Generative AI 관련 오픈소스•저작권 소송사례: Copilot, Stable Diffusion

전응준(법무법인 린 변호사)

2023. 8. 31. / 오픈소스SW 라이선스 컨퍼런스

법무법인 린

목차

0. AI 에 대한 우려 및 그 원인

1. 생성형 AI의 저작권 관련 3가지 쟁점

2. AI 학습과정에서 발생하는 저작권 문제(INPUT단계)

3. AI 출력과정에서 발생하는 저작권 문제(OUTPUT단계)

4. AI 산출물에 대한 법적 보호

발표자 소개

전응준 변호사

T. 02-3477-8686

F. 02-3477-8687

E. ejjeon@law-lin.com

- 서울대학교 계산통계학과 (현 컴퓨터공학과 및 통계학과)
- 사법연수원 제33기 (43회)
- 전) 유미 법무법인/특허법인 변호사,변리사
- 현) 서울지방변호사회 지적재산권커뮤니티 위원장
- 현) 한국지적재산권변호사협회(KIPLA) 부회장
- 현) 산업재산권분쟁조정위원회 위원
- 현) 특허청 부정경쟁방지법 제도개선위원회 위원
- 현) 무역위원회 지식재산권 자문단 위원

❑ AI 기술에 대한 우려와 불확실성

- 2023. 3.22. Future of Life에서 안전한 AI 거버넌스 확립을 위해 AI기술의 개발을 6개월 정도 중단하자는 제안을 함. Yoshua Bengio, Elon Musk, Steve Wozniak, Emad Mostaque 등이 찬성.
- Center for AI Safety는 AI가 인류의 절멸(extinction)을 가져올 수 있다고 보고 팬데믹이나 핵전쟁 위험 수준의 대응이 필요하다고 발표. Geoffery Hinton, Sam Altman 지지서명.
- 이러한 우려는, 현재 상황이 AI 기술을 통제하기 어려운 정도에까지 이르렀고 AI 모델은 예측할 수 없는 블랙박스와 같기 때문에 안전하지 못하다는 인식에 기초하고 있음

❑ 이에 대한 반대의견

- Yann LeCun, Andrew Ng은 위 모라토리엄 제안을 반대하는 견해를 피력
- AI 기술과 AI 모델의 서비스(상품)을 구분하여, AI 상품은 규제하는 것이 타당하지만 AI 기술 자체를 막는 것은 현실적이지 못함
- 현재 AI 기술의 위험성은 매우 시급하지는 않고 (서버 폐쇄 등으로) 통제가능하며 AGI(Artificial General Intelligence)의 도래는 요원함(현재 인류는 10살 아이도 쉽게 할 수 있는 '식탁을 치우고 그릇을 식기세척기에 담는 행위'를 할 수 있는 로봇을 가지고 있지 못하다고 지적)

❑ AI 산출물의 표현력 수준의 변화

(Left) "Edmond de Belamy" -The first AI-generated portrait sold at Christie's art auction in 2018.

(Right) "Théâtre D'opéra Spatial" - The winner of the digital art category at the Colorado State Fair's annual art competition in 2022.

❑ AI 산출물의 표현력 수준의 향상과 저작권의 미래

- AI 산출물의 표현력 수준이 향상되어 인간의 저작물과 구분될 수 없는 단계에 이르는 경우 기존 저작권 체계에 큰 혼란을 줄 것으로 보임
- 인간의 저작물과 구분되지 않는 AI 산출물에 대해 (사람의) 저작권을 주장할 수 있고 이 경우 대량의 유사한 저작물(?)들이 각각의 배타적 권리를 행사할 수 있음
- 저작권 보호기간(사후 70년), 보호의 범위(실질적 유사성), 보호의 강도(물권적 보호, 형사처벌) 등에 대한 전반적인 검토가 요구될 수 있음

