



SUMMARY

산업/기업

기술

산업 구글, 2017년 영구 금지명령에 근거해 미국 내 검색 결과에서 사이허브 도메인 제외**법원 명령과 플랫폼 집행의 교차점: 구글 미국 내 사이허브 검색 결과 제외**

▶ 구글은 2025년 12월 2일 미국 내 검색 결과에서 사이허브(Sci-Hub) 관련 도메인을 제외했으며, 해당 조치는 2017년 버지니아 연방법원이 발부한 영구 금지명령을 근거로 이루어졌다. 해당 명령은 발부 이후 약 7년간 실질적으로 집행되지 않았으나, 2025년 구글의 검색 결과 제외 조치를 통해 가시적인 효과를 나타냈다. 영구 금지명령의 적용 범위는 사이허브와 적극적으로 협력하거나 참여하는 주체까지이며, 검색엔진도 명시적으로 포함했다. 그러나 구글은 '적극적 협력 또는 참여' 주체로 간주되지는 않는다는 입장을 지속적으로 취해온 것으로 알려졌다.

산업 안나스 아카이브, 스포티파이 음원 대규모 스크래핑 주장... AI 학습 악용 우려 확산**안나스 아카이브의 스포티파이 음원 대규모 스크래핑에 따른 AI 학습 악용 우려**

▶ 해적판 도서·논문 공유 사이트 안나스 아카이브가 2025년 12월 스포티파이 음원 8,600만 개와 메타데이터 2억 5,600만 행을 스크래핑했다고 발표하면서 관련 우려가 확산되고 있다. 해당 데이터는 약 300테라바이트 규모로 공개된 음악 메타데이터 데이터베이스 가운데 최대 규모에 해당한다. 안나스 아카이브는 스포티파이의 인기도 지표를 기준으로 수집 우선순위를 설정하고 DRM 우회 기술을 활용해 데이터를 수집했다. 수집 데이터는 인류의 음악 유산 보존을 명분으로 토렌트 기반 네트워크를 통해 단계적으로 배포될 계획이며, 관련 업계에서는 해당 데이터가 AI 모델 학습에 활용될 가능성이 높다고 경고하고 있다.

산업 호주 TDM 예외 규정 도입의 한계와 유료 라이선스 중심의 정책적 대응 방안**강한 반발 직면한 텍스트 및 데이터 마이닝(TDM) 예외 규정 도입 제안**

▶ 호주 생산성위원회가 제안한 텍스트 및 데이터 마이닝(TDM) 예외 규정 도입 제안은 제작 생태계 훼손 우려 등 특정 단체의 강한 반발에 직면하며 정책적 한계를 드러냈다. 이는 일방적인 규제 완화만으로는 기술 혁신과 권리 보호 간 이해관계 균형을 달성하기 어렵다는 점을 시사한다. 특히 사후 탐지 중심 접근은 보상 체계와 데이터 검증 메커니즘이 부재한 상황에서 저작권자의 권익을 충분히 보호하기 어렵다는 구조적 한계를 보여준다. 이에 호주 정부는 해당 제안을 수용하지 않고, 유료 라이선스 체계 검토와 3년 유예를 통한 단계적 제도 정비로 정책 방향을 전환하였는데, 이는 데이터 전송 인프라 구축을 통해 혁신과 보호를 조화시키려는 보다 실질적인 정책 해법으로 평가된다.



SUMMARY

산업/기업

기술

산업 어도비 집단소송으로 본 AI 학습 데이터셋의 구조적 저작권 리스크**학습 데이터 공급망의 저작권 쟁점에 따른 책임 소급 및 산업적 리스크 전이**

▶ 어도비(Adobe)는 자사 AI 모델 'SlimLM' 학습에 활용된 데이터가 저작권을 침해했다는 혐의로 소송에 직면하였다. 이번 사건은 앤트로픽(Anthropic)이 해적판 도서 학습 문제로 15억 달러 규모의 기록적인 합의를 마친 직후 발생하며 업계의 법적 긴장감을 고조시켰다. 핵심은 어도비가 직접 데이터를 수집하지 않았더라도, 활용한 데이터셋이 불법 복제 서적 19만여 권을 포함한 데이터의 파생물이라는 점에 있다. 최근 주요 빅테크 기업들이 동일한 데이터셋 활용 문제로 연쇄 제소됨에 따라, 본 사례는 특정 기업의 문제를 넘어 AI 산업 전체의 구조적 리스크로 확산되었다.

기술 [사례 탐구] NFT 거래 환경에서의 이미지 저작권 검증 기술과 AI 기반 사전 탐지 접근**NFT로 거래되는 이미지의 저작권 검증 필요성 증대**

▶ NFT로 발행·유통되는 디지털 이미지 콘텐츠는 블록체인을 통해 거래 이력과 소유 기록을 확인할 수 있으나, 토큰의 소유와 이미지 저작권의 범위가 기술적으로 분리된 구조로 인해 권리 관계가 명확히 검증되지 않은 채 거래가 이루어질 가능성을 내포하고 있다. 특히 생성형 AI 확산과 함께 기존 이미지나 제3자의 저작물이 NFT로 발행·유통되는 사례가 증가하고 있으나, 현재의 NFT 거래 환경에서는 해당 이미지가 다른 저작물과 유사한지, 이미 NFT로 발행된 이력이 있는지 여부를 일반 이용자가 사전에 판단하기 어려운 구조적 한계가 있다. 이에 따라 최근에는 NFT 거래 이전 단계에서 디지털 이미지 콘텐츠의 저작권 침해 가능성을 기술적으로 점검하려는 시도가 증가하고 있으며, 관련 사례를 살펴보도록 한다.

기술 주간기술동향**AI 생성 음성의 출처 증명, Smark 기술의 부상**

▶ 최근에는 AI 모델이 음성을 생성하는 과정 자체에 저작권 정보를 식별자 형태로 삽입하는 '생성형 워터마킹' 방식이 대안으로 주목받고 있다. 이는 콘텐츠가 생성되는 시점부터 고유 정보를 갖게 함으로써, 사후에 진위를 판별하던 기존 방식에서 벗어나 출처와 소유권을 원천적으로 증명하는 선제적 보호 체계로의 전환을 의미한다. 본 보고서는 이러한 흐름 속에서 확산 모델 기반 TTS 시스템을 위해 개발된 Smark 기술을 중심으로, 이산 웨이블릿 변환을 활용한 새로운 워터마킹 사례를 심층 분석하고자 한다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

구글, 2017년 영구 금지명령에 근거해 미국 내 검색 결과에서 사이허브 도메인 제외

뉴스브리프

구글(Google)은 2025년 12월 2일 미국 내 검색 결과에서 사이허브(Sci-Hub) 관련 도메인을 제외했다. 해당 조치는 2017년 버지니아 연방법원이 발부한 영구 금지명령을 근거로 이루어졌다. 해당 명령은 실질적으로 집행되지 않았으나, 2025년 구글의 검색 결과 제외 조치를 통해 가시적인 효과를 나타냈다. 영구 금지명령의 적용 범위는 사이허브와 적극적으로 협력하거나 참여하는 주체까지이며, 검색엔진도 명시적으로 포함했다. 그러나 구글은 '적극적 협력 또는 참여' 주체로 간주되지는 않는다는 입장을 지속적으로 취해온 것으로 알려졌다. 이러한 상황 속에서 영국 로펌 위긴 LLP는 2025년 12월 2일 영구 금지명령을 근거로 구글에 사이허브 관련 도메인의 검색 결과 제외를 요청했고, 구글은 해당 요청을 반영해 미국 내 검색 결과에서 도메인을 제외했다.

구글의 사이허브 도메인 미국 검색 결과 제외 조치

• 2017년 법원 명령에 따른 2025년 검색 결과 제외 조치

- 구글(Google)은 2025년 12월 2일 미국 내 검색 결과에서 사이허브(Sci-Hub) 관련 도메인 34개를 제외하는 조치를 시행함¹⁾
- 해당 조치는 일반적인 디지털밀레니엄저작권법(Digital Millennium Copyright Act, DMCA)* 기반 삭제 요청이 아니라, 2017년 버지니아 연방법원이 발부한 영구 금지명령을 근거로 이루어짐
 - * 디지털밀레니엄저작권법(Digital Millennium Copyright Act, DMCA): 1998년 제정된 미국 저작권법으로, 온라인 서비스 제공자가 저작권 침해 콘텐츠에 대해 권리자의 삭제 요청(takedown notice)을 받을 경우 이를 신속히 제거하면 책임을 제한받을 수 있도록 한 제도
- 다만, 해당 영구 금지명령은 발부 이후 약 7년간 실질적으로 집행되지 않았으며, 2025년 말 구글의 검색 결과 제외 조치를 통해 뒤늦게 가시화됨

1) Ernesto Van der Sar, "Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order", Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>

• 지식재산권 침해에 따른 영구 금지명령 도입

- 사이허브는 수백만 건의 과학 논문과 학술 자료를 무료로 열람할 수 있도록 제공하는 플랫폼으로 지식재산권 침해 문제가 지속적으로 제기되어 왔음
- 미국 화학회(American Chemical Society, ACS)는 2017년 버지니아 연방법원에 사이허브와 운영자 알렉산드라 엘바키안(Alexandra Elbakyan)을 상대로 저작권 및 상표 침해 혐의로 소송을 제기함²⁾
- 사이허브 측이 소송에 응하지 않으면서 법원은 기본 판결을 통해 ACS의 주장을 받아들이고, 손해배상과 함께 광범위한 영구 금지명령을 발부함
- 해당 영구 금지명령은 적용 대상이 사이허브 운영자 등 직접 당사자에 그치지 않고, 사이허브와 '적극적으로 협력하거나 참여하는(in active concert or participation)' 모든 주체로 확장됨. 명령문에는 검색엔진, 호스팅 제공자, 인터넷서비스제공자, 도메인 등록기관 등이 명시적으로 포함됨

