



AI-저작권 관련 활용 가이드(안)에 대한 BSA 의견서

2023년 11월 17일

한국저작권위원회 귀하

BSA | 소프트웨어연합 (BSA)¹ 은 한국저작권위원회 (KCC)의 AI-저작권 관련 활용 가이드(안) (이하 '가이드')에 대한 의견을 제출할 수 있는 기회에 감사드립니다. BSA는 각국 정부와 세계시장을 중심으로 글로벌 소프트웨어 산업을 대변하고 있는 연합체로 오랫동안 효과적인 저작권 보호를 지지해왔습니다. BSA 회원사는 AI 제품 및 서비스 제공 뿐만 아니라 제 3자가 AI 시스템 및 애플리케이션 개발에 활용할 수 있는 도구를 제공하는 데 앞장서고 있습니다.

저작권법과 같은 맥락에서 AI는 콘텐츠 창작자를 포함한 모든 산업 분야에 혜택을 줄 수 있습니다. AI 시스템은 모델 훈련에 사용할 수 있는 대량의 데이터를 컴퓨터로 분석하여 학습합니다. 그런 다음 새로운 데이터가 포함된 쿼리가 제시되면 모델이 예측을 수행합니다. 일부 생성형 AI 모델은 여기서 한 걸음 더 나아가 텍스트나 그림, 심지어 코드와 같은 새로운 데이터를 생성할 수 있습니다. 그러나 어떤 형태의 AI를 사용하든지 저작물과 개인 데이터 보호, 영업 비밀 정보에 대한 존중은 책임감 있는 AI 개발의 초석이 되어야 합니다. 어떤 중요한 기술의 발전일지라도 예술가와 저작권자들의 희생을 수반해서는 안 됩니다.

요약

BSA는 다음과 같은 내용을 제언드립니다:

1. **첫째, AI 모델 훈련, 컴퓨터 분석은 일반적으로 데이터를 토큰으로 변환하여 다른 토큰화 된 데이터와 통계적으로 상관관계가 있는 토큰으로 변환하는 작업을 포함합니다.** 해당 데이터 중 일부는 저작물의 일부일 수 있으나, 일반적으로 이 과정에서의 데이터 사용은 저작물의 표현 내용과는 무관합니다. 예를 들어, 언어 기술을 학습하기 위해 책을 사용할 수 있는데 이는 데이터베이스 관리 방법을 개선하는 데 사용할 수 있습니다. 이러한 사용은 한국

¹ BSA의 회원사는 다음과 같습니다: Adobe, Alteryx, Altium, Amazon Web Services, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, Cloudflare, CNC/Mastercam, Dassault, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, IBM, Informatica, 주니퍼 네트워크, Kyndryl, MathWorks, Microsoft, Nikon, Okta, Oracle, Palo Alto Networks, Prokon, PTC, Rockwell, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc, 지멘스 인더스트리 소프트웨어, 스플링크, 트렌드 마이크로, 트림블 솔루션즈 코퍼레이션, 트라이넷, 트윌리오, 유니티 테크놀로지스, 워크메이, 젠테스크, 줌 비디오 커뮤니케이션즈, 줌 비디오 커뮤니케이션즈.

저작권법 제35조 제2항 및 제5항에 해당합니다². 한편, '크롤링 금지' 도구처럼 저작권자가 AI 학습을 목적으로 한 콘텐츠 사용을 원치 않음을 표시하는 자동화된 도구의 개발과 사용에 대한 업계의 자발적인 논의가 진행 중인 것으로 알고 있습니다. 보다 합의된 표준 도출을 위한 심도 있는 논의를 권장합니다.

