



SUMMARY

산업/기업

기술

## 기술 구글 딥마인드, AI 학습용 데이터 정제 기술 논문 발표

### 구글 딥마인드, 유해한 내용 담긴 데이터도 AI 학습용 데이터로 정제

▶ 구글 딥마인드가 개인정보나 유해 콘텐츠 때문에 활용할 수 없었던 데이터를 생성형 AI 학습용으로 정제하는 '생성형 데이터 정제(GDR)' 기술 논문을 발표했다. 이 기술은 사전 훈련된 생성 모델을 활용하여 문제가 있는 데이터를 훈련에 적합한 형태로 변환하는 방식으로, 완전히 새로운 데이터를 만드는 대신 실제 데이터의 문제 부분만 수정한다. GDR은 기존에 폐기되었던 데이터를 학습에 활용 가능한 자원으로 전환함으로써, 데이터 고갈 문제에 대응하고 안전하면서도 유용한 훈련 데이터의 총량을 증가시킬 수 있는 실용적 대안을 제시한다.

## 기술 AI 산출물의 기존 저작물 활용 여부 탐지 신기술, 대형 음반사와 연구소 협력으로 첫 상용화 추진

### AI 기반 음원 연구소, 음반사와 협력하여 인간 음악 저작물 활용 여부 탐지 기술 개발

▶ 스탠퍼드대학교 연구진이 설립한 AI 기반 음원 분석 연구소인 사운드페트롤은 2025년 9월 유니버설뮤직그룹, 소니뮤직과 함께 '뉴럴 핑거프린팅' 기술을 공동 개발하고, 산업 현장 적용을 추진 중이라고 발표했다. 개발된 기술은 기존의 오디오 핑거프린팅이 단순히 동일한 음원을 식별하는 수준을 넘어, AI가 음악의 멜로디·리듬·화성 등을 분석해 인간 창작물의 흔적을 찾아내는 신경망 기반 모델로 설계되었으며, 실시간 스트리밍 환경에서도 작동하기 때문에 실시간으로 기존의 저작물 활용 여부를 탐지할 수 있다.

## 산업 언론사 콘텐츠 마켓플레이스 등장, 대형 기술 기업 최초의 AI 활용 콘텐츠 보상 사례

### 마이크로소프트, 언론사 콘텐츠 마켓플레이스 시범 운영 준비 중

▶ 마이크로소프트는 2025년 9월 '퍼블리셔(언론사) 콘텐츠 마켓플레이스' 출시에 앞서 미국 내 언론사를 대상으로 시범 프로그램을 준비 중이라고 밝혔다. 이 플랫폼은 AI 기업과 언론사를 직접 연결하는 구조로, AI 기업이 언론사로부터 콘텐츠 라이선스를 직접 구매하고, 언론사는 마이크로소프트의 중앙화된 시스템을 통해 수익을 추적할 수 있도록 설계되었다. 결제는 콘텐츠의 실제 활용 빈도와 범위에 따라 산정되는 사용량 기반 결제 모델로 진행되며, 이는 기존의 선불 라이선스 계약과 차별화된다. 또한 마이크로소프트는 해당 모델을 통해 언론사들이 제기해 온 무단 데이터 스크래핑 문제를 완화하고, 콘텐츠 활용도에 비례한 공정한 보상 체계를 구축하고자 한다.



## SUMMARY

## 산업/기업

## 기술

**기술** 디지털 엔터테인먼트 환경에서의 지역별 콘텐츠 접근 제한 기술 동향**IP 필터링부터 AI 탐지까지, 진화하는 지역 차단 기술의 현황 및 과제**

▶ 디지털 엔터테인먼트 환경에서 지역 차단 기술은 지역별 라이선싱 계약 준수를 위해 필수적인 수단이 되었다. 콘텐츠 제공업체는 IP 필터링, DNS 차단, 심층 패킷 검사(DPI), AI 기반 탐지 등 다양한 기술을 활용하여 지역별 접근을 제한하고 있다. IP 필터링은 머신러닝으로 연결 패턴을 분석하며, DNS 차단은 네트워크 인프라 수준에서 우회 시도를 막는다. 심층 패킷 검사는 전체 데이터 페이로드를 분석해 무단 접근을 식별하고, AI 알고리즘은 사용자의 의심스러운 활동을 탐지한다. 그러나 활성 인터넷 사용자의 25%가 VPN을 사용하며, 암호화 기술 발전으로 탐지가 어려워지는 등 기술적 한계가 존재한다.

**산업** 엔비디아, 오디오 기반 3D 얼굴 애니메이션 생성 AI 모델 오픈소스로 공개**▶ 엔비디아, 게임 캐릭터용 AI 얼굴 애니메이션 도구 오픈소스 전환**

▶ 2025년 9월 엔비디아가 오디오 기반 3D 얼굴 애니메이션 생성 기술인 Audio2Face의 AI 모델, SDK, 훈련 프레임워크를 MIT 라이선스로 오픈소스화하면서 게임 개발자의 기술 접근성이 대폭 확대됐다. 이 기술은 음성 입력만으로 캐릭터의 립싱크와 감정 표현을 자동 생성하며, ‘오토데스크 마야’와 ‘엔리얼 엔진 5’용 플러그인 제공으로 주요 개발 환경에 즉시 통합 가능하다. 개발자는 제공된 훈련 프레임워크를 통해 자체 데이터로 맞춤형 모델을 생성할 수 있으며, 실시간 스트리밍과 사전 제작 콘텐츠 모두에 활용할 수 있다. 해당 기술은 현재 다양한 게임에 적용되어 애니메이션 작업 시간을 크게 단축한 것으로 나타났으며, 엔비디아는 디스코드 커뮤니티 운영을 통해 개발자 생태계를 구축하려는 행보를 보이고 있다.

**기술** 주간기술동향**변형 시뮬레이션형 텍스트-투-비디오 기술**

▶ 최근 영상 제작 산업에서는 픽셀 조합을 통해 시각적으로 유사한 영상을 만드는 것을 넘어, 시간의 흐름에 따른 변화나 물리 법칙이 적용된 현상을 구현하려는 시도가 본격화되고 있다. 본 보고서에서는 하나의 시뮬레이션형 비디오 생성 기술의 사례 분석을 통해, 현재의 생성형 AI 기술이 단순한 영상 제작을 지나 현실 세계의 변화를 얼마나 사실적으로 구현하는지, 또한 어떠한 원리로 기술이 작동되는지 그 현황을 살펴보고자 한다.



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 구글 딥마인드, AI 학습용 데이터 정제 기술 논문 발표

### 뉴스브리프

구글 딥마인드가 개인정보나 유해 콘텐츠 때문에 활용할 수 없었던 데이터를 생성형 AI 학습용으로 정제하는 '생성형 데이터 정제(GDR)' 기술 논문을 발표했다. 이 기술은 사전 훈련된 생성 모델을 활용하여 문제가 있는 데이터를 훈련에 적합한 형태로 변환하는 방식으로, 완전히 새로운 데이터를 만드는 대신 실제 데이터의 문제 부분만 수정한다. GDR은 기존에 폐기되었던 데이터를 학습에 활용 가능한 자원으로 전환함으로써, 데이터 고갈 문제에 대응하고 안전하면서도 유용한 훈련 데이터의 총량을 증가시킬 수 있는 실용적 경로를 제시한다. 참고로 구글은 이 기술을 1년간 비공개로 유지해 왔는데 이번 공개와 함께, GDR의 학습 데이터를 일반적인 개인정보를 포함한 데이터 유형 외 다른 유형으로 확장하는 추가 연구도 계획하고 있으며, 추가 연구 대상에는 저작권이 있는 콘텐츠도 포함되어 있다.

### 구글 딥마인드, 유해한 내용 담긴 데이터도 AI 학습 데이터로 정제하는 기술 공개

#### • 1년간 비공개로 유지했던 데이터 정제 기술 논문 발표

- 구글 딥마인드(Google DeepMind)가 유해하거나 민감한 일부 내용 때문에 활용할 수 없었던 데이터를 생성형 AI의 학습용 데이터로 정제하는 방법에 대한 논문을 발표함
- '생성형 데이터 정제: 더 나은 데이터 요청하기(Generative Data Refinement: Just Ask for Better Data)'라는 제목의 이 논문은 2025년 9월 11일 온라인 아카이브 arXiv를 통해 공개됨
- 논문이 작성된 시기는 2024년 9월이나 약 1년간 비공개로 유지되다가, 제1저자인 민치 지양(Minqi Jiang)이 메타 슈퍼인텔리전스 랩(Meta Superintelligence Labs)으로 이직한 이후 공개함<sup>1)</sup>

1) 임대준, "구글, 1년 간 공개하지 않았던 AI 학습 데이터 '정제법' 발표", 시타임스, 2025.09.16., <https://www.aitimes.com/news/articleView.html?idxno=202466>

## • 연구 배경 : 거대언어모델의 학습 데이터 고갈 문제

- 고정된 파라미터 규모에서 거대언어모델(LLM)의 성능은 주로 훈련 데이터의 품질과 양에 의해 결정됨. 스케일링 법칙\*에 따르면 LLM을 훈련시킬 때 최적의 연산 효율을 달성하기 위해서는 모델 파라미터와 훈련 토큰 모두 비례적으로 확장해야 함

\* 스케일링 법칙(scaling laws): AI 모델이 더 많은 리소스를 활용할수록 성능이 비례적으로 향상된다는 원리

- 현재 AI 모델의 훈련 데이터 세트는 웹상에서 데이터가 크롤링되는 속도보다 빠르게 확장되고 있어, 향후 10년 내 데이터 고갈이 예상됨
- AI 기존 모델의 훈련 데이터는 웹상에 존재하는 전체 데이터가 아니라, 검색엔진을 통해 공개적으로 크롤링되어 수집된 데이터 세트만을 대상으로 하고 있는데, 웹상에는 크롤링되지 않은 형태로 지속적으로 생성되는 훨씬 더 많은 활용 가능 데이터가 존재함
- 문제는, 이러한 데이터 세트에는 사용자 생성 데이터와 다양한 형태의 민감 정보가 포함되며, 이러한 범주의 데이터로 훈련할 경우 AI 모델이 개인정보, 혹은 유해 콘텐츠 및 저작권이 있는 자료까지 학습하게 되는 점임

## • 합성 데이터 생성 방식의 한계와 새로운 접근 필요성

- 이러한 문제 때문에 최근 산업계에서 AI 학습용 데이터 확장 노력은 합성 데이터(synthetic data) 생성 프로토콜 개발에 집중되고 있음
- 합성 데이터는 크롤링된 데이터를 기반으로 알고리즘과 기술을 활용해 인공적으로 재합성된 데이터를 의미하며, 이는 사전에 훈련된 모델 또는 샘플 데이터 세트로 이미 파인튜닝(Fine-tuning)\*된 모델이 자체적으로 생성하는 데이터 결과물을 의미함

\* 파인튜닝(Fine-tuning): 사전 학습된 모델을 새로운 작업에 특화시키기 위해 매개변수를 추가 학습하는 과정

- 그러나 이러한 합성 데이터만으로 접근하는 방법은 한계가 있음. 모델 파인튜닝에는 추가적인 컴퓨팅이 필요하고, 인공적으로 합성된 샘플은 모방하려는 현실 세계의 데이터와는 확연한 차이가 있음