[그림 1] 영구 금지명령에서 규정한 적용 주체 범위

ORDERED that any person or entity in active concert or participation with Defendant Sci-Hub and with notice of the injunction, including any Internet search engines, web hosting and Internet service providers, domain name registrars, domain name registries, and other service or software providers, cease facilitating access to any or all domain names, websites, and services through which Defendant Sci-Hub engages in unlawful access to, use, reproduction, and distribution of the ACS Marks or ACS's Copyrighted Works (including, but not limited to, the domain names sci-hub.ac, sci-hub.biz, sci-hub.bz, sci-hub.cc, sci-hub.cf, sci-hub.cn, sci-hub.ga, sci-hub.gq, sci-hub.io, sci-hub.hk, sci-hub.is, sci-hub.la, sci-hub.name, sci-hub.nu, sci-hub.nz,

출처: Ernesto Van der Sar, "Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order", Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>

영구 금지명령 적용 범위를 둘러싼 해석 논란과 집행 과정

• '적극적 협력' 요건에 대한 해석 논란

- 네덜란드 기술 매체 토렌트프릭크(Torrentfreak)는 영구 금지명령에 포함된 '적극적 협력 또는 참여' 요건을 두고, 검색엔진이 이에 해당하는지에 대해 해석상 논란이 존재해 왔다고 지적함²⁾
- 해당 영구 금지명령에는 사이허브와 '적극적으로 협력하거나 참여하는' 주체의 예시로 검색엔진이 언급돼 있으나, 토렌트프릭크 보도에 따르면 구글을 포함한 주요 기술 기업들은 검색엔진이 중립적 제3자 서비스에 해당하며, 반드시 '적극적 협력 또는 참여' 주체로 간주되지는 않는다는 입장을 지속적으로 취해온 것으로 알려짐²⁾

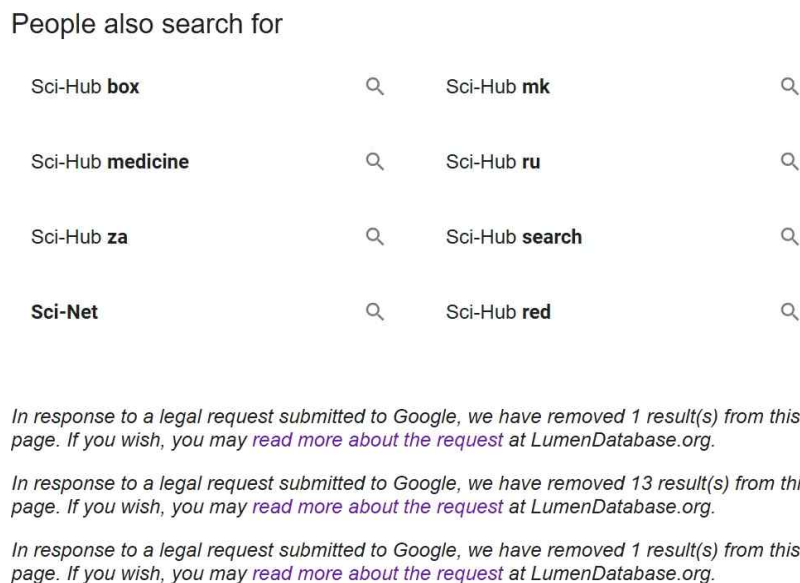
• 영구 금지명령 집행 요청과 검색 결과 제외 과정

- 이러한 해석 논란 속에서 영구 금지명령은 장기간 실질적인 집행 없이 유지되었으며, 영국 로펌 위긴 LLP(Wiggin LLP)의 집행 요청을 계기로 비로소 가시적인 집행 단계로 이어짐

²⁾ Ernesto Van der Sar, "Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order", Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>

- 토렌트프릭크에 따르면, 위긴 LLP는 2025년 12월 2일 해당 영구 금지명령을 근거로, 구글에 사이허브와 연계된 34개 도메인*을 검색에서 제외해 달라는 요청을 발송함³⁾
 - * 도메인(domain): 인터넷에서 특정 웹사이트를 식별하기 위해 사용하는 주소 체계로, 웹서비스 운영자는 도메인을 변경하거나 새로운 도메인을 등록함으로써 동일한 콘텐츠를 다른 주소에서 제공할 수 있음
- 토렌트프릭크는 위긴 LLP가 2025년 9월에도 유사한 요청을 발송해 수십 개의 사이허브 관련 도메인을 대상으로 검색 결과 제외를 요구한 사실을 확인했으며, 관련 요청은 최소 2022년부터 총 7건 이상 확인된 것으로 보도함³⁾
- 구글은 해당 요청을 바탕으로 미국 내 검색 결과에서 해당 도메인들을 제외 조치함. 그 결과 미국 구글에서 사이허브를 검색할 경우 검색 결과 하단에 '법적 요청에 의해 제거됨(Removed by legal request)'이라는 안내 문구가 표시됨
- 다만, 동일한 도메인들은 미국 외 국가의 검색 결과에서는 여전히 접근 가능한 상태로 유지되고 있음

[그림 2] 요청에 따른 사이허브 구글 검색 결과 제외 안내 화면



출처: Ernesto Van der Sar, "Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order", Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>

구글의 사이허브 검색 제외가 학술 지식 접근과 플랫폼에 미치는 영향

• 교육 기술 분야 및 디지털 콘텐츠 스타트업에 미치는 영향

- 스페인 스타트업·기술 매체 엘 에코시스템아 스타트업(El Ecosistema Startup)은 사이허브 검색 제외 사례를 통해, 교육 기술(EdTech) 분야 및 디지털 콘텐츠 기반 스타트업이 단순히 기술적 완성도나 사용자 편의성만을 고려할 것이 아니라, 지식재산권 준수와 법적 합법성을 전제로 한 사업 구조를 설계할 필요성이 있음을 시사한다고 전함⁴⁾
- 동시에, 사용자 관점에서는 정보 접근성이 급격히 제한되지 않도록, 합법성과 지속가능성을 충족하면서도 포용적이고 접근 가능한 서비스 설계 필요성도 함께 제기함

3) Ernesto Van der Sar, "Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order", Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>

4) Cristian Tala Sánchez, "Google, Sci-Hub y el bloqueo de dominios: impacto y contexto", El Ecosistema Startup, 2025.12.13., <https://ecosistemastartup.com/google-sci-hub-y-el-bloqueo-de-dominios-impacto-y-contexto/>

• 디지털 정보 유통 영역에서의 플랫폼 책임 부상

- 엘 에코시스템마 스타트업은 오랜 기간 실제로 집행되지 않았던 법원 명령을 구글이 뒤늦게 이행한 사례가 빅테크 기업이 단순한 중개자를 넘어 정보의 노출 여부를 결정하는 ‘문지기’ 역할을 수행하고 있음을 보여준다고 분석함
- 아울러 본 사례는 불법 콘텐츠로 판단된 자료에 대한 정당한 지식재산권 보호 조치인지, 아니면 연구자와 이용자의 정보 접근을 제한하는 검열에 해당하는지 그 경계가 명확하지 않다는 문제를 제기함
- 디지털 정보 유통 환경에서 플랫폼의 자율적 판단이 지식 접근 구조 전반에 미치는 영향이 확대되고 있는 만큼, 향후 플랫폼 책임과 규범 설정에 대한 논의가 중요해질 필요성이 있음을 의미함

참고문헌

- Ernesto Van der Sar, “Google Removes Sci-Hub Domains from U.S. Search Results Due to Dated Court Order”, Torrentfreak, 2025.12.17., <https://torrentfreak.com/google-removes-sci-hub-domains-from-u-s-search-results-due-to-dated-court-order/>
- Cristian Tala Sánchez, “Google, Sci-Hub y el bloqueo de dominios: impacto y contexto”, El Ecosistema Startup, 2025.12.13., <https://ecosistemastartup.com/google-sci-hub-y-el-bloqueo-de-dominios-impacto-y-contexto/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

안나스 아카이브, 스포티파이 음원 대규모 스크래핑 주장... AI 학습 악용 우려 확산

뉴스브리프

해적판 도서·논문 공유 사이트 안나스 아카이브는 2025년 12월 스포티파이 음원 8,600만 개와 메타데이터 2억 5,600만 행을 스크래핑했다고 발표하면서 관련 우려가 확산되고 있다. 해당 데이터는 약 300테라바이트 규모로 스포티파이 전체 청취의 약 99.6%를 포괄하며, 공개된 음악 메타데이터 데이터베이스 가운데 최대 규모에 해당한다. 안나스 아카이브는 스포티파이의 인기도 지표를 기준으로 수집 우선순위를 설정하고 DRM 우회 기술을 활용해 데이터를 수집했다. 수집 데이터는 인류의 음악 유산 보존을 명분으로 토렌트 기반 네트워크를 통해 단계적으로 배포될 계획이며, 관련 업계에서는 해당 데이터가 AI 모델 학습에 활용될 가능성이 높다고 경고하고 있다. 이번 사건은 AI 학습 데이터에 대한 법적·기술적·제도적 대응이 시급함을 보여준다.

