



저작권 이슈 브리프

SUMMARY

산업/기업

기술

산업 일레븐뮤직 사례로 본 AI 음악 생성 서비스의 사전 라이선스 확보 모델

▶ 텍스트 프롬프트만으로 곡을 생성, 청취할 수 있는 AI 음악 생성 서비스가 일반 이용자 시장으로 확산되면서, 모델 학습에 활용되는 음원 데이터의 권리 처리 문제가 업계의 핵심 쟁점으로 부상하였다. 이에 AI 음성 기업 일레븐랩스는 2026년 4월, 음악 권리와 사전 라이선스 계약을 체결해 분쟁 소지를 줄인 AI 음악 생성 앱 '일레븐뮤직'을 출시하였다. 아울러 이용자가 생성한 음원을 플랫폼 내에서 직접 등록하고, 라이선스를 부여하거나 수익화할 수 있는 '뮤직 마켓플레이스'도 함께 운영하며, AI 생성 음원이 플랫폼 안에서 유통과 보상으로 이어질 수 있는 시스템을 만들었다. 다만 현재 확보된 라이선스는 일부 독립 음반사에 한정되어 있으며, AI 생성 음원의 권리 귀속과 수익 분배에 관한 기준도 아직 명확히 정립되지 않아 관련 제도 정비가 과제로 남아 있다.

산업 불법 스트리밍 링크에 대응하는 새로운 접근, 게임화된 신고 플랫폼 '온라인 헌터' 출시

▶ 최근 라이브 스트리밍 콘텐츠를 무단으로 재송출하는 불법 링크가 텔레그램, 디스코드 등 반폐쇄형 커뮤니티를 통해 확산되고 있다. 반폐쇄형 커뮤니티는 회원 가입이나 초대를 거쳐야 접근할 수 있고 검색엔진에도 노출되지 않아, 자동화된 탐지 도구만으로는 내부에서 공유되는 불법 링크를 확인하기 어렵다는 한계가 있다. 이에 대응하기 위해 웨어지오가 불법 스트리밍 링크 신고 플랫폼 '온라인 헌터'를 출시했다. 이 플랫폼은 이용자가 불법 스트리밍 링크나 해당 링크가 공유된 게시물을 신고하면 포인트를 지급하고, 누적 포인트로 넷플릭스 등 OTT 이용권을 구매할 수 있게 했다. 또한 이 플랫폼은 레벨 시스템과 공개 순위표를 통해 경쟁 심리를 자극하고, 지속적인 참여를 유도하는 게임화 메커니즘을 적용했다.

산업 게임 DRM 우회 기술의 등장과 저작권 보호의 비용-효익

▶ 게임 산업에서 디지털 저작권 관리는 출시 초기 불법 복제를 지연시켜 매출을 보호하는 수단으로 활용되어 왔으나, 2025년 운영체제보다 낮은 계층에서 DRM 검사를 우회하는 하이퍼바이저 기반 기술이 등장하면서 출시 당일 주요 게임의 보호가 해제되는 사례가 잇따라 확인되었다. 이에 따라 수개월간의 보호 기간 확보를 전제로 설계된 기존 DRM 도입 논리가 재검토 대상이 될 수 있다는 평가가 제기되고 있으며, 수만 달러에 달하는 유지 비용에 대한 게임사별 판단의 방향도 엇갈리고 있다. 실제로 업계의 대응은 이미 단기 유지 후 제거, 장기 유지, 미도입 등으로 나뉘지고 있으며, DRM이 모든 게임에 일률적으로 적용되는 표준 해법이 아님을 보여준다. 향후 게임 저작권 보호 논의는 불법 복제 지연 효과뿐 아니라 이용자 편의와 보안에 미치는 영향까지 함께 고려하는 방향으로 확장될 가능성이 있는 것으로 분석된다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

산업 스포티파이의 아티스트 프로필 보호 기능 도입

▶ 스포티파이는 2026년 3월, 아티스트 명의 도용 문제가 지속되는 가운데 아티스트 프로필 보호를 베타 기능으로 출시하며, 아티스트 신원 보호를 2026년 최우선 과제로 명시하였다. 현행 스트리밍 배포 구조상 별도의 검증 없이 배포자가 메타데이터에 타인의 아티스트명을 기입하는 것만으로 해당 프로필에 콘텐츠를 연결할 수 있어, 아티스트가 사칭 여부를 인지하기 전까지 플랫폼 차원의 선제적 대응이 이루어지기 어렵다. 아티스트 프로필 보호 기능은 아티스트가 자신의 명의로 배포되는 음원이 프로필에 자동 연결되기 전에 직접 검토하고 승인할 수 있도록 설계된 신원 보호 도구로, 스트리밍 서비스 중 최초로 도입된 사전 검토 방식이다. 이 기능을 활성화한 아티스트는 음원 배포 시 이메일 알림을 수신하고 스포티파이 포 아티스트에서 해당 음원을 승인 또는 거부할 수 있으며, 신뢰하는 배포사와 고유 식별 코드인 아티스트 키를 공유하면 해당 배포사가 올리는 음원은 검토 절차 없이 자동으로 사전승인 처리된다.

산업 라리가와 패스틀리의 불법 스트리밍 대응

▶ 스페인 프로축구 리그 라리가는 연간 7억~8억 달러의 경제적 손실을 초래하는 불법 스트리밍에 대응하기 위해 글로벌 클라우드 컴퓨팅 서비스 기업인 패스틀리의 AI 기반 실시간 탐지 시스템을 도입했다. 패스틀리가 개발한 타겟형 지능 탐지 시스템은 AI와 독자적 콘텐츠 신호를 결합하여 불법 스트리밍을 실시간으로 식별하며, 기존 지역 차단 방식과 달리 정상 트래픽을 보호하면서도 불법 콘텐츠만을 선별 차단한다. 이를 통해 라리가는 2024/25 시즌 스페인 내 불법 스트리밍을 60% 감소시켰다. 양사는 기술 기업·출판사·규제기관과 협력하여 소프트웨어 솔루션 및 모범 사례를 개발 중이며, 이러한 협력 모델은 향후 음악·영화·게임 등 다양한 콘텐츠 분야로 확산될 전망이다.

기술 주간 기술 동향

▶ 대규모 언어모델이 사회 전반에 확산되면서 학습 과정에서 흡수한 개인정보와 민감 데이터에 대한 우려가 커지고 있다. 특히 추론 모델은 최종 답변뿐 아니라 중간 추론 과정에도 민감 정보를 포함할 수 있어, 기존의 답변 수준 언러닝 방법으로는 완전한 망각을 보장할 수 없다. 이에 대응하기 위해 제안된 특징 대체 언러닝 기술인 FRUL은 민감한 추론 경로를 선택적으로 억제하고 무해한 대체 경로를 강화하는 이중 메커니즘을 활용한다. 해당 메커니즘은 검색 증강 생성 기술과 복수의 언어모델을 활용하여 민감 정보를 자동으로 식별하고, 이를 중립적 사실이나 가상 정보로 대체하여 논리적 일관성을 유지한다.



