

〈OpenChain-KWVG T&L SG 22차 회의〉

AI 기술의 미국 저작권 동향

– 생성형 AI 학습의 Fair Use 적용 가능성 –

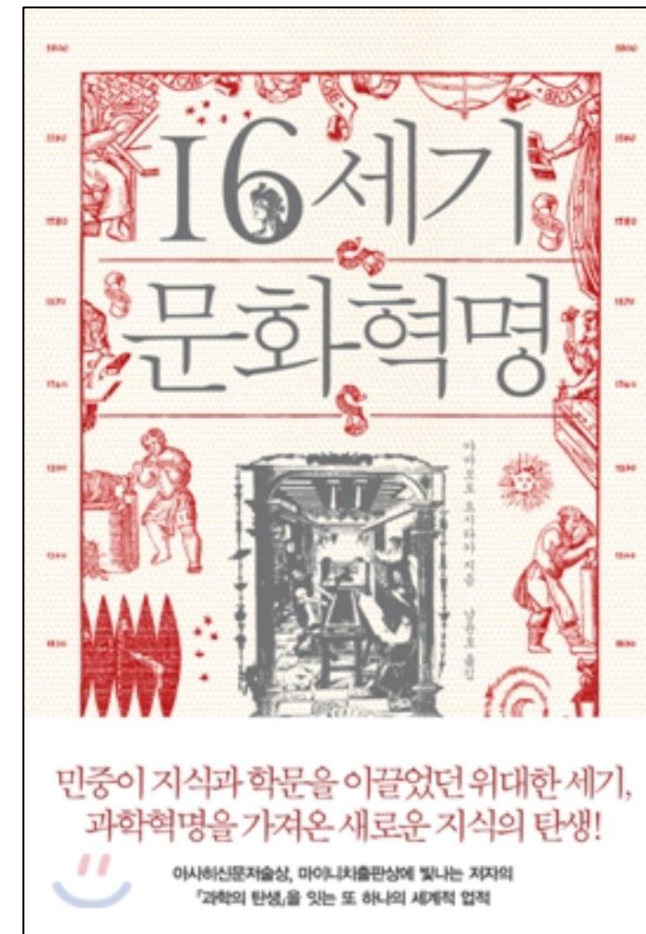
(Copyright and AI Report Part III 내용 요약)

2025. 6. 26.(목)

ETRI 오픈소스센터
박정숙

내용

- 보고서 개요
- 미국 저작권청장의 해임 및 그 영향
- 보고서 구조
- 보고서 분석
- 결론



보고서 개요

• Copyright and Artificial Intelligence

➢ ('23 초) AI와 저작권의 접점을 탐구하는 광범위한 계획 발표

➢ ('23.3) Part 1: Digital Replicas

❖ 디지털 기술을 사용하여 개인의 목소리나 외모를 사실적으로 복제하는 디지털 복제

➢ ('25.1) Part 2: Copyrightability

❖ AI 생성 콘텐츠가 포함된 작품의 저작권 등록에 관한 정책 지침 발표

❖ 1. 현행법의 적용

- ✓ 현재의 저작권법으로도 AI 관련 저작권 문제를 해결할 수 있음
- ✓ 별도의 법 개정은 필요하지 않음

❖ 2. AI 도구사용과 저작권

- ✓ AI를 도구로 사용하는 것은 저작권 보호에 영향을 미치지 않음
- ✓ 인간의 창의성을 보조하는 수단으로서의 AI 활용 인정

❖ 3. 인간 저작물의 보호

- ✓ AI 요소가 포함되어 있더라도 인간이 창작한 표현은 저작권으로 보호

❖ 4. 순수 AI 생성물

- ✓ AI가 독자적으로 생성하거나 인간의 통제가 불충분한 결과물은 보호되지 않음

❖ 현재 기술 수준에서 프롬프트만으로는 저작권 보호를 받기 어렵다고 명시: 프롬프트는 본질적인 아이디어나 지시사항의 성격을 가짐, AI시스템의 결과물에 대한 인간의 실질적 통제가 제한적임, 동일한 프롬프트도 매번 다른 결과물이 생성될 수 있음

