

저작권 이슈 브리프



COPYRIGHT ISSUE BRIEF

Weekly Report
2026. 8-3



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 음원 차트 내 AI 생성 음원 포함 적법성 논의 지속 확산

▶ AI 생성 음원의 유통이 확대되고 부정 스트리밍 문제가 이어지면서, 음악 차트가 창작 활동과 이용자의 선택을 충분히 반영하고 있는지에 대한 논의가 확산하고 있다. 이에 2026년 7월 11개 음악 기업은 적법한 AI 서비스의 활용, 사람의 실질적인 창작 기여, 재생 실적의 정당성, 관련 법률과 이용약관 준수, AI 활용 사실의 표시 등을 차트 포함 요건으로 제안하였다. 이러한 기준을 실제로 적용하려면 제작자의 AI 활용 정보 제출과 스트리밍 플랫폼의 정보 수집 및 표시 체계가 뒷받침되어야 하지만, 플랫폼마다 판단 기준과 신고 방식이 달라 공통 기준으로 활용하는 데에는 한계가 있다. 향후 해당 원칙이 차트 집계를 넘어 판매량 인증과 산업 내 성과 평가에도 활용될 가능성이 있으나, 업계가 자율적으로 제안한 기준이라는 점과 제안 기업의 이해관계를 함께 고려할 필요가 있다.

산업 빅블루마블 사례로 본 C2PA 기반 영상 출처 증명과 작업 프로세스 통합

▶ 생성형 AI로 합성 및 조작 영상 제작이 쉬워지면서, 유통 이후 진위를 판별하는 탐지 방식만으로는 생성 기술의 발전을 따라가거나 허위 영상의 확산을 막기 어렵다는 한계가 커지고 있다. 이에 C2PA는 콘텐츠의 출처와 편집 및 처리 이력을 암호학적으로 서명한 콘텐츠 자격증명에 기록해, 제작부터 재생까지 영상이 거친 경로를 확인하는 표준을 제시한다. 이러한 방식을 실제 작업 프로세스에 적용한 빅블루마블은 자사 제품 클라우드 비디오 키트의 트랜스코딩 단계에서 자격증명을 함께 생성하고, C2PA 지원 플레이어에서 이를 확인하도록 설계했다. 다만 이 방식이 확산하려면 플레이어와 유통 플랫폼의 표준 수용이 필요하며, 출처 이력만으로 적법성이나 저작권 귀속까지 확인할 수 없으므로 권리 관계는 별도로 판단해야 한다.

산업 AI 음성 워터마크 검증 API 도입에 따른 출처 확인 절차의 변화

▶ 오픈AI는 2026년 7월 31일 챗GPT 음성 대화와 API로 생성되는 오디오에 신스아이디 워터마크를 적용하고, 개발자와 조직이 자체 워크플로에 통합할 수 있는 검증 API를 함께 공개했다. 파일에 부착된 메타데이터는 압축이나 재녹음 과정에서 소실되는 반면, 워터마크는 소리 자체의 스펙트로그램에 심리음향 마스킹 원리로 심겨 유통 이후에도 신호가 남는다. 이에 따라 언론사와 플랫폼, 기업이 각자의 처리 절차 안에서 출처를 조회할 수 있게 됐으나, 검출기는 여전히 한곳에만 있어 제3자는 오디오를 제출해 판정 결과를 받는 구조다. 검증 대상 또한 오픈AI가 생성한 오디오로 한정돼 신호가 없다는 결과가 사람의 음성임을 뜻하지 않으며, 권리 귀속과 침해 여부의 판단은 법적 해석에 따라 결정된다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

산업 챗GPT, 사망한 작가의 문체 모방 요청에도 거부... 어문 저작물 보호 범위의 쟁점

▶ 2026년 7월 공개된 실험에서, 챗GPT가 유명 작가의 문체를 그대로 재현해 달라는 요청에 대해 정확한 모방은 거부하고, 해당 작가 작품의 특징을 반영한 독자적 문장을 제시하는 응답이 확인됐다. 생존 작가와 사망한 작가 모두에 대해 같은 응답이 나타났으나, 오픈AI는 응답 방식이 변경된 시점이나 사유를 밝히지 않았고 공개된 모델 사양 문서에서도 관련 규정이 확인되지 않는다. 미국 저작권법은 문체와 같은 추상적 요소를 보호 대상으로 삼지 않지만 결과물이 원작과 실질적으로 유사하면 침해가 성립할 수 있어 개별 결과물마다 판단이 필요하다. 다만 유사 특성을 반영한 글 생성은 유지되고 모델별 처리 방식도 서로 달라, 제한 대상과 허용 범위를 이용자가 확인하기 어렵다는 과제가 남는다.

산업 넷플릭스 사례로 본 영상 제작 산업 속 AI 활용의 확산 현황

▶ 영상 제작 현장에서 생성형 AI 활용이 시범 적용을 지나 1년 사이 수백 편 규모로 넓어지면서, 활용이 특정 공정에 머무르지 않고 기획과 사전 시각화 단계부터 후반 작업과 공개에 이르는 제작 전 과정에 걸쳐 이루어지는 상황이 확인되고 있다. 이와 함께 제작 도구를 개발하는 기업을 인수하고 시각 효과와 애니메이션 조직을 함께 운영하는 등 활용을 뒷받침하는 기반이 갖춰지는 한편, 제작 파트너에게 활용 계획을 미리 알리도록 요구하는 절차도 마련되어 활용 정보가 제작 관계 안에서 오가는 형태가 나타난다. 다만 활용 규모는 사업 성과 발표에서 한 차례 제시되었을 뿐 개별 작품 단위의 적용 내역은 공개되지 않아, 어떤 작품에 어떤 방식으로 활용되었는지를 밖에서 확인할 수 있는 범위는 제한된 상태에 머물러 있다.

기술 주간 기술 동향

▶ AI 생성 이미지의 출처를 증명하는 워터마크는 유통 과정에서 쉽게 지워진다는 한계를 안고 있다. 기존 기법들은 정보 한 비트를 이미지의 좁은 영역에만 심기 때문에, 공격자가 그 부분만 교란해도 워터마크가 사라진다. 확산 스펙트럼 삽입 기술인 SpreadMark는 각 비트를 이미지 전면에 걸친 무작위 패턴으로 펼쳐 넣고, 검출 시 동일한 패턴과 대조해 읽어낸다. 신호가 넓게 흩어져 있어 이를 지우려면 이미지 전체를 크게 훼손해야 한다. 실험에서 이 기술은 생성 모델로 이미지를 다시 그리는 재생성 공격과 이미지 특징을 압축해 신호를 지우는 공격 양쪽에서 탐지율을 유지한 유일한 기법으로 확인됐다. 다만 흐림 처리나 자르기처럼 패턴 위치가 어긋나는 변형에는 여전히 취약하며, 패턴 묶음이 외부에 노출될 위험도 남아 있다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

음원 차트 내 AI 생성 음원 포함 적법성 논의
지속 확산

AI 생성 음원의 차트 포함에 따른 자격 기준 논의 확산

• AI 생성 음원의 유통 증가에 따른 차트 집계 신뢰성 문제 제기

- 음악 차트는 실제 시장에서 음원이 사용되고 소비된 양을 집계하고, 그 집계된 결과에 따라 성과를 평가하고 보상을 배분하는 기준으로 기능해 왔으나, 최근 AI 생성 음원의 유통이 지속적으로 증가하면서 해당 기능의 신뢰성에 대한 문제 제기 역시 증가하고 있음
- 특히 음반사 측은 스트리밍 플랫폼에 유통되는 AI 생성 음원 다수가 권리자의 허락 없이 음원을 학습에 이용한 AI 서비스로 제작되거나 스트리밍 부정행위(streaming fraud)*와 연계된 것으로 보고 있음. 이에 차트 순위가 실제 창작 활동과 이용자의 선택을 제대로 반영하지 못할 수 있다는 우려가 커짐
- 그동안 업계는 문제가 있는 음원을 유통한 플랫폼을 상대로 법적 조치를 취하며 대응해 왔으나, 이러한 사후 조치만으로는 허락 없이 학습된 AI 서비스로 제작된 음원과 조작된 재생 실적을 차트 집계 전에 가려내기 어려워 차트 포함 자격 요건을 사전에 정해야 한다는 논의가 확산됨

