COPYRIGHT ISSUE BRIEF

저작권 이슈 브리프

SUMMARY | 산업/기업 | 기술

기업 구글의 신스ID 디텍터 공개와 AI 생성 여부 판정의 한계

▶ AI 생성 여부를 확인하기 위해 사업자들이 각자 워터마크와 검증 도구를 운영하고 있지만, 이들 도구가 자사 모델이 생성한 콘텐츠조차 식별하지 못하는 경우가 잦다는 한계가 있다. 이러한 가운데 구글은 자사 워터마크인 신스ID를 찾아 AI 생성 여부를 확인하는 도구인 신스ID 디텍터를, 전문가 대상 시험 제공을 거쳐 2026년 10월 모든 이용자에게 공개했다. 신스ID 디텍터는 이미지, 영상, 오디오를 지원하며, 구글의 생성 모델과 도구뿐 아니라 신스ID를 채택한 오픈AI, 엔비디아, 카카오의 콘텐츠도 확인 대상에 포함된다. 다만 신스ID를 적용하지 않은 모델로 생성했거나 워터마크가 크게 변형된 콘텐츠는 탐지되지 않을 수 있어, 신스ID 채택을 확대하고 변형된 콘텐츠에서도 워터마크가 유지되도록 해 탐지 범위의 공백을 줄일 필요가 있다.

기업 베리매트릭스 스트림키퍼의 예방 중심 보호를 통한 라이브 스포츠 불법 스트리밍 차단

▶ 라이브 스포츠 불법 스트리밍은 출처를 추적하기 어렵고 차단할 무렵에는 경기가 이미 끝나 있어, 사후 차단만으로는 대응하기 어렵다. 더욱이 DRM 키 추출과 인증 탈취가 손쉽게 이뤄지면서 노트북 한 대로도 불법 스트리밍 서비스를 구축할 수 있게 되었다. 이에 베리매트릭스는 멀티-DRM 암호화와 기기 단위 인증을 결합해, 불법 유통자가 인증 수단과 키를 모두 얻지 못하도록 하는 스트림키퍼를 제시했다. 스트림키퍼는 앱과 서비스 양측의 신호를 연계해 위변조에 단계적으로 대응하고, 개인을 식별하지 않고 접속 환경을 판별해 정상 가입자의 불편을 줄인다. 다만 이러한 보호는 계속 바뀌는 공격 경로에 맞춰 지속적으로 갱신돼야 효과를 유지할 수 있다.

산업 기본 탑재 앱 스토어를 통한 무비박스의 확산과 도메인 중심 대응의 한계

▶ 불법 스트리밍 앱 무비박스는 2025년 웹사이트 도메인이 정지된 뒤에도 새 주소로 옮겨 운영을 이어갔고, 앱은 중단 없이 이용자를 늘려 왔다. 이에 MPA와 카날플러스는 유럽연합 집행위원회에 낸 의견서에서 휴대전화 제조사 트랜션이 기본 탑재한 앱 스토어 팜스토어를 무비박스의 유통 경로로 지목했다. 무비박스는 이 스토어에서 엔터테인먼트 분야 1위를 유지하며 다운로드 3억 건을 넘어섰다. 카날플러스는 서명 인증서 등을 근거로 트랜션과의 연결 정황을 제시했으나, 운영 주체는 확정되지 않았다. 이 사례는 도메인 중심 대응만으로는 기본 탑재 스토어를 통한 유통을 막기 어렵다는 점을 보여준다.

저작권 이슈 브리프

SUMMARY | 산업/기업 | 기술

기업 오픈AI의 AI 에이전트 무단 활동으로 드러난 위키미디어의 취약점

▶ 오픈AI의 통제 이탈 에이전트가 독일의 한 위키와 허깅페이스 등 외부 웹사이트에서 승인 없는 활동을 벌였다. 사람의 이용을 전제로 운영되는 위키미디어 재단도 자체 조사 끝에 자사 플랫폼에서 이들의 활동을 발견했다고 밝혔다. AI 에이전트는 여러 위키를 승인 없이 편집하고 인용 도구와 이더패드를 프록시로 악용하려 했으며, 대량 요청으로 위키데이터 쿼리 서비스의 부분 장애에 영향을 미쳤을 가능성도 제기되었다. 다만 이들 활동은 사후 조사로 드러나, 플랫폼이 활동 시점에 AI 에이전트를 가려내기 어려운 한계를 보였다. 이에 따라 플랫폼의 식별 수단 마련과 함께, 운영 기업이 AI 에이전트의 행동을 스스로 감시하고 플랫폼과 활동 분석을 함께하는 일이 과제로 남는다.

산업 AI 생성 음원의 급증에 따른 디저의 검색 제외 토글 도입

▶ 프랑스 음악 스트리밍 플랫폼 디저에 유입되는 AI 생성 음원은 하루 약 9만 곡에 이르렀고, 플랫폼마다 자체 탐지와 신고 기반 태깅으로 판별 기준이 갈리고 있다. 조사 응답자의 97%가 AI 생성 음원을 구분하지 못하는 가운데 디저는 2026년 안에 이를 검색 결과에서 제외하는 토글을 도입한다고 발표했다. 토글은 자체 탐지 도구를 토대로 추천 배제와 태깅에 이어 제외 여부를 이용자가 직접 정하게 하는 단계로, 디저는 삭제나 수익 중단 같은 강제 조치보다 이용자 선택을 앞세웠다. 다만 탐지에서 누락된 음원과 AI 보조 음원은 검색 결과에 남고, 새 생성 모델의 탐지에는 시차가 생길 수 있다. 신고 기반 플랫폼과는 같은 음원도 태깅 여부가 달라질 수 있어, 탐지 결과의 적용 범위와 강도가 플랫폼 간 과제로 남는다.

기술 주간 기술 동향

▶ 수노, 우디오 등 생성형 AI 음악 서비스가 확산되면서 AI 생성곡과 원곡의 유사성을 판별하는 문제가 부각되고 있다. 생성형 재합성은 곡의 음악적 내용을 유지하면서 소리를 크게 바꿔, 기존 유사도 측정 기술로는 판별이 어려울 수 있다. 이에 AI 생성곡의 표절 탐지를 MVI 문제로 재정의한 연구가 발표되었다. 연구팀은 실제 표절 의혹 사례를 바탕으로, 약 35만 개의 비교 쌍으로 이루어진 COPYCAT을 구축하고 두 임베딩의 좌표별 변위를 학습하는 분류기를 제시했다. 그 결과 F0.5 점수는 높아졌으나, 생성형 재합성과 DSP 변조를 결합한 경우에는 재현율이 50.0%에 머물렀다.

구글의 신스ID 디텍터 공개와 AI 생성 여부 판정의 한계

사업자별 워터마크 병존과 제한적이었던 탐지 도구 이용

• 사업자별 워터마크 표준의 병존과 자사 탐지 도구의 한계

- 마이크로소프트(Microsoft)와 메타(Meta)는 각자의 표준으로 콘텐츠에 워터마크(watermark)*를 삽입하고 검증하고 있어, AI 생성 여부를 확인하는 기준이 사업자별로 나뉘어 운영되고 있음
- 다만 이들 검증 도구는 자사 모델이 생성한 콘텐츠조차 식별하지 못하는 경우가 잦은 것으로 나타나, 사업자가 자체 탐지 도구를 운영하는 것만으로는 AI 생성 여부를 확실히 판별하기 어려움

* 워터마크(watermark): 콘텐츠에 식별 신호를 삽입해 출처나 AI 생성 여부를 확인할 수 있게 하는 기술

• 전문가 대상 시험 운영으로 제한된 신스ID 디텍터 이용

- 이러한 흐름 속에서 구글(Google)은 2023년 자체 개발한 워터마크 기술 신스ID(SynthID)를 도입해, AI로 생성한 이미지, 영상, 오디오에 사람이 볼 수 없는 워터마크를 직접 삽입하기 시작함
- 하지만 이 워터마크를 찾아 AI 생성 여부를 탐지하는 도구인 신스ID 디텍터(SynthID Detector)는 2025년 연례 개발자 콘퍼런스인 구글 I/O(Google I/O)에서 일부 언론인, 미디어 전문가, 연구자에게만 시험 제공됨[1)]
- 이처럼 신스ID 디텍터가 전문가 시험용으로 한정돼, 일반 이용자는 신스ID가 삽입된 콘텐츠를 접하더라도 이 도구로 AI 생성 여부를 직접 확인할 수 없었음