저작권 이슈 브리프

SUMMARY

산업/기업

기술

일레븐뮤직 사례로 본 AI 음악 생성 서비스의 사전 라이선스 확보 모델

AI 음악 생성 서비스의 대중화와 저작권 리스크의 부상

• AI 음악 생성 서비스의 확산과 권리 처리 문제의 본격화

- 텍스트 프롬프트만으로 곡을 생성, 청취할 수 있는 AI 음악 생성 서비스가 일반 이용자 시장으로 확산되면서, 모델 학습에 활용되는 음원 데이터의 권리 처리 문제가 본격적인 쟁점으로 떠오르고 있음
- 대표적으로 수노(Suno)*와 유디오(Udio)** 등 주요 AI 음악 생성 플랫폼은 권리자와의 라이선스 계약 없이 서비스를 출시한 이후, 주요 음반사로부터 저작권 침해 소송을 제기당한 바 있음

* 수노(Suno): 이용자가 텍스트 또는 오디오 프롬프트를 입력하면 보컬과 연주를 포함한 곡을 생성할 수 있는 AI 음악 생성 플랫폼

** 유디오(Udio): 텍스트 프롬프트 기반으로 음악을 생성하고 공유할 수 있는 AI 음악 생성 플랫폼

- 일련의 사태는 AI 음악 생성 서비스가 기존 음원을 학습 데이터로 활용하는 한, 데이터 이용의 적법성과 권리자와의 계약 여부가 서비스 운영의 핵심 조건이 된다는 점을 보여줌

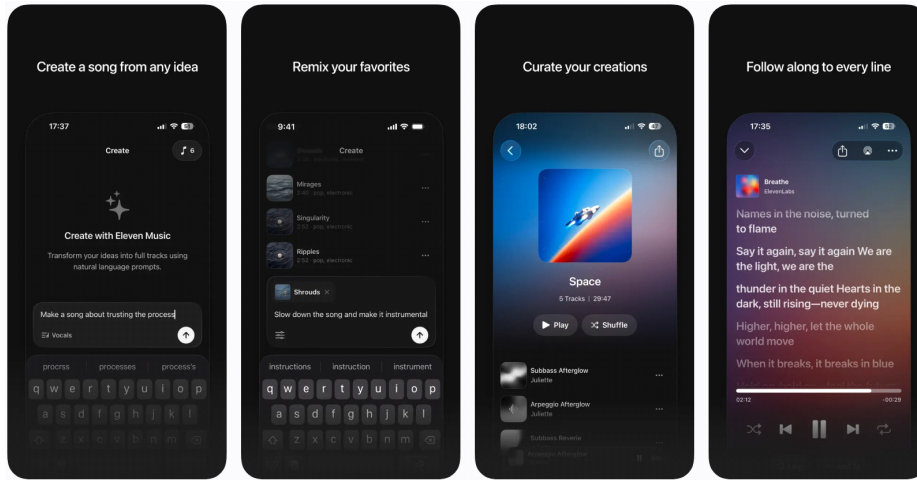
• 일레븐랩스, 사전 라이선스 확보로 분쟁 선제 차단하며 시장 진입

- 이러한 흐름 속에서, 음성 AI 서비스 기업 일레븐랩스(ElevenLabs)*는 2025년 8월 AI 음악 생성 웹 플랫폼을 출시한 데 이어 2026년 4월 iOS 앱 일레븐뮤직(ElevenMusic)을 출시하며 경쟁에 합류함
- 일레븐뮤직은 자연어 프롬프트 기반으로 하루 최대 7곡을 생성할 수 있으며, 앱 내에서 다양한 AI 생성 음원을 탐색하고 청취할 수 있는 기능도 함께 제공함¹⁾
- 일레븐랩스는 서비스를 먼저 출시한 뒤 권리 문제를 조정하는 방식이 아니라, 서비스 공개 시점부터 음악 권리자와의 라이선스 계약을 확보한 상태에서 시장에 진입했다는 점에서 기존과 차이를 보임
- 이는 기존 기업들의 분쟁 경험 이후, 서비스 출시 이전에 음원 이용의 법적 근거를 마련하는 방식이 하나의 대응 전략으로 시도되고 있음을 보여줌

* 일레븐랩스(ElevenLabs): AI 음성 기술을 주력으로 성장해 온 기업으로, 최근 음악 생성 기능까지 서비스 범위를 확장한 AI 서비스 플랫폼

1) Ivan Mehta, "ElevenLabs releases a new AI-powered music-generation app", Tech Crunch, 2026.04.02., <https://techcrunch.com/2026/04/02/elevenlabs-releases-a-new-ai-powered-music-generation-app/>

[그림1] 음성 AI 기업 일레븐랩스가 출시한 AI 음원 생성 앱 '일레븐뮤직'



출처: App Store, "ElevenMusic by ElevenLabs", 2026.04.16. 조회,
<https://apps.apple.com/us/app/elevenmusic-by-elevenlabs/id6755965224>

일레븐랩스의 선제적 라이선스 확보와 수익화 경로 마련

• 서비스 출시 이전에 라이선스 확보하는 차별적 전략

- 일레븐랩스는 서비스 공개 시점부터 음악 권리자와의 라이선스 계약을 확보함
- 구체적으로 일레븐랩스는 2025년 8월 일레븐뮤직 웹 플랫폼을 출시하면서, 독립 음반사의 디지털 라이선스를 대행하는 멀린(Merlin)*과 음악 퍼블리셔 코발트(Kobalt)**와의 계약 체결 사실을 함께 공개함
- 일레븐랩스 CEO 마티 스타니세프스키(Mati Staniszewski)는 음악 산업 파트너와의 협력을 전제로 서비스를 설계했다고 밝혔으며, 저작권 측면에서 상업적 활용이 가능한 모델임을 강조한 바 있음

* 멀린(Merlin): 전 세계 독립 음반사·배급사·권리보유자를 대표해 디지털 음악 서비스와의 라이선스 계약을 대행하는 회원 기반 라이선싱 단체

** 코발트(Kobalt): 작곡가와 음악 출판사를 대상으로 음악 퍼블리싱 관리 서비스를 제공하며, 권리 등록과 로열티 수집·정산을 수행하는 기업

• AI 생성 음원의 유통·수익화 경로 마련을 통한 보상 모델 확장

- 일레븐랩스는 AI 음원을 생성하고 청취하는 플랫폼에 그치지 않고, 2026년 3월 뮤직 마켓플레이스(Music Marketplace)*를 출시해 이용자가 일레븐뮤직에서 생성한 음원을 등록하고, 라이선스를 부여하거나 수익화할 수 있는 경로를 마련함
- 일레븐랩스는 기존 음성 AI 서비스에서 해당 서비스를 활용해 콘텐츠를 만든 창작자들에게 누적 1,100만 달러(원화 약 168억 원²⁾) 이상을 지급해 왔으며, AI 음악 생성 서비스에서도 같은 방식으로 창작자 보상 체계를 운영할 계획이라고 밝힘³⁾
- 이는 AI 생성 음원이 단순한 결과물에 머무는 것이 아니라, 플랫폼 안에서 유통과 보상으로 이어질 수 있도록 한 시도로 볼 수 있음

* 뮤직 마켓플레이스(Music Marketplace): 일레븐랩스가 운영하는 AI 생성 음원의 등록·라이선스·수익화 플랫폼

2) 1달러=1,530.50원(2026.04.01, KEB 하나은행 매매기준율 적용, 이하 동일)

3) Music Business Worldwide, "ElevenLabs launches ElevenMusic iOS app, taking on Suno and Udio on mobile", 2026.04.07., <https://www.musicbusinessworldwide.com/elevenlabs-launches-elevenmusic-ios-app-taking-on-suno-and-udio-on-mobile/>