* 스트리밍 부정행위(streaming fraud): 자동화된 계정 등을 이용해 재생 횟수를 인위적으로 늘려 차트 순위나 음원 정산 금액을 왜곡하는 행위

차트 포함 요건과 AI 활용 표시를 결합한 사전 관리 방식의 제안

• 11개 음악 기업, AI 활용 음원의 차트 포함을 제한하는 6개 자격 요건 제안

- 2026년 7월 29일 유니버설뮤직그룹(Universal Music Group), 소니뮤직엔터테인먼트(Sony Music Entertainment) 등 11개 음악 기업*은 권리자의 허락 없이 학습된 AI 서비스로 제작되거나 재생 실적이 조작된 음원을 차트에서 제외하는 차트 포함 자격 원칙을 제안함¹⁾
- 해당 원칙은 2026년 7월 10일 음악 및 창작자 단체가 제시한 'AI 생성'과 'AI 보조' 구분 표시 방안에서 한 단계 더 나아가, AI 활용 사실을 소비자에게 알리는 데 그치지 않고 자격 요건을 충족하지 못한 것으로 판단되는 음원을 차트 집계에서 제외하는 내용을 담음²⁾
- 제안된 6개 요건은 제작에 사용된 AI 서비스의 적법성과 권리자 승인 여부, 사람의 실질적인 창작 기여, 재생 실적 조작 여부 등을 확인하도록 구성되었으며, 적법한 AI 서비스를 활용해 사람이 실질적으로 제작한 음원만 차트 포함을 허용하는 데 초점을 둠

* 11개 음악 기업: 유니버설뮤직그룹, 소니뮤직엔터테인먼트, 워너뮤직그룹(Warner Music Group), 빌리브(Believe), BMG, 콩코드(Concord), 더티 히트(Dirty Hit), 글래스노트 레코드(Glassnote Records), 하이브(HYBE), 맘애파프 뮤직(Mom+Pop Music), 파티잔 레코드(Partisan Records)

1) Ariel King, "UMG, WMG, Sony Music, HYBE, Believe, BMG & More Propose New Principles for AI Song Chart Eligibility", Billboard, 2026.07.29., <https://www.billboard.com/pro/umg-wmg-sony-propose-principles-ai-song-chart-eligibility/>

2) Terrence O'Brien, "The major labels propose rules to keep AI slop off the charts", The Verge, 2026.08.01., <https://www.theverge.com/ai-artificial-intelligence/973741/ai-music-major-record-labels-charts>

[표 1] 11개 음악 기업이 제안한 차트 포함 자격 6개 요건

번호	자격 요건	확인 대상
1	음원 제작에 사용한 생성형 AI 서비스가 적법한 권한을 확보했을 것	제작 도구
2	사람이 음원 제작을 실질적으로 주도했을 것	제작 주체
3	재생 횟수나 차트 순위를 조작한 정황이 없을 것	재생 실적
4	저작권·저작권접권·인격권 관련 법률을 준수했을 것	관련 법률
5	AI 서비스 이용약관을 위반하지 않는 방식으로 음원을 공개했을 것	이용약관
6	AI를 활용한 사실을 소비자에게 표시했을 것	소비자 표시

출처: Terrence O'Brien, "The major labels propose rules to keep AI slop off the charts", The Verge, 2026.08.01.,
<https://www.theverge.com/ai-artificial-intelligence/973741/ai-music-major-record-labels-charts>

• 자격 심사를 뒷받침할 스트리밍 플랫폼의 정보 수집 및 표시 체계 확대

- 차트 집계 기관이 제안된 6개 요건을 심사하려면 음원 제작자가 AI 활용 도구와 방식을 신고하고, 스트리밍 플랫폼이 해당 정보를 수집하고 표시해 판단 근거로 제공해야 함
- 이러한 심사 기반은 플랫폼별로 구축되고 있음. 우선 프랑스 음악 스트리밍 플랫폼 디저(Deezer)가 2025년 AI 생성 음원 추적 시스템을 도입함
- 이어서 애플뮤직(Apple Music), 스포티파이(Spotify), 타이달(Tidal)도 2026년 상반기부터 AI 활용 정보 제출, 관련 정보 공개, 수익화 제외 등의 관리 방안을 차례로 도입함³⁾
- 다만 현재는 플랫폼마다 AI 활용 음원의 판단 기준과 제작자의 신고 방식이 달라, 동일한 음원도 플랫폼별로 다르게 분류될 수 있음. 제안된 원칙이 각국 차트 집계 기관의 공통 기준으로 일관되게 적용되려면 플랫폼 간 정보 수집 및 표시 방식의 차이를 조정할 필요가 있음

[표 2] 주요 스트리밍 플랫폼의 AI 활용 음원 관리 체계 발표 및 도입 현황

플랫폼	발표·도입 시점	주요 내용
디저 (Deezer)	2025년	AI 활용 음원을 자동으로 추적하는 시스템 도입
애플뮤직 (Apple Music)	2026년 3월	신규 콘텐츠에 AI 활용 정보를 담은 '투명성 태그(transparency tag)' 제출 요구
스포티파이 (Spotify)	2026년 4월	아티스트가 유통사를 통해 음원 제작 과정의 AI 활용 방식을 공개할 수 있는 도구 제공
타이달 (Tidal)	2026년 6월	AI 생성 음원을 수익화 대상에서 제외하고 유통사에 관련 정보 공개 책임 부여

출처: Ariel King, "UMG, WMG, Sony Music, HYBE, Believe, BMG & More Propose New Principles for AI Song Chart Eligibility", Billboard, 2026.07.29., <https://www.billboard.com/pro/umg-wmg-sony-propose-principles-ai-song-chart-eligibility/>

• '적법한 권한 확보'와 '사람의 실질적 주도'를 판단할 세부 기준의 필요성

- 제안된 원칙은 차트 포함 자격 요건의 기본 틀을 제시한 수준으로, 개별 요건의 구체적인 판단 기준은 향후 마련해야 할 과제로 남아 있음
- 특히 '사람이 음원 제작을 실질적으로 주도했을 것'이라는 요건은 작곡, 작사, 가창, 편집 과정에서 사람의 관여를 어느 수준까지 인정할지를 두고 차트 집계 기관별 해석이 달라질 수 있음

3) Ariel King, "UMG, WMG, Sony Music, HYBE, Believe, BMG & More Propose New Principles for AI Song Chart Eligibility", Billboard, 2026.07.29.,
<https://www.billboard.com/pro/umg-wmg-sony-propose-principles-ai-song-chart-eligibility/>

- ‘적법한 권한 확보’ 요건도 AI 음악 서비스의 학습 방식이 저작권을 침해하는지에 대한 미국 법원의 판단이 확정되지 않은 상황과 맞물려 있음. 이에 따라 법적 분쟁이 진행 중인 AI 음악 서비스를 활용해 제작한 음원의 차트 포함 자격을 어떻게 판단할지가 쟁점으로 남아 있음
- 또한 요건 충족 여부가 제작자의 자기 신고에 의존하는 만큼, 신고 내용의 사실 여부를 확인할 방법에 대한 논의도 뒤따를 것으로 보임

차트 포함 자격 기준 확대 가능성과 기업별 자율적 기준의 한계

• 판매량 인증과 산업 내 성과 평가로의 자격 기준 확대 가능성

- 제안된 원칙을 공식 채택하겠다고 밝힌 차트 집계 기관은 아직 없으나, 상당수 차트 집계 기관이 음반 산업 단체와 연계되어 있거나 음반사를 주요 고객으로 두고 있어 향후 관련 기준을 도입할 가능성이 있음
- 제안된 원칙이 차트 집계에 반영되면 음원의 판매량과 재생 실적을 인증하는 제도에도 동일한 자격 요건이 적용될 수 있음. 이에 따라 AI 활용 음원의 인정 여부가 차트 순위뿐 아니라 판매 실적 인증과 산업 내 성과 평가에도 영향을 미칠 수 있음