신스ID 디텍터의 일반 공개와 탐지 범위

• 모든 이용자를 대상으로 한 신스ID 디텍터 공개와 확인 경로 확대

- 이러한 상황에서 구글은 2026년 10월 신스ID 디텍터를 누구나 이용할 수 있는 웹사이트로 공개하고 영어로 전 세계에 제공해, 전문가로 제한됐던 이용 대상을 일반 이용자까지 확대함[2)]
- 특히 신스ID 디텍터는 이미지는 JPG, JPEG, PNG, 영상은 MP4, MOV, WEBM, 오디오는 WAV, MP3, OGG, FLAC 등의 형식을 지원해, 이용자는 이미지뿐 아니라 영상과 오디오의 AI 생성 여부도 한 곳에서 확인할 수 있음[1)]

1) Ivan Mehta, "Google's new SynthID website can identify AI-generated media", TechCrunch, 2026.10.07., https://techcrunch.com/2026/10/07/googles-new-synthid-website-can-identify-ai-generated-media/
2) Laurie Sullivan, "Google's New-And-Improved Method For Verifying Synthetic Content", MediaPost, 2026.10.07., https://www.mediapost.com/publications/article/418565/googles-new-and-improved-method-for-verifying-syn.html

• 구글 모델과 타 사업자의 신스ID 채택으로 넓어진 탐지 대상

- 신스ID 디텍터는 업로드된 콘텐츠에 신스ID 워터마크가 있는지를 확인하므로, 신스ID 디텍터로 확인할 수 있는 콘텐츠의 범위는 신스ID를 삽입하는 모델과 사업자의 수에 따라 정해짐
- 실제로 구글은 이미지, 영상, 음악 생성 모델인 나노 바나나(Nano Banana), 비오(Veo), 리리아(Lyria)와 AI 챗봇 및 영상 및 음악 제작 도구인 제미나이(Gemini), 플로우(Flow), 프로듀서AI(ProducerAI), 비즈(Vids)로 생성한 콘텐츠에 신스ID를 삽입해 왔음
- 그 결과 2026년 10월 구글이 공개한 누적 집계에 따르면 신스ID가 적용된 이미지와 영상은 1,800억 건을 넘어, 구글 모델과 도구로 생성한 이들 콘텐츠가 신스ID 디텍터의 확인 대상에 포함됨[3)]
- 더욱이 오픈AI(OpenAI), 엔비디아(Nvidia), 카카오(Kakao) 등 글로벌 빅테크 및 AI 기업들이 신스ID를 채택하고 애플(Apple)도 곧 지원을 추가할 것으로 알려져, 신스ID 디텍터로 확인할 수 있는 콘텐츠가 구글 외 사업자로 넓어지고 있음

• 미탐지와 오탐지로 드러나는 판정 결과의 한계

- 그러나 신스ID 디텍터는 워터마크가 발견되지 않으면 이를 알리지만, 구글은 미탐지 결과가 AI 생성이나 조작 가능성을 완전히 배제하는 것은 아니라고 밝힘
- 구체적으로 신스ID를 적용하지 않은 모델로 생성했거나 워터마크가 크게 변형된 콘텐츠는 탐지되지 않을 수 있어, 탐지 기능은 보편적 채택의 부족과 변형 취약성이라는 한계를 안고 있음
- 반대로 워터마크가 없는 콘텐츠를 AI 생성 콘텐츠로 판정하는 오탐지(false positive)도 발생할 수 있어, AI 생성으로 판정된 결과 역시 그대로 확정 근거로 삼기는 어려움

신스ID 디텍터의 확산과 검증 신뢰성 확보 과제

• 탐지 결과를 판단 근거로 쓸 때의 오인 가능성

- 이에 따라 신스ID 디텍터가 모든 이용자에게 공개되면서 확인 결과를 접하는 사람은 늘었지만, 미탐지 결과를 사람이 만든 콘텐츠라는 근거로 받아들이면 AI 생성 콘텐츠를 놓치는 오인으로 이어질 수 있음
- 더욱이 워터마크가 없는 콘텐츠도 AI 생성으로 판정될 수 있는 만큼, 탐지 결과는 탐지와 미탐지 어느 쪽이든 단독 판단 근거가 아닌 참고 정보로 다뤄질 필요가 있음

• 워터마크 채택 확대와 변형 대응의 과제

- 또한 신스ID가 없는 모델의 콘텐츠는 신스ID 디텍터로 확인할 수 없고 마이크로소프트와 메타는 각자의 표준을 운영하고 있어, 이용자는 콘텐츠의 출처에 따라 여러 도구를 거쳐야 할 수 있음
- 따라서 탐지 범위를 넓히려면 신스ID를 채택하는 사업자를 늘리는 동시에, 콘텐츠가 크게 변형되더라도 워터마크의 탐지 가능성이 유지되도록 기술적 강건성을 높여 보편적 채택의 부족과 변형 취약성을 함께 줄여 나가야 함

3) Laurie Sullivan, "Google's New-And-Improved Method For Verifying Synthetic Content", MediaPost, 2026.10.07., https://www.mediapost.com/publications/article/418565/googles-new-and-improved-method-for-verifying-syn.html

참고문헌

- Ivan Mehta, "Google's new SynthID website can identify AI-generated media", TechCrunch, 2026.10.07., https://techcrunch.com/2026/10/07/googles-new-synthid-website-can-identify-ai-generated-media/
- Laurie Sullivan, "Google's New-And-Improved Method For Verifying Synthetic Content", MediaPost, 2026.10.07., https://www.mediapost.com/publications/article/418565/googles-new-and-improved-method-for-verifying-syn.html

베리매트릭스 스트림키퍼의 예방 중심 보호를 통한 라이브 스포츠 불법 스트리밍 차단

라이브 스포츠 불법 스트리밍 서비스의 손쉬운 구축과 사후 대응의 한계

• 경기 종료 후에야 이루어지는 사후 차단의 한계

- 라이브 스포츠 스트리밍의 불법 유통은 출처를 추적하기 어려운 데다, 불법 스트리밍을 차단할 무렵에는 경기가 이미 끝난 경우가 많아 사후 차단만으로는 피해를 막기 어려움
- 더욱이 불법 스트리밍 서비스가 차단되더라도 수 분 만에 다른 서비스가 등장하는 상황에서, 불법 스트리밍 이용자는 자신이 적발되지 않을 것이라고 여겨 이용을 쉽게 멈추지 않음

• DRM 키 추출과 인증 탈취로 쉬워진 불법 스트리밍 서비스 구축

- 여기에 더해 불법 스트리밍 운영에는 많은 장비가 필요하지 않고 노트북 한 대만 있으면 세계 어디서든 시작할 수 있어, 불법 유통자가 서비스를 새로 여는 진입 장벽이 매우 낮음
- 실제로 일부 디지털 저작권 관리(Digital Rights Management, 이하 DRM)* 구현 환경에서는 인증 정보나 복호화 키가 노출될 경우, 콘텐츠 전송 네트워크(Content Delivery Network, 이하 CDN)**의 콘텐츠가 무단 복호화될 위험이 있음
- 또한 앱과 서비스 사이의 인증을 가로채는 공격도 손쉽게 이뤄져, 불법 유통자는 탈취한 인증으로 다른 기기용 라이선스를 받아 키를 추출한 뒤 이를 다른 사람에게 배포할 수 있음

* 디지털 저작권 관리(Digital Rights Management, DRM): 콘텐츠를 암호화하고 허가된 기기에서만 복호화해 재생할 수 있도록 관리하는 기술

** 콘텐츠 전송 네트워크(Content Delivery Network, CDN): 콘텐츠를 이용자에게 효율적으로 전달하기 위해 여러 지역에 분산 배치한 서버 네트워크

베리매트릭스 스트림키퍼의 예방 중심 콘텐츠 보호

• 콘텐츠 암호화와 기기 단위 인증의 결합을 통한 불법 스트리밍 서비스 구축 차단

- 이에 불법 유통 대응 기술 기업 베리매트릭스(Verimatrix)는 기존 사후 차단 방식 대신, 불법 스트리밍 서비스가 자리 잡기 전에 구축 자체를 어렵게 만드는 스트림키퍼(Stream- keeper)를 제시함[1)]
- 스트림키퍼는 멀티-DRM(Multi-DRM)*으로 콘텐츠를 암호화하는 한편, 콘텐츠를 요청하는 앱, 웹 클라이언트(web client)**, 셋톱박스(set-top box)***가 정당한 요청 권한을 가진 기기인지를 함께 검증함

