번	용어	설명
1	메타데이터 (metadata)	데이터를 위한 데이터로서 음악 관련해서는 곡명, 아티스트, 장르 등을 기록할 수 있음
2	XML	태그를 붙여 각 정보가 무엇을 의미하는지 함께 설명하는 마크업 언어로, 서로 다른 시스템 간에도 데이터를 일정한 형식으로 주고받기 위해 사용
3	UTF-8 인코딩 (UTF-8 Encoding)	한글, 영어 등의 언어 문자를 컴퓨터가 읽을 수 있는 형태로 변환하는 방식
4	네임스페이스 (namespace)	XML 문서에서 요소·속성 이름 충돌을 방지하기 위해 URI 기반으로 구분하는 식별 체계
5	스키마 (schema)	XML 문서의 구조, 데이터 타입, 제약 조건 등을 정의하여 문서의 유효성을 검증하는 규칙 집합