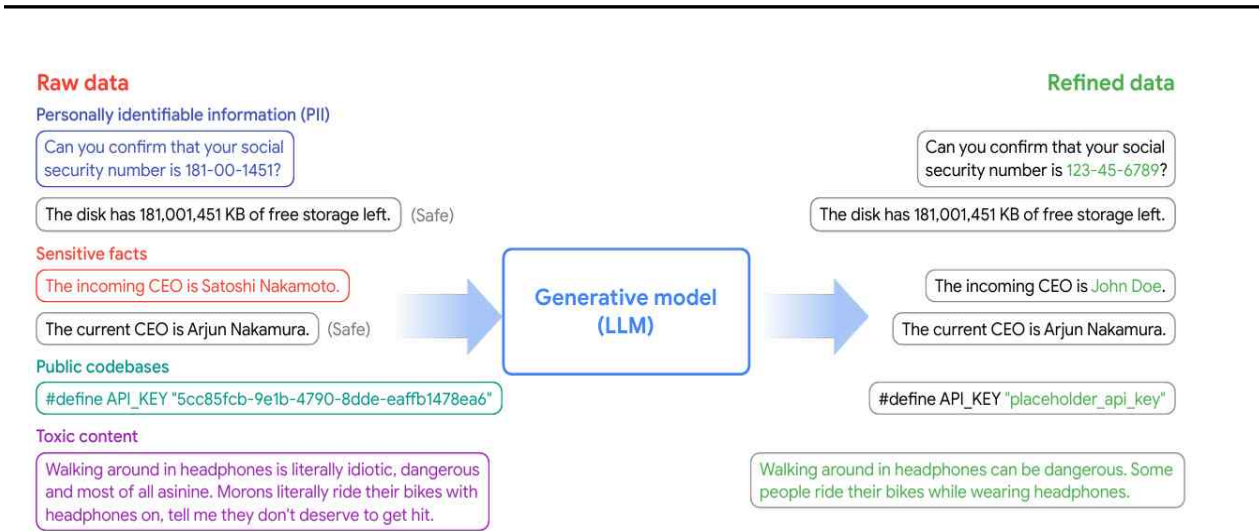
## 새로운 접근: 생성형 ‘데이터 정제 프레임워크’의 기술 구조

### • 기존 데이터를 기반으로 새롭게 만드는 정제 방식

- 구글 딥마인드가 이번에 발표한 ‘생성형 데이터 정제(Generative Data Refinement, GDR)’ 기술은 합성 데이터 생성을 위한 새로운 문제 접근법으로, 사전 훈련된 생성 모델을 활용하여 문제가 있는 데이터를 훈련에 적합한 형태로 변환함
- 이 방식에서는 완전히 새로운 데이터를 만들어내는 대신, 실제로 존재하는 데이터를 바탕으로 문제 부분만 수정하여 합성 데이터를 생성함
- 이처럼 실존하는 데이터에 기반을 둔 합성 데이터 생성 방식은 실제 데이터가 가진 내용에만 기반하여 산출물을 생성하기 때문에 더 현실적인 결과물을 만들어낼 수 있고, 대규모 데이터 세트가 본래 가지고 있던 다양성을 자연스럽게 유지할 수 있음

- GDR의 작동 원리는 다음 [그림1]과 같이 원본 데이터 세트에서 각 입력 데이터를 생성 프로세스에 통과시켜 새로운 데이터를 만들어내는 프로세스로, 이때 생성된 새로운 데이터는 사전에 설정한 기준을 충족해야 하고 이와 동시에 원본 데이터와 최대한 유사해야 함

[그림 1] '생성형 데이터 정제(GDR)'의 개요도



원본 데이터(개인식별정보, 유해 콘텐츠 등 포함) → 생성 모델(LLM) 처리 → 정제 데이터(안전한 콘텐츠로 변환)

출처: Minqi Jiang 외 4인, "Generative Data Refinement: Just Ask for Better Data", arXiv, 2025.09.11., <https://arxiv.org/pdf/2509.08653>

### • 프롬프트를 이용한 LLM 변환과 상세한 검증 절차로 작은 모델로도 효과

- 본 연구에서는 대규모 데이터 세트로 훈련된 프롬프트 기반 LLM을 생성 프로세스로 사용하는 단순한 경우에 집중함<sup>2)</sup>
- GDR에서 설정하는 기준은 매우 다양할 수 있음. 가령, “텍스트에 철자 오류가 포함되어 있는지”, 혹은 “이미지에 일출 장면이 담겨 있는지” 등의 구체적인 내용을 검증 함수로 삼아 평가할 수 있으며, 이처럼 구체적인 정보들의 조건이 평가의 기준이 될 수 있음
- GDR 역시 기존의 데이터 합성 프로세스처럼 많은 양의 컴퓨팅 용량이 필요하지만, 훨씬 경제적임. 이유는, 한 번 정제된 데이터 세트는 이후 모델 훈련에 반복적으로 재사용할 수 있어, 시간이 지남에 따라 비용 부담이 분산됨
- 또한, 더 작은 모델을 파인튜닝하여 더 큰 모델의 품질에 근접하게 만들 수 있고, 초기에 큰 모델에서 지식을 추출하여 작은 모델을 개선하는 방법도 활용할 수 있어 실제 사용에서는 비용이 더 저렴해질 수 있음

## 개인식별정보 익명화 성능 실험 결과

### • 기존의 상용화된 솔루션 보다 개인정보 탐지 정확도 높아

- 연구진은 GDR의 개인정보 제거 성능을 검증하기 위해 실제 기업들이 사용하는 상용 서비스와 비교 실험을 진행함. 비교 대상은 DIRS(Detector-based Information Removal Service)\*로, 2024년 9월 버전 기준 산업 표준급 솔루션임

<sup>2)</sup> Minqi Jiang 외 4인, "Generative Data Refinement: Just Ask for Better Data", arXiv, 2025.09.11., <https://arxiv.org/pdf/2509.08653>

\* DIRS(Detector-based Information Removal Service): Optery, Incogni, DeleteMe 등의 서비스가 대표적으로, 자동 스캔 기술을 활용하여 데이터 브로커 및 기타 온라인 출처에서 개인정보를 찾아 제거해 주는 서비스

- DIRS는 실명이나 각국의 주민등록번호처럼 특정한 개인정보 유형마다 별도의 탐지 프로그램을 갖춘 방식으로 작동되는데, 대부분 특정 키워드를 기반으로 한 정해진 규칙에 기반하여 작동되기 때문에 종종 불안정하고 텍스트 전체 맥락을 고려하지 못하는 한계가 있음
- 반면 GDR은 거대언어모델(LLM)이 학습 과정에서 익힌 방대한 문맥 이해 능력을 활용해 문장 전체의 의미를 고려하며 개인정보를 식별함. 또한 DIRS처럼 항목별 탐지기를 여러 개 운영하지 않고, 하나의 모델로 다양한 개인정보를 한 번에 처리할 수 있어 확장성과 효율성이 높음
- 연구진은 DIRS가 지원하는 108개 개인정보 카테고리에 걸쳐 2만 개 이상의 문장을 대상으로 실험을 진행함. 그 결과, GDR은 재현율 0.99, 정밀도 0.80을 기록하여 DIRS의 기록을 상회함. 이는 GDR이 실제 개인정보는 대부분 찾아내면서도, 일반 정보를 개인정보로 잘못 판단하는 경우가 적다는 것을 의미함

[표1] GDR과 DIRS의 비교 실험 결과

구분	재현율(Recall)*	정밀도(Precision)**	F-스코어(F-score)***
DIRS	0.53	0.52	0.52
GDR	0.99	0.80	0.88

\* 재현율(Recall): 실제 개인정보 중 시스템이 찾아낸 비율. 1.0에 가까울수록 개인정보를 놓치지 않고 잘 찾아냄

\*\* 정밀도(Precision): 시스템이 개인정보라고 판단한 것 중 실제로 개인정보인 비율. 1.0에 가까울수록 오탐지가 적음

\*\*\* F-스코어: 머신러닝에서의 F-스코어는 정밀도(precision)와 재현율(recall)의 조화 평균으로, 0부터 1 사이의 값을 가지며 1에 가까울수록 높은 성능을 나타냄

출처: Minqi Jiang 외 4인, "Generative Data Refinement: Just Ask for Better Data", arXiv, 2025.09.11., <https://arxiv.org/pdf/2509.08653>

### ● 맥락을 이해하는 세밀한 데이터 재작성 방식

- GDR과 DIRS의 가장 큰 차이는 ‘데이터를 직접 만들어 수정하느냐, 단순히 문제를 찾아 표시하느냐’에 있음. DIRS는 개인정보를 찾아내는 기능에만 집중한 판별형 방식으로, 탐지된 부분을 실제로 수정하려면 별도의 모듈이 따로 필요함
- 반면 GDR은 생성형 접근을 통해 문제 부분을 스스로 인식하고, 그 자리에 안전한 문장으로 자동 대체함
- 또한 DIRS는 탐지된 부분 문자열을 미리 정의된 안전한 문자열 집합에서 제공된 값으로 교체하는 옵션을 제공하지만, 주로 판별적 특성으로 인해 일반적으로 개인정보가 포함될 가능성이 있는 문서를 표시하여 학습 세트에서 제거하는 방식으로 사용됨
- 반면 GDR은 맥락적으로 관련성 있는 대체 콘텐츠를 직접 생성하여 개인정보를 교체함. 이러한 차이로 인해 GDR은 데이터를 학습에 활용할 수 있다는 점이 핵심적 장점임
- GDR은 사전 훈련된 거대언어모델(LLM)에 내재된 풍부한 세계 지식을 활용하여 문제가 되는 데이터 부분만 선택적으로 재작성함

## 유해 콘텐츠 정화 및 데이터 다양성 보존 성능

### • 인터넷상 악성 콘텐츠를 안전한 학습 데이터로 변환

- 인터넷상의 유해 콘텐츠를 그대로 AI 모델 학습에 사용하면 모델이 잘못된 가치관이나 위험한 내용을 학습할 수 있으나, 이러한 데이터에도 유용한 정보가 포함되어 있을 수 있음. 연구진은 GDR을 유해 콘텐츠 제거 작업에 적용하여, 유해성은 낮추면서도 유용한 지식은 유지할 수 있는지 검증함
- 연구진은 인종차별, 성차별, 공격적 언어가 만연한 것으로 악명 높은 웹사이트인 4chan의 /pol/ 게시판 데이터를 실험 대상으로 선정함. 400만 개 토론에서 10만 개의 대화 쌍을 추출한 데이터 세트 'pol100k'를 구성하여 실험에 활용함
- GDR을 적용한 결과, 유해성 측정 도구인 퍼스펙티브 API 점수가 원본 데이터 0.19에서 정제 후 0.13으로 크게 감소한 것으로 나타남

\* 퍼스펙티브 API(Perspective API): 구글이 건전한 대화 환경을 조성하기 위해 개발한 악성 댓글 차단 도구로, 머신러닝 모델을 사용하여 텍스트의 독성(Toxic)을 검사하고 수치로 반환하는 기능을 제공함

### • 데이터 정화 과정에서도 유용한 정보는 보존, 실제 데이터 수준의 다양성 유지

- 연구진은 데이터 정제 과정에서 'pol100k' 데이터에 포함된 세계에 관한 사실 정보를 추출하여 5천 개의 질문-답변 쌍으로 구성된 'pol5k-quiz' 데이터 세트를 생성함. 이 데이터 세트는 원본 유해 데이터에 포함된 실제 지식을 평가하기 위한 기준으로 활용됨
- 정제된 데이터로 학습한 AI 모델의 정확도는 학습 전 0.88에서 학습 후 0.92로 향상되어, 정제 과정에서도 원본 데이터의 유용한 정보가 잘 보존되었음을 확인함
- 또한 웹 규모 데이터는 매우 다양한 특성을 지니나, AI 모델에서 다양한 샘플을 생성하는 것은 어려운 과제임. 기본 모델은 복잡한 프롬프트 설계가 필요하고, 지시를 잘 따르도록 조정된 모델은 출력의 다양성이 감소하는 경향이 있음
- GDR은 각 샘플을 실제 데이터를 기반으로 재작성하는 방식으로, 실제 데이터가 지닌 다양성을 자연스럽게 유지할 수 있음. 분석 결과 정제된 'pol100k' 데이터의 다양성은 순수 합성 대화 데이터보다 훨씬 높았으며, 원본 데이터의 다양성을 약간 상회하는 수준을 보임