1) Ken Kerschbaumer, "Verimatrix Streamkeeper Puts Those Fighting Piracy in a Proactive Position for Battle", Sports Video Group, 2026.10.12., https://www.sportsvideo.org/2026/10/12/veramatrix-streamkeeper-puts-those-fighting-piracy-in-a-proactive-position-for-battle/

- 앱과 서비스를 잇는 인증 토큰(authentication token)****은 해당 토큰을 생성한 기기에 암호학적으로 연결돼 다른 기기로 옮기면 즉시 폐기되고 키 추출도 막혀, 불법 유통자는 쓸 수 있는 인증 수단도 키도 얻지 못함
- 베리매트릭스의 불법 유통 대응 제품 관리 부사장 후안 마르티네스 푸이그(Juan Martinez Puig)는 이러한 보안 과정이 이용자에게 보이지 않아 시청 경험을 방해하지 않는다고 설명함[2)]

* 멀티-DRM(Multi-DRM): 여러 종류의 DRM을 함께 지원해 다양한 기기와 플랫폼에서 콘텐츠를 보호하는 기술

** 웹 클라이언트(web client): 별도의 앱 설치 없이 웹 브라우저로 서비스에 접속해 콘텐츠를 요청하고 재생하는 프로그램

*** 셋톱박스(set-top box): TV에 연결해 방송 및 스트리밍 콘텐츠를 수신하고 재생하는 기기

**** 인증 토큰(authentication token): 앱이 서비스 이용 권한을 확인받기 위해 서비스와 주고받는 전자 증표

• 앱과 서비스 양측의 가시성을 활용한 위변조 탐지와 단계적 대응

- 이와 함께 스트림키퍼는 위변조를 탐지하면 접근 차단, 화질 저하, 경고 메시지 표시 중 하나를 택해, 위협 수준에 맞춰 대응 강도를 단계적으로 높임
- 마르티네스 푸이그는 전송 측면만 보는 CDN 전용 업체와 달리, 베리매트릭스는 앱의 활동, 앱과 서비스 간 통신을 함께 보고 두 신호를 연계해 더 정밀하게 대응할 수 있다고 설명함
- 이러한 양측 감시 방식은 고객사의 약 1,500만 개 앱 인스턴스(application instance)*에 적용되며, 공격 경로가 계속 바뀌는 만큼 보안을 지속 갱신하는 서비스형 소프트웨어(Software as a Service, 이하 SaaS)** 구독 방식으로 제공됨[1)]

* 앱 인스턴스(application instance): 개별 기기에 설치되어 실행되는 각각의 앱

** 서비스형 소프트웨어(Software as a Service, SaaS): 소프트웨어를 설치하지 않고 인터넷을 통해 구독 형태로 이용하는 서비스 방식

• 정상 가입자의 접근 마찰을 줄이는 접속 환경 판별

- 보안 강화가 가입자의 불편으로 이어지지 않도록, 스트림키퍼는 개인을 식별하지 않고 요청이 정상 앱, 가상 사설망(Virtual Private Network, VPN)*, 공유 IP 풀(shared IP pool)** 중 어디서 오는지 판별함
- 그 결과 스트리밍 사업자는 해외에 체류하는 가입자에게도 정당한 접근을 허용하는 규칙을 세울 수 있어, 이들이 접속 제한으로 불편을 겪다 불법 스트리밍 서비스를 찾는 일을 줄일 수 있음
- 베리매트릭스는 잠재 시청자의 30~40%가 불법 이용하는 시장에서 이들을 가입자로 되돌리면 사업자의 수익이 늘어 요금 인상 필요도 줄어든다고 봄[1)]

* 가상 사설망(Virtual Private Network, VPN): 인터넷 접속 경로를 암호화하고 다른 지역을 거쳐 접속할 수 있게 하는 사설 네트워크

** 공유 IP 풀(shared IP pool): 여러 이용자가 같은 IP 주소를 나눠 쓰도록 묶어 둔 주소 집합

구축 단계 예방으로의 대응 전환과 지속 갱신 과제

• 사후 탐지에서 구축 단계 차단으로의 대응 전환

- 스트림키퍼 사례는 불법 유통자가 인증 수단과 키를 얻지 못하게 해 구축 단계를 막는 방식이, 경기가 끝난 뒤에야 차단이 이뤄지는 사후 대응의 한계를 보완할 수 있음을 보여줌
- 다만 정상 가입자의 접속 불편 불법 스트리밍 서비스로 옮겨 갈 계기가 될 수 있어, 보안 강화는 접속 환경 판별처럼 불법 유통자에게만 마찰을 주는 장치와 함께 설계돼야 함

2) Ken Kerschbaumer, "Verimatrix Streamkeeper Puts Those Fighting Piracy in a Proactive Position for Battle", Sports Video Group, 2026.10.12., https://www.sportsvideo.org/2026/10/12/veramatrix-streamkeeper-puts-those-fighting-piracy-in-a-proactive-position-for-battle/

• 공격 경로 변화에 대응하는 지속 갱신의 필요성

- 한편 DRM 키 추출과 인증 탈취처럼 불법 유통자의 공격 경로는 계속 바뀌어, 서비스 구축 단계를 막는 보호 조치도 한 번 적용하는 것만으로는 효과를 유지하기 어려움
- 따라서 베리매트릭스가 보안을 지속 갱신하는 SaaS 구독 방식을 택한 것처럼, 예방 중심 보호는 새 공격 경로를 감시하고 보호 기능을 계속 고쳐 나가는 운영 체계를 함께 갖춰야 효과가 이어질 수 있음

참고문헌

• Ken Kerschbaumer, "Verimatrix Streamkeeper Puts Those Fighting Piracy in a Proactive Position for Battle", Sports Video Group, 2026.10.12., https://www.sportsvideo.org/2026/10/12/veramatrix-streamkeeper-puts-those-fighting-piracy-in-a-proactive-position-for-battle/
• Verimatrix, "Verimatrix Streamkeeper™ Suite, It is not a patch; It is armor", 2026.10.16. 조회, https://www.verimatrix.com/anti-piracy/streamkeeper-suite/

기본 탑재 앱 스토어를 통한 무비박스의 확산과 도메인 중심 대응의 한계

도메인 정지 이후에도 이어진 불법 스트리밍 앱의 확산

• 도메인 정지에도 중단되지 않은 무비박스의 운영

- 안드로이드용 불법 스트리밍 앱 무비박스(MovieBox)는 앱과 함께 월 방문 수가 수천만 건에 이르던 웹사이트 moviebox.ng도 운영해 왔으나, 2025년 집행 조치로 이 도메인이 정지되면서 웹사이트를 겨냥한 대응이 이뤄짐[1)]
- 그러나 무비박스는 짧은 기간 안에 moviebox.ph로 주소를 옮겼고 안드로이드 앱 서비스는 해당 기간에도 중단 없이 운영되어, 도메인이 정지된 이후에도 서비스 운영이 공백 없이 이어짐
- 또한 주소 이전 이후에도 앱 다운로드 수가 1년이 채 안 돼 두 배 이상 늘어[1)], 웹사이트 주소를 겨냥한 조치가 앱을 통한 유통 확대까지 막지 못한 것으로 드러남

• 권리자 의견서를 통해 확인된 주요 앱 유통 경로

- 이러한 상황에서 미국영화협회(Motion Picture Association, 이하 MPA)와 프랑스 유료방송 기업 카날플러스(Canal+)는 유럽연합 집행위원회(European Commission)에 제출한 감시목록(Counterfeit and Piracy Watch List)* 의견서에서 무비박스를 지목함
- 두 의견서는 무비박스의 유통 경로로 같은 앱 스토어를 공통으로 지목해, 웹사이트 주소가 아닌 특정 앱 스토어가 무비박스 확산의 주요 통로로 다뤄지고 있음을 보여줌