• 업계 이해관계가 반영될 가능성과 운영 기준 구체화 과제

- 제안된 원칙은 법원이나 정부가 정한 의무 기준이 아니라 11개 음악 기업들이 자율적으로 제시한 기준이므로, 실제 적용 범위와 효과는 차트 집계 기관의 채택 여부와 제작자가 제출한 정보의 정확성에 따라 달라질 수 있음
- 또한 기준을 제안한 음반사들은 AI 음악 서비스와 저작권 소송을 진행하거나 이용 계약을 협의하는 당사자이기도 하므로, 요건의 설계와 적법한 AI 서비스의 인정 범위에 자사 이해관계가 반영될 수 있다는 비판이 제기될 가능성이 있음
- 따라서 기준을 실제로 운영하려면 차트 포함을 인정할 AI 서비스의 범위, 제작자가 제출한 정보의 검증 방법, 플랫폼 간 정보 표시 방식 등을 구체화해야 함

참고문헌

- Ariel King, "UMG, WMG, Sony Music, HYBE, Believe, BMG & More Propose New Principles for AI Song Chart Eligibility", Billboard, 2026.07.29., <https://www.billboard.com/pro/umg-wmg-sony-propose-principles-ai-song-chart-eligibility/>
- Ethan Millman, "Major Record Labels Call to Disqualify AI Slop Songs From Global Charts", The Hollywood Reporter, 2026.07.29., <https://www.hollywoodreporter.com/music/music-industry-news/record-labels-call-to-disqualify-ai-slop-songs-from-charts-1236659042/>
- Chris Cooke, "Record labels come together to propose rules for when AI-generated tracks can get into the music charts", Complete Music Update, 2026.07.29., <https://completemusicupdate.com/record-labels-come-together-to-propose-rules-for-when-ai-generated-tracks-can-get-into-the-music-charts/>
- Terrence O'Brien, "The major labels propose rules to keep AI slop off the charts", The Verge, 2026.08.01., <https://www.theverge.com/ai-artificial-intelligence/973741/ai-music-major-record-labels-charts>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

빅블루마블 사례로 본 C2PA 기반 영상 출처 증명과 작업 프로세스 통합

합성 및 조작 영상의 사후 탐지 한계와 출처 증명 요구 확대

• 생성 기술의 발전 속도를 따라가기 어려운 탐지 중심 대응의 한계

- 생성형 AI의 확산으로 실제 촬영물과 구분하기 어려운 합성 및 조작 영상을 누구나 손쉽게 제작할 수 있게 되면서, 허위 정보가 영상 형태로 널리 퍼질 수 있다는 우려가 커지고 있음
- 지금까지는 개별 영상의 진위를 판별하는 탐지 기술을 중심으로 대응 방안이 논의돼 왔으나, 탐지 기술이 빠르게 고도화되는 생성 기술을 지속해서 따라가기 어렵다는 한계점이 있음
- 더욱이 탐지 방식은 문제가 된 영상이 이미 확산된 이후에야 조치할 수 있어, 방송사, 공공기관, 언론사처럼 콘텐츠 신뢰성이 중요한 기관에는 사전 대응의 필요성이 요구됨

• 제작부터 재생까지 전 과정의 출처 관리와 C2PA 표준 적용 필요성

- 이에 따라 개별 영상의 진위를 확인하는 수준을 넘어, 제작부터 재생까지 콘텐츠가 거쳐 온 전체 경로를 증명하는 출처 관리 체계를 마련하는 일이 과제로 대두됨
- 이를 해결하기 위한 기술적 수단 중 하나로 2021년 콘텐츠 출처 및 진위 검증 표준 연합체(이하 C2PA)*가 마련한 표준이 있음. C2PA 표준은 콘텐츠의 출처와 편집 및 처리 이력을 암호학적으로 서명된 메타데이터인 콘텐츠 자격 증명(Content Credentials)**에 기록하는 방식을 제시함¹⁾
- 다만 출처 증명 정보의 생성과 확인 절차가 기존 제작 및 유통 작업 프로세스와 분리되면, 대량의 영상을 처리하는 환경에서는 업무 부담이 커질 수 있음. 따라서 도입 과정에서는 증명 절차를 실제 작업 프로세스에 어떠한 방식으로 결합하는지가 매우 중요함

* C2PA(Coalition for Content Provenance and Authenticity): 디지털 콘텐츠의 출처와 생성·편집 이력을 암호학적으로 기록하고 검증하는 기술 표준을 개발하는 연합체

** 콘텐츠 크리덴셜(Content Credentials): 디지털 콘텐츠 파일에 결합되는 암호학적으로 서명된 메타데이터로, 콘텐츠의 출처와 제작에 사용된 도구, 편집 및 처리 이력을 기록함

1) Big Blue Marble, "Big Blue Marble brings C2PA Content Credentials to Cloud Video Kit", 2026.07.21., <https://www.bigbluemarble.com/en/resources/post/c2pa-content-credentials/>

빅블루마블, 영상 처리 과정에 C2PA 출처 증명 기능 통합

• 트랜스코딩 단계에서 콘텐츠 자격 증명을 함께 생성하는 방식

- 출처 증명 절차가 기존 작업 프로세스와 분리될 때 발생하는 부담을 줄이기 위해, 폴란드 영상 처리 기업 빅블루마블(Big Blue Marble)은 콘텐츠 자격 증명 생성을 통상적인 영상 변환 과정에 결합하는 방식을 제시함
- 해당 기업은 2026년 7월 C2PA 적합성 목록(Conforming Products List)*에 등재돼 자사 영상 처리 서비스가 표준 요구사항을 충족한다는 점을 확인받았으며, 자사 제품인 클라우드 비디오 키트(Cloud Video Kit)에서 유효한 콘텐츠 자격 증명을 생성할 수 있게 됨²⁾
- 이를 바탕으로 주문형 비디오(Video on Demand, VOD)에 C2PA 서명을 선택 기능으로 제공할 예정이며, 이 기능은 영상을 재생 환경별 규격으로 변환하는 트랜스코딩(transcoding)** 단계에서 콘텐츠 자격 증명도 함께 생성되도록 설계됨
- 이로써 대량의 영상을 다루는 유통 주체는 별도의 도구를 도입하거나 증명 절차를 따로 운영하지 않고도 기존 작업 프로세스 안에서 C2PA 기반 출처 증명을 적용할 수 있게 될 예정임

* 적합성 목록(Conforming Products List): C2PA가 표준 요구사항의 충족 여부를 확인해 적합 판정을 받은 제품·서비스를 공개하는 목록

** 트랜스코딩(transcoding): 영상을 다양한 기기와 네트워크 환경에서 재생할 수 있도록 형식, 해상도, 비트레이트를 변환하는 처리 과정

• 플레이어 화면에서의 출처 이력 표시와 저작자 확인 규격 연계

- 생성된 콘텐츠 자격 증명도 C2PA 지원 플레이어에서 무료 웹 검증 도구를 통해 표시되며, 시청자는 재생 화면을 벗어나지 않고 해당 영상의 출처와 편집 및 처리 이력을 확인할 수 있음
- 빅블루마블은 이 기능을 통해 방송사, 공공기관, 언론사가 영상의 제작부터 재생까지 각 단계의 이력을 증명하는 출처 추적 체계(chain of custody)를 확보할 수 있다고 설명함²⁾

* 출처 추적 체계(chain of custody): 콘텐츠의 제작부터 편집·유통·재생까지 각 단계에서 발생한 처리와 전달 이력을 연속적으로 기록·확인하는 체계

[표1] 빅블루마블의 C2PA 적합성 등재 내역

구분	내용	구분	내용
등재일	2026.7.11.	보증 수준	Level 1
적합 분류	생성자(Generator)	생성 지원 형식	영상 mp4 / 오디오 flac·m4a·mp3·wav / 이미지 jpeg·png
적용 사양 버전	C2PA 2.2	검증 지원 형식	영상 mp4

출처: C2PA, "Conformance Explorer", 2026.08.07. 조회, <https://spec.c2pa.org/conformance-explorer/>

2) Big Blue Marble, "Big Blue Marble brings C2PA Content Credentials to Cloud Video Kit", 2026.07.21., <https://www.bigbluemarble.com/en/resources/post/c2pa-content-credentials/>

