

# SW·저작권 동향리포트

〈제2023-13호〉 2023년 11월 10일

## 정책/제도

### 생성형 AI 학습데이터의 저작권 침해 문제

#### ■ 개요

- 생성형 AI 모델은 마케팅, 소프트웨어, 디자인, 엔터테인먼트 및 대인 커뮤니케이션에 상당한 영향을 미치면서 콘텐츠 제작 방식을 변화시키고 있음. 관련 플랫폼이 상용화되어 널리 사용됨에 따라 학습데이터로 사용된 저작물의 권리침해 문제가 대두됨
- (생성형 AI 콘텐츠 제작 현황) 상용화된 AI를 통해 개인들 역시 블로그 게시물, 프로그램 코드, 시, 예술작품 등의 텍스트와 이미지 생성이 가능하며, 향후 더욱 복잡한 텍스트 생성을 향해 발전하고 있으며 극적인 대본, 각본, 이야기 및 기타 문학 작품을 생성할 것으로 예상됨
- (생성형 AI의 콘텐츠 제작방식) 생성형 AI 플랫폼은 학습할 대량의 데이터가 제공되는 훈련 단계와 해당 학습을 기반으로 새로운 콘텐츠 (텍스트, 이미지, 음악 및 비디오)를 생성하는 추론단계로 구성<sup>1)</sup>
  - 기계학습 알고리즘을 사용하여 대량의 데이터 스캔, 검토 및 학습 후, 패턴과 관계를 인식, 이를 사용하여 규칙을 만들고 프롬프트가 제공될 때 판단과 예측을 내림
  - 생성형 AI 모델은 기존 데이터를 단순히 가공하거나 분석하는 것이 아니라 새롭고 창의적인 콘텐츠를 생성가능하나, 최첨단 표절기술이라는 비판도 공존
- (생성형 AI 콘텐츠 제작 사례) Stable Diffusion, Midjourney 또는 DALL·E 2 와 같은 AI 이미지 생성기는 오래된 사진과 수채화부터 연필그림과 점묘법에 이르기까지 놀라운 스타일의 시각적 개체를 생성

1) Thomas H. Davenport and Nitin Mittal, 'How Generative AI Is Changing Creative Work', Harvard Business Review blog, Nov 14, 2022, <<https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>>

- 뉴욕의 현대미술관(Museum of Modern Art) 자체 컬렉션으로 학습한 AI가 생성한 설치물들을 전시하였으며, 헤이그의 Mauritshuis는 원본이 대여 중인 동안 AI가 생성해낸 Vermeder의 진주귀걸이를 한 소녀의 변형을 전시

**[그림 1] 생성형AI가 출력한 Vermeder의 진주귀걸이를 한 소녀의 변형**



## ■ 주요 내용

### 1. 생성형 AI 학습데이터의 저작권 침해<sup>2)</sup>

- (문제 제기) 이미지 생성형 AI의 경우 적법한 이용허락 없이 저작물이 포함된 학습데이터를 이용하여 원저작물과 유사한 생성물을 만들어냄. 저작자는 자신의 저작물이 AI의 학습과 성능향상에 활용됨에도 불구하고 이에 대한 아무런 대가나 보상이 없거나 심지어 고지받지 못하는 경우 발생
- (저작권 침해의 구성) 저작권자의 동의 없이 자료를 내려받고 AI 소프트웨어에 압축 저장한 것은 저작권법에 따른 복제권 침해에 해당할 수 있음
- (공정이용 법리의 적용 가능성) 학습데이터로 사용되는 저작물에 대해 공정이용 규정을 적용하는 것은 AI 기술발전과 산업진흥의 관점에서 적절하다고 주장될 수 있으나, 공익과 사익 간의 비교형량이 필요
  - (소프트웨어 개발과정에서의 공정이용 인정 판례) Google v. Oracle America 사건<sup>3)</sup>에서는 Google이 Android의 개발을 위해 Oracle이 소유한 Java SE의 API 패키지에 포함된 선언코드를 복제한 것이 문제가 됨. 미 대법원은 구글의 이용이 본질적으로

2) Gil Appel, Juliana Neelbauer, and David A. Schweidel, 'Generative AI Has an Intellectual Property Problem', April 07, 2023, <<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>>

3) Google LLC v. Oracle America Inc., 593 U.S. \_\_\_ (2021).