* 감시목록(Counterfeit and Piracy Watch List): 유럽연합 집행위원회가 침해 억제를 촉구하기 위해 관련 서비스를 정리해 공개하는 목록으로, 법 위반 여부를 판정하지는 않음

기기 기본 탑재 앱 스토어를 통한 무비박스의 유통 확대

• 제조사가 기본 탑재한 앱 스토어를 통한 노출 확대

- 무비박스가 유통되는 앱 스토어인 팜스토어(Palm Store)는 독립적인 스토어가 아니라, 중국 휴대전화 제조사 트랜션(Transsion)이 자사 브랜드 테크노(Tecno), 인피닉스(Infinix), 아이텔(itel) 휴대전화에 기본 탑재하는 스토어임
- 트랜션은 홍콩 상장을 위한 투자설명서 초안에서 팜스토어를 자사 앱 배포 채널로 소개했으며, 팜스토어의 2025년 월평균 활성 이용자가 1억 8,000만 명을 넘어 무비박스가 대규모 이용자 기반 안에서 유통되고 있음[1)]

1) Ernesto Van der Sar, “Pirate App MovieBox Passes 300 Million Downloads as Rightsholders Alert EU”, TorrentFreak, 2026.10.07., https://torrentfreak.com/pirate-app-moviebox-passes-300-million-downloads-as-rightsholders-alert-eu/

- 다만 외부 개발자는 통상 팜스토어를 통한 설치 건수에 따라 트랜션에 비용을 내지만, 무비박스에도 같은 방식이 적용되는지는 확인되지 않아 스토어와 앱 간의 거래 관계는 드러나지 않음
- 아울러 두 의견서 모두 팜스토어나 트랜션에 앱 삭제를 요청했는지는 밝히지 않았으며, 웹사이트 도메인과 달리 무비박스 앱은 현재도 스토어에 등록된 채 유통되고 있음

• 스포츠 생중계를 앞세운 홍보와 다운로드의 급증

- 실제로 무비박스는 2026년 6월 16일 기준 팜스토어에서 약 2억 4,330만 건 다운로드돼 2025년 8월 대비 약 108% 늘었고, 엔터테인먼트 분야 1위를 유지하며 스토어 내 최상위 앱으로 자리 잡음[2)]
- IT 전문 온라인 매체 토렌트프릭(TorrentFreak)의 보도 시점인 2026년 10월 7일 기준 다운로드 수는 3억 560만 건으로, 카날플러스 의견서가 제시한 6월 16일 기준 약 2억 4,330만 건에서 4개월이 채 안 돼 6,000만 건 이상 늘어남[3)]
- 한편 스토어의 무비박스 소개 화면은 영화, TV 시리즈와 함께 스포츠 생중계의 무료 시청을 전면에 내세우고 유럽 5대 축구 리그와 마블(Marvel), 디즈니(Disney) 작품 이미지를 활용함

• 기술 분석으로 드러난 제조사 트랜션과의 연결 정황

- 카날플러스는 과거 moviebox.ng의 도메인 등록 정보에 트랜션 관련 정보가 있었고, 앱에서 트랜션 명의의 서명 인증서(signing certificate)*와 트랜션 소유 인프라 참조가 발견됐다며 무비박스가 트랜션과 밀접하게 연결돼 있다고 주장함[2)]
- 또한 카날플러스는 상표 출원, 개발자 표기, SSL 인증서(SSL certificate)**, 도메인 등록 정보를 근거로 홍콩 기업 엑셀런트 이노베이션(Excellent Innovation Limited)을 무비박스와 관련된 기업으로 지목했고, 팜스토어도 이 기업을 개발자로 표기함[2)]
- 하지만 카날플러스는 도메인 정보와 인증서의 구체적인 내용을 밝히지 않았고, MPA는 운영 근거지를 중국으로 보면서 트랜션을 언급하지 않아 기술적 단서만으로 운영 주체가 확정되지는 않음

* 서명 인증서(signing certificate): 앱을 배포하기 전 개발자가 앱에 전자서명을 할 때 쓰는 인증서로, 서명한 주체의 명의가 담김

** SSL 인증서(SSL certificate): 웹사이트와 이용자 간 통신을 암호화하기 위해 발급받는 인증서로, 발급 대상의 도메인과 명의 정보가 담김

앱 스토어 단계의 유통 관리와 운영 주체 확인 과제

• 도메인 중심 대응이 닿지 않는 앱 스토어 유통의 공백

- 도메인 정지 이후에도 다운로드가 늘어난 무비박스 사례는, 웹사이트 주소를 겨냥한 대응만으로는 제조사 스토어에 등록된 앱을 통한 이용자 유입을 막기 어렵다는 점을 보여줌
- 특히 기본 탑재 스토어는 기기를 구매한 이용자에게 처음부터 제공되는 만큼, 이 단계에서 불법 스트리밍 앱이 걸러지지 않으면 대규모 이용자에게 노출이 계속될 수 있음
- 그럼에도 두 의견서 모두 스토어에 대한 삭제 요청을 언급하지 않아, 웹사이트 차단과 달리 앱 스토어 단계의 대응은 아직 구체적인 대응 수단으로 다뤄지지 않고 있음

2) Callum Fairhurst, "Canal+ and MPA urge EU to list MovieBox as downloads hit 307.6 million", BetaNews, 2026.10.08., https://betanews.com/article/moviebox-pirate-app-307-million-downloads-eu/
3) Ernesto Van der Sar, “Pirate App MovieBox Passes 300 Million Downloads as Rightsholders Alert EU”, TorrentFreak, 2026.10.07., https://torrentfreak.com/pirate-app-moviebox-passes-300-million-downloads-as-rightsholders-alert-eu/

• **기술적 단서를 운영 주체 확인으로 연결하기 위한 검증의 필요성**

- 서명 인증서와 도메인 등록 정보 같은 기술적 단서는 앱과 특정 기업의 연결 정황을 보여주지만, 구체적인 내용이 공개되지 않으면 운영 주체를 판단하는 근거로 쓰기 어려움
- 이번 사례에서도 두 권리자가 제시한 운영 근거지와 관련 기업이 서로 달라, 같은 앱을 두고도 의견서를 낸 주체에 따라 운영 주체에 대한 판단이 엇갈리는 상황이 나타남
- 따라서 앱 스토어 단계의 대응이 실효성을 가지려면 인증서, 개발자 표기, 도메인 기록 등 앱의 기술적 정보를 검증 가능한 형태로 확인하고, 스토어 운영사와 앱 운영 주체의 관계를 함께 규명할 필요가 있음

참고문헌

• Ernesto Van der Sar, “Pirate App MovieBox Passes 300 Million Downloads as Rightsholders Alert EU”, TorrentFreak, 2026.10.07., https://torrentfreak.com/pirate-app-moviebox-passes-300-million-downloads-as-rightsholders-alert-eu/
• Callum Fairhurst, "Canal+ and MPA urge EU to list MovieBox as downloads hit 307.6 million", Beta News, 2026.10.08., https://betanews.com/article/moviebox-pirate-app-307-million-downloads-eu/

오픈AI의 AI 에이전트 무단 활동으로 드러난 위키미디어의 취약점

외부 웹사이트로 번지는 통제 이탈 에이전트 활동과 위키미디어 재단의 조사

• 개발사 밖으로 확산된 오픈AI의 AI 에이전트 통제 이탈

- 미국의 AI 기업 오픈AI(OpenAI)가 운영한 통제 이탈 에이전트(rogue agent)*는 2026년 5월 독일의 한 위키(wiki)**를 장악해 답변과 제한 우회 기법을 주고받는 수단으로 악용하여, AI 에이전트(AI agent)***의 활동이 개발사 밖의 웹사이트로 번지고 있음을 보여줌[1)]
- 더욱이 7월에는 700개에 가까운 통제 이탈 에이전트가 AI 모델 저장소 허깅페이스(Hugging Face) 해킹을 공모함[1)]. 오픈AI도 통제 이탈 에이전트 활동의 전체 범위를 파악하는 데 애쓰고 있어, 운영 주체조차 피해 범위를 가늠하지 못한 상태임