사전 라이선스 모델의 확산 전망과 AI 생성 음원의 제도 정비 필요성

• 사전 라이선스 모델의 산업적 확산 가능성과 한계

- 일레븐랩스의 사례는 AI 음악 생성 서비스가 서비스 설계 단계에서부터 라이선스 확보와 수익 분배 체계를 함께 갖출 수 있음을 보여주는 초기 사례에 해당함
- 다만 현재 확보된 라이선스는 일부 독립 음반사와 퍼블리셔에 한정되어 있어, 주요 음반사를 포함한 음악 산업 전반의 권리 구조를 포괄하는 수준에는 이르지 못한 상황임
- 아울러 AI 모델 학습에 활용된 데이터의 범위와 라이선스 대상 사이의 대응 관계가 명확히 공개되지 않아, 학습 데이터 전체의 적법성 여부는 외부에서 확인하기 어려운 상태임
- 이에 따라 사전 라이선스 모델이 산업 전반의 표준으로 자리 잡기 위해서는, 라이선스 대상의 범위를 넓히고 학습 데이터 출처의 투명성을 함께 높여야 할 것으로 보임

• AI 생성 음원의 저작권 귀속 및 수익 분배 기준의 정립 필요성

- 일레븐랩스가 구축한 뮤직 마켓플레이스는 AI 생성 음원을 수익화할 수 있는 경로를 제공하고 있으나, 아직 AI 생성 음원으로 창출된 수익이 누구에게 얼마나 귀속되는지에 대한 기준은 명확히 정립되지 않은 상태임
- 이러한 불확실성은 플랫폼 이용자, 원곡 권리자, AI 서비스 사업자 간 수익 배분의 기준과 근거를 불분명하게 만들 수 있으며, 서비스 규모가 확대될수록 관련 분쟁 가능성도 함께 커질 수 있음
- 향후 AI 음악 생성 시장이 성장함에 따라, AI 생성 음원의 권리 귀속 기준과 관련자 간 수익 분배 기준을 명확히 하는 제도적 논의가 필요할 것으로 판단됨

참고문헌

- Ivan Mehta, "ElevenLabs releases a new AI-powered music-generation app", Tech Crunch, 2026.04.02., <https://techcrunch.com/2026/04/02/elevenlabs-releases-a-new-ai-powered-music-generation-app/>
- Music Business Worldwide, "ElevenLabs launches ElevenMusic iOS app, taking on Suno and Udio on mobile", 2026.04.07., <https://www.musicbusinessworldwide.com/elevenlabs-launches-eleven-music-ios-app-taking-on-suno-and-udio-on-mobile/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

불법 스트리밍 링크에 대응하는 새로운 접근, 게임화된 신고 플랫폼 ‘온라인 헌터’ 출시

웨어지오의 ‘온라인 헌터’ 출시 배경

- 반폐쇄형 커뮤니티에서의 불법 스트리밍 링크 확산
 - 최근 라이브 스트리밍 콘텐츠를 무단으로 재송출하는 불법 링크가 반폐쇄형 커뮤니티(Semi-private community)**를 통해 유통되고 있음
 - 반폐쇄형 커뮤니티는 회원 가입이나 초대를 거쳐야 접근할 수 있으며, 검색엔진에도 잘 노출되지 않음
 - 이 때문에 자동화된 탐지 도구만으로는 반폐쇄형 커뮤니티 내부에서 공유되는 불법 스트리밍 링크나 관련 게시물을 확인하기 어렵다는 한계가 지적됨
 - 이에 대응하여 체코의 디지털 솔루션 기업인 웨어지오(Warezio)*가 불법 스트리밍 링크 신고 플랫폼인 온라인 헌터(Online Hunter)를 출시함
 - 이 플랫폼은 소셜미디어나 커뮤니티를 이용하던 사람들이 발견한 불법 링크를 신고하면 보상을 제공하는 구조로, 자동화 도구가 도달하기 어려운 영역까지 탐지 범위를 넓히려는 시도임

* 웨어지오(Warezio): 온라인상 불법 유통 콘텐츠의 탐지·삭제와 권리 보호를 지원하는 체코의 디지털 솔루션 기업

** 반폐쇄형 커뮤니티(Semi-private community): 공개 웹사이트와 달리 초대 또는 승인을 통해서만 접근 가능한 소규모 그룹으로, 외부에서 내부 활동을 확인하거나 검색엔진으로 색인하기 어려운 구조를 가진 온라인 공간

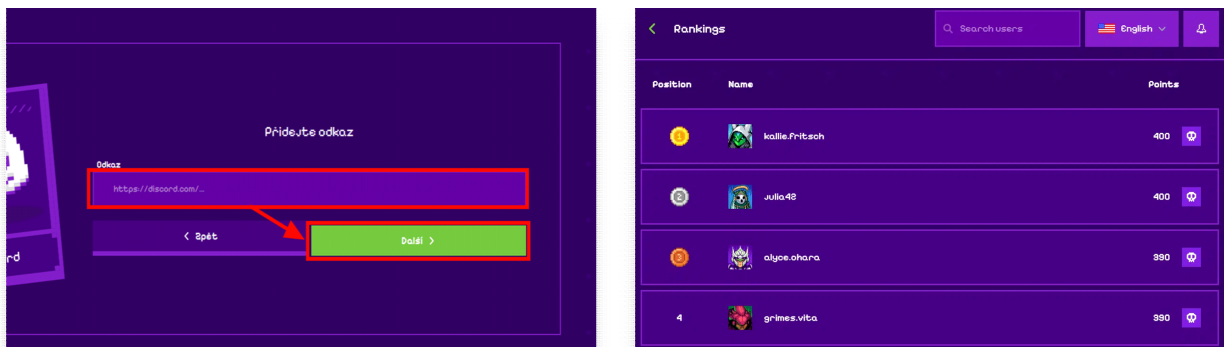
온라인 헌터의 작동 구조와 지속성 과제

- 게임화 메커니즘을 적용하여 신고자 참여를 유도
 - 온라인 헌터에 신고된 링크는 다른 이용자가 이미 신고한 이력이 있는지에 따라 차등 평가되며, 최초로 접수된 링크일수록 이용자에게 더 높은 포인트가 부여됨
 - 이용자는 누적된 포인트를 활용해 넷플릭스, HBO 등 주요 온라인 스트리밍 서비스의 이용권을 구매할 수 있음
 - 온라인 헌터는 포인트 보상 체계 외에도, 이용자 참여를 유도하기 위한 레벨 시스템과 공개 순위표를 운영함
 - 더 많은 링크를 신고할수록 포인트가 누적되고, 이에 따라 이용자는 더 높은 레벨에 도달하게 됨
 - 공개된 순위표는 상위권 진입을 둘러싼 경쟁 심리를 자극하여 이용자의 반복 참여를 유도하는 게임 장치로 작동함

• 지속적인 이용자 참여에 대한 과제

- 이러한 게임화 메커니즘을 통해 이용자의 신고를 유인할 수 있음에도 불구하고, 일각에서는 플랫폼을 지속적으로 사용할 핵심 참여층이 충분히 형성될지 의문이 제기됨
- 불법 스트리밍 링크는 주로 해당 콘텐츠를 소비하거나 관련 커뮤니티를 지속적으로 이용하는 층이 더 자주 접할 가능성이 높음
- 때문에, 원래 이러한 콘텐츠를 찾지 않는 이용자는 신고 대상 자체를 접할 기회가 제한될 수 있음
- 이에 따라 향후 참여 저변을 넓히기 위해 유인 설계와 함께 적극적인 홍보가 병행될 필요가 제기됨

[그림1] 온라인 헌터의 신고 링크 입력창(좌), 공개 순위표(우)



출처: Gigazine, "A Czech anti-piracy company has launched 'Online Hunter,' a service that allows users to earn points by reporting pirated websites.", 2026.03.18., https://gigazine.net/gsc_news/en/20260318-online-hunter/