❑ 생성형(Generative) AI의 기술적 특성

- 다양한 수준의 자율성을 가진(with varying levels of autonomy) AI가 이용자가 입력한 프롬프트(prompt)에 따라 텍스트, 이미지, 오디오, 비디오 등의 콘텐츠를 생성하도록 하는 인공지능 기술
- 생성형 모델은 주어진 원본 데이터를 학습하여 원본 데이터의 분포를 따르는, 유사한(가짜) 데이터를 생성하는 인공지능 모델(GAN, Diffusion 등)
- 생성형 AI는 학습데이터의 분포를 모방하려는 기술적 특성을 가지고 있으므로, 그 산출물이 학습데이터로 사용된 타인의 저작물과 유사할 잠재적 가능성 있음
- 다만, AI의 실제 OUTPUT이 학습데이터와 유사한 비율은 극히 적은 것으로 보임(Copilot 사건의 소장도 대략 1% 정도의 Copilot의 코드 산출물이 학습데이터 코드와 유사하다는 연구를 원용하고 있음)

❑ 모방성 방지를 위한 대책 필요

- 학습데이터에서 중복되는 데이터를 삭제
- AI의 OUTPUT 단계에서 필터링에 의해 저작권 침해물 생성 방지

❑ 생성형 AI의 산출물을 사람 저작물로 인정하는 경우 문제점

- ○ 학습데이터의 저작권자에 대한 보상없이 AI 생성물을 사람 저작권으로 보호하는 것은 형평에 어긋남
  - • AI 개발사는 AI 생성물에 대한 권리를 주장하지 않고 유료 수익 모델을 추구(투자회수 가능)
  - • AI 모델 이용자는 단지 프롬프트만을 입력할 뿐이므로 사람 저작자로 보기 어려움
- ○ AI 생성물을 사람 창작물과 동일 취급하는 경우 사람 창작자에 대한 인센티브 소멸
- ○ 생성형 AI의 산출물을 다시 학습데이터로 이용하여 AI 모델을 만드는 경우의 문제(Alpaca AI)

❑ AI 워터마크 기술의 정책적 필요성

- ○ AI 산출물에 워터마크를 삽입하여 AI 산출물임을 쉽게 탐지할 수 있다면 현재 저작권 체계는 큰 변동없이 유지될 수 있을 것으로 예상됨
- ○ 2023. 7. OpenAI, 구글, MS, 메타 등 미국 AI 기업은 생성형 AI 콘텐츠에 워터마크를 삽입하기로 합의
- ○ 그러나 현재 기술 수준은 이미지 생성 AI의 산출물에는 워터마크 적용이 가능하나 LLM 생성물에는 워터마크를 적용하기 어려운 상황임(e.g. OpenAI의 AI classifier 서비스 종료)
- ○ AI 학습데이터의 품질 유지의 관점에서도 AI 산출물을 탐지할 필요 있음

❑ EU AI Act의 EU 의회 수정안 draft(2023. 5.11)

➔ 2023. 6.14. 수정의결

○ GPT모델과 같은 파운데이션(foundation) 모델 및 그 중 생성형 AI에 대한 규제 추가(제28b조)

- 파운데이션 모델: 대규모 데이터에 의해 학습된 AI모델로서 범용적인 산출물을 얻기 위해 설계되었고 다양한 작업에 적용될 수 있는 모델
- 생성형 AI: AI가 다양한 수준의 자율성을 가지고(with varying levels of autonomy) 텍스트, 이미지, 오디오, 비디오 등의 콘텐츠를 생성하는 인공지능 기술
- 생성형 AI 에 사용되는 파운데이션 모델은 기본 의무 외 추가 의무를 부담함: ① 파운데이션 모델 성격의 생성형 AI 서비스제공자는 이용자에게 AI 시스템과 상호작용하고 있는 사실을 알려야 함. ② 위법한 콘텐츠가 생성되지 않도록 파운데이션 모델을 학습, 설계, 개발하여야 함. ③ 저작권의 보호를 받는 학습데이터의 사용에 관한 구체적인 요약을 공개하여야 함