C2PA 출처 증명의 현장 적용 의미와 확산 과제

• 영상 처리 과정 내재화의 효과와 적용 범위의 한계

- 빅블루마블 사례는 C2PA 기반 사전 증명 절차를 별도의 업무로 운영하지 않고, 트랜스코딩과 같은 통상적인 영상 처리 과정에 포함할 수 있음을 보여줌
- 다만 시청자가 콘텐츠 자격증명을 확인하려면 C2PA 지원 플레이어는 필요하므로, 확산 여부는 콘텐츠 제작 기업의 도입을 넘어 플레이어와 유통 플랫폼의 표준 수용에 영향을 받을 것으로 보임
- 또한 콘텐츠의 출처와 편집 및 처리 이력을 증명하더라도 적법성이나 저작권 귀속까지 보장되는 것은 아니므로, 권리 관계는 별도의 법적 판단이 필요함
- 이러한 한계에도 콘텐츠 신뢰성에 대한 요구가 높아지는 방송 및 미디어 산업에서는 저작자 확인 규격과 결합된 출처 증명이 영상 유통 과정의 기본 관리 수단으로 자리 잡을 가능성이 있음

참고문헌

- Big Blue Marble, "Big Blue Marble brings C2PA Content Credentials to Cloud Video Kit", 2026.07.21., <https://www.bigbluemarble.com/en/resources/post/c2pa-content-credentials/>
- Matthew Corrigan, "Big Blue Marble adds C2PA verification to Cloud Video Kit", TVBEurope, 2026.07.21., <https://www.tvbeurope.com/media-management/big-blue-marble-adds-c2pa-verification-to-cloud-video-kit>
- TVNewsCheck, "Big Blue Marble Brings C2PA Content Credentials To Cloud Video Kit", 2026.07.21., <https://tvnewscheck.com/tech/article/big-blue-marble-brings-c2pa-content-credentials-to-cloud-video-kit/>
- C2PA, "Conformance Explorer", 2026.08.07. 조회, <https://spec.c2pa.org/conformance-explorer/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

AI 음성 워터마크 검증 API 도입에 따른 출처 확인 절차의 변화

생성 단계의 보호 장치와 유통 이후 출처 확인의 한계

• 실시간 생성 음성의 적용과 생성 시점에 한정된 보호 조치

- 미국의 AI 기업 오픈AI(OpenAI)는 2026년 7월 AI 음성 모델 GPT-라이브(GPT-Live)를 공개하였으며¹⁾, 공개된 모델은 음성 입력의 처리와 출력 생성을 동시에 수행하는 방식을 적용해 말하기와 듣기, 침묵 등 대화상태 여부를 초당 여러 번 판단하도록 설계됨
- AI 모델 챗GPT(ChatGPT)의 음성 대화와 받아쓰기 기능은 매주 1억 5천만 명이 넘는 이용자가 사용하고 있으며¹⁾, 음성 대화 기능에 GPT-라이브 모델이 기본적으로 적용되나 기존 음성 모드도 일부 상황에서는 유지됨
- 생성 단계에서는 챗GPT가 제공하는 기본 음성을 사용하고 실제 사람의 목소리를 모방하지 못하도록 하는 보호 장치가 적용되나, 이는 음성을 만들어 내는 시점에 적용되는 조치로 생성 이후의 유통 단계와는 구분됨

• 메타데이터와 발급자 의존형 검출 방식의 한계

- 파일에 부착된 메타데이터(metadata) 태그는 압축과 재인코딩(Re-encoding)*을 거치거나 스피커로 재생한 뒤 다시 녹음되는 과정에서 소실되므로, 음성이 유통된 이후 단계에서 메타데이터를 통해 출처를 확인하는 데에는 어려움이 존재함
- 이에 파일 외부가 아니라 콘텐츠 자체에 신호를 심는 워터마크(watermark) 방식으로 출처를 확인했으나, 워터마크 검출은 제3자가 독립적으로 실행할 수 없어 확인을 요청하려면 발급 주체의 확인을 거쳐야 하는 한계점이 있음

* 재인코딩(Re-encoding): 이미 압축된 오디오나 비디오 같은 디지털 미디어 파일을 풀어서(Decoding) 원하는 포맷, 해상도, 코덱으로 다시 압축(Encoding)하는 과정

콘텐츠 내재형 워터마크와 검증 API를 결합한 출처 확인 구조

• 검증 API 제공과 제3자 업무 절차의 연동

- 오픈AI는 2026년 7월 31일 챗GPT 음성 대화와 API(Application Programming Interface)*를 통해 생성되는 오디오에 신스아이디(SynthID)** 워터마크를 적용하고, 공개 검증 도구***에서 해당 오디오에 자사 생성 신호가 들어 있는지 확인할 수 있게 함

1) OpenAI, "GPT-Live 소개", 2026.07.08., <https://openai.com/ko-KR/index/introducing-gpt-live/>

- 또한 개발자와 조직이 자체 워크플로에 출처 검증 기능을 통합할 수 있는 검증 API가 함께 제공되어, 이용자가 웹사이트에 파일을 올리던 방식과 달리 시스템 간 연동을 통한 검증 절차로 구성할 수 있게 됨
- 이에 따라 제출된 오디오 파일을 심사하는 언론사, 업로드된 음성을 검토하는 플랫폼, 외부로 보내는 녹음을 확인하는 기업은 각자의 처리 절차 안에서 출처 조치가 가능함

* API(Application Programming Interface): 서로 다른 소프트웨어나 시스템이 기능과 데이터를 주고받을 수 있도록 정해 놓은 연결 방식 또는 규칙

** 신스아이디(SynthID): 구글 산하의 인공지능 연구·개발 조직인 구글 딥마인드가 개발한 워터마킹 기술로, 생성된 콘텐츠에 사람이 인지하기 어려운 신호를 삽입해 해당 콘텐츠의 생성 출처를 확인할 수 있도록 함

*** 공개 검증 도구: 오픈시가 웹사이트에 공개한 확인 창구로, 이용자가 파일을 직접 올리면 해당 파일에 자사 모델이 생성한 출처 신호가 들어 있는지를 알려줌

• 유통 과정의 변형에도 신호를 유지하는 삽입 방식

- 출처 조화에 조건이 붙는 이유는 신호를 파일이 아니라 소리 안에 심기 때문인데, 신스아이디는 음성 파형을 스펙트로그램(spectrogram)*으로 바꾼 뒤 심리음향 마스킹(psychoacoustic masking)** 원리에 따라 그 안에 패턴을 심고, 수정된 스펙트로그램에서 다시 파형을 만들어 내는 구조이기 때문임
- 사람의 귀는 더 큰 소리에 가려진 신호를 인지하지 못하는데 MP3 압축도 같은 원리로 데이터를 떨어내므로, 신스아이디는 바로 그 가려지는 영역에 출처 신호를 배치하되 같은 신호를 복제한 사본을 여러 주파수 대역과 시점에 나누어 넣음
- 그 결과 MP3 압축을 거치거나 속도와 음높이를 바꾸고 스피커로 재생한 소리를 마이크로 다시 녹음해도 신호가 남는 것으로 보고되고 있으나²⁾, 출처 신호의 제거를 시도한 연구도 제시돼 성능 수치에 대해서는 의견이 분분한 상황임

* 스펙트로그램(spectrogram): 소리의 주파수 성분이 시간에 따라 어떻게 변하는지를 그림처럼 나타낸 표현으로, 파형만으로는 드러나지 않는 음역대별 세기를 확인할 수 있음

** 심리음향 마스킹(psychoacoustic masking): 큰 소리 근처의 작은 소리를 사람의 귀가 인지하지 못하는 현상으로, 음원 압축에서 버릴 데이터를 고르는 기준으로 쓰임

• 중앙화된 검출 구조와 상호운용성의 한계

- 신호는 여러 대역에 분산해 삽입하지만 이를 대조하는 정합 필터(matched filter)* 검출기는 한 곳에만 있어, 제3자는 검출을 자체적으로 실행하지 못하고 오디오를 제출해 판정 결과를 받는 구조임
- 이처럼 검출이 한 곳에 묶여 있으면 다른 사업자의 도구로는 확인할 수 없어, 유럽연합 인공지능법(EU AI Act)**이 표시 기술의 요건으로 둔 상호운용성과 어긋나는 지점이 생김
- 이에 따라 현재 유럽연합 인공지능법의 4가지 표시 기술 요건을 모두 충족하는 워터마킹 기술은 없다는 분석이 제기되며, 유럽연합의 실무 지침으로는 메타데이터 삽입과 비가청 워터마크를 함께 두는 결합 방식이 요구됨