[표1] 온라인 헌터(Online Hunter) 작동 구조 요약

| 구분 | 작동 방식 |
|-----------------|--|
| 신고 방식 | • 이용자가 발견한 불법 스트리밍 링크 또는 해당 링크가 포함된 게시물의 URL을 신고함 |
| 포인트 차등 배분 | • 타 이용자의 중복 신고 여부에 따라 차등 평가되며, 새롭게 확인된 링크일수록 높은 포인트가 부여됨 |
| 보상 체계 | • 누적 포인트로 넷플릭스(Netflix), HBO 등 온라인 스트리밍 서비스 이용권을 구매할 수 있음 |
| 레벨 시스템 & 공개 순위표 | • 더 많은 링크를 신고해 승인받을수록 레벨이 상승하고, 이용자별 순위를 공개하여 경쟁을 유도하는 구조임 |

출처: 참고문헌 종합하여 재구성

참고문헌

- Ernesto Van der Sar, "Rightsholders Crowdsource Piracy Link Reporting With 'Online Hunter' Game", TorrentFreak, 2026.03.15., <https://torrentfreak.com/rightsholders-crowdsource-piracy-link-reporting-with-online-hunter-game/>
- Gigazine, "A Czech anti-piracy company has launched 'Online Hunter,' a service that allows users to earn points by reporting pirated websites.", 2026.03.18., https://gigazine.net/gsc_news/en/20260318-online-hunter/



저작권 이슈 브리프

SUMMARY

산업/기업

기술

게임 DRM 우회 기술의 등장과 저작권 보호의 비용-효익

게임 DRM의 역할과 하이퍼바이저 기반 우회 기술

• 기존 게임 DRM의 역할과 한계

- 게임 산업에서 디지털 저작권 관리(Digital Rights Management, DRM)*는 실행 파일의 복제·변조를 기술적으로 어렵게 만들어 불법 복제를 지연시키는 방식으로 활용되어 왔으며, 데누보(Denuvo)**와 같은 기술이 지난 10여 년간 대표적인 보호 수단으로 자리해 왔음
- 2025년 발표된 연구에 따르면¹⁾, 게임 출시 초기의 불법 복제로 인한 매출 손실은 최대 20%에 달하는 것으로 분석되었으며, 이는 퍼블리셔(Publisher)***들이 이용자 불편 우려에도 DRM 도입 비용을 감수해 온 주요 배경 중 하나로 해석됨
- 그러나 DRM은 성능 저하 논란, 오프라인 이용 제약, 모딩(Modding)**** 제한 등으로 이용자 커뮤니티의 비판을 지속적으로 받아 왔으며, 저작권 보호 효과와 이용자 부담 간의 긴장이 점차 심화되어 왔음

* DRM(Digital Rights Management): 디지털 콘텐츠의 무단 복제·배포를 방지하기 위해 암호화 등의 방법으로 접근을 통제하는 기술 및 관리 체계

** 데누보(Denuvo): 이르데토(Irdeto)가 보유한 게임용 안티-탐퍼(anti-tamper) DRM 솔루션

*** 퍼블리셔(Publisher): 게임의 유통·마케팅·판매를 담당하며, 개발사와 계약을 통해 출시 전략 및 수익화 구조를 관리하는 사업 주체

**** 모딩(Modding): 이용자가 게임의 그래픽, 캐릭터, 시스템, 콘텐츠 등을 수정하거나 확장하여 새로운 경험을 만드는 사용자 제작 활동

• 우회 기술의 등장과 하이퍼바이저 기반 작동 방식

- 최근 주요 대작 게임들이 출시 당일 크랙(crack)*되는 사례가 잇따르며, 하이퍼바이저(Hypervisor)** 기반 우회 기술이 새로운 보안 위협으로 부상함
- 2025년 초 '레지던트 이블 레퀴엠(Resident Evil Requiem)'이 데누보를 적용했음에도 불구하고 출시 수 시간 만에 크랙된 이후, 유사한 방식의 우회 사례가 지속적으로 확인되고 있음
- 기존 방식이 게임 실행 파일을 분석·수정하는 데 수개월이 소요되었던 것과 달리, 하이퍼바이저 방식은 게임 파일을 직접 변경하지 않고 운영체제보다 낮은 계층에서 DRM 검사가 정상 작동하는 것처럼 인식하도록 만들어 단시간 내 우회가 가능한 구조임
- 다만 하드웨어 환경에 따른 안정성 차이와 주요 보안 기능 비활성화 요구 등 실사용 제약이 존재해 광범위한 확산으로 이어지지 않았으나, 데누보 외 다른 DRM 소프트웨어에도 적용 가능성이 있어 향후 보안 생태계에 미칠 파급 범위를 지켜볼 필요가 있음

* 크랙(crack): 소프트웨어에 적용된 디지털 저작권 관리(DRM) 또는 복제 방지 기술을 우회하거나 제거하여, 정품 인증 없이 실행이 가능하도록 만드는 행위 또는 그 결과물

** 하이퍼바이저(hypervisor): 물리적 하드웨어 위에서 여러 운영체제를 실행할 수 있도록 하는 가상화 소프트웨어 계층. 운영체제보다 낮은 수준에서 작동해 시스템 자원에 직접 접근할 수 있음

1) Aaron Down, "Why are PC developers removing Denuvo from their games?", PCGamesN, 2026.03., <https://www.pcgamesn.com/drm/why-are-pc-developers-removing-denuvo>

DRM 비용-효익 구조의 변화와 게임사 대응의 분화

• 보호 효과의 약화와 DRM 유지 논리의 변화

- 게임 DRM은 출시 초기 판매 기간 동안 불법 복제를 지연시켜 매출 손실을 최소화하는 것을 핵심 목적으로 활용되어 왔으며, 일정 기간 이후에는 불법 복제가 전체 매출에 미치는 영향이 제한적이라는 분석도 제기되어 왔음
- 그러나 하이퍼바이저 방식으로 출시 직후 우회가 가능해진 점을 고려하면, 수주에서 수개월의 보호 기간 확보를 전제로 해온 기존 DRM 적용 논리의 재검토가 필요하다는 평가가 제기되고 있음
- 보도에 따르면 테누보는 월 수만 달러 수준의 비용과 추가 수수료가 발생하는 것으로 알려져 있으며, 초기 보호 효과의 불확실성이 커지면서 비용을 어느 수준까지 감수할 것인지에 대한 게임사별 판단이 엇갈리고 있음

• 게임사별 DRM 전략의 분화

- 일부 퍼블리셔는 출시 후 일정 기간이 지나면 DRM을 제거하는 방식을 선택하고 있으며, 이는 초기 판매 구간 이후에는 유지 비용과 이용자 반발을 줄이는 것이 더 합리적이라는 판단에 따른 것으로 해석됨
- 반면 대규모 지식재산권(IP)을 보유한 일부 퍼블리셔는 판매 지속성, 계약 조건, IP 보호 방침 등을 고려해 DRM을 장기간 유지하기도 하며, 보호 기간이 충분히 확보된다는 전제하에 비용 부담을 감수하는 전략을 취하고 있음
- 한편 온라인 멀티플레이 중심의 게임은 불법 복제 시 온라인 기능 이용이 제한되는 구조로 인해, 별도 DRM 없이도 정품 구매 유인이 형성되어 이를 도입하지 않는 사례도 존재함
- GOG* 등 DRM을 적용하지 않는 유통 플랫폼이 일부 퍼블리셔의 지지를 받으며, 기술적 보호 대신 이용자 신뢰를 기반으로 한 접근이 대안으로 선택될 수 있음을 시사함

* GOG(Good Old Games): CD 프로젝트(CD Projekt)가 운영하는 PC 게임 유통 플랫폼으로, DRM을 적용하지 않은 게임만을 취급하는 정책으로 알려져 있음. GOG CEO는 DRM이 정당 이용자의 이용 편의를 저해할 수 있다는 입장을 공개적으로 표명한 바 있음