❑ 논의의 기초

- GAN, Stable Diffusion, LLM 마다 결론이 다를 수 있음
- 학습완료된 AI 시스템에는 학습데이터가 남아 있지 않다는 점, AI 개발자와 이용자가 분리되었다는 점에 유의
- INPUT 단계에서 공정이용 등의 저작권 침해가 면책되더라도 OUTPUT 단계에서 저작권 침해가 문제될 수 있음

❑ 인공지능 학습행위에서 발생하는 저작권 문제(INPUT 단계)

- 대규모 데이터 입력이 필요하므로 저작권 복제가 불가피함
- 공정이용 법리로 충분한가? TDM(Text/Data Mining) 면책 입법이 필요한가?

❑ 인공지능 출력행위에서 발생하는 저작권 문제(OUTPUT 단계)

- AI의 OUTPUT가 학습데이터의 2차적저작물(derivative works)인가? (e.g. Stable Diffusion)
- 현재 기술로는 AI 출력값을 보고 이것이 어떠한 입력값, 노드, 파라미터에 의해 생성되었는지 파악하기 어려우므로, 의거성 입증이 어려움

❑ AI 산출물에 대한 법적 보호 필요성

- 표현력이 높은 AI 산출물에 대해 법적 보호가 필요한가?

❑ 인공지능 학습과정(INPUT 단계)에서 저작권 침해

- 인공지능 학습단계 내지 TDM 과정에서 학습데이터를 수집하고 이를 인공지능 학습용 소프트웨어에 입력할 때 입력데이터의 복제가 발생
- 인공지능 학습은 데이터 전처리 (pre-processing) 을 거친 데이터 전부 내지 대다수를 학습데이터로 입력하는 경우가 많으므로 데이터에 대한 적법한 이용권한이 없다면 대규모의 저작권 침해가 발생
- 인공지능 학습은 콘텐츠의 '표현'을 습득하는 것이 아니라 데이터의 패턴, 상관관계를 분석하는 것이라는 점에서 단순히 복제권 침해로 보기 어려운 점이 있음
- 현행 공정이용 법리로 충분한지, 아니면 EU, 일본과 같이 입법적인 TDM 면책 조항이 필요한지 논의되고 있음

저작권법 전부개정안 (도종환 의원 대표발의) * 2022. 10.31. 이용호 의원 대표발의안도 제출됨

제 43 조(정보분석을 위한 복제·전송) ①컴퓨터를 이용한 자동화 분석기술을 통해 다수의 저작물을 포함한 대량의 정보를 분석(규칙, 구조, 경향, 상관관계 등의 정보를 추출하는 것)하여 추가적인 정보 또는 가치를 생성하기 위한 것으로 저작물에 표현된 사상이나 감정을 향유하지 아니하는 경우에는 필요한 한도 안에서 저작물을 복제·전송할 수 있다. 다만, 해당 저작물에 적법하게 접근할 수 있는 경우에 한정한다.

②제 1 항에 따라 만들어진 복제물은 정보분석을 위하여 필요한 한도에서 보관할 수 있다.

❑ 공정이용(fair use)

- 저작권법 제35조의5(저작물의 공정한 이용)에서 정한 고려사항
  - 문제된 이용의 목적 및 성격(제1요소)
  - 침해된 저작물의 종류 및 용도(제2요소)
  - 이용된 부분의 비중 및 중요성(제3요소)
  - 저작물의 이용이 시장에 미치는 영향(제4요소)
- 한국 대법원이 공정이용을 인정한 사례는 없음
- 미국 실무로 보면, 이용의 목적, 시장 영향 요건이 중요하고 특히 문제된 이용이 기존 저작물에 새로운 가치를 부여하는 변형적(transformative) 이용이어야 공정이용으로 볼 수 있다고 함
- Authors Guild v. Google 사건. 미국 법원은 검색을 용이하게 하기 위해 콘텐츠를 색인하고 책의 일부분(snippets)을 보여주는 행위를 매우 변형적인 행위로 판단
- Google v. Oracle 사건. 상업적인 라이선스 시장이 있더라도 문제된 이용이 저작권자의 잠재적인 손해에 비해 공공적인 이익을 증진시키는 경우에는 공정이용 인정 가능