➢ ('25.5.10) Part 3: Generative AI Training

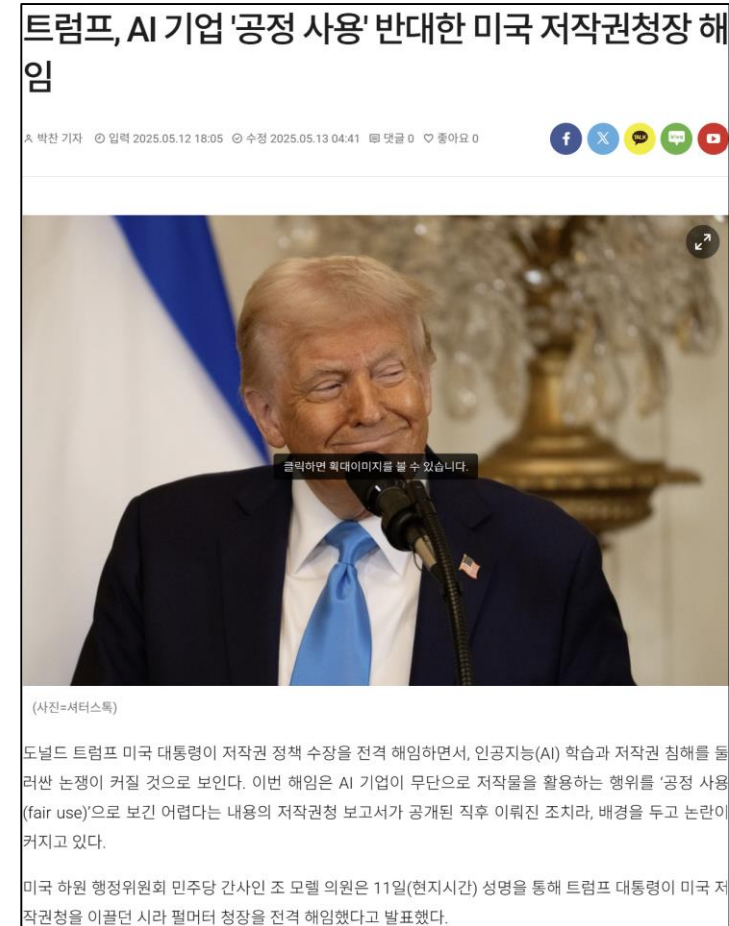
➢ ('25.5.11) 저작권청장 전격 해임

구분	주요내용
적용 대상	• 실제와 구별하기 어려운 디지털 레플리카
보호 대상	• 유명세나 상업적 노출 여부와 관계없이 모든 개인을 포함
보호 기간	• 최소한 개인의 생애 동안 지속
침해 행위	<ul style="list-style-type: none"> • 무단 디지털 레플리카의 배포 또는 제공에 의해 책임이 발생할 수 있지만 생성 자체만으로 책임이 발생해서는 안 됨 • 또한 그 책임은 해당 표현이 ① 특정 개인의 디지털 레플리카라는 사실과, ② 무단 복제물이라는 사실에 대한 실질적 인식(actual knowledge)을 요구
간접책임	<ul style="list-style-type: none"> • 간접책임의 전통적인 불법행위 원칙이 적용되어야 함 • 온라인서비스제공자에게는 안전지대(safe harbor) 체계를 포함시켜야 함
라이선스 및 양도	<ul style="list-style-type: none"> • 개인은 일정한 가이드라인 하에서 자신의 디지털 레플리카 권리를 라이선스 하고 수익을 창출할 수 있도록 해야 하지만 이를 완전히 양도할 수는 없도록 해야 함 • 미성년자는 추가적인 안전장치를 마련해야 함
수정헌법 제1조	• 언론의 자유 문제는 법에서 명시적으로 다루어야 함
구제책	• 금지명령과 급진적 손해배상을 포함한 효과적인 구제책이 제공되어야 함
주법과의 관계	<ul style="list-style-type: none"> • 연방법은 전국적으로 일관된 보호의 기반을 제공해야 하며, 주에서는 계속해서 추가적인 보호를 제공할 수 있어야 함 • 미국 저작권법 제114조(b)가 무단 음성 디지털 레플리카를 제한하는 법률에 우선하거나 충돌하지 않는다는 점을 명확히 해야 함

한국지식재산권('24.10.)