## 결론 및 시사점

### • 버려지던 데이터를 학습에 활용 가능한 자원으로 전환

- 구글 답마인드가 제시한 GDR 기법은 충분히 발전된 거대언어모델(LLM)을 활용하여 문제가 있던 데이터를 안전하게 재작성하는 방식으로, 이 과정에서 개인정보나 유해 표현 같은 바람직하지 않은 내용은 제거하면서도, 해당 데이터에 담긴 유용한 정보는 그대로 보존함
- 이는 실제 데이터에 기반하여 새로운 데이터를 만들어내는 합성 데이터 생성 방식의 한 형태로, 완전히 새로운 데이터를 생성하는 방식과는 구별됨
- 연구진이 여러 대규모 실제 데이터 세트를 대상으로 실험한 결과, GDR은 개인식별정보와 유해 콘텐츠를 안정적으로 제거하면서도 작업에 필요한 정보를 유지하는 것으로 나타남

- 또한 완전히 새로운 합성 데이터를 직접 생성하는 방식보다 더 다양한 형태의 데이터 세트를 만들어냄으로써, 안전하면서도 유용한 훈련 데이터의 총량을 증가시킬 수 있는 실용적인 대안을 제시함

#### • 향후 연구 과제와 기술 개선 방향

- 연구진은 향후 연구 방향으로 증류(distillation) 기법이나 강화학습 기반 파인튜닝을 통해 컴퓨팅 비용을 줄이고 정제 품질을 개선하는 방안을 제시함
- 증류 기법은 큰 모델의 능력을 작은 모델로 전이시켜 같은 작업을 더 적은 비용으로 수행할 수 있게 하는 기술이며, 강화학습 파인튜닝은 모델이 더 나은 결과를 생성하도록 학습시키는 방법임
- 또한 구글은 GDR을 다른 데이터 형식과 위험 유형으로 확장하는 연구도 계획하고 있다고 밝힘.<sup>3)</sup> 여기에는 저작권이 있는 콘텐츠 처리와, 여러 문서에 분산되어 있어 개별 문서 수준에서는 감지하기 어려운 개인정보 유출 문제를 다루는 코퍼스\* 수준(corpus-level) 개인정보 처리 등이 포함됨

\* 코퍼스(corpus): 체계적으로 수집되고 구성된 텍스트 모음을 뜻하며 특정 언어, 방언, 주제, 스타일 또는 시대를 대표하는 텍스트들을 포함할 수 있음. 코퍼스 수준 개인정보 유출이란 한 문서만 봤을 때는 문제가 없어 보이지만, 여러 문서를 종합하면 특정 개인에 대한 민감한 정보를 추론할 수 있게 되는 경우를 의미함

## 참고문헌

- Minqi Jiang 외 4인, "Generative Data Refinement: Just Ask for Better Data", arXiv, 2025.09.11., <https://arxiv.org/pdf/2509.08653>
- 임대준, "구글, 1년 간 공개하지 않았던 AI 학습 데이터 '정제법' 발표", AI타임스, 2025.09.16., <https://www.aitimes.com/news/articleView.html?idxno=202466>

3) Minqi Jiang 외 4인, "Generative Data Refinement: Just Ask for Better Data", arXiv, 2025.09.11., <https://arxiv.org/pdf/2509.08653>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## AI 산출물의 기존 저작물 활용 여부 탐지 신기술, 대형 음반사와 연구소 협력으로 첫 상용화 추진

### 뉴스브리프

2025년 9월, 스탠퍼드대학교 연구진이 설립한 AI 기반 음원 분석 연구소인 사운드패트롤은 유니버설뮤직그룹, 소니뮤직과 함께 ‘뉴럴 핑거프린팅’ 기술을 공동 개발하고, 산업 현장 적용을 추진 중이라고 발표했다. 개발된 기술은 기존의 오디오 핑거프린팅이 단순히 동일한 음원을 식별하는 수준을 넘어, AI가 음악의 멜로디·리듬·화성 등을 분석해 인간 창작물의 흔적을 찾아내는 신경망 기반 모델로 설계되었으며, 실시간 스트리밍 환경에서도 작동하기 때문에 실시간으로 기존 저작물의 활용 여부를 탐지할 수 있다. 유니버설뮤직과 소니뮤직은 이를 “산업 전반을 지원하기 위한 해결책”이라 평가했으며, 사운드패트롤은 향후 이 기술을 제3의 스트리밍 플랫폼과 연구기관에도 제공할 계획이다.

### AI 산출물 음악의 기존 저작물 활용 여부 탐지 기술 개발

#### • AI 산출물 음악의 저작권이 있는 음악 무단 활용 여부 탐지하는 신기술 개발

- 2025년 9월, 사운드패트롤(SoundPatrol Inc.)은 유니버설뮤직그룹(Universal Music Group, UMG), 소니뮤직(Sony Music)과 협력하여 AI 산출물 음악이 기존의 저작권이 있는 음악을 무단으로 활용하거나 유사하게 생성했는지를 탐지하기 위한 새로운 뉴럴 핑거프린팅(Neural Fingerprinting)\* 기술을 공동으로 개발하였으며, 신기술을 산업 현장에 적용할 것이라고 발표함<sup>1)</sup>

\* 뉴럴 핑거프린팅(Neural Fingerprinting): 음악의 멜로디·코드·리듬 등 구성요소를 신경망으로 수치화하여, 의미적 유사성을 분석하는 기술

- 사운드패트롤은 스탠퍼드대학교(Stanford University) 연구진이 주축이 되어 설립한 AI 기반 음원 분석 전문 연구기관으로, 주로 음악 및 오디오 분야의 저작권 보호 기술을 연구·개발하고 있으며, 최근에는 AI 산출물 음악의 기존 저작물 활용 여부를 실시간으로 탐지하는 기술 인프라 구축을 추진하고 있음

1) SoundPatrol Inc., "SoundPatrol Collaborates With Universal Music Group and Sony Music to Deploy Groundbreaking Neural Fingerprinting Technologies for Detecting Copyright Infringement in Music, Including AI-Generated Works", Financial CONTENT, 2025.09.25., <https://markets.financialcontent.com/stocks/article/bizwire-2025-9-25-soundpatrol-collaborates-with-universal-music-group-and-sony-music-to-deploy-groundbreaking-neural-fingerprinting-technologies-for-detecting-copyright-infringement-in-music-including-ai-generated-works>

## 신기술: 음악과 영상 내 기존 저작물 활용 요소 감식하는 AI 기반 탐지 기술

### • 음악과 영상에 포함된 기존 저작물 활용 요소 탐지하는 포렌식 AI 모델

- 새롭게 개발된 기술은 기존의 오디오 핑거프린팅(audio fingerprinting)\*이 단순히 동일한 음원 조각을 식별하는 데 그쳤던 한계를 넘어, 음악의 멜로디·리듬·화성 구조 등 의미적 유사성(musical semantics)\*\*을 분석할 수 있는 신경망 기반 모델을 적용함

\* 오디오 핑거프린팅(audio fingerprinting): 음원의 파형, 주파수, 비트 등 고유한 음향 특징을 수치화하여 짧은 코드 형태로 저장한 뒤, 동일하거나 유사한 음원을 빠르게 식별하는 기술

\*\* 의미적 유사성(musical semantics): 음악의 멜로디, 리듬, 화성, 구조 등 '음향적 형태'뿐 아니라 그 안에 담긴 음악적 의미 관계를 분석하여 두 곡의 유사도를 파악하는 개념

- 이는 AI가 음악의 코드, 멜로디, 리듬, 음색 등 핵심 구성요소를 수치화하여 '뉴럴 임베딩(Neural Embedding)\*' 형태로 변환하고, 이를 통해 음악의 의미적 유사성을 학습하는 구조로, 이 과정을 통해 모델은 인간이 작곡한 음악이 AI 산출물 음악 속에 어느 정도 '영향을 미쳤는지'를 정밀하게 분석할 수 있음

\* 뉴럴 임베딩(Neural Embedding): 음악의 멜로디, 코드, 리듬 등 다양한 구성요소를 인공지능이 이해할 수 있는 수치 벡터 형태로 표현하는 기술

- 예를 들어, 기존 노래의 일부 멜로디가 미세하게 변형되어 포함되어 있거나 코드 진행이 유사하게 구성된 경우도 감지할 수 있으며, 실시간 스트리밍 환경에서도 작동되기 때문에 음악이 전송·재생되는 즉시 기존의 저작물 활용 여부를 탐지할 수 있음

**[표1] AI 산출물의 기존 저작물 활용 여부 탐지 신기술 상세 특징 및 응용 범위**

구분	명칭 및 특징	상세 내용
핵심 기술 요소	뉴럴 핑거프린팅 (Neural Fingerprinting)	음악의 구조적·의미적 유사성을 분석하여 AI 산출물 속 인간 창작물의 흔적 식별
	뉴럴 서명 (Neural Signature)	인공지능 신경망이 음악을 학습할 때 각 곡의 멜로디, 리듬, 음색 등에서 추출하는 고유한 특징 패턴(feature pattern)
	뉴럴 임베딩 (Neural Embedding)	음악 데이터를 벡터 형태로 변환해 신경망이 패턴과 유사성을 정량적으로 학습할 수 있도록 함
	심층 임베딩 (Deep Embedding)	인공지능이 학습한 음악의 구성 요소(멜로디, 리듬, 코드, 음색 등)를 다차원 벡터 형태로 변환하여 정량화하는 기술
기술적 특징	실시간 감지 및 추적 (Real-time Detection)	스트리밍 환경에서 음원이 재생되는 즉시 기존 저작물 활용 여부 분석 가능
응용 범위	오디오·영상 전송 경로 감시 (Pipeline Tracking)	음악의 생성·학습·유통 전 과정을 추적해 원본 콘텐츠의 불법 사용 방지

출처: 참고문헌 종합하여 재구성

## 신기술 개발에 대한 참여 기업들의 평가와 향후 계획

### ① 유니버설뮤직그룹: “음악 산업 전체를 지원하기 위한 해결책 제시 필요해”<sup>2)</sup>

- 유니버설뮤직그룹의 회장 겸 CEO 루시안 그레이지(Sir Lucian Grainge)는 “AI는 아티스트에게 상업적·창의적 기회를 제공하지만, 동시에 그들의 창작물을 보호하기 위한 효과적인 도구가 필요하다”며 사운드패트롤과의 협력은 산업 전체를 지원하기 위한 해결책을 제시하는 데 목적이 있다”고 언급

### ② 소니뮤직: “모두에게 지속 가능하고 공정한 생태계 구축 위한 협력”<sup>2)</sup>

- 소니뮤직의 글로벌 디지털 비즈니스 총괄 데니스 쿠키(Dennis Kooker)는 “AI는 올바르게 활용될 때 창작자에게 기회를 제공하지만, 그 과정에서 예술가의 권리를 침해해서는 안 된다”며 “이번 협력은 예술가의 권리를 존중하고, 모두에게 지속 가능하고 공정한 생태계를 구축하기 위한 것”이라고 밝힘

### ③ 사운드패트롤 : “AI 산출물 음악의 저작물 활용 여부 식별할 수 있는 기술 필요”<sup>2)</sup>

- 사운드패트롤의 공동 창립자 월터 드 브라우어(Walter De Brouwer) 박사는 “생성형 AI가 음악을 혁신적으로 변화시키고 있지만, 저작권을 포기한다면 예술가들은 자신의 작품에 대한 소유권을 잃게 된다”며 “AI가 학습한 음악의 고유한 특징 정보를 스트리밍 시스템에 미리 연결해 두어야, 창작자가 자신의 음악이 어디에서 어떻게 사용되는지를 직접 관리하고 정당한 보상을 받을 수 있다”고 강조함
- 또한, 공동 창립자 마이클 오비츠(Michael Ovitz)는 “아티스트의 가장 오래된 문제는 지식재산권 보호였으며, 사운드패트롤은 뉴럴 핑거프린팅을 통해 실시간으로 저작권 침해를 식별할 수 있는 최초의 기술을 구현했다”며 “이번 협력은 모든 예술가에게 큰 승리”라고 자평함