❑ GitHub Copilot 소송(2022. 11.3. 제소)

- OpenAI는 GPT-3 모델을 이용하여 Codex라는 코드 자동생성 모델을 개발
- GitHub Copilot는 Codex를 기반으로 동작하는 AI 모델임. 주로 GitHub(MS가 인수)에 저장되어 있는 오픈소스 코드들을 학습데이터로 사용하였음
- GitHub, Microsoft, OpenAI를 상대로 class action이 제기됨
- 이 사건은 현재까지는 오픈소스 라이선스 위반 사건이므로, 공정이용 항변이 불가능함
- 최근 피고의 소각하 신청(motion to dismiss)이 기각되어 증거개시(discovery) 절차로 이행

❑ INPUT 단계에서 발생하는 라이선스 위반(원고 주장)

- 피고는 개별 오픈소스들이 채택하고 있는 라이선스(MIT, Apache, BSD, GPL 등) 의무사항을 위반함.
- 이러한 의무사항에는 공통적으로 저작권자 표시(attribution), 저작권 고지(copyright notice), 라이선스 조건 제공(license terms)이 있음
- 이에 더하여 GPL 등 상호적(reciprocal), 카피레프트적(copyleft) 성격이 강한 오픈소스 라이선스는 소스코드 수정시 동일 라이선스 조건으로 배포할 것을 요구
- Copilot 출력물에는 저작권 관련 표시가 없고 오픈소스 라이선스에 따라 배포되지도 않음

❑ 오픈소스 운동의 관점에서 본 Copilot의 문제점

- permissive 계열의 BSD, MIT, Apache 라이선스는 소스코드의 이용목적을 특별히 제한하는 바 없고 소스코드 수정 후에도 수정된 소스코드를 무료로 공개하라고 요구하지 않음(이들 라이선스는 주로 보증제한 정도를 규정한다)
- 반면 copyleft(reciprocal)적인 라이선스인 GPL은 이용목적을 제한하지 않지만 소스코드의 재배포 및 수정후 배포시 소스코드를 무상으로 공개할 것을 요구
- GPL의 관점에서 보면, GitHub가 오픈소스를 무상으로 학습(이용)하였음에도 불구하고 Copilot의 산출물을 유료로 공개하면서 이용자에게 오픈소스의 변형물을 독점적으로 이용할 수 있도록 하는 것을 부정적으로 볼 수밖에 없음

❑ SFC(Software Freedom Conservancy)의 반응

- MS(GitHub)에게 copyleft 등 FOSS 라이선스의 취지를 존중하고 오픈소스 커뮤니티와 협력할 것을 촉구하면서, 원고에게도 GPL 법집행 원칙 특히 재정적인 이득을 우선하지 않는다는 원칙을 준수하도록 요구함

❑ Stable Diffusion 소송

- Stable Diffusion은 LAION-5B 내지 LAION-Aestheics 이라는 공개데이터셋에 의해 학습되어 2022년에 출시된 text to image AI 모델임
- Getty Images가 영국 런던 및 미국 델라웨어에서 Stable AI에 대해 제소
- 미국에서 Stable AI에 대해 class action 발생

❑ INPUT 단계에서 발생하는 저작권 침해(원고주장)

- 게티 이미지의 1,200만개의 이미지가 학습데이터인 LAION-5B에 포함되어 복제됨.
- Class action의 원고들도 대체로 같은 주장임
- 다만 게티 이미지는 별도의 성명에서 자신이 보유하는 이미지들을 학습데이터로 라이선스 할 의향이 있다고 밝힘
  - Napster 사건을 의도?
- 그 외 Copyright Management Information(CMI)의 무단 제거 및 수정, 상표법 위반 등을 주장