미국 저작권청장의 해임 및 그 영향

- 미 저작권청 “Copyright and AI Part 3(Generative AI Training) 발표(’25.5.10)
 - AI 기업이 무단으로 저작물을 활용하는 행위를 ‘공정 사용(fair use)’으로 보긴 어렵다는 내용
 - AI 기업들이 주장하는 ‘공정 사용(fair use)’이 한계가 있다고 지적
 - “연구나 분석 목적은 공정 사용에 해당할 수 있지만, 시장에서 기존 저작물과 경쟁하는 콘텐츠를 대규모 상업적 생성에 사용하는 것은 공정 사용 범위를 넘어선다”라고 명시
 - 정부 개입은 시기상조, AI 기업이 저작권자에게 콘텐츠 이용 대가를 지불하는 ‘라이선스 시장’의 성장을 기대하며, 시장 실패에 대비해 ‘집단 라이선스 체계’ 등 대안을 검토할 필요가 있다고 제안
 - ‘25년 여름에 정식 버전 공개 예정
- 도널드 트럼프 미국 대통령이 저작권청장 쉐라 펄머터 전격 해임(’25.5.11)
 - 펄머터 청장은 ‘20년 트럼프 1기 행정부 당시 카를라 헤이든 국회도서관장의 지명으로 청장에 임명
 - Part 3 보고서는 향후 미국의 저작권 분쟁의 가이드라인 역할을 하게 될 중요한 잣대
 - 보고서에 AI 기업에 불리한 내용이 포함됐다는 것이 알려지며 미국 정부가 저지에 나섰으며, 자신이 해임될 것을 안 펄머터 청장이 앞당겨 이를 공개했다는 추측까지...
 - 현재 OpenAI를 포함한 여러 AI 기업은 저작권 침해 소송에 직면해 있으며, 이들은 정부에 공정 사용 범위 확대를 통한 법제화 요구
- 출처 : AI타임스(<https://www.aitimes.com>)