2. **둘째, AI 시스템을 사용한 산출물의 저작물성을 판단할 때는 산출물의 표현을 위해 사용된 도구나 기술에 상관없이 인간의 창의력이 저작물에 반영되었는지 여부를 기준으로 삼아야 합니다.** 이는 예술가가 기술의 도움으로 소프트웨어 도구를 사용하여 창작물을 만든 경우에도 일반적으로 저작권을 인정하는 기존 관행과 일치합니다. 다른 소프트웨어 애플리케이션이 오랫동안 예술가와 스토리텔러의 중요한 도구로 사용되어 온 것처럼, 생성형 AI는 창의성을 강화할 수 있습니다. 일례로, 생성형 AI도 작가를 위한 워드 프로세싱과 시각 예술가를 위한 사진 향상에 사용되며, 시청각 작품의 특수 효과를 만들고 음원 녹음을 위해 음악을 편곡하는 데에도 사용됩니다. 생성형 AI가 인간의 창의력을 향상시키는 데 사용되는 경우, 그 산출물은 저작권으로 보호되어야 합니다. AI 산출물에 창의적 요소가 포함되어 있지 않더라도 사람이 제작한 저작물과 결합된 경우, 결합된 결과물 전체를 보호할 수 없게 해서는 안 됩니다. 대신, AI 산출물 중 저작물성으로 인정할 수 없는 부분은 제외되어야 하지만, 결합된 저작물 중 저작물성이 있는 요소는 저작권으로 보호할 수 있어야 합니다.
3. **셋째, AI 시스템이 저작권을 침해하는 저작물을 생성하는 데 사용될 위험과 관련하여 저작권법은 저작권 침해 책임 유무를 충분한 유연하게 판단할 수 있습니다.** 즉, 산출물의 저작권 침해 여부는 일반적으로 사용된 기술과 관련이 없습니다. 산출물이 저작권을 침해하는 경우, 저작권을 침해하는 복제물을 제작하는 과정에서 AI가 사용되었는지 여부와 관계없이 명백한 법적 책임이 존재해야 합니다.

저작권 맥락에서의 AI

AI 머신러닝은 다양한 산업과 애플리케이션에서 사용하기 위해 개발되거나 배포된 방대한 양의 기술을 포괄합니다. 머신러닝은 훈련 데이터의 컴퓨터 분석을 통해 상관관계, 패턴, 또는 기타 메타데이터를 식별하여 향후 데이터 입력을 기반으로 예측 또는 추천을 할 수 있는 모델을

² 저작권법

제 35 조의 2(저작물 이용과정에서의 일시적 복제) 컴퓨터에서 저작물을 이용하는 경우에는 원활하고 효율적인 정보처리를 위하여 필요하다고 인정되는 범위 안에서 그 저작물을 그 컴퓨터에 일시적으로 복제할 수 있다. 다만, 그 저작물의 이용이 저작권을 침해하는 경우에는 그러하지 아니하다.

제 35 조의 5(저작물의 공정한 이용) ① 제 23 조부터 제 35 조의 4 까지, 제 101 조의 3 부터 제 101 조의 5 까지의 경우 외에 저작물의 일반적인 이용 방법과 충돌하지 아니하고 저작자의 정당한 이익을 부당하게 해치지 아니하는 경우에는 저작물을 이용할 수 있다. <개정 2016. 3. 22., 2019. 11. 26., 2023. 8. 8.>

② 저작물 이용 행위가 제 1 항에 해당하는지를 판단할 때에는 다음 각 호의 사항등을 고려하여야 한다.

1. 이용의 목적 및 성격, 2. 저작물의 종류 및 용도, 3. 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성, 4. 저작물의 이용이 그 저작물의 현재 시장 또는 가치나 잠재적인 시장 또는 가치에 미치는 영향]

개발합니다.³ 최근에는 새로운 텍스트, 이미지, 소리를 생성할 수 있는 생성형 AI 모델도 등장했습니다.