❑ 게티이미지가 주장하는 침해의 증거

AI 출력과정에서 발생하는 저작권 침해 문제

❑ AI 산출물에 의한 저작권 침해 현상

- GitHub Copilot 모델에 특정 주석문 또는 함수명을 입력하면 학습에 제공된 일부 소스코드가 거의 그대로 출력되는 현상이 발생함
- ChatGPT은 짧은 분량의 유명한 작품을 거의 그대로 산출해 주며, 긴 문단의 경우에는 내용은 유사하나 표현 및 문장 순서가 상이한 결과물을 출력해 줌(Copilot와 ChatGPT는 거의 동일한 엔진임)
- Diffusion 모델은 학습데이터를 잘 기억하기 때문에 AI 출력물에서 원본 이미지가 추출될 수 있다는 연구 있음

❑ 의거관계 입증의 어려움

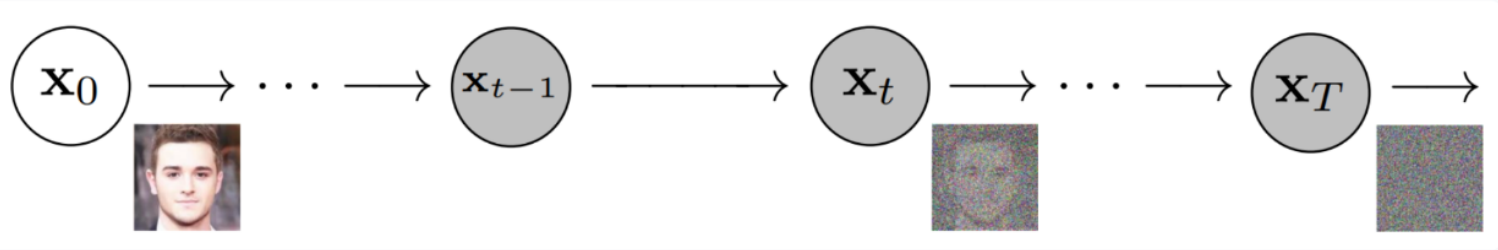
- 학습완료 모델에는 학습데이터가 남아 있지 않다는 점은 의거성 입증에 어려움을 줌
- 현재 기술수준으로는(특히 심층학습) 인공지능의 출력값이 어느 입력값, 노드, 가중치에 의해 산출되었는지 경로 파악이 불가능하므로, 이론적인 관점에서도 의거성 입증이 어려움
- 기계학습은 데이터들의 패턴, 특징을 추출하여 인식하는 것이고 특정 데이터의 표현에 의존하는 것이 아니라는 점에서 학습데이터에 저작물이 있었다고 하더라도 이론적 관점에서 의거성이 부정될 여지 있음