안나스 아카이브의 스포티파이 음원 스크래핑 사건 발생 배경

• 안나스 아카이브의 대규모 무단 스크래핑과 AI 학습 데이터 사용 우려 제기

- 해적판 도서·논문 공유 사이트 안나스 아카이브(Anna's Archive)는 2025년 12월 20일 자체 블로그를 통해 스포티파이 (Spotify) 음원을 대규모로 스크래핑했다고 공개했으며, 수집된 음원 파일과 메타데이터 전체 규모는 약 300 테라바이트(TB)에 이른다고 설명함¹⁾
- 스포티파이는 전 세계 7억 명 이상의 사용자를 보유한 글로벌 최대 음악 스트리밍 플랫폼으로, 1억 곡 이상의 음원을 서비스하고 있음. 스포티파이 측에 따르면 이번 스크래핑은 디지털 저작권 관리 (Digital Rights Management, 이하 DRM)*를 우회하는 무단 접근을 통해 이뤄짐

* 디지털 저작권 관리(Digital Rights Management, DRM): 전자책, 음원, 영상 등 디지털 콘텐츠의 불법 복제와 무단 사용을 막기 위해 접근을 제어하고 저작권을 보호하는 기술 및 시스템

1) Anna's Archive, "Backing up Spotify", Anna's Blog, 2025.12.20., <https://annas-archive.li/blog/backing-up-spotify.html>

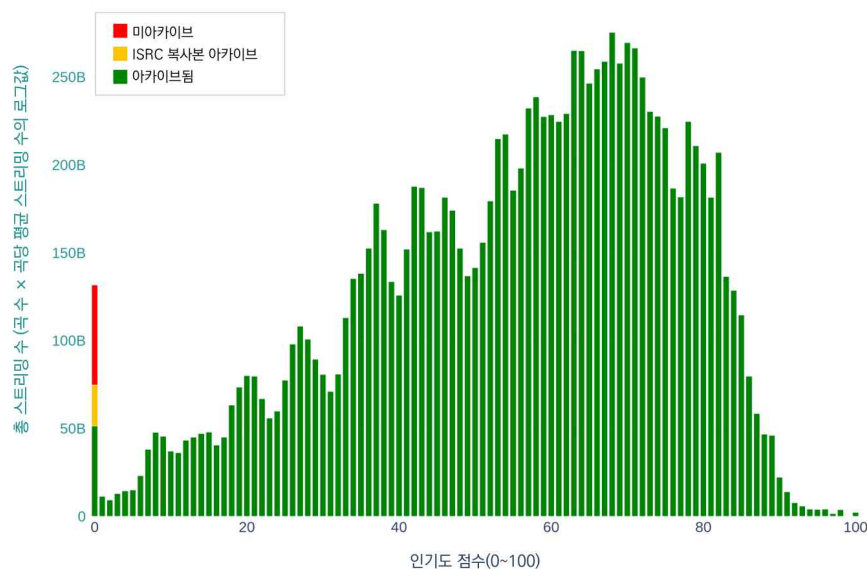
- 이번 사건은 AI 활용이 확산되는 환경에서 대규모 데이터 수집과 저작권 침해 문제가 결합된 사례로 평가됨. 특히 음원 파일과 메타데이터가 외부에 공개될 경우 AI 학습 데이터로 활용될 가능성이 제기되면서, 저작권 보호의 한계와 기술 규제 필요성을 둘러싼 논의가 확산되고 있음

안나스 아카이브의 스크래핑 데이터 구조와 수집 방법론

• 스크래핑 방법론과 우선순위 설정 체계

- 안나스 아카이브는 스포티파이의 인기도 지표를 기준으로 트랙 수집 우선순위를 설정함. 해당 인기도 지표는 0에서 100 사이의 값으로, 총 재생 횟수와 최근 재생 빈도를 알고리즘으로 산출한 수치이며, 현재 소비 비중이 높은 곡일수록 높은 점수가 부여되는 구조임
 - 수집 과정에서는 DRM 우회 기술이 활용되었으며, 인기도가 높은 곡부터 순차적으로 수집하는 방식이 적용하여 대중적 소비가 집중된 음원을 우선적으로 확보하는 전략을 채택함
 - 동일한 국제 표준 음원 코드(ISRC)*를 가진 곡이 복수로 존재하는 경우, 인기가 가장 높은 버전만을 보관해 중복을 제거하여 저장 공간을 효율적으로 관리하는 동시에 음원의 대표성을 유지함
- * 국제 표준 음원 코드(International Standard Recording Code, ISRC): 각 음원 녹음물에 고유하게 부여되는 국제 표준 식별자로, 음원 유통 과정에서 저작권 관리, 사용량 집계, 로열티 정산을 위해 활용됨
- 인기도가 0인 곡의 경우에는 원곡자의 팔로워 수와 앨범 인기도를 결합한 보조 우선순위 지표를 적용해 수집 여부를 판단함. 이에 따라 청취 빈도는 낮지만 일정 수준 이상의 부차적 지표를 충족하는 곡 역시 선별적으로 수집 대상에 포함됨

[그림1] 총 스트리밍 수 기준 아카이브된 트랙 규모 추정치



출처: Anna's Archive, "Backing up Spotify", Anna's Blog, 2025.12.20., <https://annas-archive.li/blog/backing-up-spotify.html>

• 스크래핑된 데이터의 구성 요소2)

- 안나스 아카이브가 수집한 데이터는 메타데이터와 음원 파일로 구성됨. 메타데이터에는 2억 5,600만 개의 트랙 정보와 1억 8,600만 개의 고유 ISRC가 포함되어 있으며, 공개적으로 알려진 음악 메타데이터 데이터베이스 가운데 최대 규모에 해당함

2) Anna's Archive, "Backing up Spotify", Anna's Blog, 2025.12.20., <https://annas-archive.li/blog/backing-up-spotify.html>

- 음원 파일은 총 8,600만 개로, 스포티파이 전체 청취의 약 99.6%를 포괄하는 수준임. 인기도 지표에 따라 인기가 높은 곡은 원본 OGG Vorbis* 160kbit/s 형식으로 수집되었으며, 인기가 적은 곡은 OGG Opus** 75kbit/s로 재인코딩되어 저장 용량을 절감함

* OGG Vorbis: 오픈소스 기반의 손실 압축 오디오 코덱으로, 중·고음질 음원 스트리밍에 주로 사용됨

** OGG Opus: 저비트레이트 환경에 최적화된 오디오 코덱으로, 음질 대비 파일 크기를 크게 줄일 수 있어 대규모 데이터 저장에 적합함

- 전체 데이터 용량은 약 300테라바이트(TB)로, 압축된 토렌트 형태로 보관됨. 안나스 아카이브는 해당 데이터가 2025년 7월 이전 발매된 곡을 중심으로 수집된 만큼, 이후 업로드된 콘텐츠는 포함되지 않았을 가능성이 있다고 밝힘

안나스 아카이브의 배포 전략과 AI 학습 악용 우려

• 안나스 아카이브가 제시한 음악 보존 명분과 데이터 배포 방식

- 안나스 아카이브는 인류의 음악 유산을 자연재해, 전쟁, 예산 삭감 등으로 인한 소실 위험으로부터 영구적으로 보호한다는 명분 아래, 인류의 지식과 문화 보존을 목적으로 한 음악 보존 아카이브 구축을 핵심 목표로 제시함
- 이에 관해 안나스 아카이브는 기존 음악 아카이브가 인기 음원에 편중되고 고음질 위주로 수집되면서 음악 전체를 포괄적으로 보존하기 어렵고, 전 세계 음악을 대표할 권위 있는 토렌트 목록 역시 부재하다는 점을 이번 아카이빙의 이유라고 설명함
- 안나스 아카이브는 수집된 데이터를 토렌트 기반 분산형 P2P 네트워크를 통해 단계적으로 배포할 계획이며, 충분한 저장 공간만 확보되면 누구나 미러링이 가능한 완전 개방형 아카이브를 지향함
- 2025년 12월 메타데이터가 공개되었으며, 이후 음원 파일이 인기도 순서에 따라 순차적으로 배포될 예정임. 이와 함께 앨범 아트, 원본 음원 재구성을 위한 패치 파일 역시 공개 계획에 포함됨

• AI 학습 데이터 악용 가능성에 따른 업계 반응

- 스포티파이는 사건 공개 직후 제3자의 일부 음원 무단 접근과 불법 스크래핑 관여 계정을 확인해 이를 차단했다고 밝힘. 아울러 반저작권 공격에 대응하기 위한 신규 보안 장치를 도입하고, 의심 행위를 상시 모니터링하는 한편 저작권 보호를 위해 업계 파트너들과 협력하고 있다고 설명함
- AI 스타트업 서드 체어(Third Chair)의 창업자인 요아브 짐머만(Yoav Zimmerman)은 기술적으로 이미 누구나 개인용 무료 음악 스트리밍 구축과 현대 음악의 대규모 학습이 이미 가능해졌으며, 이는 법적 책임과 집행의 위험이 따른다고 분석함³⁾
- 작곡가이자 저작권 옹호 활동가인 에드 뉴턴-렉스(Ed Newton-Rex)는 유출 음원이 AI 모델 학습에 활용될 가능성을 경고하며, 해적판 콘텐츠를 활용한 AI 학습이 업계 전반에서 빈번히 발생하고 있는 만큼 정부가 AI 기업의 학습 데이터 출처 공개를 의무화해야 한다고 주장함⁴⁾
- 한편 메타(Meta Platforms, Inc.)는 과거 리브젠(LibGen)으로 알려진 대규모 해적판 도서 아카이브를 AI 모델 학습에 활용한 전례가 있으며, 미국 법원 문서에 따르면 메타의 CEO 마크 저커버그(Mark Zuckerberg)는 사내 AI 임원진의 우려에도 불구하고 해당 데이터 활용을 승인한 것으로 나타남⁴⁾