AI의 응용

머신러닝 맥락에서 컴퓨터 분석을 통해 도출된 인사이트, 예측 및 기타 결과를 설명하기 위해 널리 알려진 몇 가지 예시는 다음과 같습니다:

- 기상 조건, 실시간 연료 소비, 항공기 운항 데이터, 인근 항공 교통 상황, 공항 혼잡도 및 기타 다양한 데이터 요소의 컴퓨터 분석을 기반으로 한 자동화된 비행 관리 및 항공 교통 관제.⁴
- 종양학, 자가 면역 질환, 파킨슨병 및 알츠하이머병 분야의 조기 진단, 예방 및 치료를 위한 화학적 및 세포 이상 식별.⁵
- 보안 운영에 생성형 AI의 통합으로 보안 이상 징후를 효과적으로 식별하고 대응 및 잠재적인 위협 탐지와 완화.
- 위성 데이터, 기상 관측소 데이터, 지형 정보 및 다양한 IoT 및 센서 데이터의 컴퓨터 분석을 기반으로 한 예측 기후 모델링.⁶
- 운송 기록, 계량기 검침, 연료 구매 기록, 대기 오염 추적, 발전소 및 기타 시설의 시각적 모니터링, 기타 데이터 소스의 컴퓨터 분석을 기반으로 탄소 추적 및 완화 개선.⁷
- 재난 계획과 생존 가능성 파악을 위한 해안 취약 지역 매핑을 통한 사이클론 위험 지도를 생성 및 사이클론 대피소, 학교, 보건 시설 및 기타 인프라에 대한 투자 계획을 지원하는 컴퓨터 분석.⁸

³ BSA, *Confronting Bias: BSA's Framework to Build Trust in AI (BSA AI Bias Framework)* <https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai>; 참조: 미국표준기술연구소, NIST Risk Management Framework (2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> ("인공지능 시스템을 주어진 목표 집합에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 추천 또는 결정 등의 산출물을 생성할 수 있는 엔지니어링 또는 기계 기반 시스템으로 정의")

⁴ M. Durgut, *인공 지능과 항공 교통 관제*, Aviationfile.com (2023년 1월), <https://www.aviationfile.com/artificial-intelligence-and-air-traffic-control/#:~:text=One%20of%20the%20primary%20applications%20is%20to%20help.make%20informed%20decisions%20on%20routing%20and%20scheduling%20of%20flights;>

⁵ Hunter 외, *조기 암 진단에서 인공지능의 역할*, 14(6) *Cancers* 1524 (2022), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8946688/>; Stafford 외, *자가면역질환에서 인공지능과 머신러닝의 적용에 대한 체계적인 검토*, 3 *NPJ - Digital Medicine* 30 (2020), <https://www.nature.com/articles/s41746-020-0229-3>; Diogo 외, *머신러닝을 이용한 알츠하이머병의 조기 진단*, 14 *Alzheimers Research and Theory* 107 (2022), <https://alzres.biomedcentral.com/articles/10.1186/s13195-022-01047-y>

⁶ Schneider 외, *기후 모델링 및 예측을 발전시키기 위한 AI 및 컴퓨팅 활용*, 13 *자연 기후 변화* 887(2023), <https://www.nature.com/articles/s41558-023-01769-3>; 세계 경제 포럼, *지구를 구하는 데 도움이 되는 머신러닝의 역할*(2021), <https://www.weforum.org/agenda/2021/08/how-is-machine-learning-helping-us-to-create-more-sophisticated-climate-change-models/>.

⁷ Global Data Alliance, *국경 간 데이터 전송 및 환경 지속 가능성*(2023)(내부 인용 생략), <https://globaldataalliance.org/wp-content/uploads/2023/04/04192023gdacbdtsustainability.pdf>.

⁸ Global Data Alliance, *국경 간 데이터 전송 및 환경 지속 가능성*(2023).