### ● 제3의 스트리밍 플랫폼과 연구기관에도 제공 계획

- 유니버설뮤직그룹과 소니뮤직은 이번에 개발한 신기술을 실제로 사용할 계획<sup>3)</sup>이며, 제3의 스트리밍 플랫폼과 연구기관에도 제공할 예정<sup>4)</sup>임
- 신규 개발된 기술은 스트리밍 인프라에 직접 통합되어, 음악이 전송·재생되는 순간 실시간으로 기존의 저작물 활용 여부를 탐지할 수 있기 때문에, 사운드패트롤은 이 뉴럴 핑거프린팅 기술을 제3의 스트리밍 플랫폼과 연구기관에도 제공할 계획임
- 또한, 해당 시스템은 생성형 AI 모델의 학습·생성·유통 과정 전반에서 음악의 진위성과 출처를 추적할 수 있도록 설계되어 있으며, 음악 산업 전반의 투명한 데이터 관리 체계 구축에 기여할 것으로 기대됨

2) Businesswire, "SoundPatrol Collaborates With Universal Music Group and Sony Music to Deploy Groundbreaking Neural Fingerprinting Technologies for Detecting Copyright Infringement in Music, Including AI-Generated Works", 2025.09.25., <https://www.businesswire.com/news/home/20250925867965/en/SoundPatrol-Collaborates-With-Universal-Music-Group-and-Sony-Music-to-Deploy-Groundbreaking-Neural-Fingerprinting-Technologies-for-Detecting-Copyright-Infringement-in-Music-Including-AI-Generated-Works>

3) Daniel Tencer, "Universal and Sony Music partner with new platform to detect AI music copyright theft using 'groundbreaking neural fingerprinting' technology", MusicBusinessWorldwide, 2025.09.25., <https://www.musicbusinessworldwide.com/universal-and-sony-music-partner-with-new-platform-to-detect-ai-music-copyright-theft-using-groundbreaking-neural-fingerprinting-technology/>

4) Universal Music, "SOUNDPATROL COLLABORATES WITH UNIVERSAL MUSIC GROUP AND SONY MUSIC TO DEPLOY GROUNDBREAKING NEURAL FINGERPRINTING TECHNOLOGIES FOR DETECTING COPYRIGHT INFRINGEMENT IN MUSIC, INCLUDING AI-GENERATED WORKS", 2025.09.25., <https://www.universalmusic.com/soundpatrol-collaborates-with-universal-music-group-and-sony-music-to-deploy-groundbreaking-neural-fingerprinting-technologies-for-detecting-copyright-infringement-in-music-including-ai-generated-wor>

## 신기술 개발 협력의 산업적 의미

- **저작물 활용 여부 탐지 기술의 진보와 산업 적용 첫 사례**
  - 대형 음반사인 유니버설뮤직그룹과 소니뮤직이 AI 음악으로 인한 무단 저작권 침해에 대응하기 위해, 연구기관과 함께 기존의 저작물 활용 여부를 탐지하는 기술을 산업에 도입한 첫 사례임
  - 또한, 새롭게 개발된 기술은 기존의 오디오 핑거프린팅이 단순히 동일한 음원 조각을 식별하는 데 그쳤던 한계를 넘어, 음악의 멜로디·리듬·화성 구조 등 의미적 유사성을 분석할 수 있는 신경망 기반 모델을 적용했다는 점에서 기술적 진보임
  - 새로운 저작물 활용 여부 탐지 기술은 생성형 AI 모델의 학습·생성·유통 과정 전반에서 음악의 진위성과 출처를 추적할 수 있도록 설계된 모델이기 때문에 향후 음악 산업 전반의 투명한 데이터 관리 체계 구축에 기여할 것을 기대해 볼만한 기술적 사례임

## 참고문헌

- SoundPatrol Inc., "SoundPatrol Collaborates With Universal Music Group and Sony Music to Deploy Groundbreaking Neural Fingerprinting Technologies for Detecting Copyright Infringement in Music, Including AI-Generated Works", Financial CONTENT, 2025.09.25., <https://markets.financialcontent.com/stocks/article/bizwire-2025-9-25-soundpatrol-collaborates-with-universal-music-group-and-sony-music-to-deploy-groundbreaking-neural-fingerprinting-technologies-for-detecting-copyright-infringement-in-music-including-ai-generated-works>
- Businesswire, "SoundPatrol Collaborates With Universal Music Group and Sony Music to Deploy Groundbreaking Neural Fingerprinting Technologies for Detecting Copyright Infringement in Music, Including AI-Generated Works", 2025.09.25., <https://www.businesswire.com/news/home/20250925867965/en/SoundPatrol-Collaborates-With-Universal-Music-Group-and-Sony-Music-to-Deploy-Groundbreaking-Neural-Fingerprinting-Technologies-for-Detecting-Copyright-Infringement-in-Music-Including-AI-Generated-Works>
- Ashley King, "UMG, Sony Music Sign Major AI Copyright Detection and Fingerprinting Alliance", 2025.09.25., Digital Music News, <https://www.digitalmusicnews.com/2025/09/25/umg-sony-music-ai-fingerprinting/>
- Daniel Tencer, "Universal and Sony Music partner with new platform to detect AI music copyright theft using 'groundbreaking neural fingerprinting' technology", MusicBusinessWorldwide, 2025.09.25., <https://www.musicbusinessworldwide.com/universal-and-sony-music-partner-with-new-platform-to-detect-ai-music-copyright-theft-using-groundbreaking-neural-fingerprinting-technology/>
- Andre Paine, "UMG & Sony work with AI research lab SoundPatrol to protect artists from copyright infringement", MusicWeek, 2025.09.25., <https://www.musicweek.com/labels/read/umg-sony-work-with-ai-research-lab-soundpatrol-to-protect-artists-from-copyright-infringement/092731>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 언론사 콘텐츠 마켓플레이스 등장, 대형 기술 기업 최초의 AI 활용 콘텐츠 보상 사례

### 뉴스브리프

마이크로소프트는 2025년 9월 ‘퍼블리셔(언론사) 콘텐츠 마켓플레이스’ 출시에 앞서 미국 내 언론사를 대상으로 시범 프로그램을 준비 중이라고 밝혔다. 이 플랫폼은 AI 기업과 언론사를 직접 연결하는 구조로, AI 기업이 언론사로부터 콘텐츠 라이선스를 직접 구매하고, 언론사는 마이크로소프트의 중앙화된 시스템을 통해 수익을 추적할 수 있도록 설계되었다. 결제는 콘텐츠의 실제 활용 빈도와 범위에 따라 산정되는 사용량 기반 결제 모델로 진행되며, 이는 기존의 선불 라이선스 계약과 차별화된다. 또한 마이크로소프트는 해당 모델을 통해 언론사들이 제기해 온 무단 데이터 스크래핑 문제를 완화하고, 콘텐츠 활용도에 비례한 공정한 보상 체계를 구축하고자 한다. 시범 운영은 코파일럿이 첫 구매자로 참여하는 형태로 시작되며, 향후 다른 AI 제품으로의 확대가 계획되어 있다.

### 마이크로소프트, 언론사 콘텐츠 마켓플레이스 시범 운영 준비 중

#### • 사용량 기반 보상 체계의 양면 플랫폼

- 2025년 9월, 마이크로소프트(Microsoft)는 ‘퍼블리셔(언론사) 콘텐츠 마켓플레이스(Publisher Content Marketplace, PCM)’ 출시에 앞서, 미국 내 언론사를 대상으로 시범 프로그램을 준비 중이라고 밝힘<sup>1)</sup>
- 이 마켓플레이스는 AI 개발사와 언론사가 참여하는 양면 플랫폼 구조로, AI 기업이 언론사로부터 콘텐츠 라이선스를 직접 구매할 수 있도록 설계됨. 언론사는 마이크로소프트가 관리하는 중앙화된 시스템을 통해 수익을 추적할 수 있음

\* 양면 플랫폼(two-sided platform): 공급자와 수요자 등 두 집단을 연결하여 상호 간의 거래나 상호작용을 촉진하는 플랫폼 구조로, 한쪽 이용자의 참여가 다른 쪽 이용자의 가치를 높이는 네트워크 효과를 가짐

1) Sara Fischer, Kerry Flynn, "Scoop: Microsoft looks to build AI marketplace for publishers", Axios, 2025.09.23., <https://www.axios.com/2025/09/23/microsoft-ai-marketplace-publishers>

[표1] 언론사 콘텐츠 마켓플레이스의 양면 플랫폼 구조

참여자	역할	주요 활동
언론사	공급자	콘텐츠 제공, 수익 추적
마이크로소프트	운영자	중앙 시스템 관리
AI 기업	수요자	콘텐츠 라이선스 구매 및 활용

출처: Dave Ritchie, "Microsoft Unveils Publisher Marketplace for AI Content Compensation", WPN, 2025.09.23., <https://www.webpronews.com/microsoft-unveils-publisher-marketplace-for-ai-content-compensation-2/>

- 마이크로소프트는 제휴사들을 대상으로 모나코(Monaco)에서 개최된 ‘파트너 서밋(Invite-only Partner Summit)\*’을 통해 “언론사들은 자신이 보유한 지식재산권의 품질에 상응하는 보상을 받을 자격이 있다”는 메시지를 발표하며 보상 원칙을 강조함<sup>2)</sup>

\* 파트너 서밋(Invite-only Partner Summit): 마이크로소프트가 협력사 및 산업 관계자를 대상으로 신제품 발표와 사업 방향을 공유하는 연례 파트너십 행사

## 공정한 보상 체계를 위한 언론사 콘텐츠 마켓플레이스의 사용량 기반 결제 모델

### • 사용량 기반 결제 모델 도입과 선불 라이선스 계약과의 차이점

- 언론사들은 AI 학습 데이터와 관련하여 무단 데이터 스크래핑 문제를 지속적으로 제기하고 있으며, 이에, AI 기업들은 콘텐츠 접근에 대한 보상을 선지급하는 선불 라이선스 계약(prepaid license)을 주로 체결해 왔음
- 마이크로소프트는 2024년 코파일럿 데일리(Copilot Daily)를 출시하며 로이터(Reuters), 악셀 슈프링어(Axel Springer), 허스트 매거진(Hearst Magazines), USA 투데이 네트워크(USA Today Network), 파이낸셜 타임스(Financial Times) 등과 선불 라이선스 계약을 체결한 바 있음<sup>2)</sup>
- 이후, 마이크로소프트 대형 기술 기업 최초로 사용량 기반 결제 모델(usage-based payment model)을 도입함으로써 언론사들이 제기해 온 무단 스크래핑에 대한 불만에 대응하고자 함
- 이 모델은 콘텐츠가 AI 제품에 활용된 사용량을 기반으로 수익이 결정되는 구조이기 때문에 언론사는 제공한 콘텐츠에 대해 공정한 보상을 받을 수 있음
- 또한, 새로 도입된 모델은 선불 라이선스 계약과 달리 콘텐츠 활용도에 따라 지속적인 보상이 이루어지기 때문에 언론사에 동적인 수익원을 제공함

[표2] 사용량 기반 결제 모델과 선불 라이선스 계약 비교

사용량 기반 결제 모델	구분	선불 라이선스 계약
콘텐츠 사용량에 따라 지속 지급	결제 방식	계약 시 선불 지급
사용량 증가 시 수익 증가	수익 구조	고정 금액
지속적	보상 지속성	일회성

출처: Dave Ritchie, "Microsoft Unveils Publisher Marketplace for AI Content Compensation", WPN, 2025.09.23., <https://www.webpronews.com/microsoft-unveils-publisher-marketplace-for-ai-content-compensation-2/>

2) Sara Fischer, Kerry Flynn, "Scoop: Microsoft looks to build AI marketplace for publishers", Axios, 2025.09.23., <https://www.axios.com/2025/09/23/microsoft-ai-marketplace-publishers>