보고서 구조

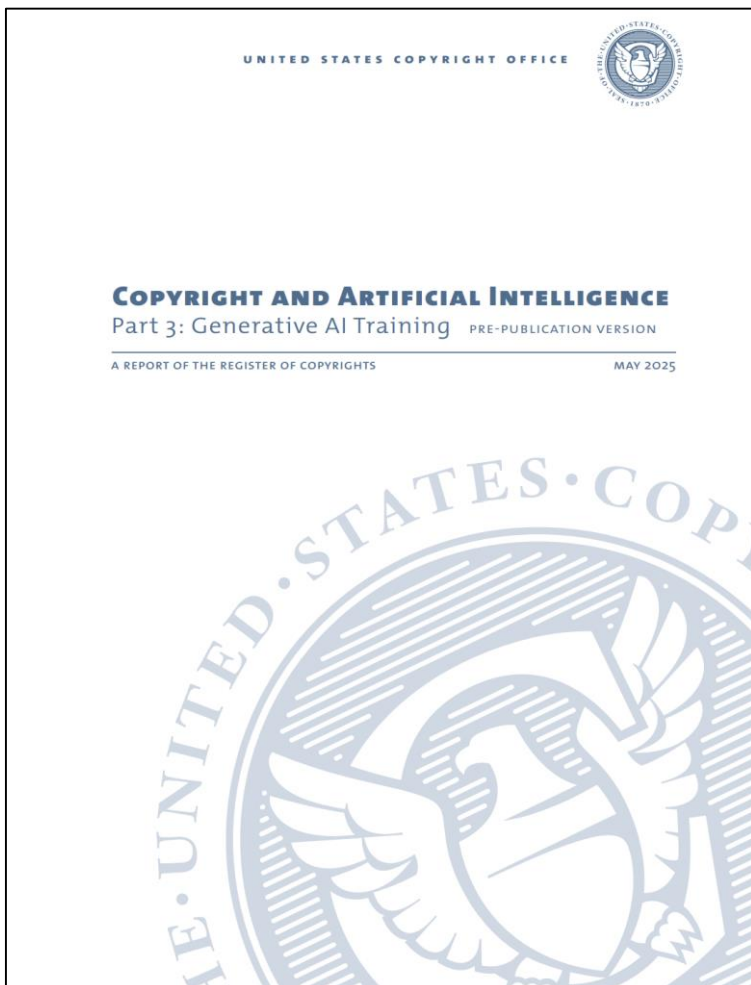


TABLE OF CONTENTS	
I. INTRODUCTION.....	1
II. TECHNICAL BACKGROUND.....	4
A. Machine Learning.....	4
B. Generative Language Models.....	6
C. Training Data.....	9
1. Data Characteristics.....	9
2. Acquisition and Curation.....	13
D. Training.....	17
1. Training Phases.....	17
2. Memorization.....	19
E. Deployment.....	21
III. PRIMA FACIE INFRINGEMENT.....	26
A. Data Collection and Curation.....	26
B. Training.....	27
C. RAG.....	30
D. Outputs.....	31
IV. FAIR USE.....	32
A. Factor One.....	35
1. Identifying the Use.....	36
2. Transformativeness.....	37
3. Commerciality.....	48
4. Unlawful Access.....	51
B. Factor Two.....	53
C. Factor Three.....	54
1. The Amount Used.....	55
2. Reasonableness in Light of Purpose.....	55
3. The Amount Made Available to the Public.....	57
D. Factor Four.....	61

1. Lost Sales.....	62
2. Market Dilution.....	64
3. Lost Licensing Opportunities.....	66
4. Public Benefits.....	71
E. Weighing the Factors.....	74
F. Competition Among Developers.....	74
G. International Approaches.....	76
V. LICENSING FOR AI TRAINING.....	85
A. Voluntary Licensing.....	85
1. Feasibility of Voluntary Licensing.....	86
2. Ability to Provide Meaningful Compensation.....	92
3. Possible Legal Impediments to Collective Licensing.....	94
B. Statutory Approaches.....	95
1. Compulsory Licensing.....	95
2. Extended Collective Licensing.....	99
3. Opting Out.....	101
C. Analysis and Recommendations.....	103
VI. CONCLUSION.....	107

보고서 - 서론

- 저작권의 공정이용 원칙 적용에 초점을 맞춘 수십 건의 소송 계류 중
 - Getty Image v. Stability AI 등
- ('23.8) 저작권청이 질의 통지서(NOI) 발표
 - AI 모델 학습과 관련한 기술적 질문
 - 공정 이용 원칙 적용에 대한 법적 질문
 - 기존 또는 잠재적 라이선스 계약에 대한 사실 관계 질문
- 10,000건 이상의 의견이 접수됨
- 보고서 내용
 - 생성형 AI 시스템 개발에 있어 저작권 있는 저작물 활용에 대한 검토
 - 특정 사건의 법적 판단 대신, 관련 사실과 정책 고려사항을 식별가능한 분석 틀 제시

보고서 - 기술적 배경

- 기계 학습
- 생성 언어 모델
- 데이터 학습
- 학습
- 배포

보고서 - 일견상 침해(Prima Facie Infringement)

- 일견상 침해
 - 특정 주장이나 주장이 성립하기 위한 최소한의 입증 요건이 충족되었음을 의미
- 저작권법은 저작권자에게 자신의 작품을 복제, 배포, 공개적으로 공연, 공개적으로 전시할 수 있는 독점적 권리와 파생 작품을 준비할 수 있는 권리 부여
- 침해의 요소
 - 유효한 저작권의 소유
 - 작품의 독창적인 구성 요소의 복사
- 저작권으로 보호되는 자료를 사용하여 생성 AI 시스템을 만들고 배포할 때 라이선스 또는 기타 방어 수단이 없는 경우 하나 이상의 권리를 침해할 수 있는 여러 행위 필요
- 침해 발생 단계
 - 데이터 수집 및 선별
 - 학습
 - RAG(Retrieval-Augmented Generation)
 - 출력물