- 예측 타이핑 및 기타 오피스 생산성 솔루션 (예: "sin"을 입력하면 "...cerely yours"를 제안하는 '자동 완성' 기능),⁹ 또는 음향 효과나 특수 효과 생성으로 창작자와 영화 제작자의 새로운 작품 개발 지원.¹⁰
- 생성형 AI는 대화형 인터페이스를 통해 위의 응용 사례와 같은 광범위한 산업과 애플리케이션에서 AI를 더 쉽고 유연하게 활용하여 협업할 수 있게 합니다. 대화형 인터페이스를 통해 사용자들은 AI 시스템의 데이터와 산출물에 대해 질문하고 추론할 수 있습니다.

활용 사례는 다양하지만 각각의 훈련 요소는 매우 유사하며, 이에 대해서는 아래에서 자세히 서술하겠습니다. 모델이 저작물성이 없는 요소를 학습할 수 있도록 하는 것이 학습의 목적입니다.

AI 개발 수명 주기

AI 개발 수명 주기에는 일반적으로 다음과 같은 단계로 이루어집니다:

- 프로젝트 기획: 먼저, AI 개발팀은 시스템이 해결하고자 하는 '문제'를 명확히 정의하고 시스템이 예측하고자 하는 구조와 대상 변수를 매핑합니다. 특정 작업을 위해 훈련된 모델 (좁은 의미의 AI 애플리케이션)의 경우, 소비자의 심박수를 분석하여 뇌졸중이나 심장 질환의 위험 (즉, 목표 변수)을 예측하여 불규칙성을 모니터링하는 피트니스 앱이 그 예가 될 수 있습니다. 그러나 기초 모델은 다양한 애플리케이션을 지원하기 위해 개발될 것입니다.
- 원시 데이터(Raw data) 식별: 둘째, AI 개발팀은 추후 변환 및 구조화될 '원시 데이터'의 범위를 식별합니다. 데이터 소스는 머신러닝 AI의 잠재적 애플리케이션만큼이나 다양하며, 기계 간 데이터 (예: 위성 전송 데이터) 및 국제 무역 통계, 출판 자료, 블로그 게시물, 웹사이트 댓글, 채팅방 로그 등이 포함됩니다. '원시 데이터'는 '지저분한' 경우가 많기 때문에 데이터를 사용 가능한 형태로 변환하기 위해서는 아래에 설명과 같은 상당한 작업이 필요합니다. 특정 애플리케이션을 위한 좁은 범위의 AI 시스템을 개발하거나 특정 작업을 위한 기초 모델을 미세 조정하기 위한 데이터는 특정 작업과 연관됩니다. 기초 모델 훈련에는 더 광범위하고 다양한 데이터가 필요합니다. 대규모 AI 모델을 훈련하기 위해서는 방대한 규모의 데이터가 필요하며, 이러한 규모를 확보하기 위해서는 인터넷에서 데이터를 분석할 수 있어야 합니다.
- 데이터 세트 준비: AI 개발팀은 '원시 데이터'를 기계가 의미적으로 이해하고 모델을 훈련하는 데 사용할 수 있도록 수정할 수 있습니다. 이 과정에서 팀은 필요에 따라 데이터를 수정, 정리, 정규화합니다. 데이터는 일반적으로 '토큰화'를 통해 의미론적, 구조적으로 변환되며, 텍스트나 데이터를 더 작은 단위 (또는 '토큰')로 분해하여 컴퓨터 분석을 하는데 사용됩니다.

⁹ S. Ashraf, *자동 완성의 신비화*, 데이터 과학을 향하여(2020), <https://towardsdatascience.com/index-48563e4c1572>; A. Wickramarachchi, *기계 학습: 단어 임베딩 및 예측*, 데이터 과학을 향하여(2020), <https://towardsdatascience.com/machine-learning-word-embedding-and-predicting-e603254e4d7b>.