* 정합 필터(matched filter): 신호를 심을 때 사용한 패턴을 미리 알고 있는 상태에서 그 패턴이 남아 있는지 대조하는 검출 방식으로, 잡음이 섞이거나 압축을 거친 뒤에도 신호를 찾아낼 수 있음

** 유럽연합 인공지능법(EU AI Act): 인공지능 시스템을 위험 수준에 따라 구분해 규율하는 유럽연합의 법으로, 생성형 인공지능 산출물에 대한 표시 및 고지 요건을 함께 담고 있음

²⁾ Shannon Harwood, "GPT-Live Voice Gets SynthID Watermarks One Day Before EU AI Act Enforcement", TechTimes, 2026.08.01., <https://www.techtimes.com/articles/322592/20260801/gpt-live-voice-gets-synthid-watermarks-one-day-before-eu-ai-act-enforcement.htm>

시사점: 출처 확인 범위의 한계와 권리 판단의 분리

• 출처 신호의 검증 범위와 권리 판단

- 이번 검증 API는 오디오에 오픈AI가 심은 출처 신호가 들어 있지만 판별하므로, 다른 상용 음성 AI 모델이나 오픈소스(open source)* 모델의 산출물은 해당 검증 API의 확인 대상이 되지 못하고 신호가 없다는 결과도 사람의 음성임을 뜻하지 않음
- 누구나 내려받아 쓸 수 있는 오픈소스 음성 모델은 사업자를 거치지 않아 신호가 심기지 않는데, 유럽연합 인공지능법은 이러한 모델도 완성된 시스템인 경우 표시 의무 대상에 해당하므로 의무는 있으나 표시할 수단이 없는 상태가 됨
- 또한 출처 신호는 어떤 시스템이 오디오를 만들었는지 알려주는 정보에 그치므로, 저작권자나 실연자** 등 권리 귀속과 침해 여부의 판단은 법적 해석에 따라 결정된다고 분석됨

* 오픈소스(open-source): 소프트웨어의 설계도에 해당하는 소스코드를 공개해 누구나 정해진 라이선스 범위에서 열람·사용·수정·배포할 수 있도록 한 방식

** 실연자: 노래·연주·연기·무용 등을 통해 저작물을 실제로 표현하거나 전달하는 사람

참고문헌

- Shannon Harwood, "GPT-Live Voice Gets SynthID Watermarks One Day Before EU AI Act Enforcement", TechTimes, 2026.08.01., <https://www.techtimes.com/articles/322592/20260801/gpt-live-voice-gets-synthid-watermarks-one-day-before-eu-ai-act-enforcement.htm>
- OpenAI, "GPT-Live 소개", 2026.07.08., <https://openai.com/ko-KR/index/introducing-gpt-live/>
- Lukas Willecke, "Transparency Obligations for AI-Generated Content: The Code of Practice Adequacy Decision and the Final EU Commission Guidelines on Article 50 AI Act", Reed Smith, 2026.07.28., <https://www.reedsmith.com/our-insights/blogs/viewpoints/102nbz0/transparency-obligations-for-ai-generated-content-the-code-of-practice-adequacy/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

챗GPT, 사망한 작가의 문체 모방 요청에도 거부... 어문 저작물 보호 범위의 쟁점

특정 작가의 문체 모방 요청에 대한 응답 방식의 변화

• 사망한 작가까지 확대된 문체 모방 거부 응답

- 2026년 7월 28일 공개된 실험 결과에 따르면, AI 모델 챗GPT(ChatGPT)는 특정 작가의 문체를 그대로 재현해 달라는 요청에 대해 정확한 모방은 거부하되 해당 작가 작품에서 폭넓게 나타나는 특징을 반영한 독자적 문장을 제시함¹⁾
- 실험에서는 스티븐 킹(Stephen King), J.K. 롤링(J.K. Rowling) 등 생존 작가와 찰스 디킨스(Charles Dickens), 어니스트 헤밍웨이(Ernest Hemingway) 등 사망한 작가가 함께 대상이 되었으며, 두 경우 모두에서 같은 방식의 응답이 나타남
- 한편 사망한 작가인 애거사 크리스티(Agatha Christie)를 대상으로 한 또 다른 실험에서는 해당 작가의 작품이 여전히 저작권 보호를 받고 있다는 점이 정확한 모방을 거부한 사유로 제시됨²⁾

• 공식 정책으로 확인되지 않은 응답 기준

- 같은 달 앞서 공개된 외부 분석에서는 생존 작가에 대한 요청만 거부되고 사망한 작가에 대한 요청에는 응답이 이루어진 것으로 확인돼, 짧은 기간 사이에 처리 결과가 달라진 것으로 나타남¹⁾
- 챗GPT를 개발한 오픈AI(OpenAI)는 이번 변화에 관해 별도의 설명을 내놓지 않았으며 언론의 논평 요청에도 답하지 않아, 이번 변화가 공식 정책에 따른 것인지는 확인되지 않은 상태임
- 오픈AI가 공개한 모델 사양 문서(Model Spec)*에서는 텍스트에 대해 저작권 자료의 복제나 문체 모방을 명시적으로 금지하는 규정이 확인되지 않은 반면, 이미지 생성 모델인 달리3(DALL·E 3)에 대해서는 생존 예술가의 스타일을 따른 이미지 요청을 거부하도록 설계되었다고 명시됨

* 모델 사양 문서(Model Spec): 오픈AI가 자사 모델의 응답 원칙과 허용·제한 범위를 정리해 공개한 문서

문체 비보호 원칙과 유사성 판단 사이에서 작동하는 ‘응답 제한’

• 정확한 모방 제한과 유사 특성 제공 사이의 경계

- 챗GPT는 특정 작가의 문체를 정확히 모방할 수는 없다고 응답하면서도, 해당 작가 작품에 공통적으로 나타나는 요소를 반영해 비슷한 느낌을 살린 문장은 함께 제공함

1) Kyle Orland, "ChatGPT starts blocking direct requests to copy an author's style", Ars Technica, 2026.07.28., <https://arstechnica.com/ai/2026/07/chatgpt-stops-cloning-famous-writers-voices-but-may-capture-a-similar-feeling/>

2) Mariella Moon, "ChatGPT is now refusing requests to write in famous authors' styles", Engadget, 2026.07.28., <https://www.engadget.com/2224867/chatgpt-refusing-to-write-in-famous-authors-styles/>

- 이번 실험에서는 특정 작가를 지목한 문체 모방은 제한되나, 유사 특성을 반영한 글 생성은 허용되고 있어 응답 결과만으로는 구체적인 제한의 범위를 확인하기 어려움
- 챗GPT의 이러한 응답 방식에 대해서는 이용자에게 해당 작가의 특징을 살린 글을 함께 제안한다는 점에서 모방 요청에 대한 완전한 거부로 보기는 어렵고 면책 문구에 가깝다는 평가가 제기됨³⁾
- 한편 오픈AI는 작가들이 제기한 다수의 저작권 소송에 대응 중이며, 이 중 한 소송에서는 저작권 보호를 받는 기존 텍스트와 유사한 문장을 생성하는 챗GPT의 능력도 쟁점으로 언급됨

• 동일 요청에 대한 모델별 처리 방식의 차이

- AI의 문체 모방에 관한 외부 분석에서는 해당 요청에 단서 없이 응답하는 모델, 거부한 뒤 다른 방식을 제안하는 모델, 응답하되 단서를 함께 제시하는 모델로 처리 방식이 구분됨
- 이 중 거부 방식을 택한 모델도 요청을 그대로 차단하기보다는 다른 방향의 결과물을 제안했고, 요청에 응한 모델 가운데 일부도 모방 쟁점을 의식한 단서를 함께 제시해, 양쪽 모두 모방 요청을 그대로 처리하지는 않은 것으로 나타남