공정 이용(Fair Use)

- 공정 이용

- 1976년 저작권법 제107조에 명문화된 판결 원칙
- 특정 사용이 공정한지 판단할 때 고려하는 비배타적 요소
 - ❖ 사용의 목적 및 특성, 즉 상업적 성격인지 비영리 교육 목적인지 여부
 - ❖ 원 저작물의 성격
 - ❖ 저작물 전체에 비해 사용된 부분의 양과 중요성
 - ❖ 저작물의 사용이 잠재적 시장 또는 가치에 미치는 영향

- 공정이용 요소별 분석

- 요인1: 용도 식별, 변형성, 상업성, 불법적인 접근
- 요인2
- 요인3: 사용된 양, 목적에 따른 합리성, 대중에게 공개된 금액
- 요인4: 판매손실, 시장 희석, 라이선싱 기회 손실, 공공 혜택
- 요인 평가
- 개발자 간의 경쟁
- 타 국가들의 접근 방법

공정 이용(Fair Use) – 요인 1(이용 목적 및 성격)

- 침해 혐의가 있는 사용이 추가적 목적이거나 다른 성격을 가지는지 여부에 초점
- 용도 식별
 - AI가 사용한 방식이 연구인지, 상업적 서비스인지, 단순 참조인지에 따라 달라짐
 - 이용 유형: 학습 자체, 검색 기반 생성, 결과물의 생성 및 배포, 미세조정이나 모델 공유
- 변형성
 - 원 저작물을 새로운 목적이거나 의미로 사용했는지의 여부
 - AI 학습에 대한 견해는 변형적이라는 주장과 변형적이지 않다는 주장이 공존
 - 예) Warhol 판결: 새로운 표현이나 의미만으로는 변형성이 인정 안 됨
- 상업성
 - 상업성의 정도와 공익성의 균형을 따짐
 - AI 학습이 R&D의 일환이므로 공정이용에 해당한다는 의견도 있음
- 불법적인 접근
 - 불법적 접근시(접근 제한된 콘텐츠 사용 등) 공정이용 성립 안 됨

공정 이용(Fair Use) – 요인 2(원 저작물의 성격)

- 사실 기반 콘텐츠보다는 창의적 표현물에 더 강한 보호 부여)
 - 백과사전, 뉴스 기사, 공공기록 등 사실적·정보 중심 저작물은 공정이용에 가까움
 - 시, 소설, 미술작품 등의 창작적인 저작물은 공정이용 판단에 불리
- 저작물이 공개된 것인지 미공개 작품인지의 여부도 고려됨
- 저작권이 있는 작품의 특성은 "일부 작품이 다른 작품보다 의도된 저작권 보호의 핵심에 더 가깝다는 점을 인식해야 함"
 - 소설, 영화, 미술, 음악과 같이 더 창의적이거나 표현력이 풍부한 작품을 사용하는 경우 사실적이거나 기능적인 작품(컴퓨터 코드 등)을 사용하는 경우보다 공정 사용으로 간주될 가능성이 낮음
 - 작품의 미공개 특성도 공정 사용 결정에 영향을 미칠 수 있음
- 훈련 데이터 세트에는 창의성이 덜하거나 보호받을 수 없는 자료가 포함되어 있더라도 일반적으로 표현적 작품이 포함되기 때문에 공정 사용에 불리하게 작용하는 경우가 많다고 주장
 - 데이터 세트에 미공개 작품이 포함될 수 있다고 지적했지만 대부분의 작품은 이미 출판되었을 것이므로 공정 사용 주장을 적당히 뒷받침
 - 여러 사람은 두 번째 요인이 전반적인 공정 사용 균형에 실질적인 역할을 거의 하지 않는다고 지적
- 생성 AI 모델은 표현적 저작물과 기능적 저작물, 그리고 출판된 저작물과 출판되지 않은 저작물 모두에 대해 정기적으로 학습되므로, 사실 관계는 모델과 해당 저작물에 따라 달라짐
 - 언어 모델은 소설처럼 매우 창의적인 저작물뿐만 아니라 컴퓨터 코드나 학술 논문처럼 사실적이거나 기능적인 내용을 더 많이 담고 있는 저작물에도 학습되는 경우가 많음
 - 관련 저작물이 표현적이거나 이전에 출판되지 않은 저작물인 경우, 두 번째 요인은 공정 사용에 불리하게 작용