¹⁰ D. Nelson, *AI 연구자들이 영화 및 기타 미디어를 위한 음향 효과를 생성하는 프로그램을 설계하다*, Unite AI 웹사이트(2022), <https://www.unite.ai/ai-researchers-design-program-to-generate-sound-effects-for-movies-and-other-media/>

분석용 데이터의 신뢰성, 품질 및 적합성을 개선하기 위해 추가 처리가 필요할 수 있으며, 이는 전체 데이터 세트에서 누락값, 중복값, 이상값 및 일관되지 않은 서식과 같은 품질 문제를 해결하는 데 도움이 됩니다. 대규모 AI 모델은 일반적으로 학습 데이터에 레이블을 지정할 필요가 없는 자체 감독 방법으로 학습되며, 일반적으로 인터넷에서 수집한 방대한 규모의 훈련 데이터에 따라 모델을 확장할 수 있습니다

- **모델 정의:** 입력 데이터가 적절하게 처리된 후 AI 개발팀은 시스템의 기본 아키텍처를 설정해야 합니다. 여기에는 알고리즘이 미래를 예측하는데 있어 규칙의 기초가 되는 패턴과 관계를 찾을 때 평가할 훈련 데이터의 변수 (즉, '특징')를 식별하는 것이 포함됩니다. 또한 시스템을 구동할 알고리즘 모델 유형 (예: 선형 회귀, 로지스틱 회귀, 심층 신경망)을 선택하는 것도 포함됩니다.¹¹ 데이터가 준비되고 알고리즘이 선택되면 팀은 향후 데이터 입력에 대해 예측할 수 있는 기능 모델을 생성하도록 시스템을 훈련합니다.
- **모델 검증, 테스트 및 수정:** 모델을 훈련시킨 후, AI 개발팀은 모델이 의도한 대로 작동하는지 확인하기 위해 모델을 검증하고, 시스템의 산출물이 예상 매개변수 내에 있는지, 예기치 않은 오류나 의도하지 않은 편향이 없는지 테스트해야 합니다. 검증 및 테스트 결과에 따라 위험을 완화하기 위해 모델을 수정하고 개선해야 하는 경우도 있습니다.

AI 모델 훈련을 위한 저작물 사용

위에서 설명한 바와 같이, 컴퓨터 분석은 일반적으로 수백만 또는 수십억 개의 토큰화된 데이터 세트로 구성된 대규모 훈련 데이터 세트에 적용됩니다. 모델 훈련 방식에 따라 인터넷에서 접근 가능한 데이터가 토큰화된 요소로 변환되는 원시 데이터 세트의 일부로 수집될 수 있습니다.

인터넷에서 무료로 제공되는 콘텐츠의 상당 부분은 잠재적으로 저작권 보호의 대상이 될 수 있으며, '독창성'을 입증하기 위한 임계값이 낮고 저작물이 등록되지 않았더라도 저작물이 생성되면 저작권이 자동으로 발생하기 때문에 이 원시 데이터에는 저작물이 포함될 수 있습니다. 그러나 중요한 점은 저작권 보호가 사실, 아이디어, 수학적 개념에는 적용되지 않기 때문에 온라인에 있는 모든 소재가 저작권의 대상이 되는 것은 아니라는 것입니다.¹²

컴퓨터 분석에는 (1) '훈련 데이터' 세트를 생성하기 위해 필요한 복제와 (2) AI 모델을 훈련시키는 계산 과정에서 부수적으로 발생하는 일시적 복제, 두 가지가 포함될 수 있으며 이는 잠재적으로 저작권법에 저촉될 수 있습니다. 그러나 각각의 경우에서 복제본은 일반적으로 공개되어 있지도, 공람되지도 않습니다. 대신 이러한 복제본은 작품에서 유래한 기본 데이터에 대한 비저작성 정보, 즉

¹¹ BSA AI Bias Framework (2022).

¹² TRIPS 협정 제 9.2 조는 저작권 보호가 아이디어, 절차, 작동 방법 또는 수학적 개념이 아닌 표현에까지 확대된다고 규정하고 있습니다. 즉, 저작권 보호는 저작물에 포함된 정보나 아이디어에는 적용되지 않으며, 그러한 정보나 아이디어가 표현된 원래의 방식만을 보호합니다. 따라서 누구나 저작물에 포함된 정보를 자유롭게 사용할 수 있으며, 새로운 저작물을 창작할 목적도 포함됩니다.