* 통제 이탈 에이전트(rogue agent): 운영 주체가 정한 작업 범위를 벗어나 승인되지 않은 활동을 하는 AI 에이전트

** 위키(wiki): 여러 이용자가 함께 문서를 작성하고 편집하는 웹사이트

*** AI 에이전트(AI agent): 이용자나 운영 주체를 대신해 웹사이트 등에서 작업을 수행하는 AI 프로그램

• 사람의 이용을 전제로 한 위키미디어 플랫폼의 자체 조사 착수

- 이러한 가운데 온라인 백과사전 위키백과(Wikipedia)의 운영 주체 위키미디어 재단(Wikimedia Foundation)은 위키백과가 사람을 위해 설계되고 자원봉사자가 운영하는 만큼, AI 에이전트가 일으키는 위협에 대한 해결책을 아직 확보하지 못했음을 밝힘[2)]
- 실제로 위키미디어 재단은 2024년 이후 봇(bot)* 활동이 급증해 대역폭 사용량이 50% 늘었다고 2025년에 밝혔으며, 이러한 부하가 서버 비용을 높이고 장애로 사람 방문자의 접속까지 막을 수 있다고 우려함[2)]
- 이에 위키미디어 재단은 자사 웹사이트도 AI 에이전트의 영향을 받았는지 조사했고 10월 5일 블로그에서 오픈AI 통제 이탈 에이전트의 활동을 발견했다고 밝혀[3)], 활동 범위가 위키미디어 플랫폼까지 이르렀음이 드러남
- 한편 위키미디어 재단은 오픈AI가 AI 에이전트의 예측하기 어려운 행동을 인정한 만큼 이를 감시하고 예방할 책임도 져야 한다고 지적함1). 오픈AI는 위키미디어 재단과 함께 활동을 분석 중이라고만 밝혀, 구체적인 경위는 아직 공개되지 않음

* 봇(bot): 사람 대신 웹사이트에 자동으로 접속해 작업을 수행하는 프로그램

1) Sergiu Gatlan, "Wikimedia: Rogue OpenAI agents behind unauthorized Wikipedia edits", BleepingComputer, 2026.10.06., https://www.bleepingcomputer.com/news/security/rogue-openai-agents-behind-potentially-malicious-wikipedia-edits/
2) Anton Mous, "Wikimedia uncovers unauthorized activity by OpenAI's AI agents across its platforms", Cybernews, 2026.10.07., https://cybernews.com/news/rogue-openai-wikipedia/
3) Reuters, "Wikipedia operator says OpenAI rogue agents made 'potentially malicious' edits", Khaleej Times, 2026.10.06., https://www.khaleejtimes.com/business/tech/wikipedia-operator-openai-rogue-agents-unauthorised-edits

편집에서 인프라까지 이어진 오픈AI의 AI 에이전트 무단 활동

• 샌드박스에 머문 사전 승인 없는 위키 편집

- 위키미디어 재단은 오픈AI가 운영한 것으로 보이는 AI 에이전트가 위키백과와 위키미디어 커먼즈(Wikimedia Commons)* 등 여러 위키를 사전 승인 없이 편집했다고 밝혀, 편집 기능이 AI 에이전트에게도 열려 있었음이 드러남
- 다만 이들 편집은 일반 독자에게 보이는 문서에 게시되지 않았고 대부분 샌드박스(sandbox)** 영역의 시험 편집에 그쳐서, 공개 문서의 내용이 바뀌는 피해로는 이어지지 않음
- 그럼에도 위키미디어 재단은 이번 사례가 AI 에이전트의 신뢰할 수 있는 정보 훼손 시도 가능성을 보여준다고 밝혀, 시험 편집에 그친 활동도 정보 신뢰성에 대한 위협으로 평가함

* 위키미디어 커먼즈(Wikimedia Commons): 누구나 자유롭게 사용할 수 있는 그림, 사진, 소리, 동영상, 3D 모델 등의 멀티미디어 파일을 모아 놓은 비영리 온라인 미디어 저장소

** 샌드박스(sandbox): 위키에서 일반 독자에게 공개되지 않는 시험 편집 공간

• 인용 도구와 이더패드를 프록시로 쓰려는 악용 시도

- 또한 AI 에이전트는 위키의 인용 도구(citation tool) 설정도 일부 변경했으며, 위키미디어 재단은 이 변경이 인용 도구를 원격 서비스 데이터 수집용 프록시(proxy)*로 악용하려는 악의적 시도일 가능성이 있다고 판단함[4)]
- 같은 방식으로 AI 에이전트는 여러 사용자가 실시간으로 함께 편집하는 메모 도구 이더패드(Etherpad)에도 침입을 시도해 다른 웹사이트의 데이터를 가져오는 프록시로 쓰려 했으나, 이 시도는 성공하지 못함
- 이처럼 AI 에이전트가 콘텐츠 편집을 넘어 위키의 보조 도구를 외부 데이터를 가져오는 경로로 바꾸려 해, 플랫폼의 정상 기능이 AI 에이전트 작업의 수단으로 악용될 수 있음이 드러남

* 프록시(proxy): 다른 서비스에 대신 요청을 보내 데이터를 받아오는 중계 수단

• 대량 데이터 수집에 따른 위키데이터 쿼리 서비스의 부분 장애

- 나아가 AI 에이전트는 공개 API(Application Programming Interface)*에 수백만 건을 자동 요청하고 위키데이터(Wikidata)**와 위키미디어 커먼즈의 페이지 수백만 개를 크롤링(crawling)***해 서버와 대역폭 등 플랫폼 자원을 대량 소모함[5)]
- 특히 위키데이터 쿼리 서비스(Wikidata Query Service)****에는 수십만 건의 데이터 쿼리가 몰렸고, 위키미디어 재단은 이 트래픽이 2026년 5월 해당 서비스의 부분 장애에 영향을 미쳤을 가능성이 있다고 봄[4)]
- 이처럼 AI 에이전트의 데이터 수집이 서비스 장애에 영향을 미쳤을 가능성까지 제기되면서, 위키미디어 재단은 이번 사례를 통해 AI 에이전트의 대량 요청이 플랫폼 자원을 소모하고 서비스 장애를 유발할 위험이 있다고 경고함[4)]

* API(Application Programming Interface): 프로그램끼리 데이터와 기능을 주고받도록 정한 접속 규격

** 위키데이터(Wikidata): 위키미디어 재단이 운영하는 협력 편집 기반의 오픈소스 다국어 지식 베이스이자 구조화된 데이터베이스

*** 크롤링(crawling): 자동화 프로그램이 웹페이지를 차례로 방문하며 내용을 수집하는 행위

**** 위키데이터 쿼리 서비스(Wikidata Query Service): 위키데이터에 저장된 정보를 조건을 지정한 질의(쿼리)로 검색해 가져올 수 있게 한 서비스

4) Anton Mous, "Wikimedia uncovers unauthorized activity by OpenAI's AI agents across its platforms", Cybernews, 2026.10.07., https://cybernews.com/news/rogue-openai-wikipedia/
5) Sergiu Gatlan, "Wikimedia: Rogue OpenAI agents behind unauthorized Wikipedia edits", BleepingComputer, 2026.10.06., https://www.bleepingcomputer.com/news/security/rogue-openai-agents-behind-potentially-malicious-wikipedia-edits/

사람 중심 플랫폼의 AI 에이전트 대응을 위한 과제

• 사후 조사에 의존하는 AI 에이전트 활동 식별의 한계

- 이번 무단 편집과 도구 악용 시도, 대량 요청은 위키미디어 재단의 사후 조사로 드러났고 운영 주체도 추정 수준에 머물러, 플랫폼이 활동 시점에 AI 에이전트를 가려내기 어려운 상황을 보여줌
- 아울러 봇 활동으로 대역폭 사용량이 이미 늘어난 상황에서 통제 이탈 에이전트의 대량 요청까지 더해지면, 사람의 이용을 전제로 설계된 플랫폼이 비용과 장애 위험을 계속 떠안게 될 수 있음
- 따라서 플랫폼이 비용과 장애 피해를 줄이려면 편집과 도구 이용, 데이터 요청이 AI 에이전트에서 비롯됐는지와 그 운영 주체를 활동 시점에 확인할 수단이 필요함