[표1] 문체 모방 요청에 대한 AI 모델별 처리 방식

모델명	요청 처리	부가 대응
챗GPT	정확한 모방 거부	유사 특성을 반영한 문장 제시
제미나이(Gemini)	요청 수용	별도 단서 없음
퍼플렉시티(Perplexity)	요청 거부	다른 방식의 결과물 제안
클로드(Claude)	요청 수용	모방 쟁점에 대한 단서 부기
코파일럿(Copilot)	요청 수용	모방 쟁점에 대한 단서 부기

출처: 참고문헌 종합하여 재구성

시사점: 문체 모방 제한의 불명확성과 기준 공개 과제

• 기준 비공개에 따른 제한 범위의 불명확성

- AI 문체 모방 요청과 관련해서는, 정확한 모방과 유사 특성 반영을 구분하는 모델별 세부 기준이 공개되지 않아 이용자와 관리자가 어떤 요청이나 결과물이 제한 대상인지 미리 확인하기 어려운 상태가 지속될 것으로 분석됨
- 또한 결과물의 침해 여부는 원작과의 실질적 유사성에 따라 개별적으로 판단되는 만큼, 문체 모방 요청에 대한 응답 제한만으로 침해 가능성을 차단하기에는 한계가 있음
- 이와 별개로, 이번에 확인된 응답 제한은 이용자의 요청을 처리하는 단계에서 이루어지는 조치이므로, AI 학습 단계의 저작물 이용을 둘러싼 기존 쟁점과는 구분해서 볼 필요가 있음
- 다만 이번 응답 제한은 학습 데이터를 둘러싼 작가들의 소송에 대응하는 맥락에서 해석되고 있으나, 조치 자체는 이용자의 요청을 처리하는 단계에서 이루어지는 것이므로 두 층위는 구분해 볼 필요가 있음

3) Mariella Moon, "ChatGPT is now refusing requests to write in famous authors' styles", Engadget, 2026.07.28., <https://www.engadget.com/2224867/chatgpt-refusing-to-write-in-famous-authors-styles/>

• 모델별 기준 차이와 처리 기준 공개 과제

- 전미작가협회(Authors Guild)*는 AI를 사용해 다른 작가의 문체를 의도적으로 모방하지 말라고 권고해 왔으나, 이는 작가를 대상으로 한 지침이어서 AI 서비스를 운영하는 사업자가 어떤 요청을 제한할지에 대한 기준은 별도의 과제로 남아 있음
- 동일한 요청이라도 모델에 따라 처리 결과가 달라지는 만큼, 문체 모방 요청의 제한 대상과 허용 범위에 대한 사업자별 처리 기준의 공개가 향후 검토 과제로 남을 것으로 보임

* 전미작가협회(Authors Guild): 미국의 출판 작가를 대표하는 전문 단체로, 저작권 보호와 공정한 계약 보상, 표현의 자유 등 작가의 권익 보호 활동을 수행함

참고문헌

- Kyle Orland, "ChatGPT starts blocking direct requests to copy an author's style", Ars Technica, 2026.07.28., <https://arstechnica.com/ai/2026/07/chatgpt-stops-cloning-famous-writers-voices-but-may-capture-a-similar-feeling/>
- tks24, "ChatGPT、有名作家の文体を模倣するリクエストを拒否するよう仕様変更。単なる免責事項との指摘も", INTERNET Watch, 2026.07.29., <https://internet.watch.impress.co.jp/docs/yajiuma/2128785.html>
- Mariella Moon, "ChatGPT is now refusing requests to write in famous authors' styles", Engadget, 2026.07.28., <https://www.engadget.com/2224867/chatgpt-refusing-to-write-in-famous-authors-styles/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

넷플릭스 사례로 본 영상 제작 산업 속 AI 활용 확산 현황

영상 제작 전 과정으로 확대된 생성형 AI 활용

• 시범 적용에서 실제 작품 적용으로 이어진 활용 범위의 변화

- 미국의 엔터테인먼트 기업이자 OTT(Over-the-top)* 사업자인 넷플릭스(Netflix)는 2026년 7월 공개한 2분기 실적 보고서를 통해 2026년 약 300편의 프로그램 제작 과정에서 생성형 AI를 활용했다고 발표함¹⁾
- 2025년 공개된 아르헨티나의 SF 시리즈 '디 에테르나우츠(The Eternauts)'의 일부 장면에만 한정해 적용되었던 활용 범위가 1년 만에 수백 편으로 확대되며, 시범 적용 단계를 지나 제작 현장의 일반적인 작업 방식으로 자리 잡은 양상이 나타남
- 넷플릭스가 밝힌 AI 활용 범위는 기획과 사전 시각화 단계부터 후반 작업까지 이어지는 제작 전 과정에 걸쳐 있는 것으로 나타났으며, 실제 작업은 후반 작업에 집중된 것으로 알려짐

* OTT(Over-the-top): 케이블이나 위성 기반 공급자를 거치지 않고 인터넷을 통해 시청자에게 직접 콘텐츠를 배포하는 디지털 서비스

• 비용과 제작 일정의 제약 완화를 중심으로 제시된 활용 동기

- 넷플릭스는 기존 제작 방식보다 짧은 시간과 낮은 비용으로 더 높은 품질의 결과물을 확보할 수 있다는 점을 활용 확대의 근거로 제시하며, 도구 도입을 인력 대체가 아닌 제작 역량 보완으로 설명함
- 미국 독립 전쟁을 다룬 다큐멘터리 '디 아메리칸 익스페리먼트(The American Experiment)'에서는 17분 분량의 영상에 생성형 AI가 적용되었으며, 제작 기간은 기존 방식 대비 2배 빠르게 진행됐고, 비용은 절반으로 줄여 제작된 것으로 확인됨
- 또한, 인도의 스포츠 스릴러 시리즈 '글로리(Glory)'에서는 경기장 관중 규모를 확대하는 방식으로 AI가 활용되었으며, 이로 인해 기존에는 예산 문제로 제외될 수 있었던 중요한 장면을 포함시켜 제작을 완료할 수 있었음

도구 확보와 고지 절차, 사업 영역 확대로 나타난 AI 활용 사례

• 자체 촬영 소재를 입력으로 삼는 도구 방식과 제작 기반의 내부 확보

- 넷플릭스는 2026년 3월 AI 기반의 영상 제작 도구를 개발해 온 인터포지티브(InterPositive)를 6억 달러(약 8,647억 원)²⁾에 인수하며, 콘텐츠 제작 전반에서 창작자가 사용할 도구를 확보하려는 움직임을 보이고 있음

1) Corbin Bolies, "About 300 Netflix Titles Used Generative AI This Year, Company Reveals", Variety, 2026.07.16., <https://variety.com/2026/biz/news/about-300-netflix-programs-used-ai-this-year-q2-earnings-1236812914/>

2) 1달러=1,441.10원(2026.08.01, KEB 하나은행 매매기준율 적용, 이하 동일)

- 해당 도구는 텍스트 프롬프트로 영상을 생성하는 방식과 달리, 제작진이 촬영한 원본 영상을 입력 자료로 삼아 조명 조정, 시각 효과 수정, 편집 작업을 처리하는 등 기존 촬영본의 완성도를 높이는 용도에 초점이 맞춰짐
- 또한, 넷플릭스는 시각 효과를 연구하는 조직과 자체 애니메이션 조직을 함께 운용하며 인터포지티브와 함께 세 조직이 제작 과정에서 연계하도록 함

• 제작 파트너에게 요구되는 사전 고지와 제한적인 대외 공개

- 넷플릭스는 제작 파트너에게 생성형 AI 사용 계획을 사전에 알리도록 요구하며, 새로운 도구가 계속 등장해 성능과 유의 사항이 서로 다르다는 점을 이러한 요구의 근거로 제시함
- 또한, 넷플릭스는 AI 도구 활용이 투명하고 책임 있는 방식으로 이루어질 때 유용하다는 입장을 웹사이트를 통해 공개했으며, 고지 요구 역시 전 세계 규모의 제작을 지원하고 업계에서 통용되는 방식에 맞추기 위한 것으로 설명함
- 다만, 약 300편이라는 활용 규모는 사업 성과 발표를 통해 한 차례 공개되었으나, 어떤 작품에 어느 단계에서 적용되었는지는 별도로 제시되지 않아 개별 작품 단위의 확인이 어려운 상태임
- 이에 따라 AI 도구 활용 계획은 제작 파트너에서 넷플릭스로 전달되고 활용 규모는 넷플릭스에서 투자자에게 전달되는 형태가 되어, 활용 정보가 오가는 범위가 제작과 사업 관계 안에 놓인 것으로 나타남