AI 모델 생성에 정보를 제공하고 향후 데이터 입력을 기반으로 예측을 가능하게 하는 상관관계 및 패턴을 식별하기 위한 기술 프로세스의 필수적인 부산물입니다. 이러한 비표현적 복제는 a) 저작권법 제 35 조 제 2 항에 따른 일시적 복제이며 b) 저작권법 제 35 조 제 5 항에 따른 저작권이 보호하고자 하는 저작자의 정당한 이익을 부당하게 해치지 않는 경우로 정의됩니다.

더불어 컴퓨터 분석은 표현적 콘텐츠의 저작물을 소비하는 것이 아닙니다. 오히려 이러한 분석에는 전체 토큰화된 데이터 세트를 통한 확률, 상관관계, 추세 및 기타 패턴에 대한 수학적 계산이 포함됩니다. 이러한 분석은 전체 데이터 세트에 분포된 수학적 패턴 (예: 다른 토큰과 관련된 특정 토큰의 관계)만을 이해하려고 하는 데, 수학적 패턴 자체는 저작권법에 의해 보호되는 표현적 콘텐츠가 아닙니다.

AI 훈련 데이터의 컴퓨터 분석은 공정한 이용입니다.

저작권법 제 35 조 제 2 항 및 제 5 항에 규정된 예외 조항은 대량의 정보를 분석하여 패턴, 상관관계 및 기타 메타데이터를 식별해 향후 데이터 입력에 대한 예측을 수행하는 AI 모델을 개발하기 위한 저작물의 사용을 보장합니다.

다시 말해, AI 기반 예측 타이핑 모델과 같은 자연어 처리 모델을 만들려는 AI 개발자는 공개적으로 사용 가능한 텍스트 기반 소재를 사용하여 훈련 데이터베이스를 만들 수 있습니다. 이 과정에서 AI 개발자는 해당 텍스트를 표현의 목적으로 복제하지 않습니다. 오히려 수천 개의 구문, 비유, 직유, 은유, 문법 패턴, 일반적인 언어 형식 및 표현에 나타나는 문자와 단어 간의 상관관계, 패턴, 관계에 대한 보호되지 않은 정보를 추출하기 위한 목적으로만 텍스트를 복제합니다. 수천 또는 수백만 개의 글에 등장하는 문자, 단어, 구문이나 이들 사이의 수학적 패턴은 저작권 보호 대상이 아닙니다.

마찬가지로 이미지 생성 모델의 경우, 모델 개발자는 단어가 태그된 방대한 양의 이미지 (예: '고양이'로 태그된 고양이 사진과 '강아지'로 태그된 강아지 사진)를 가져옵니다. 시간이 지남에 따라 모델은 특정 패턴이 '고양이'의 특징임을 학습하고, 입력된 이미지가 고양이인지 아닌지, 이미지가 만화인지 사진인지 등을 인식하는 방법을 학습합니다. 다시 말해서 머신은 표현을 목적으로 특정 고양이 이미지를 재현하는 것이 아니라 기본적인 사실/통계적 상관관계 등 고양이가 무엇인지 이해하기 위해 이미지를 분석하는 것입니다.

또한 데이터 세트에 적용된 컴퓨터 분석의 최종적인 용도는 원본 콘텐츠의 변형적 이용입니다. 예측 타이핑 소프트웨어의 자동 완성 기능은 AI 훈련 데이터 전체와 해당 데이터 세트 내의 단일 작업에서 모두 구별되는 소프트웨어 코드 형태의 새로운 창작물입니다. 일반적으로 사용된 모든 기능은 원본 데이터에서 찾을 수 있는 저작물과는 거리가 있습니다.