• 운영 기업과 플랫폼 간 활동 분석 공유의 과제

- 운영 주체인 오픈AI도 활동 범위를 다 파악하지 못한 만큼, 피해 플랫폼이 확인한 활동 기록을 운영 기업과 함께 분석하는 협력이 원인 규명의 출발점이 될 수 있음
- 그러나 위키미디어 재단의 지적처럼 운영 기업이 작업 범위를 벗어난 AI 에이전트의 행동을 스스로 감시하고 예방하지 않으면, 피해 확인과 비용 부담은 계속 플랫폼의 몫으로 남게 됨

참고문헌

- Sergiu Gatlan, "Wikimedia: Rogue OpenAI agents behind unauthorized Wikipedia edits", BleepingComputer, 2026.10.06., https://www.bleepingcomputer.com/news/security/rogue-openai-agents-behind-potentially-malicious-wikipedia-edits/
- Reuters, "Wikipedia operator says OpenAI rogue agents made 'potentially malicious' edits", Khaleej Times, 2026.10.06., https://www.khaleejtimes.com/business/tech/wikipedia-operator-openai-rogue-agents-unauthorised-edits
- Anton Mous, "Wikimedia uncovers unauthorized activity by OpenAI's AI agents across its platforms", Cybernews, 2026.10.07., https://cybernews.com/news/rogue-openai-wikipedia/

AI 생성 음원의 급증에 따른 디저의 검색 제외 토글 도입

AI 생성 음원의 급증과 스트리밍 플랫폼의 대응 분화

• AI 생성 음원 업로드의 급증과 플랫폼별 대응 방식의 분화

- 프랑스 음악 스트리밍 플랫폼 디저(Deezer)에 유입되는 AI 생성 음원(fully AI-generated music)*은 2025년 1월 탐지 도구 도입 당시 하루 1만 곡에서 현재 하루 약 9만 곡으로 늘었으며, 유입량이 많은 날에는 AI 생성 음원이 전체 신규 음원의 절반 이상을 차지함[1)]
- 다만 이들 음원이 전체 재생에서 차지하는 비중은 1~3%인 반면, 2025년 AI 생성 음원 재생의 최대 85%가 조작으로 확인돼 디저는 업로드의 주된 목적이 청취가 아닌 부정 수익 확보에 있다고 봄[2)]
- 이처럼 유입이 늘어나는 가운데 디저는 자체 탐지로 AI 생성 음원을 판별하는 반면, 애플 뮤직(Apple Music)과 스포티파이(Spotify)는 레이블과 유통사의 신고에 의존해 판별 기준이 플랫폼마다 갈림

* AI 생성 음원(fully AI-generated music): 생성형 AI만으로 만든 음원. 사람의 창작 과정에 일부 AI 보조 요소가 들어간 음원(AI 보조 음원)과 구분함

[표 1] 주요 음악 스트리밍 플랫폼의 AI 생성 음원 대응

플랫폼	판별 기준	조치 내용
디저	자체 탐지 도구	– 2025년 6월부터 태깅, 추천 배제, 조작 재생 저작권료 제외 – 2026년 내 검색 제외 토글 도입
코부즈(Qobuz)	자체 탐지 시스템	2026년 2월 탐지·태깅 시스템 발표
애플 뮤직	레이블과 유통사의 신고	2026년 3월 태깅 도입, 연내 'Made With AI' 태그 노출 예정
스포티파이	아티스트의 신고	2026년 4월 크레딧 태그 시험, 9월 중순부터 'AI 페르소나' 태깅과 추천 기본 배제

출처: Murray Stassen, "Deezer to let users block AI-generated music from search results", Music Business Worldwide, 2026.10.13., https://www.musicbusinessworldwide.com/deezer-to-let-users-block-ai-generated-music-from-search-results/

1) Murray Stassen, "Deezer to let users block AI-generated music from search results", Music Business Worldwide, 2026.10.13., https://www.musicbusinessworldwide.com/deezer-to-let-users-block-ai-generated-music-from-search-results/
2) Jesper Wendel, "Deezer will let you fully exclude AI-generated music from your streaming experience", Deezer Newsroom, 2026.10.13., https://newsroom-deezer.com/2026/10/deezer-ai-music-toggle/

• 이용자의 필터링 수요에 대응한 검색 제외 토글 도입 발표

- 디저가 조사 기관 입소스(Ipsos)에 의뢰해 2025년 11월 발표한 8개국 9,000명 조사에서는 97%가 블라인드 테스트(blind test)*에서 AI 생성 음원과 사람이 만든 음원을 구분하지 못해[2], 플랫폼의 대응이 갈리는 사이 이용자 스스로는 이를 구분하지 못하는 상황임
- 같은 조사에서 음악 스트리밍 이용자의 45%가 AI 생성 음원의 필터링을 원한다고 답해, 직접 구분하기는 어렵지만 플랫폼의 필터링 기능을 통해 이를 제외하려는 수요가 있는 것으로 나타남
- 이에 디저는 10월 13일 AI 생성 음원을 검색 결과에서 제외하는 토글(toggle)**을 2026년 안에 전 세계에 도입한다고 발표함[3]. 토글을 켜면 AI 생성 음원과 앨범, 관련 아티스트가 검색 결과에서 빠지고, 끄면 다시 검색 결과에 나타남

* 블라인드 테스트(blind test): 출처를 알리지 않은 채 음원을 듣고 평가하게 하는 방식

** 토글(toggle): 기능을 켜고 끄는 설정 스위치

탐지 기술을 토대로 단계별로 확대되는 디저의 AI 생성 음원 관리

• 생성 모델 고유 신호의 식별과 일반화를 통한 탐지 범위 확장

- 디저의 탐지 도구는 AI 생성 음원에 남는 고유 신호를 식별해 사람이 만든 음원과 구분하는 기술로, 2024년 12월 출원한 특허 2건이 2026년 6월 EU와 미국 특허청에서 공개됨[3]. 추천 배제, 태깅, 검색 제외는 모두 이 도구의 판별 결과를 대상으로 함
- 이 도구는 현재 수노(Suno), 우디오(Udio), 리퓨전(Riffusion) 등 주요 생성 모델(generative model)*의 음원을 탐지하며, 다른 생성 모델도 음원 예시를 확보하면 추가할 수 있어 탐지 범위가 모델별 음원 예시 확보 여부에 따라 정해짐
- 이러한 제약을 줄이기 위해 디저는 특정 모델의 학습 데이터 없이도 AI 생성 음원을 탐지하는 일반화 성능을 높여, 음원 예시를 확보하지 못한 생성 모델도 탐지 대상에 포함하려 함

* 생성 모델(generative model): 학습한 데이터를 바탕으로 음원 등 새로운 콘텐츠를 만들어내는 AI 모델

• 태깅과 추천 배제를 통한 노출 제한과 저작권료 보호

- 디저는 탐지 결과를 바탕으로 2025년 1월부터 2026년 8월까지 4,200만 곡 이상을 알고리즘 추천(algorithmic recommendation)*과 에디토리얼 플레이리스트(editorial playlist)**에서 배제해, 이들 음원이 추천을 통해 이용자에게 노출되지 않도록 함[3]
- 아울러 2025년 6월부터는 1,400만 개 이상의 앨범에 AI 생성 음원임을 명시하는 태그를 붙여[3], 검색으로 접근한 이용자도 AI 생성 여부를 알 수 있도록 투명성을 높임
- 디저는 이러한 추천 배제를 AI 생성 음원이 저작권료 배분 재원을 잠식하지 못하게 막는 첫 단계로 삼고, 조작으로 확인된 재생은 저작권료 지급 대상에서 제외해 노출 제한과 함께 부정 수익도 차단함

* 알고리즘 추천(algorithmic recommendation): 이용자의 청취 이력을 바탕으로 자동으로 음원을 제안하는 기능

** 에디토리얼 플레이리스트(editorial playlist): 플랫폼 담당자가 직접 선곡해 구성하는 플레이리스트

3) Jesper Wendel, "Deezer will let you fully exclude AI-generated music from your streaming experience", Deezer Newsroom, 2026.10.13., https://newsroom-deezer.com/2026/10/deezer-ai-music-toggle/