## 언론사 콘텐츠 마켓플레이스의 산업적 의의와 향후 과제

### • 대형 기술 기업 최초의 보상 마켓플레이스 사례

- 프로라타.ai(ProRata.ai), 톨빗(TollBit) 같은 소규모 스타트업들이 마켓플레이스 구축을 시도했으나 충분한 데이터를 확보하지 못해 언론사에 의미 있는 보상을 제공하지 못했음<sup>3)</sup>
- 이러한 가운데, 마이크로소프트는 언론사를 위한 AI 마켓플레이스를 구축한 최초의 대형 기술 기업이 됨. 미국의 정치·기술 전문 온라인 미디어 액시오스(Axios)는 이를 AI 시대에 콘텐츠 기업을 위한 지속가능한 비즈니스 모델을 구축하는 단계라고 평가함<sup>3)</sup>
- 언론사 콘텐츠 마켓플레이스를 통해 기술 대기업과의 협상에 어려움을 겪어 온 소규모 언론사도 AI 기반 수익에 접근할 수 있는 환경이 조성될 것이 기대됨

### • 언론사 콘텐츠 마켓플레이스의 향후 계획과 과제

- 마이크로소프트의 코파일럿이 언론사 콘텐츠 마켓플레이스의 첫 번째 구매자로 지정됨. 제한된 수의 미국 언론사와 함께 시범 운영을 시작한 후, 파트너와 협력하여 도구, 정책, 가격 모델을 개발할 계획임<sup>3)</sup>
- 시범 운영 결과를 바탕으로, 마이크로소프트는 향후 자사 제품을 넘어 다른 AI 기업이 참여할 수 있는 개방형 마켓플레이스로 발전시키는 방안을 모색하고 있음
- 현재 언론사 콘텐츠 마켓플레이스의 구체적인 일정이나 지급률은 공개되지 않음. 또한, 콘텐츠 사용량 산정의 기준과 국제 언론사 참여 범위 등은 해결해야 할 과제가 존재할 것으로 보임
- 미국의 비즈니스·기술 전문 온라인 미디어 웹 프로 뉴스(Web Pro News, WPN)에 따르면, 업계 관계자들은 언론사 콘텐츠 마켓플레이스의 시범 운영이 조만간 시작될 것으로 보고 있으며, 마이크로소프트가 초기 운영 결과를 바탕으로 향후 개선 방안을 마련할 것으로 예상한다고 전함<sup>4)</sup>

## 참고문헌

- Dave Ritchie, "Microsoft Unveils Publisher Marketplace for AI Content Compensation", WPN, 2025.09.23., <https://www.webpronews.com/microsoft-unveils-publisher-marketplace-for-ai-content-compensation-2/>
- Michael Hoglund, "Microsoft's new 'Publisher Content Marketplace' looks to compensate AI content creators and publishers", Windows Central, 2025.09.24., <https://www.windowscentral.com/artificial-intelligence/microsoft-new-publisher-content-marketplace>
- Sara Fischer, Kerry Flynn, "Scoop: Microsoft looks to build AI marketplace for publishers", Axios, 2025.09.23., <https://www.axios.com/2025/09/23/microsoft-ai-marketplace-publishers>

3) Sara Fischer, Kerry Flynn, "Scoop: Microsoft looks to build AI marketplace for publishers", Axios, 2025.09.23., <https://www.axios.com/2025/09/23/microsoft-ai-marketplace-publishers>

4) Dave Ritchie, "Microsoft Unveils Publisher Marketplace for AI Content Compensation", WPN, 2025.09.23., <https://www.webpronews.com/microsoft-unveils-publisher-marketplace-for-ai-content-compensation-2/>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 디지털 엔터테인먼트 환경에서의 지역별 콘텐츠 접근 제한 기술 동향

### 뉴스브리프

디지털 엔터테인먼트 환경에서 지역 차단 기술은 지역별 라이선싱 계약 준수를 위해 필수적인 수단이 되었다. 콘텐츠 제공업체는 IP 필터링, DNS 차단, 심층 패킷 검사(DPI), AI 기반 탐지 등 다양한 기술을 활용하여 지역별 접근을 제한하고 있다. IP 필터링은 머신러닝으로 연결 패턴을 분석하며, DNS 차단은 네트워크 인프라 수준에서 우회 시도를 막는다. 심층 패킷 검사는 전체 데이터 페이로드를 분석해 무단 접근을 식별하고, AI 알고리즘은 사용자의 의심스러운 활동을 탐지한다. 그러나 활성 인터넷 사용자의 25%가 VPN을 사용하며, 암호화 기술 발전으로 탐지가 어려워지는 등 기술적 한계가 존재한다. 또한 사용자는 지역 차단 기술로 인한 화질 저하, 버퍼링 증가, 개인정보 보호 위험 등의 과제에 직면하고 있다. 업계는 5G 네트워크와 블록체인 기반 스마트 계약 등 신기술을 통해 더 정밀한 지역 차단을 추구하면서도, 기업의 필요와 소비자 권리를 모두 존중하는 균형점을 모색하고 있다.

### 디지털 국경의 등장과 지역별 콘텐츠 접근 제한의 배경

#### • 디지털 콘텐츠의 지역 차단 기술 현황

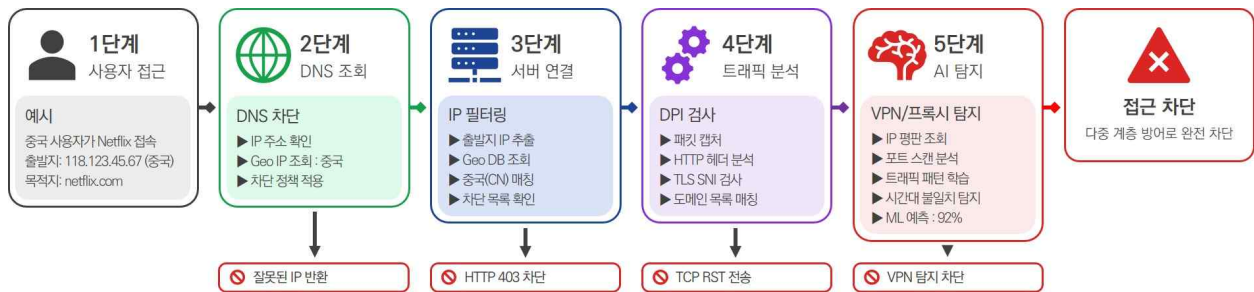
- 디지털 엔터테인먼트 환경은 국경을 넘어선 접근 제한과 함께 계속 진화하고 있음. 미디어 기업과 온라인 플랫폼은 지역 규제 및 라이선싱 계약을 준수하기 위해 지역 차단 도구를 사용하고, 디지털 경계는 지리적 위치에 따라 접근성이 다르게 적용됨
- 이로 인해, 스트리밍 서비스 이용자가 서버에 접속하면 “이 콘텐츠는 귀하의 국가에서 이용할 수 없습니다” 라는 메시지를 자주 보게 됨
- 지역 차단 기술은 지속적으로 발전하고 있는데, 현재 대부분의 시스템은 IP(Internet Protocol) 주소, 브라우저 지문 인식, 가상 사설망(Virtual Private Network, 이하 VPN) 식별을 통해 접속을 감지하고 차단함. 이는 온라인 미디어 배포 시 지역별 권리 계약을 준수하기 위해 설계됨

### • 지역별 콘텐츠 접근 제한의 필요성

- 이러한 제한은 글로벌 디지털 미디어 유통 구조에서 비롯됨. 콘텐츠 제공업체는 전 세계 이용자를 대상으로 서비스를 운영하면서도, 각 국가의 규제와 법적 요건을 동시에 충족해야 함. 이 때문에 스트리밍, 방송, 게임 등 대부분의 온라인 엔터테인먼트 플랫폼이 지역별 차단 기술을 적용하고 있음
- 특히 라이선스 계약이 이러한 구조의 중심적인 역할을 함. 권리자는 국가나 지역 단위로 콘텐츠 판매권을 나누어 독점 계약을 맺고, 이를 통해 수익을 극대화함. 따라서 플랫폼은 계약상 허용된 지역에서만 해당 콘텐츠를 제공해야 함

## 진화하는 지역 차단 기술의 작동 원리

[그림1] 다층적 지역 차단 기술의 작동 단계



출처: 참고문헌 종합하여 재구성

### • DNS 기반 지역 차단의 작동 원리

- 도메인 이름 시스템(Domain Name System, 이하 DNS) 기반 지역 차단은 네트워크 인프라 단계에서 작동하는 방식으로, 사용자의 접속 요청을 가로채거나 수정해 지역 제한을 적용함. DNS 서버는 사용자가 입력한 웹사이트 주소를 IP로 변환하는 과정에서 해당 지역이 허용 범위에 있는지를 확인함
- 허가되지 않은 지역에서 접속이 시도될 경우, DNS 확인자는 잘못된 IP 정보를 반환하거나 차단 안내 페이지로 연결하며, 이러한 방식은 네트워크 구조 자체에서 이루어지기 때문에 일반 사용자가 우회하기 어렵다는 특징이 있음

### • 다층적 IP 필터링 시스템의 고도화

- IP 주소 필터링은 지역 차단의 가장 기본적인 방식으로, 사용자가 스트리밍 콘텐츠에 접속하면 서비스는 해당 사용자의 IP 주소를 확인해 위치를 파악하고, 접근 가능한 지역인지 대조함
- 2025년 현재 이 기술은 한층 정교해졌음. 콘텐츠 제공업체는 머신러닝 알고리즘을 활용해 비정상적인 접속 패턴을 감지하고, 허용·차단 목록을 자동으로 관리함으로써 지역별 라이선스 계약에 맞춰 맞춤형 콘텐츠를 제공하고 있음
- 영국의 비티(BT), 버진 미디어(Virgin Media), 스카이(Sky), 톡톡(TalkTalk) 등 주요 인터넷 서비스 제공업체(Internet Service Provider, ISP)는 일상적으로 무단 스트리밍 서비스를 차단하고 있음. 특히 스포츠 경기 기간과 같이 이용자가 몰리는 시기에는 이러한 차단이 더욱 강화됨

• DPI 기술의 작동 구조

- 현재 스트리밍 환경에서 가장 정교한 지역 차단 기술은 심층 패킷 검사(Deep Packet Inspection, DPI)를 꼽을 수 있음. 기본 연결 정보만 확인하는 방식과 달리 DPI는 전체 데이터 페이로드를 분석함. 콘텐츠 제공업체는 이를 통해 트래픽을 세밀하게 조사할 수 있음
- DPI 시스템은 정상적인 스트리밍 트래픽과 우회 시도를 구별하는데, 이는 트래픽 패턴을 분석하여 암호화된 VPN 연결을 식별하고, 위치 정보를 속이는 프로토콜 이상 징후를 감지하는 형태임
- 이 시스템은 사용자가 고급 회피 방법을 사용해도 약 99%의 정확도로 무단 접근을 식별할 수 있으며, DPI 기술은 네트워크가 혼잡할 때 정식 비디오 패킷에 우선순위를 부여함<sup>1)</sup>

• AI 기반 VPN 및 프록시 탐지 알고리즘

- AI와 머신러닝이 발전하면서 AI 알고리즘을 통해 단 몇 초 만에 국가 간 브라우징 위치\*가 급변하는 등 의심스러운 활동을 감지할 수 있기 때문에 기업들은 인터넷 트래픽을 더 상세히 분석할 수 있게 됨
- \* 브라우징 위치: 인터넷 브라우저에서 사용자가 현재 보고 있는 웹페이지의 주소(URL) 또는 그 웹페이지의 위치
- 이러한 감지 시스템은 사용자 활동과 기기 정보를 분석하여 VPN이나 프록시 사용 패턴을 찾아내며, 머신러닝은 플랫폼이 새로운 우회 방법에 실시간으로 대응하도록 지원함
- 무단 접근 시도의 약 38%가 VPN 기술을 사용하며, 감지 시스템은 약 90%의 정확도로 VPN 사용을 식별함. 이는 권리가 요구하는 국경 간 제한을 유지하는 데 중요함<sup>1)</sup>