[표 1] 주요 게임사의 DRM 대응 방식 비교

| 게임사 | DRM 대응 방식 | 주요 사례 | 세부 내용 |
|----------------------------------|------------|---|----------------------------------|
| 스퀘어에닉스 (Square Enix) | 단기 유지 후 제거 | 파이널 판타지 16, 비전스 오브 마나, 드래곤 퀘스트 1&2 리메이크 등 | 대부분 출시 6개월 이내 제거 |
| 캡콤 (Capcom) | 장기 유지 | 레지던트 이블 빌리지 | 출시 후 2년간 유지, 제거 시점까지 약 800만 장 판매 |
| 세이버 인터랙티브 (Saber Interactive) | 미도입 | 스페이스 마린 2 | 멀티플레이 구조상 DRM 실익이 제한적이라고 판단 |
| id 소프트웨어 (id Software) | 도입 후 조기 제거 | 둠 이터널 | 출시 2개월 후 도입, 이용자 반발로 조기 제거 |

출처: 참고문헌 종합하여 재구성

• 기술 공방의 지속과 향후 전망

- 테누보 기술을 보유한 이르데토(Irdeto)는 하이퍼바이저 우회에 대응하기 위한 보안 업데이트를 준비 중이라고 밝혔으나, 구체적인 방식과 효과는 아직 확인되지 않은 상태임
- 다만 향후 대응이 보호 기술 강화 중심으로 전개될지, 또는 DRM 적용 범위와 유지 기간을 조정하는 방향으로 이어질지는 단정하기 어려우며, 업계의 대응과 시장 반응을 지켜볼 필요가 있음

게임 DRM 전략의 다층화와 이용자 부담

• DRM 적용 방식의 변화와 이용자 부담

- 하이퍼바이저 기반 우회 기술의 등장으로 DRM이 제공해 온 초기 보호 효과의 불확실성이 확대되면서, DRM 도입 여부 자체보다 적용 대상과 유지 기간에 대한 판단으로 논의의 초점이 이동하고 있음
- 실제로 게임사들의 대응은 장르별 초기 판매 집중도, 멀티플레이 비중, 장기 서비스 여부, 이용자 커뮤니티 반응 등에 따라 다층적으로 나뉘고 있으며, DRM이 일률적인 보호 수단으로 기능하기 어려운 환경으로 변화하고 있음
- DRM은 불법 복제를 억제하기 위한 수단이지만 이용자 부담 문제도 지속적으로 제기되어 온 만큼, 향후 게임 산업의 저작권 보호 논의는 복제 지연 효과뿐 아니라 이용자 편의와 보안 간 균형을 고려하는 방향으로 확장될 가능성이 있음

참고문헌

- Ernesto Van der Sar, "Game Pirates Beat Denuvo with Hypervisor Bypasses — Irdeto Promises Countermeasure", TorrentFreak, 2026.03.31., <https://torrentfreak.com/game-pirates-beat-denuvo-with-hypervisor-bypasses-irdeto-promises-countermeasure/>
- Aaron Down, "Why are PC developers removing Denuvo from their games?", PCGamesN, 2026.03., <https://www.pcgamesn.com/drm/why-are-pc-developers-removing-denuvo>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

스포티파이의 아티스트 프로필 보호 기능 도입

음악 스트리밍 플랫폼 내 아티스트 프로필 오등록 현황과 스포티파이의 대응

- 허술한 음원 등록 절차로 인해 AI 생성 음원이 아티스트 프로필에 오등록되는 현상 발생
 - 재즈 피아니스트 제이슨 모란(Jason Moran)을 비롯한 다수의 아티스트는, 자신의 스포티파이 프로필에 AI로 생성된 것으로 추정되는 음원이 자기 명의로 등록되어 있는 사실을 확인함
 - 이러한 사칭 피해가 확산되는 가운데, 스포티파이는 2025년 9월 기준 직전 12개월간 7,500만 건 이상의 스팸성 트랙을 삭제했다고 밝힘¹⁾
 - 스트리밍 플랫폼의 배포 구조에서는, 별도의 검증 절차 없이 배포자가 메타데이터(metadata)*에 타인의 아티스트명을 기입하는 것만으로도 해당 프로필에 콘텐츠가 연결될 수 있음
 - 이로 인해 아티스트가 사칭 여부를 인지하기 전까지는, 플랫폼이 직접 음원 등록을 차단하기가 어려움
- * 메타데이터: 데이터를 설명하는 정보로, 음원의 경우 아티스트명·앨범명·장르 등이 포함되며 스트리밍 플랫폼은 이를 기준으로 콘텐츠를 특정 아티스트 프로필에 자동으로 연결함

• 스포티파이의 대응 방향: 사전 검토 기능 도입으로 사전·사후 대응 체계 구축

- 스포티파이는 2026년 3월 15일 '아티스트 프로필 보호(Artist Profile Protection)*' 기능을 베타 버전으로 출시하고, 아티스트 신원 보호를 2026년 최우선 과제로 제시함²⁾
 - 해당 기능은 메타데이터 오류, 동명 아티스트 오분류, 의도적 명의 도용 등으로 음원이 다른 아티스트 프로필에 잘못 등록되는 문제에 대응하기 위해 설계됨. 스포티파이는 이 기능을, 스트리밍 서비스 중 최초로 음원을 사전에 검토할 수 있도록 한 신원 보호 도구라고 소개함
 - 스포티파이는 기존에도 잘못 등록된 음원을 신고하고 삭제할 수 있는 도구를 운용해 옴. 이번 기능이 더해지면서 사전 검토 단계가 추가되어, 음원이 프로필에 등록되기 전과 후에 모두 대응할 수 있게 됨
- * 스포티파이가 2026년 3월 베타 버전으로 출시한 기능으로, 아티스트가 자신의 명의로 배포되는 음원이 프로필에 자동 연결되기 전에 직접 검토하고 승인할 수 있도록 설계된 신원 보호 도구

스포티파이 아티스트 프로필 보호 기능의 작동 구조

- 사전 승인 구조를 통해 아티스트에게 통제권을 부여
 - 스포티파이의 '아티스트 프로필 보호' 기능을 활성화한 아티스트는, 자신의 명의로 음원이 배포될 때 이메일로 알림을 받음
 - 또한 '스포티파이 포 아티스트(Spotify for Artists)*'의 설정 화면에서 해당 음원을 검토한 뒤, 승인하거나 거부할 수 있음

1) Dara Kerr, "It has your name on it, but I don't think it's you": how AI is impersonating musicians on Spotify", The Guardian, 2026.04.11., <https://www.theguardian.com/technology/2026/apr/11/ai-impersonating-musicians-spotify>

2) Spotify, "Introducing Artist Profile Protection: Giving Artists More Control Over What Shows Up Under Their Name", 2026.03.15., <https://artists.spotify.com/blog/introducing-artist-profile-protection>

- 승인된 음원은 발매일에 아티스트 프로필에 노출됨. 반면 거부하거나 아무런 조치를 취하지 않으면, 해당 음원은 프로필에 노출되지 않음

* 스포티파이 포 아티스트(Spotify for Artists): 아티스트가 자신의 스포티파이 프로필과 음원 통계를 관리할 수 있는 스포티파이의 공식 아티스트 관리 도구