상업적임에도 불구하고 구글이 복제코드를 재구성하여 스마트폰이라는 새로운 환경을 위한 창의적이고 혁신적인 시스템을 제작한점, 새로운 시스템에 의해 새로운 제품 및 프로그래밍 기회가 생성된 점 등을 고려하며 진보적이고 변형적인 이용이라고 인정

## 2. 최근 사례

- (국내 사례) 네이버웹툰에 연재되는 ‘신과 함께 돌아온 기사왕님’은 앞서 작가가 작화 일부에 AI를 활용한 것이 알려지며 웹툰 작가들의 거센 반발을 샀으며, 카카오페이지의 한 웹소설도 표지 디자인이 생성형 AI를 활용해 만들어진 것으로 뒤늦게 알려짐
- 지난 6월, 네이버웹툰 도전 만화 페이지에 “도둑질로 만든 AI 웹툰을 반대한다”는 게시물로 작가들의 보이콧이 이루어짐<sup>4)</sup>

**[그림 2] 생성형 AI를 활용한 작업에 반대하는 작가들의 네이버웹툰 보이콧**



- (Stability AI 사건) 런던에 본사를 둔 Stability AI는 텍스트 입력에서 이미지를 생성하는 AI 기반 시스템인 Stable Diffusion과 이미지 생성기 DreamStudio를 2022년 8월 출시. 저작권이 있는 저작물을 허락 없이 학습데이터로 사용한 것에 대해 소송이 제기되어 진행 중에 있으며, 알려진 것은 크게 2건임.
- (Andersen et al. v. Stability AI et al.) 2023년 1월 13일 원고(Sarah Andersen, Kelly McKernan, and Karla Ortiz)는 DreamUP의 배후에 있는 Midjour

4) 윤정민, "내 작품 도둑질 그만"...네이버웹툰서 'AI웹툰 보이콧' 운동 쇄도', NEWSIS, 2023.06.06., <[https://www.newsis.com/view/?id=NISX20230604\\_0002327215&cID=13006&pID=13100](https://www.newsis.com/view/?id=NISX20230604_0002327215&cID=13006&pID=13100)>

ey Inc, DeviantArt Inc, 그리고 Stable Diffusion을 출시한 Stability A.I.Ltd  
 사를 상대로 소를 제기<sup>5)</sup>

[그림 3] DreamUP을 사용하여 AI로 만든 이미지



- 원고는 피고들의 AI 플랫폼(text-to-image platforms)이 자신들의 동의 없이 작품을 훈련데이터로 사용. 불법적으로 스크랩하고 ‘콜라주’ 함으로써 저작권 및 불공정경쟁법을 위반하였다고 주장
- 특히 이 플랫폼 서비스는 독일 회사가 만들고 Stability AI가 의뢰한 LAION-Aesthetics dataset에 의존하는데, 이 데이터 세트의 47%가 Shutterstock 및 Getty Images와 같은 스톡 사진 사이트, Pinterest를 포함한 쇼핑 사이트, Flickr 및 Tumblr와 같은 사용자 생성 콘텐츠를 호스팅하는 웹사이트의 이미지로 구성되어 있음. 원고들은 자신들의 작품 역시 LAION 데이터세트에 사용된 것으로 주장
- (Getty Images (US) Inc. v. Stability AI Inc.) 2023년 2월 3일, 스톡 포토(stock photo) 플랫폼 업체인 게티 이미지(Getty Images)가 이미지 생성형 AI인 Stable Diffusion을 개발한 Stability AI를 상대로 자사 보유의 이미지 1,200만 개 이상이 무단으로 사용된 것에 대하여 저작권 침해를 원인으로 소송을 제기<sup>6)</sup>

### 3. 기술적 보호 조치의 다각화

- (AI 학습 취약점 이용한 디지털 독) 인공지능이 창작자의 허가 없이 접근해 정보를 학습할 경우 인공지능 모델의 신뢰도와 정확성이 치명적 손상을 입게 되는, 새로운 저작권 보호 도구, ‘나이트 셰이드’가 개발됨<sup>7)</sup>

5) Min Chen, ‘Artists and Illustrators Are Suing Three A.I. Art Generators for Scraping and ‘Collaging’ Their Work Without Consent’, ARTNET NEWS, Jan 24, 2023, <<https://news.artnet.com/art-world/class-action-lawsuit-ai-generator-s-deviantart-midjourney-stable-diffusion-2246770>>; 소장은 <<https://stablediffusionlitigation.com/pdf/00201/1-1-stable-diffusion-complaint.pdf>>

6) Blake Brittain, ‘Getty Images lawsuit says Stability AI misused photos to train AI’, REUTERS, Feb 7, 2023, <<https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/>>; 소장은 <<https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlkmwnve/GETTY%20IMAGES%20AI%20LAWSUIT%20complaint.pdf>>