[표1] 지역 차단 기술의 종류별 작동 방식 및 주요 특징

기술명	작동 방식	주요 특징
IP 필터링	사용자의 IP 주소를 확인하여 지리적 위치를 파악하고 승인된 지역인지 대조함	- 머신러닝 알고리즘으로 연결 패턴의 이상을 감지 - 허용 목록과 차단 목록으로 지역별 라이선싱 계약에 따라 콘텐츠를 제공하고 비허가 지역 접속자를 차단
DNS 차단	도메인 해석 요청을 가로채고 수정하여 지역 경계를 적용함	- 네트워크 인프라 수준에서 작동하여 일반 사용자가 우회하기 어려움 - 전문 기술 지식이 없는 일반 사용자의 우회 시도를 약 95%까지 차단
심층 패킷 검사 (DPS)	데이터 페이로드를 분석하여 트래픽을 세밀하게 조사함	- 트래픽 패턴을 분석하여 암호화된 VPN 연결을 식별하고 위치 정보를 속이는 프로토콜 이상 징후를 감지 - 고급 회피 방법을 사용해도 약 99%의 정확도로 무단 접속을 식별
AI 기반 VPN 탐지	사용자 활동과 기기 정보를 분석하여 VPN이나 프록시 사용 패턴을 찾아냄	- 머신러닝으로 새로운 우회 방법에 실시간으로 대응하도록 지원 - 약 90%의 정확도로 VPN 사용을 식별

출처: XtreamWatch, "Best IPTV Providers: Unveiling the Complete Truth About Geo-Blocking in 2025", 2025.10.10. 조회, <https://xtreamwatch.com/iptv-tv-providers-unveiling-the-complete-truth-about-geo-blocking-in-2025/>

1) XtreamWatch, "Best IPTV Providers: Unveiling the Complete Truth About Geo-Blocking in 2025", 2025.10.10. 조회, <https://xtreamwatch.com/iptv-tv-provider-s-unveiling-the-complete-truth-about-geo-blocking-in-2025/>

## 현재 지역 차단 기술의 취약점 및 과제

### • 지역 차단 기술의 근본적 취약점

- 활성 인터넷 사용자의 약 25%가 개인정보 보호나 지역 제한 우회를 위해 VPN을 이용하는데, 알려진 VPN 서버를 차단하면 이용자는 다른 서버로 즉시 전환해 다시 접속하는 사례가 빈번함.<sup>2)</sup> 전용 IP는 개별 사용자처럼 보이기 때문에 차단 목록에 잘 올라가지 않음
- 스마트 DNS 방식은 주로 IPv4 환경에서 동작해 IPv4와 IPv6를 혼용하는 네트워크에서 문제를 일으키는 경우가 있으며, 일부 네트워크는 15~30초마다 IP를 빈번히 변경해 차단과 우회 양쪽 모두를 더 복잡하게 만들기도 함
- 콘텐츠 제공업체는 암호화된 트래픽을 직접 검사할 수 없어 우회 시도를 간접 신호로 탐지해야 하고, 사용자가 추가 보안 계층을 적용하면 탐지 난이도가 더 올라감. 특히 모바일 네트워크에서는 위치 추적 정확도가 낮아 정상 이용자가 오탐으로 차단되는 일이 발생함

### • 사용자가 직면하는 실질적 과제<sup>2)</sup>

- 지역 차단이 적용되면 인터넷 속도와 관계없이 화질이 제한됨. HD 스트림은 5~8Mbps, 4K 콘텐츠는 최소 20Mbps의 인터넷 속도를 요구하나, 지역 차단이 활성화되면 충분한 속도를 확보해도 화질 저하가 발생함
- 차단 지역의 사용자가 우회 방법을 사용하면 버퍼링과 지연이 늘어남. 1/10,000 미만의 작은 패킷 손실도 화면에서 픽셀화나 노이즈로 나타남. 인기 있는 생중계 때 많은 사용자가 동시에 우회 접속하면 네트워크 혼잡이 심화됨
- 한편 지역 제한을 우회하려는 사용자는 개인정보 보호 및 보안 위협에 직면함. 많은 제3자 스트리밍 어플에는 개인 데이터를 위협하는 스파이웨어나 트래커가 포함된 것으로 나타남. 연구에 따르면 접근 제어를 설정하는 기업의 49%가 설정 시 보안보다 사용자 경험을 우선시하여 취약점이 발생함

## 시사점: 지역 차단 기술의 발전 방향과 기업-소비자 간 균형

### • 지역 차단 기술 발전과 산업 전망

- 최근 산업 전문가들은 컴플라이언스 자동화와 분산형 라이선스 추적 기술을 개발하고 있음. 일부 플랫폼은 블록체인 기반 스마트 계약으로 권리 집행을 자동화하는 실험을 진행 중임. 5G 네트워크가 확산되면 더 정밀한 위치 데이터를 활용할 수 있어 지역 차단 기술이 더 정확해질 전망이다
- 유럽연합(European Union, EU)의 디지털 단일 시장(Digital Single Market) 전략은 회원국 간 콘텐츠 접근을 통일하려 함. 유럽 소비자가 여행할 때 국경을 넘어 콘텐츠를 이용할 수 있도록 하는 논의가 계속됨. 향후 지역 차단은 가상현실, 의료, 교육 등 새로운 산업 분야로 확대될 가능성이 있음

### • 산업과 소비자 경험 사이의 균형

- 국경 간 콘텐츠 접근은 디지털 콘텐츠 제공업체에 지속적인 과제임. 플랫폼은 지역별 법적 의무와 긍정적인 사용자 경험 사이에서 균형을 찾아야 함. 정교한 기술은 더 정밀한 차단을 가능하게 하지만 여행 중인 정상 사용자에게는 불편을 야기할 수 있음

<sup>2)</sup> Nexott, "IPTV TV Providers: The Truth Behind Geo-Blocking in 2025", 2025.10.10. 조회, <https://hexott.net/blog/iptv-tv-providers-the-truth-behind-geo-blocking-in-plus/>

- 일부 전문가들은 완전히 접근을 차단하는 대신 위치에 따라 다른 서비스나 콘텐츠를 제공하면서 모든 사용자에게 일정 수준의 접근을 허용하는 방안을 제안함
- 지역 차단의 미래는 기업의 요구와 소비자 권리를 모두 존중하는 지점을 찾는 데 달려 있음. 업계는 사용자 신뢰를 구축하면서 콘텐츠를 보호하는 방법을 모색하고 있음

## 참고문헌

- Advanced Television, “Cross-border content access in the digital entertainment era”, 2025.09.26., <https://www.advanced-television.com/2025/09/26/cross-border-content-access-in-the-digital-entertainment-era/>
- Charlotte Noah, “What are GEO Restrictions and How do I bypass them?”, Binge VPN, 2024.09.19., <https://bingevpn.com/blog/what-are-geo-restrictions-how-do-i-bypass-them/>
- Nexott, “IPTV TV Providers: The Truth Behind Geo-Blocking in 2025”, 2025.10.10. 조회, <https://nexott.net/blog/iptv-tv-providers-the-truth-behind-geo-blocking-in-plus/>
- XtreamWatch, “Best IPTV Providers: Unveiling the Complete Truth About Geo-Blocking in 2025”, 2025.10.10. 조회, <https://xtreamwatch.com/iptv-tv-providers-unveiling-the-complete-truth-about-geo-blocking-in-2025/>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 엔비디아, 오디오 기반 3D 얼굴 애니메이션 생성 AI 모델 오픈소스로 공개

### 뉴스브리프

2025년 9월 엔비디아가 오디오 기반 3D 얼굴 애니메이션 생성 기술인 Audio2Face의 AI 모델, SDK, 훈련 프레임워크를 MIT 라이선스로 오픈소스화하면서 게임 개발자의 기술 접근성이 대폭 확대되었다. 이 기술은 음성 입력만으로 캐릭터의 립싱크와 감정 표현을 자동 생성하며, '오토데스크 마야'와 '언리얼 엔진 5'용 플러그인 제공으로 주요 개발 환경에 즉시 통합 가능하다. 개발자는 제공된 훈련 프레임워크를 통해 자체 데이터로 맞춤형 모델을 생성할 수 있으며, 실시간 스트리밍과 사전 제작 콘텐츠 모두에 활용할 수 있다. 해당 기술은 현재 다양한 게임에 적용되어 애니메이션 작업 시간을 크게 단축한 것으로 나타났으며, 엔비디아는 디스코드 커뮤니티 운영을 통해 개발자 생태계를 구축하려는 행보를 보이고 있다.

### 엔비디아, 게임 캐릭터용 AI 얼굴 애니메이션 도구 오픈소스 전환

#### • Audio2Face 모델과 SDK 무료 공개로 개발자 접근성 확대

- 엔비디아(Nvidia)는 2025년 9월 24일 자사의 오디오 기반 3D 얼굴 애니메이션 생성 기술인 Audio2Face\*를 구성하는 AI 모델, SDK\*\*, 훈련 프레임워크, 플러그인을 오픈소스로 공개함<sup>1)</sup>

\* Audio2Face: 오디오(음성) 입력만으로 3D 캐릭터의 얼굴 애니메이션과 립싱크를 자동으로 생성하는 AI 기술로, 게임이나 영상 제작 시 캐릭터의 입 모양과 표정을 수작업으로 만들 필요 없이 음성 파일만 입력하면 자동으로 애니메이션이 생성됨

\*\* SDK(Software Development Kit): 개발자가 특정 소프트웨어나 하드웨어를 활용해 응용 프로그램을 만들 수 있도록 제공되는 도구 모음. Audio2Face SDK는 개발자가 Audio2Face 기술을 자신의 게임이나 애플리케이션에 통합할 수 있게 해주는 프로그래밍 도구와 문서를 포함함

- Audio2Face SDK는 컴퓨터나 클라우드 환경에서 얼굴 애니메이션을 제작하고 실행하는 데 필요한 라이브러리와 문서로 구성되어 있으며 C++와 Python 프로그래밍 언어로 작성된 소스 코드가 MIT 라이선스\*로 제공됨

\* MIT 라이선스: 소프트웨어를 자유롭게 사용, 복사, 수정, 배포할 수 있도록 허용하는 오픈소스 라이선스. 상업적 목적으로도 제한 없이 사용할 수 있어 기업들이 선호하는 라이선스 중 하나임

1) Emma Roth, "Nvidia is letting anyone use its AI voice animation tech", The Verge, 2025.09.26., <https://www.theverge.com/news/785981/nvidia-audio2face-ai-voice-animation-open-source>

[표 1] Audio2Face 오픈소스 패키지 구성

패키지	용도	라이선스
Audio2Face SDK	컴퓨터 또는 클라우드에서 얼굴 애니메이션 제작 및 실행을 위한 라이브러리	MIT
Autodesk Maya 플러그인	Maya 내에서 오디오 기반 얼굴 애니메이션 생성	MIT
Unreal Engine 5 플러그인	UE5 내에서 오디오 기반 얼굴 애니메이션 생성	MIT
훈련 프레임워크	자체 데이터로 Audio2Face 모델 생성	Apache
Audio2Face 모델	회귀 모델(2.2), 확산 모델(3.0)	NVIDIA Open Model License

출처: Ike Nnoli, "NVIDIA Open Sources Audio2Face Animation Model", NVIDIA developer Technical Blog, 2025.09.24., <https://developer.nvidia.com/blog/nvidia-open-sources-audio2face-animation-model/>

• **훈련 프레임워크 제공으로 맞춤형 모델 개발 환경 조성**

- 엔비디아는 Audio2Face 훈련 프레임워크를 함께 공개해 개발자가 자체 데이터를 활용하여 모델을 조정하고 특정 용도에 맞게 맞춤 설정할 수 있도록 하며, 판타지 게임에 맞춰 과장된 얼굴 애니메이션을 만드는 등의 활용이 가능함
- 훈련을 시작할 수 있도록 샘플 데이터가 함께 제공되며, 오디오 파일, 3D 모델의 얼굴 표정을 만드는 블렌드셰이프(blendshape)\* 데이터, 3D 모델의 움직임을 프레임별로 기록한 애니메이션 지오메트리 캐시(animated geometry cache) 파일, 3D 모델의 형태와 구조 정보를 담은 지오메트리 파일(geometry files), 3D 객체의 위치·회전·크기 등의 변환 정보를 담은 변환 파일 등을 활용할 수 있음