3) Mitchell Peteres, "Spotify Music Library Scraped by Pirate Activist Group", Billboard, 2025.12.21., <https://www.billboard.com/business/streaming/spotify-music-library-leak-1236143970/>

4) Dan Milmo, "Activist group says it has scraped 86m music files from Spotify", The Guardian, 2025.12.22., <https://www.theguardian.com/technology/2025/dec/22/activist-group-says-it-has-scraped-86m-music-files-from-spotify>

시사점: 음원 저작권 보호를 위한 법적·기술적 과제

• 법적·기술적 대응 방안의 필요성

- 이번 사건은 저작권 침해가 AI 학습 데이터 수집과 결합되는 양상을 보여주며, 해적판 콘텐츠가 AI 모델 학습에 활용될 가능성을 현실적인 문제로 부각시킴. 대규모 음원 데이터의 무단 유출은 저작권 보호를 위한 법적·기술적·제도적 대응이 시급함을 드러냄
- 법적 측면에서는 AI 기업이 활용하는 학습 데이터의 출처 공개 의무화가 핵심 과제로 제기됨. 구체적 사례로 영국 정부는 저작권과 AI의 관계에 대한 정책 검토를 진행 중이며, 관련 정책 제안을 2026년 3월 18일까지 발표할 계획임
- 기술적 측면에서는 DRM 기술 강화와 스크래핑 탐지 시스템 고도화 필요성이 제기됨. 스포티파이는 신규 보안 장치를 도입했으나 대규모 스크래핑을 사전에 차단하는 데에는 한계가 있었던 것으로 평가됨. 이에 따라 개별 플랫폼 차원의 기술적 방어만으로는 완전한 보호가 어려움이 확인됨
- 제도적 측면에서는 저작권자 동의를 기반으로 한 AI 학습 체계 구축이 필요함. 영국 정부가 저작권자의 명시적 거부가 없으면 AI 기업의 무단 활용을 허용하는 방안을 검토하자, 정책 자문 과정에서 거의 모든 응답자가 창작자 우려를 지지하며 정당한 보상 구조 마련을 촉구한 바 있음

참고문헌

- Dan Milmo, “Activist group says it has scraped 86m music files from Spotify”, The Guardian, 2025.12.22., <https://www.theguardian.com/technology/2025/dec/22/activist-group-says-it-has-scraped-86m-music-files-from-spotify>
- Mitchell Peteres, “Spotify Music Library Scraped by Pirate Activist Group”, Billboard, 2025.12.21., <https://www.billboard.com/business/streaming/spotify-music-library-leak-1236143970/>
- Anna’s Archive, “Backing up Spotify”, Anna’s Blog, 2025.12.20., <https://annas-archive.li/blog/backing-up-spotify.html>
- Keran Mahadik, “How hacktivists scraped 300TB of Spotify music, and why it matters for AI”, The Indian Express, 2025.12.24., <https://indianexpress.com/article/technology/tech-news-technology/how-hacktivists-scraped-300tb-spotify-music-why-it-matters-ai-10434412/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

호주 TDM 예외 규정 도입의 한계와 유료 라이선스 중심의 정책적 대응 방안

뉴스브리프

호주 생산성위원회가 제안한 텍스트 및 데이터 마이닝(TDM) 예외 규정 도입 제안은 제작 생태계 훼손 우려 등 특정 단체의 강한 반발에 직면하며 정책적 한계를 드러냈다. 이는 일방적인 규제 완화만으로는 기술 혁신과 권리 보호 간 이해관계 균형을 달성하기 어렵다는 점을 시사한다. 특히 사후 탐지 중심 접근은 보상 체계와 데이터 검증 메커니즘이 부재한 상황에서 저작권자의 권익을 충분히 보호하기 어렵다는 구조적 한계를 보여준다. 이에 호주 정부는 해당 제안을 수용하지 않고, 유료 라이선스 체계 검토와 3년 유예를 통한 단계적 제도 정비로 정책 방향을 전환하였는데, 이는 데이터 전송 인프라 구축을 통해 혁신과 보호를 조화시키려는 보다 실질적인 정책 해법으로 평가된다.

호주 생산성위원회의 TDM 예외 규정 도입 제안과 철회

• 생산성위원회의 TDM 예외 규정 도입 제안

- 호주 생산성위원회(Productivity Commission)는 2025년 8월 AI 모델 학습을 위해 저작권 자료를 활용할 수 있도록 텍스트 및 데이터 마이닝(TDM) 예외 규정을 도입하는 방안을 제안함
- 생산성위원회는 이미 상당수 호주 데이터가 AI 학습에 활용되고 있다는 점을 강조하며, 이를 통해 호주 특화 AI 모델 개발이 가능해지고, 경제 전반의 생산성 향상에도 긍정적 영향을 미칠 수 있다는 입장을 표명함

• 제작 업계의 강력한 반발과 정부의 제안 철회

- 음악 산업 단체들은 해당 제안이 '생산성 향상'을 명분으로 디지털 해적(digital piracy)* 행위를 합법화하려는 시도라고 비판하며 집단적으로 반발함

*디지털 해적(digital piracy): 저작권자의 허가 없이 디지털 콘텐츠를 무단으로 복제 및 배포하거나 사용하는 행위

- 미국출판사협회(Association of American Publisher, AAP)는 AI 학습용 데이터셋으로 저작물을 허가 없이 사용하는 것은 저작권자 및 출판사의 직업과 생계를 위협하는 행위라고 밝힘
- 이러한 상황 속에서, 호주 정부는 지난 10월 AI 학습을 목적으로 저작물을 활용할 권한을 부여하지

않겠다는 입장을 밝히며, 생산성위원회의 제안을 사실상 철회함

• 3년 유예 후 타당성 검토로 정책 방향 전환

- 생산성위원회는 2025년 12월 발표한 최종 보고서에서 기존 TDM 예외 규정 제안을 철회하고 향후 3년의 유예기간을 거쳐 호주 저작권 체계 및 AI 기술 영향을 종합적으로 검토할 것이라고 발표함
- 다만, 정체된 국가 생산성을 회복하기 위해 AI 등 디지털 기술의 적극적 활용이 필요하다는 점을 재차 강조하면서, 호주 내 기술 혁신 환경 조성의 필요성에 대해 강조함

기술 기업과 제작 업계 간 이해관계 충돌

• 기술 기업의 투자 유치 논리와 경제적 주장

- 호주기술위원회(Tech Council of Australia, TCA)는 AI 모델 학습을 위한 저작물 활용 제약을 해결하는 것이 호주를 글로벌 AI 투자 허브로 만드는 방안이며, 이를 통해 수십억 달러의 해외 투자 유치가 가능할 것이라고 강조함
- 또한, 생산성위원회는 국가 생산성이 향상될 경우, 2035년까지 근로자의 실질 임금 수준이 연간 최소 14,000달러(한화 약 2,000만 원) 증가할 수 있다고 분석한 바 있는데¹⁾, 이는 AI와 같은 디지털 혁신이 단순한 기술 도입을 넘어 국민의 소득 증대와 삶의 질 향상으로 이어질 수 있음을 시사함.

• 제작자·미디어 단체의 저작권 보호 요구

- 호주 일부 작가들은 대규모 AI 시스템이 책·예술 작품·공연물 등 디지털화 가능한 모든 저작물을 대상으로 한 ‘대규모 무단 전유’(massive unauthorised appropriation)*에 기반해 구축되고 있다며 강하게 비판함

* 대규모 무단 전유(massive unauthorised appropriation): 저작권자의 허락이나 정당한 보상 없이 타인의 저작물을 가져와 사용하면서 그 가치와 이익을 사실상 독점하는 행위

- 또한, 호주노동조합평의회(Australian Council of Trade Unions, ACTU)는 생산성위원회의 제안이 저작물에 대한 ‘무분별한 도용’을 정당화하는 통로가 되어 제작자들의 생존권을 위협할 수 있다며 우려를 표명함

[표1] 호주 TDM 예외 규정 도입에 대한 주요 이해관계자별 입장

구분	기술 기업 및 경제 단체	제작자	정부
주요 입장	AI 투자 허브 도약 및 혁신 강조	저작권 보호 및 생존권 사수	공정 보상 및 단계적 규제 설계
핵심 논거	글로벌 투자 유치를 위한 규제 완화 국가 생산성 향상 및 실질 임금 증대	무단 전유를 통한 디지털 해적 행위 제작자의 직업적 생계 위협	무상 면제 철회 및 유료화 검토 데이터 전송 인프라 구축 추진

출처: Patrick Commins, "Proposal to allow use of Australian copyrighted material to train AI abandoned after backlash", The Guardian, 2025.12.19., <https://www.theguardian.com/australia-news/2025/dec/19/proposal-australian-copyrighted-material-train-ai-abandoned-after-backlash>