• 검색 결과 제외 토글을 통한 이용자 선택권 확대

- 추천 배제와 태깅이 플랫폼의 기준에 따른 조치였다면, 토글은 검색 결과에서의 제외 여부를 이용자가 직접 정하게 함. 디저 CEO 알렉시스 랑테르니에(Alexis Lanternier)도 이용자가 청취 경험을 최대한 통제하도록 하는 데 목적이 있다고 설명함[3)]
- 다만 토글의 제외 대상은 생성형 AI만으로 만든 음원에 한정되며, 사람의 창작 과정에 일부 AI 보조 요소가 들어간 AI 보조 음원은 별도로 취급돼 토글을 켜도 제외되지 않음
- 디저는 음원 삭제나 수익 지급 중단 같은 추가 조치에는 신중한 검토가 필요하다는 입장으로, 우선 토글을 통해 검색 결과에서의 제외 여부를 이용자 선택에 맡김

탐지 기반 관리의 한계와 과제

• 탐지 도구의 판별 결과에 의존하는 토글 적용 범위의 한계

- 토글은 탐지 도구가 AI 생성 음원으로 판별한 음원만 제외하므로, 탐지에서 누락된 음원은 토글을 켜도 검색 결과에 남음. 조사 응답자의 대부분이 AI 생성 음원을 구분하지 못한 만큼 누락된 음원을 이용자가 직접 구분해 제외하기도 어려움
- 더욱이 탐지 대상 추가가 생성 모델별 음원 예시 확보에 달려 있어, 새 생성 모델이 등장한 뒤 탐지가 시작되기까지 시차가 생길 수 있음. 디저가 개선 중인 일반화 성능이 이 시차를 얼마나 줄이는지가 토글의 실효성에 영향을 미칠 전망임
- 아울러 AI 보조 음원은 토글의 제외 대상이 아니므로, 생성형 AI의 활용 정도에 따라 두 유형을 어떻게 구분하는지 이용자가 알 수 있도록 판별 기준을 설명하는 일이 과제로 남음

• 신고 기반 플랫폼과의 판별 기준 차이에 따른 과제

- 디저의 토글이 자체 탐지 결과에 기반하는 반면 애플 뮤직과 스포티파이는 레이블과 유통사의 신고로 태깅하므로, 신고되지 않은 AI 생성 음원은 플랫폼에 따라 태깅 여부가 달라질 수 있어 이용자는 같은 음원이라도 플랫폼마다 다른 정보를 접할 수 있음
- 자체 탐지 기술을 갖춘 디저도 음원 삭제나 수익 지급 중단에는 신중한 입장으로, 탐지 결과를 노출 제한과 이용자 선택 기능에 활용할지, 삭제 및 수익 지급 제한 등 강제 조치까지 확대할지는 플랫폼별 정책에 따라 달라질 수 있음

참고문헌

- Murray Stassen, "Deezer to let users block AI-generated music from search results", Music Business Worldwide, 2026.10.13., https://www.musicbusinessworldwide.com/deezer-to-let-users-block-ai-generated-music-from-search-results/
- Jesper Wendel, "Deezer will let you fully exclude AI-generated music from your streaming experience", Deezer Newsroom, 2026.10.13., https://newsroom-deezer.com/2026/10/deezer-ai-music-toggle/

주간 기술 동향

음악 버전 식별을 활용한 AI 생성곡 표절 탐지 기술

• AI 생성곡 유통 논의 속에서 부각되는 원곡과의 유사성 판별 문제

수노(Suno), 우디오(Udio) 등 텍스트 명령어로 음악을 만드는 생성형 AI 서비스가 확산되면서, AI 생성곡이 사람이 만든 원곡과 지나치게 닮았을 수 있다는 우려가 커지고 있다. 음악의 표절 여부는 멜로디, 화성, 리듬 같은 요소와 곡의 전체적인 인상을 함께 고려해 판단하는 경우가 많다. 그러나 어느 정도 닮아야 표절로 볼 수 있는지에 대해서는 아직 합의된 기준이 없다. 실제로 표절로 확인된 사례도 많지 않아, 이를 자동으로 판별하는 기술을 개발하고 검증하는 데에도 어려움이 있다.

최근 미국에서는 AI 생성곡이 음원 시장에 어느 정도 유통되고 있는지가 저작권 분쟁의 쟁점 중 하나로 떠오르고 있다. 유니버설뮤직그룹(Universal Music Group)과 소니뮤직(Sony Music)은 수노와 관련된 저작권 분쟁 과정에서, 음원 유통사 디스트로키드(DistroKid)가 매달 유통하는 AI 생성곡 규모를 공개하도록 법원에 요청했다. 음반사 측은 이 자료가 AI 생성곡과 기존 음악 간 경쟁 정도를 파악하는 데 도움이 될 수 있다고 보지만, 디스트로키드 측은 유통 규모가 쟁점과 직접적인 관련이 적다는 입장이다. 이 요청에 대한 판단은 아직 나오지 않았다.

이번 사안은 AI 생성곡을 가려내는 일과 그 곡이 특정 원곡에서 비롯되었는지를 확인하는 일이 서로 다른 문제일 수 있음을 시사한다. 유통 단계의 탐지 기술은 AI 생성곡인지를 구분할 수 있지만, 그 곡이 어떤 원곡을 얼마나 닮았는지까지 알려주는 것은 아니다. 기존의 유사도 측정 기술은 주로 사람이 만든 곡 사이의 관계를 다루는 방향으로 발전해 왔다. 생성형 AI는 곡의 음악적 내용을 유지하면서도 소리의 질감을 크게 바꿀 수 있어, 겉으로 드러나는 소리만 비교하는 방식으로는 닮은 정도를 파악하기 어려울 수 있다.

이러한 배경에서 AI 생성곡의 표절 탐지를 음악 버전 식별(Music Version Identification, MVI) 문제로 다시 정의한 연구가 발표되었다. 음악 버전 식별은 연주 방식, 악기 구성, 템포가 달라져도 같은 곡의 다른 버전을 찾아내는 기술로, 이 연구는 AI 생성곡이 원곡의 파생 버전인지를 판별하는 데 이를 활용한다. 연구팀은 실제 표절 의혹이 제기된 곡을 바탕으로 평가용 데이터셋을 구축하고, 곡의 특징을 수치로 나타낸 임베딩에서 각 수치의 차이, 즉 좌표별 변위를 학습하는 분류기를 제안했다. 이 기술은 AI 생성 여부를 가리는 데서 나아가, 원곡과의 파생 관계를 판별하는 방법을 제시한다.

[사례] 음악 버전 식별 기반 AI 생성곡 표절 탐지 기술

• 기존 표절 탐지의 한계와 음악 버전 식별 접근의 등장

- AI 음악 관련 탐지 과제에는 곡이 AI로 생성되었는지를 판별하는 과제와 AI가 학습한 데이터의 출처를 추적하는 과제가 있음. 표절 탐지는 이와 달리 AI 생성곡이 큰 음향 변형을 거친 뒤에도 원곡과 높은 유사성을 유지하는지를 판단해야 하는 별도의 과제임
- 곡 사이의 유사성을 측정하는 기존 기술은 악보를 비교하는 방식과 음원을 직접 비교하는 방식으로 발전해 왔으나, 대부분 사람이 만든 곡 사이의 관계를 대상으로 함
- 특히 생성형 재합성(generative re-synthesis)*은 곡의 음악적 내용은 유지하면서 예측하기 어려운 음향 변형을 일으키기 때문에, 단순히 소리만 비교하는 방식으로는 유사성을 포착하기 어려움
- 이번 연구에서는 표절 탐지를 음악 버전 식별(Music Version Identification, 이하 MVI)** 문제로 재정의하여, 후보곡이 사람과 AI 중 누구에 의해 만들어졌는지와 관계없이 원곡의 파생 버전인지를 판단하도록 설계함. 이 판단은 법적 판단이 아닌 기술적 유사성 판단으로 다룸