공정 이용(Fair Use) – 요인 3(이용된 양과 중요성)

- 사용된 양
- 목적에 따른 합리성
 - 사용된 데이터가 목적에 비추어 필수적이고 합리적인 범위 내에 있는지의 여부
- 대중에게 공개된 범위
 - 원작의 어느 정도를 재현하거나 공개적으로 노출시키는지
 - 특정 작품의 핵심 문장을 그대로 재현하거나 사실상 복제본에 가까운 출력물을 생성한다면 저작권 침해로 간주될 수 있음

공정 이용(Fair Use) – 요인 4(시장에 미치는 영향)

• 판매 손실

- AI가 생성한 콘텐츠가 원저작물과 직접적으로 경쟁하거나 대체 효과를 일으키는 경우, 저작권자의 판매 손실 발생 가능
- 예: AI가 소설 작가의 문체를 흉내내 완성된 문학작품 생성, 기존 아티스트의 스타일을 복제한 이미지 생성

• 시장 희석

- AI 시스템이 다양한 분야의 창작자들로부터 콘텐츠를 대량 수집하고 유사 출력을 생성함으로써, 콘텐츠의 희소성과 경제적 가치 감소
- 창작자들의 장기적 생계에 영향을 미치며, 새로운 콘텐츠 창작 유인을 약화시키는 결과 초래

• 라이선싱 기회 손실

- AI 학습이 사전 라이선스를 통해 가능했던 시장을 우회했다면, 이는 저작권자의 수익화 기회 상실로 평가
- 일부 개발자들은 라이선스를 구매하거나 협력사를 통해 고품질 데이터를 수집하고 있지만, 동시에 많은 기업들이 비공식 또는 무단 데이터를 이용해 경쟁 우위를 점하고 있음

• 공공 혜택

- AI 시스템이 교육, 과학, 연구 등 공공 목적에 기여한다면, 시장 영향의 일부가 상쇄. 저작권자의 피해보다 명백하고 직접적이어야 공정 이용 판단에 실질적으로 기여

공정 이용(Fair Use) – 요인 평가

- 법원은 "저작권의 목적에 비추어" 법적 요소들을 종합적으로 검토해야 함
 - 각 요소가 균형에 얼마나 기여하고 어떤 방향으로 기여하는지는 특정 사건의 사실과 상황에 따라 달라짐
 - 공정이용 조건 첫 번째와 네 번째 요소가 상당한 비중을 차지할 것으로 예상
 - AI 훈련에서 저작물을 다양하게 활용하는 방식은 다른 방식보다 더 큰 변화를 가져올 것
 - AI 시스템이 출력하는 양, 속도, 정교함, 훈련에 사용되는 방대한 저작물을 고려할 때, 시장에 미치는 영향은 전례 없는 규모
- 생성 AI는 다양한 용도와 영향을 포함하기 때문에 소송 결과를 미리 예측할 수 없음
 - 생성 AI 학습을 위해 저작권이 있는 저작물을 사용하는 경우 일부는 공정 사용으로 간주, 일부는 그렇지 않을 것으로 예상
 - 저작물의 일부를 결과물에 복제할 수 없는 비상업적 연구 또는 분석 목적의 사용은 공정 사용으로 간주될 가능성이 높음
 - 라이선스를 합리적으로 확보할 수 있는 경우, 시장에서 경쟁하는 제한 없는 콘텐츠를 생성하기 위해 불법 복제 소스에서 표현적 저작물을 복제하는 것은 공정 사용으로 간주될 가능성이 낮음
- 생성형 AI는 다수의 요소(예: 상업성, 전체 사용, 시장 대체 가능성)가 공정 이용에 불리한 방향으로 해석될 수 있는 구조라는 우려도 존재