❑ 실질적 유사성 판단의 어려움

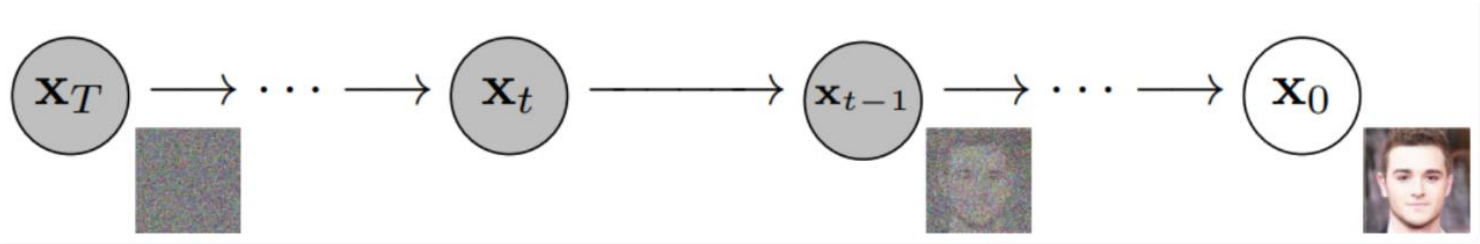
- 언어모델은 본래 암기를 하는 알고리듬이 아니나 GPT 모델의 경우 대규모의 파라메타를 가지고 있어서(GPT3은 1750억개 보유) 마치 타인의 저작물 내지 개인정보를 기억하는 것과 같은 외양을 띨 수 있음
- 생성형 AI에서 입력되는 프람프트(prompt)는 지시어, 유도어와 같은 역할을 한다고 볼 수도 있어서 간접책임의 성립이 문제됨
  - e.g. 특정 화가 풍으로 그려 달라는 프롬프트. 그러나 스타일이 유사하다는 점만으로는 저작권 침해가 인정되기 어려움
- 그러나 검토자마다 상이한 의견을 제시할 가능성 있음.
  - 예를 들어 Stable Diffusion의 경우 latent image interpolation 알고리듬의 성격상 이론적으로 실질적 유사성이 부정된다고 볼 여지 있음
- GitHub Copilot 사건의 원고 소장에서도 Copilot의 산출물의 1% 정도에서만 학습데이터로 제공된 오픈소스가 발견된다는 연구결과를 인용하고 있음
  - 학습데이터와 유사한 산출물의 빈도가 매우 적기 때문에 이를 두고 저작권 침해라고 할 수 있는지 문제됨

❑ Stable Diffusion의 기술적 내용

○ diffusion 테크닉(2015)



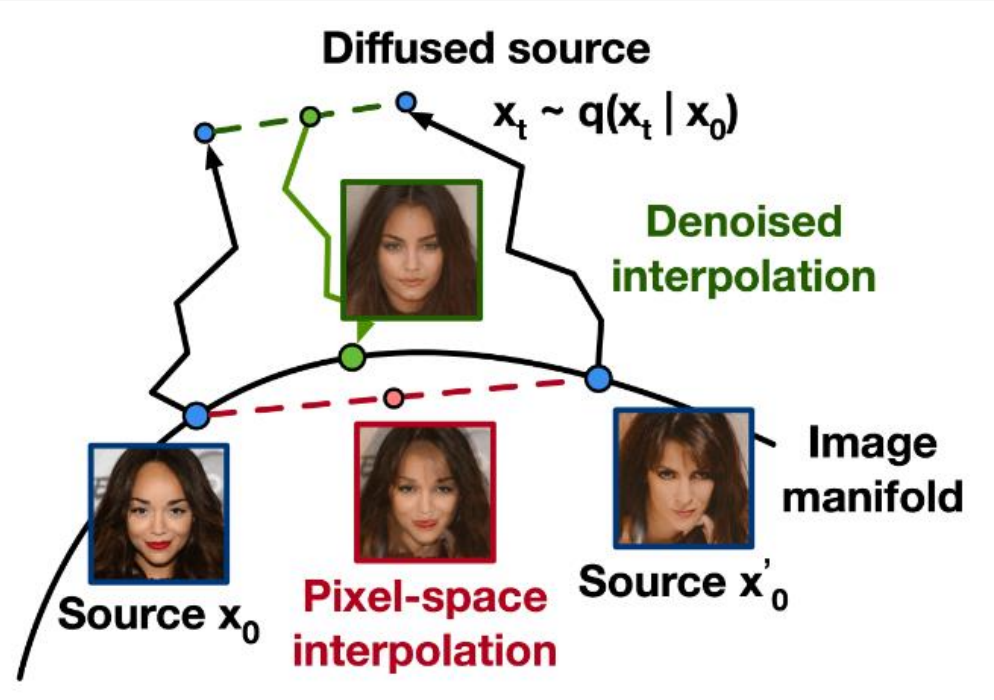
- AI는 이미지에 노이즈가 가해지면서 최종적으로 diffused 되는 과정을 학습



- 역순으로 AI는 노이즈 이미지에 대해 denoising 하는 과정을 학습하여 최종적으로 원본 이미지를 생성

❑ Stable Diffusion의 기술적 내용

- Interpolation with latent images(2020)



- 픽셀 공간 Interpolation된 왼쪽 그림(붉은색 박스)과 노이즈가 제거된 오른쪽 그림(초록색 박스)이 저작권법 관점에서 '실질적으로 유사'한지 문제됨.