\* 블렌드셰이프(blendshape): 3D 모델의 얼굴 표정을 만들기 위해 사용되는 기술. 기본 얼굴 형태에서 웃는 얼굴, 화난 얼굴 등 다양한 표정 형태를 미리 만들어두고, 이들을 섞어가며 자연스러운 표정 변화를 표현함

• **주요 게임 엔진 플러그인 배포로 통합 장벽 완화**

- 3D 애니메이션 제작 도구인 오토데스크 마야(Autodesk Maya)와 게임 엔진 언리얼 엔진 5(Unreal Engine 5)용 플러그인이 함께 배포됨
- 개발자는 오토데스크 마야나 언리얼 엔진에서 작업하다가 오디오만 입력하면 캐릭터 얼굴 애니메이션이 바로 생성할 수 있음
- 오토데스크 마야용 플러그인은 2.0 버전으로 MIT 라이선스가 적용되며, 언리얼 엔진 5.4, 5.5, 5.6 버전에서 사용할 수 있는 2.5 버전이 제공됨. 특히 언리얼 엔진용 플러그인은 사실적인 디지털 인간을 만드는 도구인 메타휴먼(Metahuman)과 함께 쓸 수 있는 샘플이 포함됨

## 오디오 기반 실시간 얼굴 애니메이션 생성 기술 구조

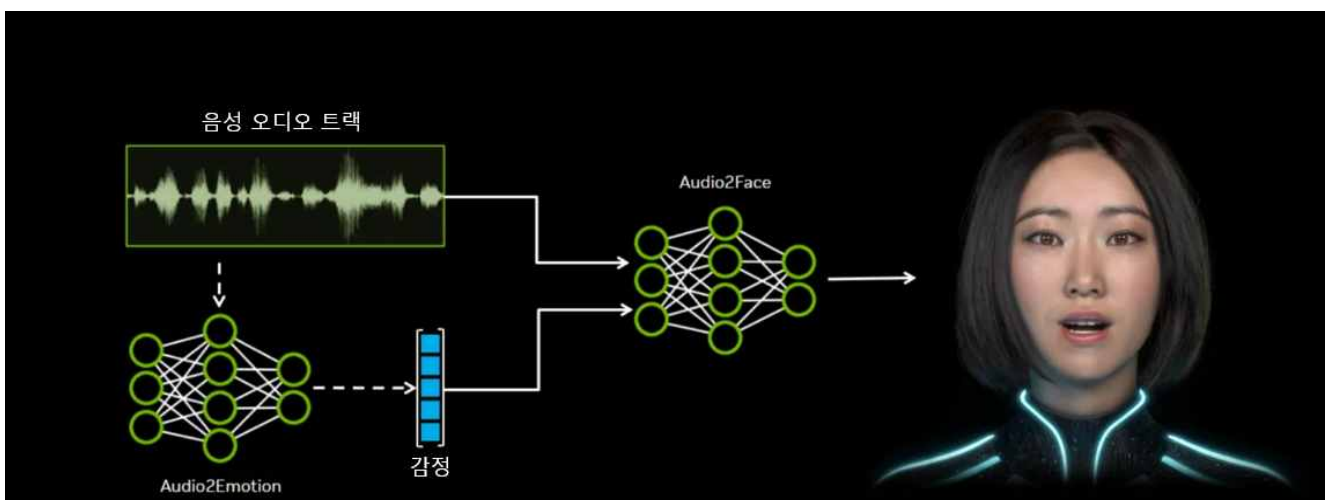
### • 음성 특징 분석을 통한 립싱크와 감정 표현 동시 구현

- Audio2Face는 음성의 특징을 분석해 애니메이션 데이터를 만들고, 이를 3D 아바타의 얼굴 표정과 입술 움직임에 적용하는 방식으로 작동함. 음성 특징에는 발음되는 소리인 음소와 말투나 억양 같은 음성 패턴이 포함됨
- 이 기술은 단순히 입 모양만 맞추는 립싱크를 넘어, 음성에 담긴 감정이나 억양을 분석해 이를 얼굴 표정으로 표현할 수 있음
- 생성된 애니메이션 데이터는 캐릭터의 얼굴 포즈에 적용되어 정확한 립싱크와 감정 표현을 전달할 수 있음
- 이는 대규모 언어 모델과 음성 모델을 활용해 사용자와 자연스러운 대화를 나눌 수 있는 지능형 3D 아바타를 만드는 생성형 AI 기술의 한 부분으로 작동함

### • 사전 제작 콘텐츠와 실시간 스트리밍 환경 모두 지원

- Audio2Face가 만든 애니메이션 데이터는 미리 준비된 대사가 있는 콘텐츠를 위해 사전에 렌더링하거나, AI가 즉석에서 대화하는 캐릭터를 위해 실시간으로 생성할 수 있음. 개발자는 게임 중간에 나오는 스토리 영상 같은 사전 제작 콘텐츠와 실시간 스트리밍용 3D 캐릭터 시스템 모두에 이 기술을 적용할 수 있음
- 실시간으로 얼굴 애니메이션과 립싱크를 제공해 생성형 AI로 움직이는 사실적인 디지털 캐릭터 제작을 빠르게 만들. 스트리밍되는 오디오를 얼굴 블렌드셰이프로 바꿔 컴퓨터나 클라우드에서 즉시 립싱크와 얼굴 애니메이션을 구현함
- 엔비디아가 제공하는 NVIDIA In-Game Inferencing(NVIGI) SDK는 게임과 애플리케이션에서 컴퓨터에 설치된 AI 모델을 실행하는 간편하고 빠른 방법을 제공함. NVIGI는 GPU, NPU, CPU 등 다양한 하드웨어에서 작동해 컴퓨터의 성능을 최대한 활용할 수 있도록 함

[그림 1] Audio2Face 작동 원리



출처: Ike Nnoli, "NVIDIA Open Sources Audio2Face Animation Model", NVIDIA developer Technical Blog, 2025.09.24., <https://developer.nvidia.com/blog/nvidia-open-sources-audio2face-animation-model/>

## 게임 개발 적용 확대와 오픈소스 생태계 구축

### • 출시 게임 적용 사례와 개발 효율화

- Audio2Face는 오픈소스화 이전부터 F1 25(Codemasters), Alien: Rogue Incursion Evolved Edition(Survios), Chernobylite 2: Exclusion Zone(The Farm 51) 등 다수의 게임에 적용되어 왔음
- The Farm 51의 크리에이티브 디렉터 Wojciech Pazdur는 Audio2Face 통합으로 수많은 시간의 애니메이션 작업을 절약했으며, 원작에서 불가능했던 아이디어가 가능해졌다고 밝힘<sup>2)</sup>
- 오픈소스 기술 공개는 개발자, 학생, 연구자가 최신 코드를 학습하고 그 위에 새로운 기능을 구축할 수 있게 하며, 커뮤니티가 새로운 기능을 추가하고 다양한 용도에 맞게 기술을 최적화하는 피드백 루프가 형성됨
- Audio2Face SDK와 플러그인은 MIT 라이선스로, 훈련 프레임워크는 Apache 라이선스로 제공되어 상업적 활용이 자유로움
- 엔비디아는 디스코드(Discord)에서 Audio2Face 개발자 커뮤니티를 운영하며, 커뮤니티 피드백을 통해 기술이 다양한 용도로 개선되는 선순환 구조를 만들고자 함

## 참고문헌

- Emma Roth, "Nvidia is letting anyone use its AI voice animation tech", The Verge, 2025.09.26., <https://www.theverge.com/news/785981/nvidia-audio2face-ai-voice-animation-open-source>
- Ike Nnoli, "NVIDIA Open Sources Audio2Face Animation Model", NVIDIA Developer Technical Blog, 2025.09.24., <https://developer.nvidia.com/blog/nvidia-open-sources-audio2face-animation-model/>
- NVIDIA Developer ACE for Games, "NVIDIA ACE for Games", NVIDIA Developer, 2025.10.10.조회, <https://developer.nvidia.com/ace-for-games#section-get-started-with-nvidia-ace>
- Kosta Andreadis, "NVIDIA's AI-powered Audio2Face animation model is now open source for all developers", TweakTown, 2025.09.29., <https://www.tweaktown.com/news/107951>

2) Ike Nnoli, "NVIDIA Open Sources Audio2Face Animation Model", NVIDIA developer Technical Blog, 2025.09.24., <https://developer.nvidia.com/blog/nvidia-open-sources-audio2face-animation-model/>



# 저작권 이슈 브리프

SUMMARY

산업/기업

기술

## 주간 기술 동향

### 변형 시뮬레이션형 텍스트-투- 비디오 기술

- 시간의 흐름에 따른 변화와 상호작용을 구현하는 변형 시뮬레이션형 텍스트-투-비디오 기술<sup>1)</sup>

최근 영상 제작 산업에서는 픽셀 조합을 통해 시각적으로 유사한 영상을 만드는 것을 넘어, 시간의 흐름에 따른 변화나 물리 법칙이 적용된 현상을 구현하려는 시도가 본격화되고 있다. 기존의 텍스트나 이미지로부터 동영상을 생성하는 AI 기술이 기존 창작 영역을 넘어 물리적 세계를 이해하고 모방하는 단계로 진입하는 있는 것이라고 볼 수 있다. 현실 시계에서 마치 꽃이 피거나 건물이 시간의 흐름에 따라 변화하는 모습처럼, 기존의 AI 모델들이 구현하기 어려웠던 '시간에 따른 변화'를 구현하는 기술이 개발되고 있다.

기존의 텍스트-투-비디오 모델들은 대부분 단일 장면에서 제한된 움직임만 표현할 수 있었으며, 이는 기존 이미지 생성 모델의 구조를 단순히 시간 축으로 확장하는 방식에 의존했기 때문이다. 이러한 접근은 장면 간의 연속적인 변화나 객체 간의 복잡한 상호작용을 구현하기 어렵게 만들었고, 결과적으로 물리적 세계의 자연스러운 변화를 반영하지 못하는 한계를 드러냈다.

이러한 한계를 극복하기 위해 최근 개발되는 기술들은 '변형 시뮬레이터'라는 개념을 지향하며, 현실 세계의 물리 법칙과 인과관계를 데이터로부터 학습하여 시간의 흐름에 따라 사물이 변화하거나 상호작용하는 과정을 영상으로 재현하는 데 초점을 맞추고 있다. 이는 AI가 단순한 콘텐츠 제작 도구를 넘어, 현실의 변화를 이해하고 표현하는 기술로 확장되고 있음을 보여준다. 실제 타임랩스 영상 데이터셋을 학습하여 자연의 변화 과정을 모방하거나, 방대한 비디오 데이터를 통해 물리적 상호작용을 내재화하려는 시도가 이를 뒷받침한다.

본 보고서에서는 이러한 시뮬레이션형 비디오 생성 기술의 사례를 분석하고자 시간의 흐름에 따른 점진적 변화, 즉 타임랩스 영상 생성에 특화된 '매직타임' 기술을 알아보도록 하겠다. 이를 통해 생성형 AI가 단순한 영상 제작을 넘어, 현실 세계의 변화를 사실적으로 구현하는 방향으로 발전하고 있음을 살펴보고자 한다.