공정 이용(Fair Use) – 개발자 간의 경쟁

- AI 기업 간 경쟁에 대한 라이선스의 영향에 대한 우려를 무시할 수는 없지만, 이러한 우려가 공정 사용 분석에 영향을 미치지 않는다고 판단
 - 자금력이 풍부한 사람들에게 라이선스는 항상 더 쉬울 것이며, 라이선스를 받을 작품이 많을수록 그 효과는 더 커짐
- 더 광범위한 경쟁 문제가 걸려 있는 경우, 독점금지법과 이를 집행할 권한을 가진 기관을 통해 더 적절하게 처리 가능

공정 이용(Fair Use) – 타 국가들의 접근 방법

- EU, 싱가포르: TDM 도입. 권리자가 명시적으로 옵트아웃 가능
- 일본: AI 학습을 포함한 모든 비표현적 활용에 대해 광범위한 예외 인정
- 영국: 연구목적 TDM은 인정, 상업적 목적에는 제한적
- 미국: 현재까지 구체적 AI 학습 외에는 존재하지 않고 공정이용 기준에 의존
- 이스라엘
- 한국
 - 생성 AI와 저작권에 대한 가이드 발표('23)
 - 관련 법원 판례가 여러 개 축적될 때까지 공정사용 적용 가능성은 불분명
- 중국
- 브라질(보류 중)
 - AI 회사가 교육에서 자신의 작품을 사용하는데 대해 권리자에게 보상하도록 요구
- 스페인(철회)
 - AI 모델 개발에서 보호되는 작품의 대량 사용을 위한 ECL 수립 제안, 철회

AI 학습을 위한 라이선싱

- 합법적인 가장 확실한 방법은 라이선스 계약 체결
- 자발적 라이선싱
- 법률적인 접근법
 - 강제적인 라이선싱
 - ECL(Extended Collective Licensing, 확장된 집단 라이선싱)
 - 옵트 아웃(Opt-out)
- 분석 및 권장 사항

AI 학습을 위한 라이선싱 – 자발적 라이선싱

- 자발적 라이선싱
 - 당사자들이 저작물의 특정 용도에 맞춰 조건 설정
 - 집단 자발적 라이선스 계약은 종종 여러 저작권자의 승인을 받아 협상하고 로열티를 징수 및 분배하는 제3자 기관이 관리
- 자발적 라이선스 실행가능성
 - 사례: OpenAI와 AP 통신사 간 기사 사용 계약, Shutterstock, Getty Images 등 이미지 플랫폼과의 제휴, 음악 데이터 사전 승낙 계약 등
 - 거래비용 매우 높음, 데이터 출처 식별 어려움, 권리자 간 힘의 불균형
 - AI 학습 전반을 정당화하기 어려움
- 의미있는 보상을 제공하는 능력
 - AI 훈련이 저작권 소유자에게 긍정적인 영향: 각 저작물의 기여도 평가, 사용 빈도 추적, 수익배분 체계
 - 입출력간 관계가 투명하지 않아 정확한 기여도 기반 보상 구조 설계가 난이
- 집단 라이선스에 가능한 법적 장애물
 - 반독점법 위반 소지, 대표성 문제
- 강제성 없는 집단 라이선스는 구조적으로 취약