1) Patrick Commins, "Proposal to allow use of Australian copyrighted material to train AI abandoned after backlash", The Guardian, 2025.12.19., <https://www.theguardian.com/australia-news/2025/dec/19/proposal-australian-copyrighted-material-train-ai-abandoned-after-backlash>

제작자 권리와 디지털 혁신의 조화를 위한 향후 발전 방향

• 유료 라이선스 체계 도입을 통한 혁신-보호 상생 모델 구축

- 호주 정부는 기존의 자발적 협약에 의존하던 접근을 넘어 저작권법에 근거한 유료 라이선스 체계 도입 가능성을 본격적으로 검토하기 시작함
- 이는 AI 산업 발전을 이유로 저작권 보호를 후퇴시키기보다는, 정당한 보상을 전제로 한 공존 모델이 장기적으로 더 지속 가능하다는 정책적 판단을 반영한 것으로 해석됨
- 상기 관련, 가디언 오스트레일리아(The Guardian Australia), 뉴스코프(News Corp) 등 주요 언론 및 호주음반산업협회(Australian Recording Industry Association, ARIA)를 비롯한 제작자 단체들은, 콘텐츠 사용 조건과 대가를 직접 결정할 수 있는 공정한 가치 교환 구조야말로 호주의 저작권을 보호하는 핵심 장치라고 평가함

• 기술 중립적 규제 원칙과 증거 기반의 신중한 접근법 확립

- 호주 정부는 AI에 특화된 신규 규제를 성급히 도입하기보다는, 기존의 기술 중립적 규제 체계로 해결이 불가능한 경우에 한해 법적 개입을 허용하는 ‘최후의 수단(last resort)’ 원칙을 정립함
- 이러한 태도는 AI 기술의 발전 경로와 사회적 영향에 대한 불확실성을 고려하여, 혁신을 저해하지 않으면서도 실질적인 규제 공백을 최소화하려는 증거 기반 정책 설계 의지를 보여줌

• 데이터 활용 인프라 구축을 통한 데이터 확보 전략의 패러다임 전환

- 호주 정부는 AI 학습 데이터 확보를 위해 저작권 면제와 같은 법적 특례에 의존하는 방식에서 벗어나, 표준화된 데이터 전송 및 공유 체계를 구축함으로써 산업 전반의 데이터 수요를 충족시키려는 대안적 정책 방향을 제시하고 있음
- 특히 개인과 소상공인이 자신의 데이터를 안전하고 통제 가능한 방식으로 공유할 수 있는 데이터 유통 인프라를 구축하고, 이를 소매 유통 부문부터 단계적으로 확대함으로써 AI 산업에 필요한 실질적이고 고품질의 학습 자원을 안정적으로 공급하는 전략을 추진함
- 짐 차머스(Jim Chalmers) 호주 재무장관은 상기 권고 사항들의 차기 예산안 반영을 검토할 것이라고 언급하며, 향후 AI를 포함한 디지털 기술이 국가 생산성 회복에 중요한 역할을 할 수 있음을 시사함²⁾

참고문헌

- Patrick Commins, “Proposal to allow use of Australian copyrighted material to train AI abandoned after backlash”, The Guardian, 2025.12.19., <https://www.theguardian.com/australia-news/2025/dec/19/proposal-australian-copyrighted-material-train-ai-abandoned-after-backlash>
- Kamyia Pandey, “Australian Government Says No to Free AI Training on Copyrighted Content”, MediaNama, 2025.10.28., <https://www.medianama.com/2025/10/223-australia-blocks-ai-copyrighted-content/>
- Dan Jarvis Bardy, “Labor rules out giving tech giants free rein to mine copyright content to train AI”, The Guardian, 2025.10.26., <https://www.theguardian.com/technology/2025/oct/27/labor-rules-out-giving-tech-giants-free-rein-to-mine-copyright-content-to-train-ai>

2) Patrick Commins, “Proposal to allow use of Australian copyrighted material to train AI abandoned after backlash”, The Guardian, 2025.12.19., <https://www.theguardian.com/australia-news/2025/dec/19/proposal-australian-copyrighted-material-train-ai-abandoned-after-backlash>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

어도비 집단소송으로 본 AI 학습 데이터셋의 구조적 저작권 리스크

뉴스브리프

어도비는 자사 AI 모델 'SlimLM' 학습에 활용된 데이터가 저작권을 침해했다는 혐의로 집단소송을 제기당했다. 이번 소송은 앤트로픽(Anthropic)이 해적판 도서 학습 문제로 15억 달러 규모 합의를 마친 직후 발생하며 업계의 긴장감을 고조시켰다. 이번 소송의 핵심은 어도비가 직접 데이터를 수집하지 않았더라도, 사용된 데이터셋이 불법 복제 서적 19만여 권이 포함된 데이터의 파생물이라는 점이다. 최근 빅테크 기업이 동일한 데이터셋 활용 문제로 연쇄 제소됨에 따라, 어도비 사례는 특정 기업의 문제를 넘어 AI 산업 전체의 구조적 리스크로 확장되었다. 관련하여 학습 데이터 공급망 전반의 합법성을 입증하는 '데이터셋 계보' 관리가 기업의 핵심 생존 전략이 될 것으로 전망된다.

AI 학습 데이터 저작권 분쟁의 구조적 심화

• 생성형 AI 확산에 따른 무단 학습 논쟁의 재점화

- 생성형 AI 기술의 급격한 확산과 모델 고도화 과정에서, 방대한 양의 저작물을 학습 데이터로 활용함에 따른 무단 이용 및 저작권 침해 여부가 글로벌 AI 산업의 핵심 쟁점으로 부상함
- 특히 대형언어모델(LLM) 및 파생 모델들이 성능 향상을 위해 대규모 텍스트 데이터셋을 구축하는 과정에서, 불법 복제 자료의 포함 여부와 그에 따른 책임 소재가 지속적으로 문제 제기되고 있음

• 글로벌 AI 기업 대상 저작권 침해 소송 및 법적 리스크 확산

- 어도비(Adobe)는 2025년 12월 자사의 소형언어모델(Small Language Model, SLM)*인 'SlimLM' 학습 과정에서 작가들의 저작권을 침해했다는 혐의로 캘리포니아 연방법원에 소송을 제기당함¹⁾

* 소형언어모델(SLM): 대형언어모델보다는 작은 규모이지만, 모바일 기기 내 문서 지원 등 특정 작업에 최적화되어 효율성이 높은 AI 모델

1) Blake Brittain, "Adobe sued for allegedly misusing authors' work in AI training", Reuters, 2025.12.17., <https://www.reuters.com/legal/government/adobe-sued-allegedly-misusing-authors-work-ai-training-2025-12-17/>

- 이번 소송은 2025년 9월 AI 기업 앤트로픽(Anthropic)이 약 50만 권의 도서 저작권자들에게 총 15억 달러(약 2조 1,900억 원)²⁾를 지급하기로 합의한 사건 직후 발생함³⁾
- 앤트로픽 합의안은 소송의 기폭제가 되어 어도비뿐만 아니라 애플(Apple), 세일즈포스(Salesforce) 등 공개되어 있는 유사한 데이터셋을 활용한 주요 기업들에 대한 연쇄적인 법적 대응으로 이어지고 있음

[표1] 주요 AI 기업 저작권 소송 현황 비교

기업명	소송제기 시점	쟁점 데이터셋	현재 상태	비고
앤트로픽 (Anthropic)	2024년 초	Books3, LibGen, Pirate Library Mirror	15억 달러 합의 (2025.09)	저작권 사건 역사상 최대 합의금
애플 (Apple)	2025년 10월	Books3, RedPajama, Pile	소송 진행중	Apple Intelligence 및 Open ELM 모델 관련
세일즈포스 (Salesforce)	2025년 10월	RedPajama	소송 진행중	xGen AI 모델 학습 데이터 쟁점
어도비 (Adobe)	2025년 12월	SlimPajama(Books3 포함)	소송 진행중	SlimLM 모델 학습 과정에서 침해 주장

출처: 참고문헌 종합하여 재구성

학습 데이터셋 계보와 책임 소급의 쟁점

• 오픈 데이터셋의 저작권 문제 부상과 파생 책임의 쟁점

- 작가 엘리자베스 라이언(Elizabeth Lyon)은 2025년 12월 17일 어도비가 'SlimLM' 학습에 자신의 저작권을 침해한 해적판 데이터를 활용했다며 캘리포니아 연방법원에 집단소송을 제기함
- 어도비의 SlimLM 학습에 사용한 'SlimPajama' 데이터셋은 'Red Pajama'의 파생본이며, Red Pajama가 'Books3'를 복제하여 구축되었다는 계보가 밝혀지면서 데이터 획득 경로의 합법성이 핵심 쟁점이 됨
- 'Books3'는 약 196,640권 이상의 불법 복제 서적을 포함한 '그림자 도서관' 데이터셋으로, AI 모델 성능 고도화를 위해 다수의 빅테크 기업이 활용해 옴⁴⁾
- 원고는 기업이 직접 자료를 수집하지 않고 제삼자의 오픈소스·가공 데이터셋을 활용했더라도, 원천 데이터의 저작권 쟁점에 대한 법적 책임은 모델 개발사에 소급 적용된다는 논리를 강조함