- Stable Diffusion은 Latent image의 interpolation에 의해 새로운 이미지를 생성하므로 이론적으로도 실질적 유사성이 부정될 수 있는 여지 있음

❑ AI의 산출물(OUTPUT)에 대한 법적 보호 필요성

- 투하자본 회수 내지 산업진흥적 관점뿐만 아니라 학습데이터를 제공한 저작자를 배려하고 인간에 의한 창작을 고취해야 한다는 헌법적 내지 근본적 관점에서 검토되어야 할 것으로 생각됨
- 법적 보호가 필요하다는 견해에서는 저작권으로 보호, 저작인접권으로 보호, 독자적인 권리로 보호 등의 방안을 제안
- ChatGPT, Stable Diffusion 등 최근 생성형 인공지능을 종래 사진기, 워드프로세서, 포토샵과 같은 창작 보조 도구와 동일한 선상에서 바라볼 것인지의 문제이기도 함
- 만약 AI를 창작의 도구로 보고 AI 산출물에 대한 저작권법적 관점의 창작성을 논의한다면, 단계별로 프롬프트 입력단계의 창작성, 출력 단계의 창작성, 출력물 수정 단계의 창작성으로 구분하여 검토할 수 있음
- 표현력이 높은 AI 산출물을 창작성 있는 저작물로 보는 경우 저작자는 누구로 할 것인지 의견이 분분함
- AI 개발자에게 저작권을 부여하자는 견해, AI를 머신러닝하도록 한 자에게 저작권을 부여하자는 견해, AI 모델을 이용하여 산출물을 출력한 자에게 저작권을 부여하자는 견해 등 다양한 견해 존재

AI 산출물에 대한 법적 보호

- **USCO, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence(2023.3.16.)**
- **기본 입장**
  - AI 표현물(expressive materials)로 인정될 수준의 생성물이라고 하더라도 저작권법과 판례에 따라 사람 저작물로 등록될 수 없음
    - '표현물'은 만약 사람이 만들었다고 하면 저작물로 평가될 수 있는 수준의 AI 산출물을 말함
    - 사람의 창작물과 기계의 산출물을 구분하여 사람 창작물에 대한 적절한 인센티브를 부여
    - AI를 창작의 보조 수단 내지 도구를 사용한 경우에는 사람에 의한 창작을 인정
- **사람의 창작 부분과 AI의 콘텐츠를 구분하여 표시**
  - AI 콘텐츠가 포함되었다는 취지를 기술하고 Author Created 필드에 사람 창작 부분을 특정
  - 사람 창작 부분과 AI 콘텐츠를 창작적으로 배열(arrange)한 경우, Author Created 필드에 선택, 조정, 배열에 관한 사람과 AI의 기여 부분을 특정
- **정정신청 내지 보완등록**
  - 저작물등록이 진행중이거나 이미 완료된 경우에는 정정신청 내지 보완등록을 통하여 사람 창작 부분과 AI 산출 부분을 특정
  - 이를 위반한 경우 USCO가 등록을 취소하거나 침해소송법원이 등록효력을 무시할 수 있음

❑ ZARYA OF THE DAWN 사례

- Kashtanova가 Midjourney로 창작한 만화책을 USCO에 저작물 등록한 사례
- 최종적으로 USCO는 전체 만화책에 대한 저작물 등록을 취소하였으나, Kashtanova는 각 그림의 전체적인 배열 순서 및 대사에 대해서는 저작권을 행사할 수 있음

감사합니다.

麟 법무법인 린