마지막으로, 이러한 기능은 본질적으로 저작권이 보호하고자 하는 어떤 저작물과도 경쟁하지 않습니다. 따라서 AI 개발자는 제 35 조 2 항 및 5 항의 예외조항에 의거하여 합법적으로 액세스할 수 있는 텍스트, 이미지 또는 기타 데이터를 사용하여 AI 훈련 데이터베이스를 구축할 수 있습니다.

AI 산출물의 저작물성

AI 산출물의 저작물성은 표현을 돕기 위해 사용된 도구나 기술에 관계없이 인간의 창의성이 저작물에 반영되었는지 여부를 판단의 기준으로 삼아야 합니다. 다른 소프트웨어 애플리케이션이

오랫동안 예술가와 스토리텔러의 중요한 도구로 사용되어 온 것처럼 생성형 AI는 창의성을 강화하는 도구가 될 것입니다. 일례로, 생성형 AI는 작가를 위한 워드 프로세싱과 시각 예술가를 위한 사진 향상에 사용되며, 시청각 작품의 특수 효과를 만들고 음원 녹음을 위한 음악을 편곡하는 데에도 사용됩니다. 생성형 AI가 인간의 창의력을 향상시키는 데 사용될 경우, 그 산출물은 저작권으로 보호되어야 합니다.

저작물성에 관한 논의는 인간 창의성의 정도에 대한 면밀한 검토로 이어질 것이며, 모든 범주의 창작영역에 도구로서의 AI 사용이 점차 증가하고 있는 상황에서 AI의 사용이 저작물성을 제한으로 이어진다면 AI 솔루션 선택을 현저히 위축시킬 것입니다.

AI 산출물 중 인간의 창의성 요건을 충족하는 저작물은 계속해서 저작권 보호를 받을 수 있어야 합니다. 대부분의 경우 AI 시스템은 인간 저작자와 창작자의 창의적 비전을 실행하기 위한 도구로 사용됩니다. 예를 들어, AI 사용으로 사진작가는 지루한 이미지 편집 과정을 자동화하고,¹³ 건축가는 설계를 보강하여 에너지 효율성을 높이고,¹⁴ 영화 제작자는 애니메이션 캐릭터의 움직임을 더욱 실감나게 표현할 수 있습니다.¹⁵ 이러한 사례처럼 사용자의 창의성 기여를 고려하면 산출물이 저작물이 되어야 한다는 결론에 쉽게 도달할 수 있습니다.

생성형 AI의 사용이 저작물성 판단의 기준이 되어서는 안 됩니다.

AI 산출물의 저작권 침해에 대한 잠재적 책임

저작권법은 AI 산출물의 저작권 침해 여부를 충분히 유연하게 판단할 수 있습니다. AI 산출물의 저작권 침해 여부는 사용된 기술과 관련이 없어야 하며, AI 산출물이 저작권을 침해하는 경우 저작권을 침해한 작업물을 제작하는 과정에서 AI가 사용되었는지 여부와 관계없이 명확한 책임을 져야 합니다. 이는 작업물이 어떻게 생성되었는지 보다 저작물이 저작권을 침해했는지 여부에 중점을 두는 기존 법과도 일치합니다.

현행 저작권법은 AI가 생성한 콘텐츠로 인해 저작권 침해 발생하는 경우를 포함하여 저작권자를 저작권 침해로부터 보호합니다.

저작권자는 자신의 권리가 침해되었을 때 발생한 모든 손해에 대해 완전하고 효과적인 구제 수단을 제공받아야 합니다. 이 원칙은 AI 시스템으로 생성된 산출물 뿐만 아니라 다른 방식으로 생성된 산출물에도 동일하게 적용됩니다. 저작권 침해가 발견될 경우 예술가와 창작자에게 발생한 모든 손해를 완전히 보상하는 효과적인 구제 조치를 갖추는 것이 중요합니다.