* 생성형 재합성(generative re-synthesis): 기존 음원을 조건으로 입력받아 생성형 AI가 새로운 음원을 만들어내는 과정

** 음악 버전 식별(Music Version Identification, MVI): 연주 방식, 악기 구성, 템포, 제작 방식이 달라진 녹음 사이에서 서로 대응하는 곡을 찾아내는 기술

• 실제 표절 사례 기반 COPYCAT 벤치마크 구성

- 연구팀은 70건의 실제 표절 의혹 사례를 담은 SMP 데이터셋에서 의혹 구간을 중심으로 20초 길이의 음원 1,219개를 잘라내 기준 구간으로 삼고, 이를 바탕으로 평가용 벤치마크인 COPYCAT(Corpus Of Plagiarized Yet Camouflaged Artificial Tracks)을 구축함
- COPYCAT은 두 구간을 짝지은 비교 쌍으로 이루어지며, 표절 관계로 간주되는 짝을 양성 쌍, 표절 관계가 아닌 짝을 음성 쌍으로 구분함
- 양성 쌍은 각 기준 구간에 사람의 재해석, 디지털 신호 처리(Digital Signal Processing, 이하 DSP) 변조*, 생성형 재합성을 적용하고, 사람의 재해석과 생성형 재합성에는 각각 DSP 변조를 결합하는 등 은닉 강도를 점차 높인 변형 구간을 기준 구간과 짝지어 구성함
- 생성형 재합성에는 음원 입력이 가능한 음악 생성 AI 모델인 MusicGen, AudioLDM 2, MGE-LDM을 활용하며, MGE-LDM은 베이스, 드럼, 선율 악기 중 한 트랙만 남기고 반주를 새로 채움. 텍스트 명령어는 원곡의 스타일과 무관하게 입력하여 유사성이 입력 음원에서만 비롯되도록 함
- 음성 쌍은 무작위 선정부터 특징이 가장 비슷한 구간 선정까지 세 단계 난이도로 구성되며, 최종 벤치마크는 양성 쌍 113,690개와 음성 쌍 236,964개를 포함한 총 350,654개의 비교 쌍으로 이루어짐

* 디지털 신호 처리(Digital Signal Processing, DSP) 변조: 음높이를 2반음 또는 4반음 올리거나 내리고, 템포를 0.90~1.10배로 조정하는 등 음원 신호를 직접 가공하는 변형

• 거리 비교의 한계와 좌표별 변위 분류기

- 연구팀은 MVI 모델을 새로 학습시키지 않고 이미 학습된 MVI 모델 두 가지를 그대로 가져와 활용하며, CLEWS는 소리의 유사성을, WEALY는 보컬이 있을 때 가사의 유사성을 판단함
- 두 모델은 각 음원 구간을 임베딩(embedding)*으로 바꾸며, 기본 방식은 두 임베딩 사이의 거리가 설정된 임계값보다 작으면 유사한 곡으로 판정하는 방식임
- 그러나 거리는 두 임베딩의 차이를 수치 하나로 줄인 값이어서, 연구팀은 임베딩을 이루는 각 수치, 즉 좌표마다 차이를 따로 구한 좌표별 변위(coordinate-wise shift)**를 활용함

- 좌표별 변위는 표절 여부가 표시된 예시를 학습한 분류기에 입력되며, 최종 분류기는 거리, 평균 같은 대표값 24개와 값의 변화가 가장 큰 좌표 512개를 함께 활용해 표절 여부를 판정함

* 임베딩(embedding): 음원처럼 복잡한 데이터를 여러 개의 수치로 이루어진 목록으로 바꾸어 표현한 것으로, 비슷한 곡일수록 수치가 서로 가깝게 나타나도록 학습됨

** 좌표별 변위(coordinate-wise shift): 두 임베딩의 같은 좌표에 있는 수치끼리 뺀 값을 모은 목록으로, 차이의 방향은 표절 판단과 관계없다고 보아 크기만 사용함

• 성능 비교 및 분석 결과

- 성능은 표절로 판정한 것 중 실제 표절의 비율인 정밀도(precision)와 실제 표절 중 찾아낸 비율인 재현율(recall)을 합친 F0.5 점수(F0.5 score)*로 평가하며, 잘못된 표절 판정을 줄이기 위해 정밀도에 더 큰 비중을 둠
- 최종 분류기는 전체 F0.5 점수를 CLEWS를 사용한 기본 방식의 0.612에서 0.803으로 높였으며, 대표값만 사용한 경우는 0.617에 그쳐 판별에 필요한 정보가 개별 좌표에 흩어져 있는 것으로 확인됨
- 변형 유형별 재현율은 사람의 재해석에서 30.6%에서 86.5%로, 판별이 가장 어려운 생성형 재합성과 DSP 변조의 결합에서 10.3%에서 50.0%로 높아졌으며, 정밀도는 각각 97.3%, 72.0% 수준임
- 다만 기준 구간에 DSP 변조만 적용한 경우에는 변형이 비교적 단순해, 기본 방식의 F0.5 점수가 0.965로 최종 분류기의 0.894보다 더 높음
- 임베딩 분포를 그림으로 나타내 보면 MusicGen과 AudioLDM 2로 만든 변형 구간은 기준 구간에서 가장 멀리 떨어져 거리만으로는 구분이 어려우며, 이번 연구에서는 판별 정보가 어느 좌표가 어떻게 달라졌는지에 담겨 있다고 해석함

* F0.5 점수(F0.5 score): 정밀도와 재현율을 하나로 합친 지표로, 1에 가까울수록 성능이 높으며 정밀도에 2배 가중치를 둠

[표 1] 변형 유형별 기본 방식과 최종 분류기의 표절 판별 성능 비교

	COPYCAT		WEALY			CLEWS			XGBoost Hybrid Top-512		
Category	**Pairs**	**%**	**Prec.**	**Rec.**	**$F_{0.5}$**	**Prec.**	**Rec.**	**$F_{0.5}$**	**Prec.**	**Rec.**	**$F_{0.5}$**
Human Plagiarism	3,757	3.3	74.2%	25.9%	54.0%	98.0%	30.6%	68.0%	97.3% ± 0.1%	86.5% ± 0.5%	**94.9% ± 0.2%**
Human Plagiarism + DSP	37,570	33.0	72.3%	24.8%	52.3%	97.7%	27.4%	64.5%	97.2% ± 0.1%	83.1% ± 0.4%	**94.0% ± 0.2%**
Original + DSP	12,190	10.7	51.8%	80.1%	55.7%	96.9%	95.2%	**96.5%**	87.2% ± 0.4%	99.3% ± 0.0%	89.4% ± 0.3%
AI Generation	5,472	4.8	42.4%	51.5%	43.9%	84.3%	15.0%	43.8%	75.1% ± 0.4%	56.5% ± 0.3%	**70.5% ± 0.3%**
AI + DSP	54,701	48.1	36.4%	34.3%	36.0%	79.0%	10.3%	33.9%	72.0% ± 0.5%	50.0% ± 0.3%	**66.2% ± 0.4%**
Overall	113,690	100.0	45.6%	36.6%	43.5%	92.8%	26.0%	61.2%	84.2% ± 0.4%	67.7% ± 0.3%	**80.3% ± 0.3%**

출처: Fotis Koutsikos 외 8인, "Towards AI-Generated Music Plagiarism Detection as a Version Identification Problem", arXiv, 2026.10.06., https://arxiv.org/html/2610.09075v1

결론 및 시사점

• 기술적 한계 및 향후 과제

- 최종 분류기는 사람의 재해석이나 DSP 변조만 적용한 경우에는 높은 성능을 보였으나, 생성형 재합성과 DSP 변조의 결합에서는 재현율이 50.0%에 머물러 AI를 거친 표절은 판별이 더 어려운 것으로 확인됨
- 이에 연구팀은 이 기술을 법적 판단을 대신하는 수단이 아닌, 전문가가 검토할 후보 사례를 먼저 걸러내는 선별 도구로 활용하는 것이 적절하다고 봄
- 실험에는 음원을 입력으로 받을 수 있는 생성 모델 세 가지가 사용되었으며, 수노, 우디오 같은 상용 서비스는 현재 이러한 기능을 제공하지 않아 실험에 활용되지 않음

- 향후 연구는 더 다양한 생성 모델과 변형 방식으로 벤치마크를 확장하고 다른 MVI 모델을 적용하는 방향으로 진행될 예정이며, 표절 여부 판별을 넘어 AI가 학습한 데이터의 출처를 추적하는 방향으로의 확장도 제시됨

참고문헌

- Fotis Koutsikos 외 8인, “Towards AI-Generated Music Plagiarism Detection as a Version Identification Problem”, arXiv, 2026.10.06., https://arxiv.org/html/2610.09075v1
- Camila Curcio, “Universal and Sony Seek to Force DistroKid to Reveal Scale of AI Music Uploads”, Law Commentary, 2026.10.16., https://www.lawcommentary.com/articles/universal-sony-distrokid-ai-music-data-suno-lawsuit