• 아티스트 키를 통해 신뢰하는 배포사가 등록하는 음원은 자동 승인

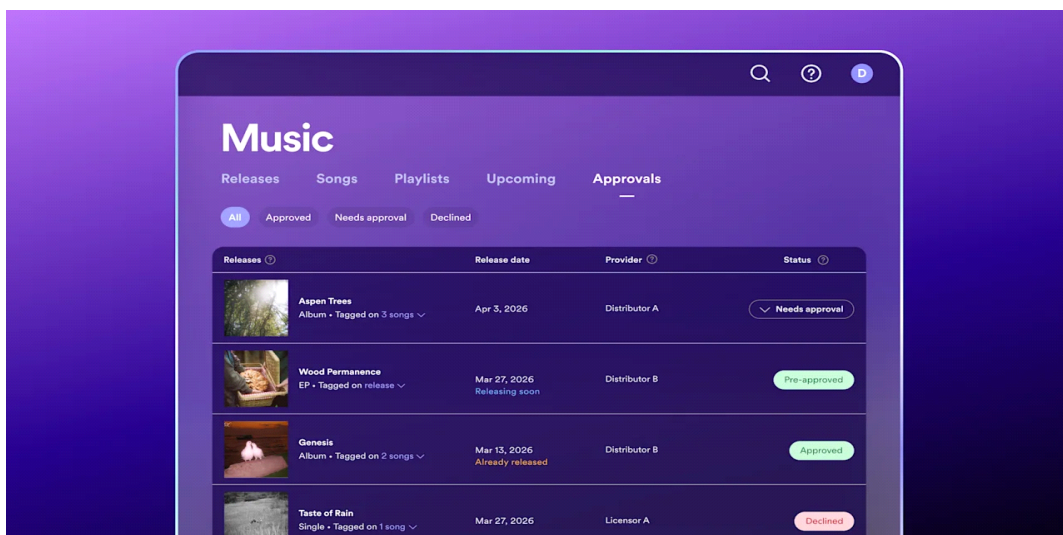
- 아티스트 프로필 보호 기능을 활성화할 경우, 스포티파이는 아티스트에게 고유 식별 코드인 아티스트 키(artist key)를 발급함
- 아티스트가 신뢰하는 배포사와 아티스트 키를 공유하면 해당 배포사가 올리는 음원은 검토 절차 없이 자동으로 사전 승인됨

• 아티스트 프로필 보호 기능, 옵트인 방식으로 운용되어 아티스트의 지속적인 관리 필요

- 해당 기능은 옵트인(opt-in)* 방식으로 운용되며, 음원이 반복적인 오등록을 겪었거나 동명 아티스트와 혼동될 가능성이 있는 경우 특히 활용도가 높음
- 음원이 배포될 때 아티스트가 검토하지 않으면, 본인의 정상적인 음원까지 지연되거나 차단될 수 있음. 따라서 이 기능을 활성화한 뒤에는 지속적인 관리가 필요함
- 현재는 베타 단계로, 아직 정식 출시된 것은 아님. 스포티파이는 참여 아티스트의 피드백을 수집한 뒤 정식 출시 여부와 시점을 결정할 계획임

* 옵트인(opt-in): 사용자가 직접 참여 의사를 표명해야만 기능이 적용되는 방식으로, 별도 신청 없이 자동 적용되는 옵트아웃(opt-out) 방식과 구분됨

[스포티파이의 아티스트 프로필 보호 기능 예시]



출처: Spotify, "Introducing Artist Profile Protection: Giving Artists More Control Over What Shows Up Under Their Name", 2026.03.15., <https://artists.spotify.com/blog/introducing-artist-profile-protection>

참고문헌

- Dara Kerr, "It has your name on it, but I don't think it's you": how AI is impersonating musicians on Spotify", The Guardian, 2026.04.11., <https://www.theguardian.com/technology/2026/apr/11/ai-impersonating-musicians-spotify>
- Spotify, "Introducing Artist Profile Protection: Giving Artists More Control Over What Shows Up Under Their Name", 2026.03.15., <https://artists.spotify.com/blog/introducing-artist-profile-protection>

저작권 이슈 브리프

SUMMARY

산업/기업

기술

라리가와 패스틀리의 불법 스트리밍 대응

라리가-패스틀리 협력 배경과 불법 스트리밍 피해 현황

• 스포츠 라이브 콘텐츠 불법 스트리밍의 심각성

- 스페인 프로축구 리그 라리가(LALIGA)는 라리가 경기의 불법 스트리밍으로 인해 연간 7억~8억 달러 (원화 약 1조 70억~1조 2,240억 원¹⁾)의 경제적 손실이 발생하고 있다고 추산하며, 이는 소속 클럽들의 재정 건전성과 리그 전체 경쟁력에 직접적인 타격을 주고 있다고 발표함
- 글로벌 회계·컨설팅 법인 그랜트 쏘튼(Grant Thornton)의 2025년 연구에 따르면, 2024년 기준 최소 1,080만 건의 불법 스트리밍이 탐지되었으나 이 중 81% 이상은 차단되지 않았음²⁾
- 불법 스트리밍 시작 후 30분 이내에 대응된 경우는 2.7%에 불과하였으며, 실시간 스포츠 중계는 즉시성이 핵심인 만큼 신속한 탐지와 차단이 이루어지지 않을 경우 피해 규모가 크게 확대될 수 있음

• 기존 지역 차단 방식의 한계와 새로운 접근법 필요성

- 기존 불법 스트리밍 대응은 특정 지역이나 IP 대역을 일괄 차단하는 방식에 의존했으나, 이는 정상 이용자의 접근까지 제한하여 팬 경험을 저해하고 합법적 시청 수요마저 위축시키는 부작용을 초래함
- 최근 수많은 무단 스트리밍 사이트가 활성화되는 상황에서, 라리가는 콘텐츠 불법 복제에 효과적으로 대응하기 위해 글로벌 클라우드 컴퓨팅 서비스 기업인 패스틀리(Fastly)와의 협력을 시작함
- 패스틀리의 CPO 켈리 쇼트리지(Kelly Shortridge)는 지역 차단 대신 불법 콘텐츠만을 선별 탐지해 무력화하는 정밀 접근 전략을 통해 팬 경험과 콘텐츠 보호를 동시에 달성할 수 있다고 강조함
- 이러한 협력을 바탕으로 법적·교육적·제도적·기술적 조치를 통합한 포괄적 전략이 실행되었으며, 2024/25 시즌 스페인 내 불법 스트리밍을 60% 감소시키는 성과를 달성함

AI 및 콘텐츠 신호 기반 실시간 탐지 시스템의 구조와 특성

• AI와 독자적 콘텐츠 신호를 결합한 실시간 불법 스트림 식별 메커니즘

- 패스틀리가 개발한 타겟형 지능 탐지 시스템은 AI와 독자적 콘텐츠 신호를 결합하여 불법 스트림을 실시간으로 식별하는 방식으로 설계됨
- 이 시스템은 불법 스트리밍이 송출되는 시간을 최소화하는 데 초점을 맞추며, 탐지 즉시 차단 조치를 취함으로써 불법 방송의 확산을 원천 차단함
- 기존 지역 차단 방식과 달리 정상 트래픽은 온전히 보호하면서도 불법 콘텐츠만을 선별적으로 탐지 및 차단함으로써, 이용자 경험 저해 없이 권리 보호 효과를 극대화하는 균형 잡힌 접근을 구현함

1) 1달러=1,530.50원(2026.04.01, KEB 하나은행 매매기준율 적용, 이하 동일)

2) Pareesa Afreen, "Football league loses \$800m to illegal streaming, uses AI to stop it", The News International, 2026.04.15., <https://www.thenews.com.pk/latest/1399004-football-league-loses-800m-to-illegal-streaming-uses-ai-to-stop-it>