시사점: 제작 단계 활용의 확대 속도와 확인 가능한 범위

• 빠른 속도로 증가하는 활용 규모와 작품 범위

- 생성형 AI를 활용한 작품은 2025년 한 편에서 1년 뒤 약 300편이라는 활용 규모가 공개되었으며, 2026년 3월 제작 도구 기업 인수와 7월 활용 규모 공개가 잇따른 것으로 나타남
- 2025년 첫 사례에서는 기존 방식보다 약 10배 빠르게 작업이 완료된 것으로 확인되었고, 2026년 사례에서는 2배 속도와 절반 비용이 제시되어 활용 결과가 수치로 제시된 사례는 두 시점에 한정됨
- 다만, 현재까지 넷플릭스가 AI를 활용한 것이 확인되는 작품은 사례로 언급된 세 편에 한정되고, AI 활용 작품의 전체 목록을 공개하지 않고 있음
- 또한, 개별 작품 단위로 활용 여부와 적용 범위를 확인할 수 있는 자료는 제시되지 않은 상태임

참고문헌

- Plataforma Media, "Netflix reveals it used AI in 300 productions in 2026", 2026.07.17., <https://www.plataformamedia.com/en/2026/07/17/netflix-ai-used-300-productions-investor-letter>
- Corbin Bolies, "About 300 Netflix Titles Used Generative AI This Year, Company Reveals", Variety, 2026.07.16., <https://variety.com/2026/biz/news/about-300-netflix-programs-used-ai-this-year-q2-earnings-1236812914/>
- BigGo Finance, "Netflix Reveals AI Touched Nearly 300 Titles in 2026, Slashing Costs in Half on Key Scenes", 2026.07.17., <https://finance.biggo.com/news/b4b32fd6-c6e2-4351-abfb-6153dc8f6075>

저작권 이슈 브리프

SUMMARY

산업/기업

기술

주간 기술 동향

이미지 전면에 신호를 분산하는 확산 스펙트럼 워터마킹 기술, SpreadMark

• 워터마크 제거 공격의 확산과 이미지 전면 분산 삽입 기술의 부상

생성형 AI가 만들어낸 이미지와 영상이 실물과 구분하기 어려운 수준에 이르면서, 콘텐츠에 출처 표시를 남기는 일은 선택이 아닌 기본 요건으로 자리 잡고 있다. 이에 따라 주요 생성형 AI 서비스 기업들은 생성형 AI가 만들어낸 콘텐츠에 기계가 판독할 수 있는 워터마크를 삽입하는 체계를 갖추기 시작했다. 워터마크가 남아 있으면 부여된 식별 정보를 통해 생성 출처 등을 확인하는데 활용할 수 있고, 이는 딥페이크 대응과 저작권 추적 모두의 출발점이 된다. 문제는 이 워터마크가 실제 유통 환경에서 얼마나 오래 살아남는가에 있다.

현재 통용되는 워터마킹 기술은 요구되는 수준을 충분히 충족하지 못한다는 평가를 받고 있다. 워터마킹 기술이 갖춰야 할 조건으로는 유효성, 상호운용성, 강건성, 신뢰성이 꼽히는데, 어떤 단일 기술도 이를 모두 만족하지 못한다는 지적이 이어지고 있다. 각 기술이 얼마나 견고한지를 비교할 공통 평가 기준도 아직 정립되는 과정에 있는 것으로 알려져 있다. 이러한 상황에서 기술의 신뢰도를 객관적으로 판단하기는 쉽지 않은 것으로 보인다.

특히 문제가 되는 것은 이미지가 여러 과정을 거치는 동안 워터마크가 흐려지거나 지워진다는 점이다. 파일에 서명된 이력 정보를 기록하는 메타데이터 방식은 일상적인 편집 과정만으로도 손쉽게 제거된다. 이미지 픽셀에 직접 신호를 심는 비가시 워터마크 방식은 일상적인 편집에는 상대적으로 잘 견디지만, 강도 높은 이미지 처리 앞에서는 여전히 훼손된다. 최근에는 두 방식과 기록 보존을 병행하는 다층 구성이 권고되고 있으나, 이는 서로의 취약점을 보완하도록 배치한 것일 뿐 개별 기술이 가진 약점이 해결되었음을 뜻하지는 않는다.

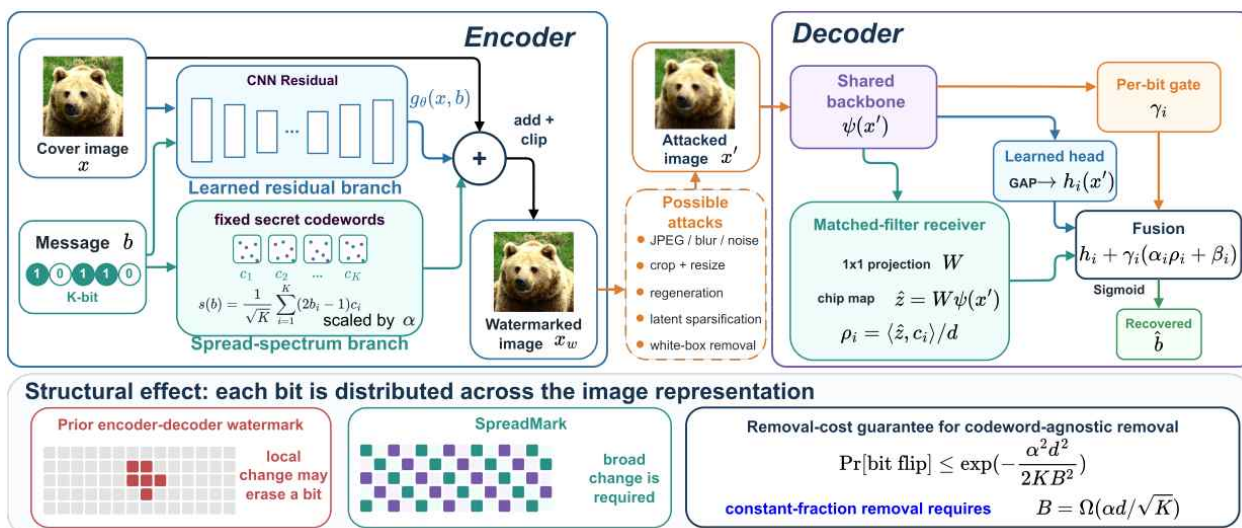
이러한 배경에서 기존 워터마크 기술의 강건성 부족이 어디에서 비롯되는지에 대한 진단이 제기되고 있다. 기존 방식은 정보 한 비트를 이미지의 좁은 영역에만 실어 놓기 때문에, 공격자가 그 부분만 교란해도 워터마크가 사라진다는 것이다. 확산 스펙트럼 삽입 기술인 SpreadMark는 각 비트를 이미지 전면에 걸친 무작위 패턴으로 펼쳐 삽입하고, 검출 단계에서 동일 패턴과의 상관관계를 계산해 읽어내는 접근법이다. 신호가 넓게 분산되어 있을수록 이를 지우는 데 필요한 훼손의 규모가 커진다는 원리를 통해, 의도적 제거 공격에 대응할 수 있는 가능성을 제시하고 있다.