• 학습 데이터셋의 '오염'에 따른 산업계 공통 리스크 전이

- 어도비가 직면한 리스크는 개별 기업의 관리 부실을 넘어, AI 성능 고도화를 위해 산업계가 공통으로 공유해온 데이터셋 자체가 오염 상태였음을 방증함
- 특히 'Books3'와 같이 불법 복제물을 포함한 원천 데이터가 여러 파생 데이터셋으로 복제 및 확산되면서, 이를 인지하지 못하고 활용한 기업에게 법적 리스크가 전이되는 구조적 한계가 드러남
- 이러한 데이터셋 오염 문제는 기업이 직접 데이터를 수집했는지 여부와 상관없이, 최종 모델의 정당성을 위협하는 '공통 리스크 자산'으로 작용하며 산업 전반의 신뢰도를 저하시키는 요인이 됨

2) 1달러=1,460.60원(2025.12.26, KEB 하나은행 최초 매매기준율 적용)

3) Scott Clark, "Inside Anthropic's \$1.5B Generative AI Copyright Lawsuit Settlement", VKTR, 2025.09.07., <https://www.vktr.com/ai-news/inside-anthropic-15b-generative-ai-copyright-lawsuit-settlement/>4) Rose Esfandiari, "Authors Take Page from Anthropic in Alleging Apple Infringed Works by Training AI on Pirated Books", IPWatchdog, 2024.10.14., <https://ipwatchdog.com/2025/10/14/authors-take-page-anthropic-alleging-apple-infringed-works-training-ai-pirated-books/>

• AI 학습 데이터 책임 구조의 변화와 기업의 주의 의무 강화

- 이번 소송들은 ‘공개된 데이터셋은 사용해도 안전하다’는 기존의 통념을 무너뜨리고 있으며, 데이터 획득 경로에 대한 검증 책임을 직접 학습 단계에서 최종 모델 활용 단계까지 확장시키고 있음
- 기업들은 이제 제삼자로부터 정제된 데이터를 공급받더라도 그 원천까지 거슬러 올라가 합법성을 확인해야 할 엄격한 의무를 안게 되었으며, 이를 증명하지 못할 경우 막대한 배상 책임뿐만 아니라 모델 폐기라는 극단적 위험에 노출됨
- 결과적으로 AI 모델의 기술적 성능만큼이나 데이터의 ‘합법적 계보’를 관리하는 역량이 기업의 핵심 컴플라이언스 요소로 부상하고 있으며, 관련 검증 및 관리 비용의 급격한 상승이 예상됨

시사점: AI 데이터 거버넌스의 패러다임 전환

• 데이터셋 계보 중심의 합법성 검증 체계 확립 필요

- 데이터셋을 기반으로 한 소송은 오픈소스나 공개 데이터셋의 활용이 안전하다는 기존의 인식을 붕괴시키며, 기업의 데이터 검증 책임을 직접 학습 단계에서 최종 모델 활용 단계까지 확장시키는 계기가 됨
- 이에 따라 향후 AI 기업들은 데이터 수집-정제-학습 전 과정에서 저작권 침해 요소를 사전에 차단하는 기술적 구축과 더불어, 원천 데이터의 출처를 투명하게 관리하는 ‘데이터셋 계보’ 확보가 필수적임
- 특히 소송의 쟁점이 ‘직접 복제’에서 ‘파생 활용’으로 넓어지는 추세에 발맞추어, 정당한 대가를 지불하는 합법적 라이선스 시장과의 협력을 강화하고 권리자의 옵트아웃(Opt-out)권*을 실질적으로 보장하는 등 투명한 데이터 거버넌스 수립이 기업의 핵심 경쟁력으로 판단됨

* 옵트아웃(Opt-out)권: 저작권자가 자신의 저작물이 AI 학습 데이터로 활용되는 것을 사전에 또는 사후에 거부할 수 있는 권리를 의미하며, 최근 AI 학습 데이터의 투명성 논의와 함께, 저작권자의 선택권을 보장하기 위한 핵심 제도로 주목받고 있음

참고문헌

- Blake Brittain, "Adobe sued for allegedly misusing authors' work in AI training", Reuters, 2025.12.17., <https://www.reuters.com/legal/government/adobe-sued-allegedly-misusing-authors-work-ai-training-2025-12-17/>
- Lucas Ropek, "Adobe hit with proposed class-action, accused of misusing authors' work in AI training", TechCrunch, 2025.12.17., <https://techcrunch.com/2025/12/17/adobe-hit-with-proposed-class-action-accused-of-misusing-authors-work-in-ai-training/>
- Scott Clark, "Inside Anthropic's \$1.5B Generative AI Copyright Lawsuit Settlement", VKTR, 2025.09.11., <https://www.vktr.com/ai-news/inside-anthropics-15b-generative-ai-copyright-lawsuit-settlement/>
- Rose Esfandiari, "Authors Take Page from Anthropic in Alleging Apple Infringed Works by Training AI on Pirated Books", IPWatchdog, 2024.10.14., <https://ipwatchdog.com/2025/10/14/authors-take-page-anthropic-alleging-apple-infringed-works-training-ai-pirated-books/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

[사례 탐구] NFT 거래 환경에서의 이미지 저작권 검증 기술과 AI 기반 사전 탐지 접근

뉴스브리프

NFT로 발행·유통되는 디지털 이미지 콘텐츠는 블록체인을 통해 거래 이력과 소유 기록을 확인할 수 있으나, 토큰의 소유와 이미지 저작권의 범위가 기술적으로 분리된 구조로 인해 권리 관계가 명확히 검증되지 않은 채 거래가 이루어질 가능성을 내포하고 있다. 특히 생성형 AI 확산과 함께 기존 이미지나 제3자의 저작물이 NFT로 발행·유통되는 사례가 증가하면서, 거래 이후 저작권 분쟁으로 이어지기도 한다. 그러나 현재의 NFT 거래 환경에서는 해당 이미지가 다른 저작물과 유사한지, 이미 NFT로 발행된 이력이 있는지 여부를 일반 이용자가 사전에 판단하기 어려운 구조적 한계가 있다. 이에 따라 최근에는 NFT 거래 이전 단계에서 디지털 이미지 콘텐츠의 저작권 침해 가능성을 기술적으로 점검하려는 시도가 증가하고 있다. 하나의 참고 사례로 국내 블록체인·웹3 보안 기업 엘에스웨어가 개발한 'NFT-EYES' 기술을 조명해 보도록 한다.

NFT로 거래되는 이미지의 저작권 검증 필요성 증대

- NFT로 발행·유통되는 디지털 콘텐츠 권리 관계 명확화 필요성 증대
 - 최근 생성형 AI와 블록체인 기술의 확산으로 디지털 이미지 콘텐츠가 대체불가능토큰(Non-Fungible Token, NFT) 형태로 발행·거래되는 사례가 증가하고 있음
 - NFT는 블록체인 기술을 활용해 디지털 자산의 거래 이력을 기록하고 소유자를 확인할 수 있도록 설계되었으나, 토큰의 거래와 이미지 저작권의 보호는 본질적으로 다른 영역임
 - NFT 구매 시 이미지의 저작권까지 함께 이전되는 것으로 이해되는 경우도 있으나, 실제로는 다수의 NFT 거래에서 이미지에 대한 저작권은 저작권자에게 그대로 유지되며, 구매자는 블록체인 상의 토큰 소유 기록을 보유하는 구조임
 - 이러한 구조적 특성으로 인해, 기존 이미지나 제3자의 저작물이 NFT로 발행·유통되는 사례가 발생할 수 있으며, 권리 관계가 명확하지 않은 상태에서 거래가 이루어지는 문제도 제기되고 있음

• NFT로 거래되는 콘텐츠 타 저작물 유사성과 권리 관계 판단하기 어려워

- 문제는 NFT 거래 과정에서 해당 이미지가 다른 저작물과 유사한지, 이미 다른 NFT로 발행된 이력이 있는지, 또는 권리 관계가 명확한지 여부를 일반 이용자가 사전에 확인하기 어렵다는 점임
- 이로 인해 NFT 시장에서는 거래 이후 저작권 분쟁이 발생하거나, 구매자가 의도치 않게 저작권 침해 가능성이 있는 콘텐츠를 보유·유통하게 되는 위험이 있음
- 이 때문에 최근 NFT로 발행·유통되는 디지털 이미지 콘텐츠의 저작권 침해 가능성을 거래 이전 단계에서 기술적으로 점검할 수 있는 수단의 필요성이 점차 커지고 있음

[사례 탐구] NFT 이미지 저작권 침해 탐지 최신 기술

• NFT 거래 과정에서 안전한 저작물인지 여부 검증하는 기술

- 최근 국내 블록체인·웹3 보안 솔루션 개발 기업인 엘에스웨어(LSWARE)가 NFT로 발행·거래되는 디지털 이미지 콘텐츠의 저작권 침해 가능성을 사전에 점검하기 위한 AI 기반 기술인 'NFT-EYES'를 베타 서비스 형태로 공개함
- NFT-EYES는 NFT 거래 과정에서 이미지가 안전한 저작물인지 여부를 기술적으로 확인할 수 있도록 설계된 저작권 침해 탐지 기술임