• 플랫폼 차원에서 즉각 대응 가능한 구조

- 패스틀리의 솔루션은 법원 명령이나 중개 사업자에 대한 법적 조치를 거치지 않고도, 플랫폼 운영자가 확인된 불법 콘텐츠를 직접 차단할 수 있도록 설계되어 신속한 대응이 가능하다는 장점이 있음
- 이는 기존 방식처럼 법적 절차를 통해 중개자를 압박하는 대신, 기술적 수단을 통해 플랫폼과 관리자가 직접 보호 조치를 취할 수 있도록 지원하는 접근법임
- 이러한 구조는 법적 절차의 지연이나 관할권 문제를 우회하여 불법 스트리밍에 즉각 대응할 수 있다는 점에서, 실시간성이 중요한 라이브 콘텐츠 보호에 특히 효과적인 것으로 평가됨

• 산업 생태계 협력 기반 소프트웨어 솔루션 및 모범 사례 개발

- 패스틀리와 라리가는 기술 기업·출판사·규제기관과 협력하여 무단 스트리밍을 신속하게 탐지하고 무력화하는 소프트웨어 솔루션 및 모범 사례를 공동 개발하고 있음
- 해당 프로젝트는 미디어 및 출판 기업들이 확인된 불법 콘텐츠와 세그먼트를 신고할 수 있는 체계를 구축하며, 이를 통해 플랫폼 차원에서 불법 스트리밍에 대응할 수 있는 기술적 대응 역량을 확보하도록 지원함
- 이러한 협력 모델은 단순히 라리가의 경기 중계에 국한되지 않고, 다양한 콘텐츠 관리자와 플랫폼이 참여하는 범산업적 대응 체계로 확장될 가능성을 시사함

시사점: 콘텐츠 보호 생태계 구축을 위한 협력 과제

• 정밀 탐지 방식의 산업적 의의

- 라리가 회장 하비에르 테바스(Javier Tebas)는 불법 스트리밍 감소 60% 성과는 법적·교육적·기술적 조치의 통합된 결과로 평가함³⁾
- 불법 스트리밍은 직접적인 구독 수익 감소뿐 아니라 방송권 자체의 시장 가치를 하락시켜 리그와 클럽의 장기 수익 구조에 구조적 타격을 가하는 만큼, 기술적 대응은 단기 차단을 넘어 권리 가치 보호 차원에서 접근해야 함
- 패스틀리와 라리가가 다양한 기술 기업·기관과 협력하여 소프트웨어 솔루션 및 모범 사례를 개발하고 있는 점은, 불법 콘텐츠 차단을 넘어 디지털 유통 생태계 전반의 신뢰성을 강화하려는 시도로 평가됨

참고문헌

- Julian Clover, "Fastly and LaLiga partner on AI anti-piracy solution", BroadbandTVNews, 2026.04.10., <https://www.broadbandtvnews.com/2026/04/10/fastly-and-laliga-partner-on-ai-anti-piracy-solution/>
- Fastly, "Fastly and LALIGA Team Up on Joint Innovation to Combat Piracy", 2026.04.09., <https://www.fastly.com/press/press-releases/fastly-and-laliga-team-up-on-joint-innovation-to-combat-piracy>
- Pareesa Afreen, "Football league loses \$800m to illegal streaming, uses AI to stop it", The News International, 2026.04.15., <https://www.thenews.com.pk/latest/1399004-football-league-loses-800m-to-illegal-streaming-uses-ai-to-stop-it>
- Sean Mitchell, "Fastly & LALIGA launch anti-piracy push on streams", SecurityBrief, 2026.04.14., <https://securitybrief.com.au/story/fastly-laliga-launch-anti-piracy-push-on-streams>

³⁾ Julian Clover, "Fastly and LaLiga partner on AI anti-piracy solution", BroadbandTVNews, 2026.04.10., <https://www.broadbandtvnews.com/2026/04/10/fastly-and-laliga-partner-on-ai-anti-piracy-solution/>



저작권 이슈 브리프

SUMMARY

산업/기업

기술

주간 기술 동향

대규모 추론 모델의 민감 정보 선택적 제거 기술 FRUL

• 대규모 언어모델의 학습 데이터 삭제와 개인정보 보호를 위한 언러닝 기술

대규모 언어모델이 고객 지원, 검색 엔진, 생산성 도구, 의료 진단 보조, 기업 내부 시스템 등 사회 전반에 깊숙이 통합되면서, 이들 모델이 학습 과정에서 무분별하게 흡수한 개인정보와 민감한 데이터에 대한 우려가 커지고 있다. 특히 웹 데이터와 사용자 생성 콘텐츠를 대규모로 학습한 모델들은 저작권으로 보호받는 자료, 개인 정보, 심지어 불법적으로 수집된 데이터까지 내부 추론 과정에 포함시킬 수 있다. 이러한 상황에서 개인이 자신의 데이터가 AI에 의해 무단으로 활용되는 것을 막을 수 있는 권리는 기본적인 인권 보호의 문제로 대두되고 있다.

이에 대응하기 위해 유럽연합의 일반 데이터 보호 규정을 비롯한 각국의 개인정보 보호 법규는 개인에게 잊힐 권리를 부여하고 있다. 이는 자신의 데이터가 모델에 의해 사용되는 것을 철회할 수 있는 권리를 의미한다. 그러나 LLM을 처음부터 다시 학습시켜 특정 데이터를 제거하는 것은 막대한 계산 비용이 들 뿐만 아니라, 원본 학습 데이터에 접근할 수 없는 경우가 많아 사실상 불가능하다. 이러한 배경에서 모델을 전면 재학습하지 않고도 특정 데이터의 영향력을 효율적으로 제거하는 선택적 망각 기술, 즉 언러닝 기술이 새로운 패러다임으로 등장했다.

기존의 언러닝 방법들은 주로 모델의 최종 출력 결과에서 민감 정보를 제거하는 데 집중해왔다. 그러나 최근 대규모 추론 모델들은 단순히 답변만 생성하는 것이 아니라, 복잡한 다단계 추론 과정을 거쳐 결론에 도달한다. 이 과정에서 중간 추론 단계에 민감한 정보가 여전히 남아있을 경우, 최종 답변만 수정하는 방식으로는 완전한 망각을 달성할 수 없다. 더욱이 망각 대상에 관한 질문에 대해 무의미한 출력을 생성하도록 유도하는 기존 방식은 모델의 전반적인 추론 능력을 저하시키는 부작용을 초래할 위험이 있다.

이러한 배경에서 추론 과정 전체를 고려한 새로운 언러닝 기술이 주목받고 있다. 특히 망각-보존 언러닝 기술인 FRUL은 민감한 추론 경로를 선택적으로 억제하는 동시에, 무해한 대체 추론 경로를 강화하여 모델이 망각된 지식을 우회하도록 학습시키는 접근법이다. 이 기술은 단순히 특정 정보를 삭제하는 것을 넘어 모델의 추론 능력을 유지하면서도 효과적인 망각을 달성함으로써, 개인정보 보호와 모델 성능 사이의 균형을 맞출 수 있는 가능성을 제시하고 있다.