1) Shy Cohen, "New AI text-to-video system masters the art of time-lapse video generation", AOL, 2025.09.25., <https://www.aol.com/articles/ai-text-to-video-system-masters-180700440.html>

## 기존 텍스트-투-비디오 기술의 한계와 세계 모델의 등장

### • 시뮬레이션형 텍스트-투-비디오 기술 특징

- 변형 시뮬레이터 개념이 적용되지 않은 기존 텍스트-투-비디오(Text-to-Video)\* 기술은, 대부분 2D 이미지 생성 모델을 시간 축으로 확장하는 방식을 택함. 이는 사전 학습된 이미지 생성 모델의 공간적 특징 추출 능력은 활용하되, 시간적 변화를 표현하기 위한 별도의 모듈을 추가하는 구조로 이루어짐

\* 텍스트-투-비디오(Text-to-Video): 사용자가 입력한 문장이나 단어(텍스트)를 기반으로, 인공지능이 그 내용을 시각적으로 해석해 동영상을 자동 생성하는 기술

- 이러한 접근 방식은 객체의 외형이나 배경의 시각적 품질은 우수하게 생성할 수 있으나, 물리적 세계의 동적인 변화를 담아내는 데에는 명백한 한계를 보이며, 생성된 영상은 대부분 단순한 카메라 움직임이나 제한된 공간 내에서의 미미한 움직임을 표현하는 수준에 그치는 경우가 많았음

### • 물리 및 논리적 세계 구현의 한계

- 이러한 한계를 극복하기 위한 대안으로 제시된 것이 '세계 모델(World Model)\*'이라는 접근법임. 이는 기존 영상 생성 기술처럼 픽셀 단위의 시각적 유사성에만 집중하는 것을 넘어, 물리 법칙과 시공간적 인과관계를 모델 내부에 내재화하여 현실의 동적인 변화 자체를 시뮬레이션하는 것이 중점임

\* 세계 모델(World Model): 인공지능이 외부 세계의 물리 법칙, 인과관계, 시공간적 변화를 내적으로 학습하여 스스로 환경을 예측·시뮬레이션할 수 있도록 하는 개념으로, 단순한 데이터 재현을 넘어 현실의 동적인 변화를 모사하기 위한 AI 모델 구조

- 이러한 기술적 지향점은 생성형 AI의 역할을 단순한 콘텐츠 제작 도구에서 한 단계 더 발전시키는 것을 의미함

## 사례 탐구: 물리적 변화 시뮬레이션에 특화된 비디오 생성 모델 사례, '매직타임'

### • 기본 개념인 '변형 시뮬레이터'<sup>2)</sup>

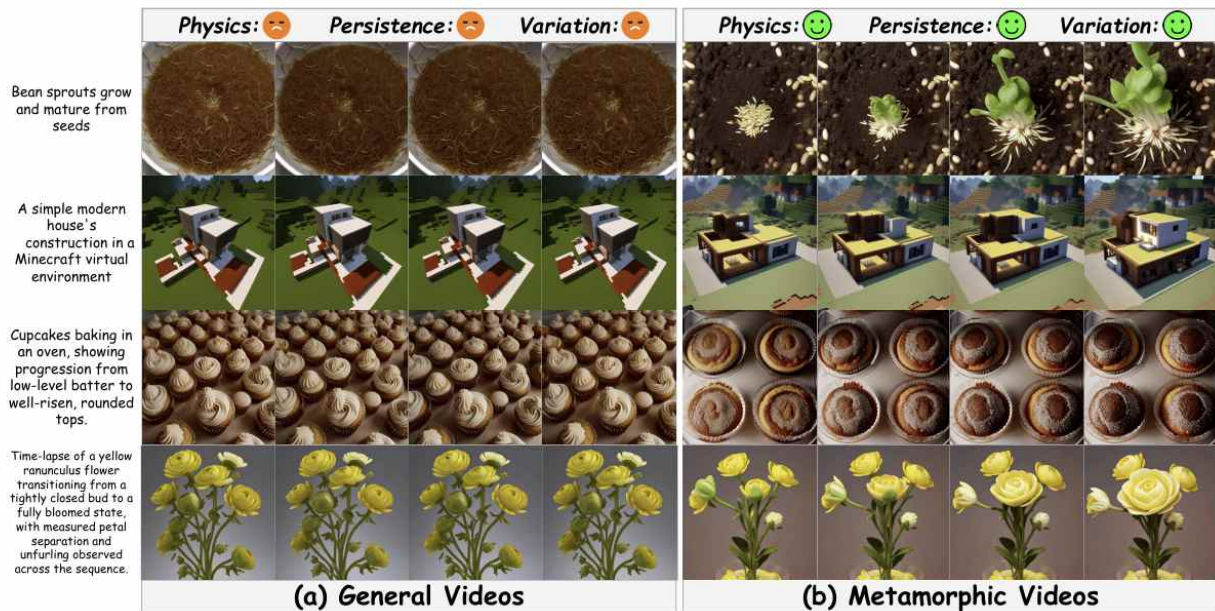
- 매직타임(MagicTime)은 기존 텍스트-투-비디오 모델이 단일 장면의 제한된 움직임만 묘사하고 물리적 세계의 동적인 변화를 담지 못한다는 문제 인식에서 출발한 기술임. 단순한 영상 생성을 넘어, 시간의 흐름에 따른 객체의 자연스러운 변형 과정을 시뮬레이션하는 것을 목표로 설정함
- 이를 위해 기술의 근간을 실제 세계의 물리적 변화를 압축적으로 담고 있는 '타임랩스(Time-lapse)' 영상 데이터에 둬, 꽃이 피거나, 빵이 부풀어 오르는 것과 같은 실제 데이터를 학습함으로써, 모델이 변화의 내재적 패턴과 물리 법칙을 이해하도록 유도하는 방식을 채택함
- 이러한 접근법을 통해 매직타임은 스스로를 '변형 시뮬레이터(Metamorphic Simulator)'로 정의하며, 텍스트 프롬프트에 따라 특정 객체나 환경이 시간에 따라 어떻게 변화할지를 예측하고 시각화하는 능력을 갖추게 됨. 이는 AI가 현실 세계의 지식을 내재화하는 하나의 경로를 제시한 것임
- 이 기술의 개념은 고품질의 실제 타임랩스 녹화 영상을 학습 데이터로 활용하여, 기존 AI가 구현하지 못했던 자연스럽고 역동적인 변형 과정을 영상으로 재현하는 데 있음. 이는 픽셀을 그리는 것을 넘어, 변화의 '과정' 자체를 생성하는 데 중점을 둔 것임

2) Shenghai Yuan 외 8인, "MagicTime: Time-lapse Video Generation Models as Metamorphic Simulators", arXiv, 2025.04.06., <https://arxiv.org/html/2404.05014v2>

• 주요 기술 요소 및 작동 방식

- 전체 비디오 프레임을 균일하게 학습하는 대신, 변화가 가장 두드러지는 순간들을 동적으로 식별하여 해당 부분에 학습을 집중시키는 방식으로 작동하며, 이 방식은 모델이 느리면서도 중요한 변화의 패턴을 포착하여 시뮬레이션 정확도를 높임
  - 사전에 학습된 대규모 텍스트-투-비디오 모델을 기반으로, 타임랩스 데이터로부터 추출한 '변화의 패턴'을 주입하는 적응형 미세조정 과정을 거치는 것이 특징임. 이 접근법은 기존 모델이 가진 영상 생성 능력을 유지하면서, 물리적 변화를 시뮬레이션하는 새로운 능력을 추가하는 역할을 수행함
  - 기존 캡셔닝 모델\*이 변형 과정의 세부 사항을 포착하지 못하는 한계를 극복하고자, GPT-4V와 같은 고성능 대규모 멀티모달 모델을 활용하여 학습 데이터의 품질을 높임. 이를 통해 영상의 시간적 흐름과 객체의 세부사항까지 묘사하는 텍스트 프롬프트를 생성하여 모델의 이해도를 높임
- \* 캡셔닝 모델(Captioning Model): 이미지나 동영상과 같은 시각 데이터를 인공지능이 분석하여, 그 안에 담긴 객체, 상황, 배경 등을 자연스러운 문장(텍스트)으로 자동 생성 및 요약하는 기술
- 영상의 여러 프레임에서 개별적으로 생성된 텍스트 캡션들을 자연스럽게 연결하고 통합하여, 전체 변형 과정을 하나의 일관된 서사로 재구성하는 기술을 적용함. 이 텍스트 융합 과정은 최종적으로 생성되는 영상의 서사적 완성도에 영향을 미침

[그림 1] 일반 영상과 매직타임의 타임랩스 영상 차이



출처: Shenghai Yuan 외 8인, "MagicTime: Time-lapse Video Generation Models as Metamorphic Simulators", arXiv, 2025.04.06., <https://arxiv.org/html/2404.05014v2>

• 성능 및 산출물의 특징<sup>3)</sup>

- 매직타임은 '꽃이 피어나는 모습', '오븐 속 빵이 부풀어 오르는 과정' 등 점진적이고 유기적인 변화를 담은 타임랩스 영상 생성에서 높은 수준의 성능을 보임. 이는 기존 모델들이 생성하던 경직된 움직임과 차별화됨

3) Shenghai Yuan 외 8인, "MagicTime: Time-lapse Video Generation Models as Metamorphic Simulators", arXiv, 2025.04.06., <https://arxiv.org/html/2404.05014v2>

- 현재 공개된 오픈소스 버전의 경우, 512x512 픽셀 해상도에서 초당 8프레임으로 재생되는 2초 길이의 클립을 생성할 수 있음. 아키텍처를 개선한 버전은 최대 10초까지 영상 길이를 확장할 수 있어, 보다 긴 변화 과정을 담아낼 가능성을 보여줌
- 또한 매직타임이 생성한 짧은 영상 클립들은 중간 장면을 채워주는 별도의 AI 프레임 보간(Frame Interpolation, 인터플레이션)\* 모델을 통해 부드럽게 연결할 수 있음. 이 방식은 마치 애니메이션의 핵심 장면들 사이를 중간 그림으로 채워 하나의 긴 영상을 만드는 것과 같은 원리임. 이를 통해 짧은 단위의 시뮬레이션 결과를 연결하여 장기적인 변화 과정을 표현할 수 있음

\* AI 프레임 보간(Frame Interpolation, 인터플레이션): 인접한 영상 프레임들 사이의 움직임을 추정해(예: 옵티컬 플로우, 딥러닝 기반 모션 추정) 중간 프레임을 생성하는 기술. 재생 시간을 늘리거나(슬로모션), 프레임레이트를 높여 부드러운 영상을 만드는 데 사용되며, 해상도를 올리는 업스케일링과는 목적과 처리 대상이 다름

## 기술적 시사점

### • 패러다임의 전환과 기술적 의의

- 매직타임과 같은 사례는 텍스트-투-비디오 기술의 발전 방향이 단순한 시각적 재현을 넘어, 실제 세계의 물리적 지식을 학습하고 시간에 따른 변화 과정을 시뮬레이션하는 단계로 나아가고 있음을 보여준다는 점에서 중요한 의미를 가짐
- 이 기술은 향후 과학 연구의 반복 실험 주기를 단축하고 아이디어 탐색을 가속화하는 시뮬레이션 도구로 발전할 잠재력을 지녔으며, 각종 분야의 실험을 직접 수행하기 전 가상으로 결과를 예측해 보는 연구 보조 도구로서의 활용이 기대됨

### • 한계점 및 향후 과제

- 현재 모델의 성능은 학습에 사용된 데이터의 규모가 제한적이어서 모든 범주의 변형 영상을 생성하지는 못함. 이는 향후 시뮬레이션 기술의 범용성 확보를 위해 다양한 데이터셋 구축이 핵심 과제를 시사함
- 또한 현재 공개된 버전이 생성하는 영상의 해상도(512x512)와 길이(최대 10초)는 상업적 콘텐츠 제작 등에 활용되기에는 여전히 제한적인 수준임. 따라서 고해상도, 장편 영상 생성을 위한 모델의 확장성과 연산 효율성 개선이 실용화의 핵심 관건이 될 것임

## 참고문헌

- Shy Cohen, "New AI text-to-video system masters the art of time-lapse video generation", AOL, 2025.09.25., <https://www.aol.com/articles/ai-text-video-system-masters-180700440.html>
- Shenghai Yuan 외 8인, "MagicTime: Time-lapse Video Generation Models as Metamorphic Simulators", arXiv, 2025.04.06., <https://arxiv.org/html/2404.05014v2>