[사례] 이미지 전면 분산 삽입 워터마킹 기술, SpreadMark

• 기존 워터마킹 기술의 취약 구조와 SpreadMark 기술의 개발 배경

- 현재 이미지 워터마킹 산업에서의 주된 방식은 워터마크를 심는 인코더(encoder)와 이를 읽어내는 디코더(decoder)를 한 쌍으로 학습시키는 구조로, 인코더는 원본 이미지에 육안으로 구별되지 않는 정보를 삽입하고, 디코더가 변형을 거친 이미지에서도 그 정보를 복원하도록 설계됨
- 그러나 인코더는 워터마크가 눈에 띄지 않고 판독만 되면 충분하도록 학습되었기 때문에 신호를 넓게 퍼뜨릴 이유가 없어 좁은 영역에 주로 심어지게 되며, 이 경우 공격자가 해당 부위만 교란해도 워터마크가 지워지는 약점이 생김
- SpreadMark 기술은 이러한 취약성이 신호를 좁게 심는 설계에서 비롯된다고 보고, 각 비트를 이미지 전면에 최대한 넓게 펼쳐 삽입하는 방향으로 설계 원리를 전환한 결과물임

[그림 1] SpreadMark의 워터마크 삽입 및 검출 구조



출처: Wei Song 외 6명, "SpreadMark: Robust Image Watermarking via Spread-Spectrum Embedding", arXiv, 2026.08.04., <https://arxiv.org/pdf/2608.03165>

• 확산 스펙트럼 방식의 삽입 원리

- 확산 스펙트럼(spread spectrum)은 하나의 정보를 넓은 대역에 흩뿌려 보낸 뒤 수신 측에 미리 공유된 패턴과 대조해 복원하는 원리로, SpreadMark는 이를 신경망 기반 워터마킹 구조에 적용함
- SpreadMark는 워터마크로 넣을 정보의 각 비트마다 이미지와 같은 크기의 무작위 패턴을 하나씩 미리 만들어 인코더와 디코더가 동일하게 보관함. 이 패턴끼리는 서로 거의 겹치지 않으므로 디코더가 특정 비트만 골라 읽어낼 수 있음
- 인코더는 자신이 만든 미세한 변화 위에서 비트값이 1이면 해당 패턴을 그대로 두고, 0이면 패턴을 뒤집어 더하는 방식으로 모든 비트의 패턴을 겹쳐 올림. 이 과정을 거치면 정보 한 비트가 이미지의 특정 부위가 아닌 전체 화소에 걸쳐 실리게 됨
- 인코더는 겹쳐 올리는 패턴의 개수가 늘어도 화소당 신호 강도가 일정하게 유지되도록 보정함. 이때 신호를 얼마나 강하게 넣을지는 학습을 통해 자동으로 정해지며, 이 세기가 워터마크의 판독 정확도와 화질 사이의 균형점을 결정함

• 정합 필터 기반 검출 구조

- 디코더는 검사할 이미지에서 원본 이미지의 성분을 걷어내고 워터마크 신호만 남긴 형태로 변환한 뒤, 이를 삽입 때 사용한 것과 동일한 패턴 묶음과 대조하여 각 비트값을 판정함
- 이러한 대조 방식을 정합 필터(matched filter)라 하며, 이미 알고 있는 패턴과 관측된 신호가 얼마나 일치하는지를 계산하는 기법으로 잡음이 섞인 신호에서 약한 성분을 찾아내는 데 유리함
- 디코더는 정합 필터 외에 합성곱 신경망(convolutional neural network)*을 활용한 두 번째 판독 방식을 함께 두어 비트값을 두 갈래로 추정하며, 비트마다 두 방식의 반영 비중을 조절하는 장치를 통해 상황에 따라 더 신뢰할 만한 쪽에 무게를 실음
- 두 판독 방식은 서로 다른 상황에 대응함. 정합 필터는 패턴의 위치가 그대로 보존되는 압축이나 잡음 환경에서 판독을 주도하고, 합성곱 신경망 방식은 이미지가 잘리거나 크기가 바뀌어 패턴 위치가 어긋난 경우에 대안 역할을 함
- 연구팀은 이미지에서 중요한 특징만 남기고 나머지를 걷어내는 과정에서 워터마크가 함께 지워지는 공격을 학습 단계에서 미리 재현하여 디코더가 이를 견디도록 훈련시킴. 또한, 평가에는 학습에 쓴 것과 다른 방식으로 만든 공격을 적용해 특정 공격에만 맞춰진 성능이 아님을 확인함

* 합성곱 신경망(convolutional neural network): 이미지의 작은 영역부터 점차 넓은 범위로 패턴을 단계적으로 추출해 나가는 신경망 구조로, 이미지 인식과 처리 분야에서 널리 사용됨

• 공격 유형 분류 및 이론적 근거

- 워터마크 제거 시도 중 두 가지는 워터마크의 존재 여부와 무관하게 이루어짐. 압축이나 흐림 처리(blur), 자르기(crop) 같은 일반적인 이미지 편집이나, 생성형 AI로 이미지를 다시 만드는 재생성(regeneration)이 이에 해당함
- 나머지 두 가지는 워터마크 자체를 겨냥함. 디코더의 내부 계산 정보와 삽입 패턴까지 활용해 정교하게 설계한 적대적 제거(adversarial removal)와, 이미지에서 중요한 특징만 남기는 과정에서 워터마크를 함께 지우는 잠재공간 희소화(latent-space sparsification)임
- 연구팀은 공격자가 패턴 묶음의 내용을 알지 못한다는 전제 아래, 워터마크를 지우는 데 필요한 훼손의 양이 신호가 퍼진 정도에 비례해 늘어난다는 점을 수식으로 제시함. 다만 이는 이상적인 조건을 가정한 분석으로 실제 시스템의 안전성을 보장하지는 않는다고 명시함

• 실험 설계 및 공격 유형별 성능 결과

- 연구팀은 SpreadMark와 공개된 기존 9개 기법을 동일한 조건에 놓고 네 가지 공격에 노출시키는 방식으로 성능을 비교함. 실험에서는 워터마크의 존재를 판별해내는 탐지율과 삽입한 정보를 정확히 읽어낸 비율인 비트 정확도를 지표로 삼았으며, 화질 저하 정도도 함께 측정함
- 실험에는 일반 사진으로 구성된 두 종류의 공개 이미지 모음을 사용하되, 평가용 이미지 1,000장은 학습 단계에서 배제함. 이는 학습에 쓴 이미지로 평가를 할 경우, 결과가 실제보다 부풀려지기 때문임
- 연구진이 설정한 데이터셋 공격 조건에서 실험 결과 기존 9개 워터마킹 기법 중 네 가지 공격을 모두 견디는 워터마킹 기법은 없었으며, 특히 재생성 공격에서는 대부분의 기법이 비트 정확도 0.5 수준으로 떨어져 사실상 워터마크가 소실됨
- 반면, SpreadMark는 재생성 공격에서 탐지율 0.77, 잠재공간 희소화에서 1.00을 기록해 실험한 두 공격 조건 모두에서 높은 탐지율을 유지한 유일한 기법으로 확인됨. 또한, 화질 면에서는 함께 재학습한 일부 기법들보다 원본에 가까웠으나, 이미 배포된 기법들에는 미치지 못하는 중간 수준으로 확인됨

결론 및 시사점

• 기술적 한계 및 개선 과제

- SpreadMark는 ‘흐림 처리’와 ‘자르기’와 같은 일반적인 이미지 편집 환경에서 탐지율이 각각 0.12와 0.24에 그쳤음. 패턴을 이미지의 고정된 위치에 펼쳐 놓는 구조여서, 이미지가 흐려지거나 잘려 패턴 위치가 어긋나면 대조 자체가 어려워지기 때문임
- 디코더의 내부 정보와 패턴을 활용한 적대적 제거에서는 탐지율이 0까지 떨어졌으나, 워터마크를 더 강하게 삽입하면 0.98 수준으로 회복됨. 다만 이 경우 화질 저하가 커지므로, 강건성과 화질 사이의 균형점을 어디에 둘 것인지가 실제 도입 시의 판단 사항이 됨
- 이 방식은 하나의 패턴 묶음이 모든 이미지에 재사용되며 외부에 알려지지 않는다는 전제 위에 설계되었으나, 신호를 넓게 펼치는 특성상 워터마크가 삽입된 이미지 한 장에 더 많은 패턴 정보가 드러날 수 있음. 연구팀은 패턴을 주기적으로 교체하거나 여러 개를 동시에 운용하는 방안을 후속 과제로 제시함

참고문헌

- Wei Song 외 6명, "SpreadMark: Robust Image Watermarking via Spread-Spectrum Embedding", arXiv, 2026.08.04., <https://arxiv.org/pdf/2608.03165>