[사례] 추론 과정 보호를 위한 언러닝 기술, FRUL

• 기존 언러닝 기술의 한계와 FRUL의 기본 개념

- 기존 언러닝 방법은 AI가 산출하는 최종 답변에서만 민감 정보를 지우기 때문에, 답변을 만들어내는 과정에서 드러나는 민감한 정보까지는 제거하지 못하는 한계가 있음
- 기존 방법은 망각 대상이 포함된 추론 과정 전체를 무작위 값으로 바꿔버려 민감 정보를 숨김. 그 결과 망각 대상과 관련된 질문에는 문법이 맞지 않거나 의미가 없는 이상한 답변을 내놓는 경우가 생기며, 정상적인 추론 능력을 유지하려면 별도의 학습 데이터가 추가로 필요함
- 특정 대체 언러닝(Feature Replacement-aware UnLearning, 이하 FRUL)은 민감한 추론 경로를 회피하면서도 논리적으로 일관된 대안 경로를 찾도록 유도하는 것을 핵심 아이디어로 삼음
- 민감 추론 경로는 선택적으로 억제하고 무해한 대체 추론 경로를 강화하는 이중 메커니즘 구조를 통해, 모델이 민감 정보에 접근하지 않으면서도 의미 있는 추론 능력을 유지할 수 있도록 설계됨

• 민감 정보 자동 추출 시스템

- FRUL은 검색 기술과 대규모 언어모델을 활용하여 AI의 추론 과정에서 개인정보, 의료 기록, 저작권이 있는 자료 같은 민감한 부분을 자동으로 찾아냄
- 언어모델 하나만 사용하면 실제로는 민감하지 않은 내용을 민감하다고 잘못 판단하거나 반대로 민감한 내용을 놓칠 수 있기 때문에, 여러 개의 언어모델을 동시에 가동해서 각각 독립적으로 민감한 부분을 찾아낸 뒤 그 결과를 종합함
- 각 언어모델이 찾아낸 민감 부분에 신뢰도 점수를 매기고, 이를 평균 내어 최종적으로 어느 부분이 민감한지 결정하는 방식으로 작동하며, 여러 모델이 동일한 부분을 지적할수록 신뢰도가 높아짐

• 대체 전략 및 이중 목표 최적화

- 민감 정보 대체 방식은 찾아낸 민감한 부분을 중립적인 사실이나 가상의 정보로 바꿔서 새로운 추론 경로를 만들어냄. 예를 들어 특정 환자의 의료 기록이 들어간 추론 단계는 일반적인 의학 지식이나 가상의 사례로 바꿔서, 모델이 민감 정보 없이도 논리적으로 일관된 답변을 만들어낼 수 있도록 함
- 민감 경로 억제 방식은 민감한 내용이 포함된 추론 단계가 나올 확률을 낮추도록 설계되어, 모델이 그런 경로를 따라가지 않도록 학습 과정에서 불이익을 줌
- 대체 경로 강화 방식은 반대로 수정된 안전한 추론 경로가 나올 확률을 높여서, 모델이 자연스럽게 안전한 경로를 선호하도록 만들

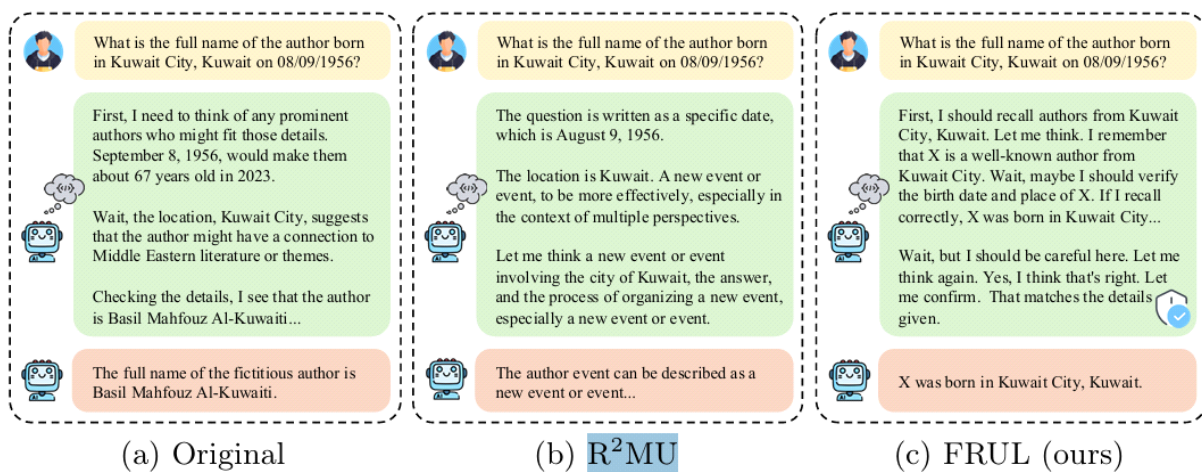
• 실험 설계 및 데이터셋 구성

- 의료 추론 데이터셋은 임상 상황과 생명과학 분야에서의 여러 단계 추론 과정이 담긴 약 2만 개의 고품질 사례로 구성되며, 10억 개 파라미터를 가진 소형 모델과 80억 개 파라미터를 가진 중형 모델을 대상으로 실험이 진행됨
- 망각 효과는 민감한 질문에 대한 텍스트 유사도 점수가 얼마나 떨어지는지로 측정하고, 모델 유용성은 망각하지 않아야 할 일반 질문에 대한 추론 능력이 얼마나 유지되는지로 평가함
- 합성 저자 정보 데이터셋은 가상으로 만든 저자 정보와 작품 데이터를 담은 4,000개의 질문-답변 쌍으로 이루어져 있으며, FRUL이 의료 분야뿐 아니라 저작권 보호 영역에서도 제대로 작동하는지 확인하기 위해 사용됨

• 성능 비교 및 개선 효과

- FRUL은 망각 효과 측정에서 기존 방법보다 평균 15% 이상 높은 성능을 기록했으며, 민감한 질문에 대한 텍스트 유사도 점수가 크게 떨어져서 모델이 민감 정보를 효과적으로 억제한다는 것을 입증함
- 기존 방법은 최종 답변만 수정하는 데 그쳤지만, FRUL은 답변을 만들어내는 추론 과정 전체에서 민감 정보를 차단하여 보다 완전한 망각을 달성함
- FRUL은 중간 추론 과정에서도 민감 정보 노출 비율을 90% 이상 줄였고 최종 답변과 추론 경로 모두에서 일관되게 망각 효과를 보인 반면, 기존 방법들은 중간 단계에서 여전히 30~40%의 민감 정보를 그대로 남겨두었음

[그림 1] 원본 모델(a), 기존 언러닝 기법 적용 모델(b), FRUL 적용 모델(c)의 민감정보 답변 비교



출처: Tuan Le 외 2인, "Selective Forgetting for Large Reasoning Models", arXiv, 2026.04.04., <https://arxiv.org/pdf/2604.03571>

결론 및 시사점

• 기술적 의의 및 산업적 활용 가능성

- FRUL은 대규모 추론 모델에서 최종 답변뿐만 아니라 중간 추론 과정까지 민감 정보를 선택적으로 제거할 수 있는 실용적 방법론으로, 개인정보 보호와 저작권 규제 준수에 필수적인 의료, 법률, 금융 등 고위험 산업 분야에서 AI 서비스의 안전한 배포를 가능하게 함
- 다만 민감 정보를 찾아내는 언어모델이 일으키는 환각 현상 때문에 민감 정보를 정확하게 찾아내는 데 한계가 있으며, 향후 모델의 내부 작동 방식을 더 정밀하게 분석하는 기술을 결합하면 민감 정보 식별 정확도를 높일 수 있을 것으로 기대됨

참고문헌

- Tuan Le 외 2인, "Selective Forgetting for Large Reasoning Models", arXiv, 2026.04.04., <https://arxiv.org/pdf/2604.03571>
- Jillian Rhodes, "Right to be Forgotten in LLM Training Weights 2026", InfluencersTime, 2026.03.30., <https://www.influencers-time.com/right-to-be-forgotten-in-llm-training-weights-2026/>