AI 학습을 위한 라이선싱 – 법률적인 접근법

- 강제 라이선스나 ECL 등 라이선스에 대한 법적 접근 방식에 대한 지지 거의 없음
- 강제 라이선싱
 - 법률에 의해 제정, 저작권자의 동의없이 저작물 이용하도록 허용
 - 특정 용도, 사용자, 저작물에 적용
 - 로열티 지불 및 관련서류 제출과 같은 특정 법률 및 규제 요건을 준수해야 함
 - 저작권청의 입장: 시장이 제대로 기능하지 않는 예외적인 경우에만 시행되어야 함
- ECL
 - 한 집단이 권리자의 명시적 위임없이도 전체 집단을 대표해 계약 가능
 - 일부 유럽 국가에서 채택된 방식, 대규모 디지털 이용에 효과적
 - 해당 유형의 모든 저작권 소유자가 별도로 협상하지 않는 한 해당 유형의 모든 저작권 소유자에게 구속력 있음
- Optout
 - 저작권자가 자신의 저작물에 대한 AI 학습을 보류하도록 요청 가능
 - 저작권자들은 거부
 - Robots.txt는 인정받고 존중받는 정도에서만 효과적이며, 생성적 AI 수집에는 한계 있음

AI 학습을 위한 라이선싱 – 분석 및 권장 사항

- 현재 AI 학습 관련 저작권 문제는 법적 불확실성이 매우 높고, 시장 자원은 공정성과 실효성의 문제가 큼
- 저작권 소유자와 창작자는 AI에 대해 자발적 집단 라이선스 개발에 관심
- AI 학습을 위한 강제 실시권 제도가 상당한 단점을 초래할 수 있음
 - 성급한 도입은 유연하고 창의적인 시장 기반 솔루션 개발을 저해할 위험도 있음
- 자발적 라이선스가 실행 불가능한 분야는 ECL이 덜 개입적인 접근 방식 사용
- 옵트아웃 가능성
- 저작권청의 입장
 - 현 시점에서는 정부의 개입없이 라이선스 시장이 계속 발전하도록 허용할 것을 권고
 - 특정 맥락에서 특정 유형의 저작물에 대한 시장 실패가 나타날 경우 ECL과 같은 집중적인 개입 고려

결론

- 현행법을 관점에서, 생성형 AI 개발의 여러 단계에서 저작권자의 권리를 침해하면서 저작물을 사용하는 것이 포함됨
- 핵심 질문은 이런 “일견상의 침해” 행위가 공정 사용으로 인정될 수 있을지의 여부
- 공정 사용 결정에는 모든 관련 상황을 고려하여 여러 법적 요소의 균형을 고려해야 함
 - 공정성의 정도는 어떤 저작물이 사용되었는지, 어떤 출처에서 사용되었는지, 어떤 목적으로 사용되었는지, 그리고 결과물에 어떤 통제가 적용되는지에 따라 달라지며, 이 모든 요소는 시장에 영향을 미칠 수 있음
 - 연구 등 비영리 목적은 결과물이 학습에 사용되는 표현 저작물을 대체할 가능성은 낮음
 - 기존 시장에서 저작물과 경쟁하는 표현 콘텐츠 제작을 위해 방대한 저작물을 상업적으로 사용하는 것은 기존의 공정 사용 범위를 넘어섬
- 공정하다고 할 수 없는 용도의 경우, 지속적인 혁신을 지원하기 위해서는 실용적인 해결책이 필수적
 - AI 훈련을 위한 개별 및 집단 라이선스 계약이 특정 분야에서 빠르게 등장하고 있지만, 현재까지 이용 가능 여부는 일관되지 않음
 - 자발적 라이선스의 급속한 증가와 법 개정에 대한 이해관계자의 지지 부족을 고려할 때, 현 시점에서 정부의 개입은 시기상조라고 판단
 - 라이선싱 시장은 지속적으로 발전하여 초기 성공 사례를 가능한 한 빨리 더 많은 분야로 확대해야 함
 - ❖ 시장 실패를 해결하기 위해 확대된 집단 라이선싱과 같은 대안적 접근 방식을 고려해야 함

참고문헌

- 디지털새싹, “미국 저작권청(U.S. Copyright Office)의 2025년 AI 저작권 보고서 살펴보기”, 2025.2.11, <https://www.xn--2z1bz5tdvbiwlf4j.com/131>.
- 야마모토 요시타카, 〈16세기 문화혁명〉, 동아시아 출판사, 2010년.
-

감사합니다