• 플랫폼에 분산돼 이루어지는 NFT 거래 특성 고려하여 특정 플랫폼에 한정되지 않고 저작권 검증

- 사용자는 NFT로 발행하려는 이미지나 거래 대상 이미지를 시스템에 업로드하면, 해당 이미지가 블록체인 상에 이미 존재하는 다른 NFT 이미지와 유사한지 여부를 자동으로 분석할 수 있음
- 이 기술의 핵심은 AI 기반 이미지 유사도 분석 방식으로, 기존의 단순한 파일 비교나 해시값 대조 방식과 달리, 이미지의 색상, 형태, 구성 요소 등 시각적 특징을 수치화해 분석함
- 이를 통해 동일 이미지를 그대로 복제한 경우뿐 아니라, 일부 수정·편집·변형이 가해진 이미지에 대해서도 유사성을 탐지할 수 있도록 설계됨
- 현재 베타 버전에서는 이더리움(Ethereum) 블록체인 기반 NFT와 글로벌 NFT 마켓플레이스 오픈씨(OpenSea)를 중심으로 탐지가 이뤄지고 있는데, 이용자는 별도의 전문 지식 없이 이미지를 업로드하는 것만으로, 해당 이미지와 유사한 NFT가 이미 발행된 사례가 있는지를 확인할 수 있음
- NFT-EYES는 아이즈프로토콜(EYES Protocol)*의 웹3 보안·신뢰 인프라 위에서 작동하며, NFT 거래 과정에서 저작권 검증 기능을 하나의 절차로 연결하는 방식으로 작동되므로, NFT 창작·전시·거래가 개별 플랫폼에 분산돼 이루어지는 현실을 고려하여 특정 플랫폼에 한정되지 않고 저작권 검증이 가능함

* 아이즈프로토콜(EYES Protocol): 엘에스웨어가 자체 개발한 웹3 보안 인프라로, NFT 디지털 이미지 콘텐츠의 저작권 침해 탐지를 위한 서비스인 NFT-EYES를 운영하고 있음

- **이미지 파일을 그대로 대조하는 방식이 아니라, 이미지가 어떤 모습으로 보이는지를 기준으로 비교**
 - NFT-EYES는 이미지 파일을 그대로 대조하는 방식이 아니라, 이미지가 어떤 모습으로 보이는지를 기준으로 비교하는 방식으로 작동됨
 - 사용자가 이미지를 업로드하면, AI가 해당 이미지의 색감, 형태, 전체적인 구성과 같은 시각적 특징을 분석해 하나의 비교용 정보로 변환하는데, 이렇게 변환된 정보는 블록체인에 등록된 다른 NFT 이미지들의 정보와 비교되어, 서로 비슷한 이미지가 이미 존재하는지 여부를 확인함
 - 실제 NFT 시장에서는 같은 이미지가 해상도를 바꾸거나 색을 조금 조정하는 식으로 변형돼 유통되는 경우가 많아, 단순히 파일이 같은지만 확인하는 방식으로는 한계가 존재하는데, NFT-EYES 기술 작동 방식은 파일 이름이나 설명 문구가 달라져 있거나, 이미지가 일부 수정된 경우에도 여전히 비슷한 이미지인지 판단할 수 있음
 - 이러한 방식은 ‘블록체인에 기록된 토큰’과 ‘실제 이미지 콘텐츠’로 나뉘어 있다는 구조를 전제로 작동되는데, 블록체인은 누가 어떤 토큰을 가지고 있는지를 기록하는 역할을 담당하고, NFT-EYES는 그 토큰이 가리키는 이미지가 저작권 측면에서 문제가 없는지를 별도로 확인하는 역할을 함

한계점과 기술적 시사점

- **블록체인 계층과 AI 분석 계층을 분리·연결해 NFT 구조의 기술적 한계 보완**
 - NFT 거래에서는 토큰의 소유 기록과 이미지 저작권의 범위가 기술적으로 분리돼 있으며, 스마트 계약 또한 거래 조건이나 수익 배분을 자동화하는 수단에 그쳐 이미지 저작권의 적법성을 직접 보증하지는 못하는 한계가 있어 NFT 시장에서는 이미지 콘텐츠의 권리 상태가 충분히 검증되지 않은 채 거래가 일어나기도 함
 - NFT-EYES 기술 개발 사례는 저작권 문제를 블록체인 내부 규칙만으로 해결하려는 접근에서 벗어나, 콘텐츠 자체를 분석하는 별도의 기술 계층을 결합했다는 점에서 의의가 있음
 - 다시 말해, 토큰의 진위와 소유 이력을 관리하는 블록체인 계층과, 이미지 콘텐츠의 유사성과 출처를 분석하는 AI 분석 계층을 분리·연결함으로써, NFT 구조의 기술적 한계를 보완하는 것이 기술적 진보로 볼 수 있음
 - 이는 NFT 저작권 보호가 스마트 계약이나 메타데이터 설계만으로는 충분하지 않으며, 실제 콘텐츠를 대상으로 한 인식·분석 기술이 필수적으로 결합되어야 함을 시사함
 - 특히 이미지의 일부 수정이나 변형이 빈번한 NFT 시장의 특성을 고려할 때, 단순 식별자 기반 관리 방식보다 AI 기반 유사도 분석이 저작권 침해 가능성을 보다 현실적으로 포착할 수 있음을 보여줌
 - 다만 이러한 기술 역시 법적 저작권 판단을 대체하는 수단은 아니며, 최종적인 권리 귀속과 침해 여부 판단은 법적 해석에 따라 결정됨

참고문헌

- 김현민 기자, "“AI로 NFT 저작권 침해 잡는다”...엘에스웨어 'NFT-EYES' 베타 서비스 공개", 전자신문, 2025.12.23., <https://www.etnews.com/20251222000014>
- Nexa Desk, "What Does NFT Ownership Really Mean? Rights, Risks, & AI Threats, 30 Outlook, 2025.12.18., <https://www.outlookindia.com/xhub/blockchain-insights/what-does-nft-ownership-really-mean-rights-risks-ai-threats>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

주간 기술 동향

AI 생성 음성의 출처 증명, Smark 기술의 부상

• 확산 모델 기반 음성 콘텐츠 생성 AI와 워터마킹 기술

최근 텍스트-음성 변환 확산 모델과 같은 생성형 AI 기술이 인간과 거의 구별하기 힘든 수준의 고품질 음성을 생성하며 다양한 산업에 혁신을 가져오고 있다. 그러나 이러한 기술 발전의 이면에는 정치인을 사칭한 가짜 뉴스 제작, 교묘한 보이스 피싱 범죄, 그리고 원본 음성 데이터를 무단으로 학습시켜 권리를 침해하는 등 심각한 사회적, 법적 위험이 동시에 커지고 있다. 특히 음성 자산이 무단으로 복제되고 악용되면서, 기존 법체계만으로는 이를 효과적으로 규제하기 어려운 기술적 공백이 발생하고 있다.

특히 미세한 잡음에서 시작해 점진적으로 목표 음성을 완성해 나가는 확산 모델(Diffusion Model)은 기존 방식보다 훨씬 자연스럽게 표현력이 풍부한 결과물을 만들어낼 수 있어 기술의 오남용 가능성을 증폭시키고 있다. 이렇게 생성된 음성 콘텐츠가 어떤 AI 모델로부터 만들어졌는지, 또 그 모델이 학습한 데이터의 소유권은 누구에게 있는지 증명하지 못한다면, AI 기술이 만든 콘텐츠의 신뢰성은 물론 창작 생태계 전반이 위협받을 수밖에 없다.

이러한 문제를 해결하기 위해 오래전부터 오디오 워터마킹 기술이 연구되어 왔지만, 생성된 오디오 파일에 후처리 방식으로 정보를 삽입하는 전통적 기법들은 압축이나 잡음 추가 같은 간단한 공격에도 쉽게 훼손되는 견고성의 한계를 보였다. 또한 딥러닝 기술을 활용한 최신 워터마킹 기법들도 특정 유형의 AI 모델에만 적용 가능하거나 원본 음질을 저해하는 문제가 있어 범용적인 저작물 보호 솔루션으로 자리 잡기에는 기술적 과제가 남아있는 실정이다.

이에 최근에는 AI 모델이 음성을 생성하는 과정 자체에 권리 정보를 식별자 형태로 삽입하여 음질 저하를 최소화하고 외부 공격에 대한 저항성을 높이는 '생성형 워터마킹' 방식이 새로운 대안으로 주목받고 있다. 이는 콘텐츠가 생성되는 시점부터 고유 식별 정보를 갖게 함으로써, 사후에 진위를 판별하던 기존 방식에서 벗어나 출처와 소유권을 원천적으로 증명하는 선제적 보호 체계로의 전환을 의미한다. 본 보고서는 이러한 흐름 속에서 확산 모델 기반 TTS 시스템을 위해 개발된 Smark 기술을 중심으로, 이산 웨이블릿 변환을 활용한 새로운 워터마킹 사례를 심층 분석하고자 한다.