현행 저작권법은 저작권 침해 문제를 해결하기에 충분합니다. 대부분의 AI 사용 사례에서 AI 시스템의 생성물은 저작권과 전혀 관련이 없습니다. 그러나 다른 기술과 마찬가지로 AI도 저작권을 침해하는 자료를 생성하는 데 사용될 수 있습니다. 이러한 경우, 저작권 침해 책임은 생성물이 '실제 복제'에 기반하고 저작물과 '실질적으로 유사'하며, 저작물 간의 '질적 및 양적으로 유의미한'

¹³ <https://theblog.adobe.com/adobes-general-counsel-makes-the-case-for-ai/>.

¹⁴ <https://www.autodesk.com/redshift/machine-learning-in-architecture/>.

¹⁵ <https://theblog.adobe.com/state-of-ai-in-animation/>.

유사성을 고려할 때 '보통의 합리적인 사람이라면 두 저작물을 구별하지 못할 것'이라는 점이 입증되어야 전제될 수 있습니다.

특정 사실을 표현할 수 있는 방법이 하나뿐이거나 몇 가지 밖에 없어 저작권 침해 주장이 '약한' 경우에는 일반적인 법적 규범에 따라 산출물과 저작물 사이의 '실질적 유사성'을 증명해야 합니다. 생성형 AI의 많은 사용자는 새로운 도구를 실험하는 소규모 기업이 될 것이므로, 법 집행 결정은 이러한 점을 고려해야 합니다. 2 차적 저작물을 포함한 다른 적절한 경우에도 책임이 발생할 수 있습니다.

또한 원고는 AI 서비스 제공 업체를 상대로 저작권 침해 소송을 제기할 수 있습니다. 이 경우에도 현행 저작권법이 책임을 평가하고 배분하는 데 충분할 것입니다. 원고가 직접적인 침해가 발생했음을 입증하면 법원은 서비스 제공 업체가 저작권 침해를 '간접적으로' 또는 '기여적으로' 도왔거나 방조한 것인지 여부를 평가할 것입니다.

충분히 큰 데이터 세트에서 훈련된 AI 모델은 저작권을 침해하는 산출물을 생성할 가능성이 낮습니다.

훈련 데이터와 AI 시스템의 출력을 혼동하지 않는 것도 중요하지만, 사용자의 침해 의도가 없는 경우 훈련에 사용된 데이터가 많을수록 시스템이 특정 입력의 복제나 파생물을 생성할 가능성이 줄어든다는 점 역시 유의해야 합니다. 충분히 큰 데이터 세트에 훈련된 잘 설계된 AI 모델에서 컴퓨터 분석은 절대 (또는 극히 드문 경우에만) 특정한 저작권의 보호를 받는 저작물과 '사실상 동일'하거나 '실질적으로 유사한' 산출물을 생성해서는 안 됩니다.

일부 AI 개발자와 배포자는 변화하는 디지털 환경에서 저작물을 지원하기 위해 이미 예술가와 창작자들과 협력하고 있으며, 사용할 수 있는 프롬프트를 제한하는 등 AI 시스템의 오용을 제한하기 위한 조치를 취하고 있습니다.

마치며

BSA의 의견이 KCC의 AI-저작권 관련 활용 가이드(안)를 마련함에 있어 도움이 될 수 있기를 희망합니다. 위에서 설명한 바와 같이, AI 훈련 데이터 세트 내용에 대한 비소비적 컴퓨터 분석은 한국 저작권법상 '공정한 이용'에 해당합니다. 또한 BSA는 AI 훈련 프로세스와 침해 위험을 최소화하기 위한 여러 이해관계자의 노력을 지지합니다. 또한 침해가 발생하는 경우 창작자들을 옳이 보호하는 것에도 강한 지지를 보냅니다.

중요한 사안에 대한 의견을 제출할 수 있는 기회를 주셔서 다시 한번 감사드립니다. 본 의견서와 관련하여 문의 사항이 있으시거나 추가로 도움이 필요하시다면 언제든지 말씀 부탁드립니다.

감사합니다.

Tham Shen Hong
아태지역 정책 매니저