생성형 오디오 워터마킹의 기술적 특징과 한계

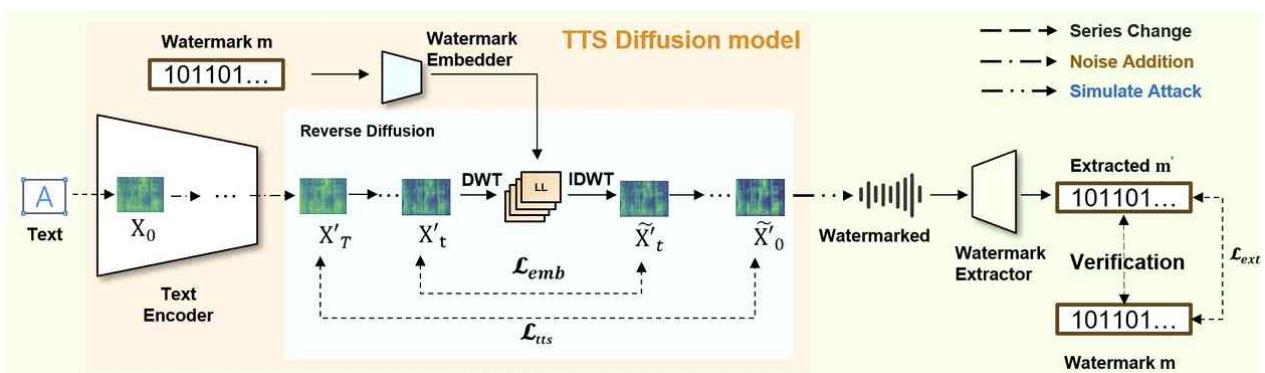
- 생성형 워터마킹은 AI 모델의 생성 과정에 직접 개입하여 식별 정보를 콘텐츠의 구조적 일부로 내재화하는 선제적 보호 방식으로, 후처리 방식 대비 압축이나 변환 공격에 대한 견고성이 향상되며 콘텐츠 출처를 원천적으로 증명하는 강력한 수단을 제공함
- 하지만 음질 저하를 막는 비가청성과 외부 공격을 견디는 견고성은 상충 관계에 있어 본질적인 기술적 난제를 안고 있으며, 알려진 공격 외에 새로운 적대적 공격에는 취약할 가능성이 존재함
- 또한, AI 모델 개발사마다 각기 다른 워터마킹 기술을 적용할 경우 상호 운용성이 보장되지 않아, 생성된 콘텐츠의 출처를 보편적으로 검증할 수 있는 표준화된 탐지 시스템의 부재가 문제로 지적됨

[사례] 이산 웨이블릿 변환 기반 TTS 확산 모델 워터마킹 기술

• Smark 기술 개발 배경 및 개요

- Smark는 확산 모델 기반의 텍스트-음성 변환(Text-to-Speech, 이하 TTS) 기술로 생성된 오디오에 권리 정보, 생성 모델 등의 식별 정보를 삽입하여, 향후 발생 가능한 법적 분쟁이나 딥페이크 악용 사례에서 콘텐츠의 출처와 소유권을 명확히 증명하는 것을 핵심 목적으로 하는 기술임
- 기존 후처리 방식이 생성하는 가청 왜곡이나 압축 등 단순한 변환에도 워터마크가 쉽게 손상되는 취약성 문제를 해결하기 위해, Smark는 생성 단계에서 식별 정보를 내재화하는 접근법을 채택함
- 잡음으로부터 음성을 복원하는 확산 모델의 역방향 프로세스 중 특정 단계에 워터마크를 삽입함으로써, 최종 결과물에 식별 정보가 분리 불가능한 형태로 자연스럽게 포함되도록 정교하게 설계됨
- Smark 기술은 높은 비가청성과 다양한 신호 처리 공격에도 정보를 보존하는 강력한 견고성이라는 상충 관계의 두 목표를 동시에 달성하는 것을 핵심 과제로 설정함

[그림] TTS 확산 모델의 DWT 기반 워터마크 삽입 및 추출 파이프라인



출처: Yichuan Zhang 외 2인, "Smark: A Watermark for Text-to-Speech Diffusion Models via Discrete Wavelet Transform", arXiv, 2025.12.21., <https://arxiv.org/pdf/2512.18791>

• Smark 작동 원리 및 특징

- Smark의 원리는 확산 모델의 지정된 노이즈 제거 단계에서, 중간 결과물인 멜 스펙트로그램*(Mel Spectrogram)에 이산 웨이블릿 변환**(Discrete Wavelet Transform, 이하 DWT)을 적용하여 신호를 여러 주파수 대역으로 분해하는 것에서 시작됨

* 멜 스펙트로그램(Mel Spectrogram): 소리 신호를 시간에 따른 주파수 스펙트럼으로 시각화할 때, 인간의 청각이 저주파에 더 민감하게 반응하는 특성을 반영하여 주파수 축을 멜 스케일(Mel scale)로 변환한 데이터 표현 방식

** 이산 웨이블릿 변환(Discrete Wavelet Transform, DWT): 디지털 신호를 여러 해상도의 주파수 대역으로 분해하여, 특정 시간대의 주파수 특성을 동시에 분석하고 처리할 수 있게 하는 신호 처리 기술

- DWT를 통해 분해된 스펙트로그램은 음성의 핵심 정보를 담고 있어 청각이 민감하게 반응하는 저주파 대역과, 상대적으로 정보량이 적어 청각적 민감도가 낮은 고주파 대역으로 구분됨.
- 권리 정보를 담은 이진 비트 시퀀스 형태의 워터마크를 청각적 왜곡 발생 가능성이 현저히 낮은 고주파 대역의 상세 계수에 선택적으로 삽입하여 원본 음질에 미치는 영향을 최소화함
- 워터마크가 삽입된 수정된 계수들은 이산 웨이블릿의 역변환을 통해 다시 멜 스펙트로그램 형태로 재구성되며, 이후 남은 확산 모델의 표준 노이즈 제거 과정을 거쳐 최종적인 오디오 파형으로 완성됨
- 이러한 생성형 삽입 방식은 워터마크가 단순한 부가 정보가 아닌 오디오 신호의 근본적인 특성으로 자리 잡게 하여, 후처리 방식으로는 달성하기 어려운 본질적인 견고성을 확보할 수 있음

● 실험 결과 및 사용 사례

- GradTTS와 WaveGrad 등 구조가 상이한 여러 TTS 변환 확산 모델을 대상으로 한 실험에서, Smark는 모델 종류에 구애받지 않고 일관적으로 높은 음질과 워터마크 탐지 성능을 보이며 기술의 범용적 적용 가능성을 검증함
- 객관적 음질 평가에서 PESQ(Perceptual Evaluation of Speech Quality)* 점수 약 4.54, STOI (Short-Time Objective Intelligibility)** 점수 1.00에 근접한 값을 달성하여, 워터마크가 삽입된 음성이 원본과 청각적으로 거의 구별 불가능한 수준임을 정량적 수치로 확인함
 - * PESQ(Perceptual Evaluation of Speech Quality): 인간 청각 심리 모델을 기반으로 음성 품질의 저하 정도를 정량적으로 측정하는 평가 지표
 - **STOI(Short-Time Objective Intelligibility): 처리된 음성의 명료도가 원본 대비 얼마나 잘 보존되었는지를 평가하는 지표
- MP3 압축, 잡음 추가, 필터링 등 10종의 공격 시뮬레이션에서 평균 99% 이상의 높은 탐지 정확도를 기록함으로써, 실제 콘텐츠 유통 과정에서 발생할 수 있는 데이터 변형에 대한 높은 저항력을 입증함
- 이와 같은 강력한 성능은 향후 AI가 생성한 오디오 콘텐츠의 권리 증명, 딥페이크 음성의 출처 추적, 방송용 콘텐츠의 불법 복제 방지 등 다양한 산업 분야에서 실용적으로 활용될 수 있음을 시사함

● 견고성, 비가청성 등 성능 평가

- 비가청성 평가는 객관적 지표(PESQ, STOI)와 주관적 청취 테스트(Mean Opinion Score, MOS)를 병행하여 이루어졌으며, 모든 평가에서 기존 워터마킹 기술인 WavMark나 GROOT보다 우수한 점수를 획득하여 음질 저하를 최소화했음을 증명함
- 견고성 평가는 오디오 파일의 형식 변환, 재샘플링, 노이즈 추가 등 현실에서 빈번하게 발생하는 다양한 유형의 공격을 포함하여 다각도로 진행되었으며, Smark는 대부분의 조건에서 안정적인 워터마크 추출 성능을 보임
- 결과적으로 Smark는 오디오 워터마킹 기술의 핵심 난제인 비가청성과 견고성 간의 상충 관계를 효과적으로 해결하여, 두 지표 모두에서 높은 수준의 균형을 달성한 것으로 최종 평가됨

결론 및 시사점

- Smark와 같은 생성 단계 워터마킹 기술은 이산 웨이블릿 변환을 활용하여 비가청성과 견고성의 균형을 달성했으며, AI 생성 콘텐츠의 법적 분쟁 및 딥페이크 악용 문제에 대한 선제적인 기술적 해결책을 제시함
- AI가 생성한 모든 디지털 음성에 출처와 소유권을 증명하는 원천 기술을 제공함으로써 콘텐츠의 신뢰성을 확보하고, 향후 미디어, 엔터테인먼트 등 다양한 산업에서 창작자의 지식재산권을 보호하는 핵심 인프라로 자리 잡을 잠재력을 가짐

참고문헌

- Joseph Gedeon, "AI scammer posing as Marco Rubio targets officials in growing threat", The Guardian, 2025.07.08., <https://www.theguardian.com/us-news/2025/jul/08/marco-rubio-ai-impostor>
- Yichuan Zhang 외 2인, "Smark: A Watermark for Text-to-Speech Diffusion Models via Discrete Wavelet Transform", arXiv, 2025.12.21., <https://arxiv.org/pdf/2512.18791>