- 지난 23일 ‘엠아이티(MIT) 테크놀로지 리뷰’에 따르면, 시카고대 벤 자오 교수는 예술가가 창작물에 눈에 보이지 않는 ‘디지털 독’을 첨가할 수 있는 ‘나이트셰이드’를 개발
- 방대한 양의 데이터를 수집해 학습하는 생성 인공지능 모델의 보안 취약점을 이용하는 구조로 창작자가 작품을 온라인에 공개하면서 나이트셰이드를 이용해 독성 코드를 심으면, 해당 데이터를 수집한 인공지능 모델 안에서 독이 나머지 데이터로 퍼져나가 전체 모델이 오염되는 방식. 한번 오염된 데이터는 삭제하기 매우 어렵기 때문에 인공지능 모델은 치명적 손상을 입게 됨
- 스테이블 디퓨전을 대상으로 테스트한 결과, 개 이미지에 독을 심어 공개한 뒤 이를 학습한 인공지능에게 “개 이미지를 만들라”라고 요구했더니, 팔다리가 많은 괴물 같은 생물을 만들어내고 더 많은 독을 주입했더니 고양이를 그려냄

[그림 4] ‘나이트셰이드’를 이미지 생성AI 스테이블 디퓨전에 테스트한 결과



\*출처 : [https://www.hani.co.kr/arti/economy/it/1114108.html?\\_fr=mt2](https://www.hani.co.kr/arti/economy/it/1114108.html?_fr=mt2)

- (AI 작성 이미지 식별 기술) 구글 딥마인드가 AI로 생성된 이미지에 보이지 않는 워터마크를 삽입하는 신스ID(SynthID)라는 도구를 공개<sup>8)</sup>
  - 일반적인 워터마크와는 달리 눈에 보이지 않아 원본 이미지를 훼손하지 않으며,

7) 구본권, ‘콘텐츠 바다에 ‘저작권 독물’ 살포...AI는 개 대신 고양이를 그렸다’, 한겨레 2023.10.30., <[https://www.hani.co.kr/arti/economy/it/1114108.html?\\_fr=mt2](https://www.hani.co.kr/arti/economy/it/1114108.html?_fr=mt2)>  
 8) AI 리포터, ‘구글, AI 생성 이미지에 워터마크 심는 '신스ID' 공개’, Digital Today, 2023.11.8., <<https://www.digitaltoday.co.kr/news/articleView.html?idxno=486557>>

인공지능만 식별할 수 있는 투명한 워터마크

- 초기 신스ID는 버텍스 AI(Vertex AI) 플랫폼과 이미지 생성 AI인 이메진(Imagen)을 사용하는 고객에게 제공될 예정이나, 구글은 신스ID를 비디오 및 텍스트 등 여러 종류의 미디어에도 확장하고 온라인 표준으로 만드는 것을 목표로 하고 있음
- 구글뿐 아니라 메타, 오픈AI, 구글 및 기타 주요 AI 기업은 최근 자사의 AI 기술에 대한 보호 및 안전 시스템을 강화하고 있으며, 일부 회사는 암호화된 메타데이터를 사용해 AI 생성 콘텐츠에 태그를 달기 위한 C2PA 프로토콜 프로젝트에 참여하고 있음
- **(AI Content Detector)** AI가 작성한 텍스트를 탐지하여 디지털 표절을 잡아내는 플랫폼들이 등장하고 있음. 인터넷과 다양한 데이터베이스에서 중복되는 콘텐츠를 찾아내 독창성을 확보할 수 있도록 도와주며, 영어뿐만 아니라 스페인어, 프랑스어 등 다양한 언어에서 콘텐츠 감지가 가능<sup>9)</sup>
  - 언어마다 차이는 있겠으나 AI 글 감별 정확도는 98% 이상으로 알려짐
  - 로이터통신, 뉴욕타임스, 블룸버그 등 언론사들은 인공지능 봇의 정보수집을 차단한다고 밝힘<sup>10)</sup>
  - 현재 출시된 AI 작성물 탐지 프로그램은 Original.AI(구글), GPT Zero, Zero GPT, HuggingFace GPT-2 출력 검출기 데모, AI Detector(Content at Scale), AI Content Detector(Copyleaks), Kazan SEO 등이 있음

#### 4. 국내외 정책논의

- **(해외 논의)** 영국, 독일, 일본 등에서 저작권 관련법의 개정을 통하여 정보 분석을 위한 복제·전송을 명시적으로 허용하고
  - **(EU DSM저작권지침)** 「디지털 단일시장 저작권 지침(Directive of the European Parliament and of the Council on copyright in the Digital Single Market, 이하 DSM 저작권지침)」을 제정하여 데이터마이닝에 관한 통일적 기준을 마련하는 등 데이터 관련 산업의 활성화와 저작물의 공정한 이용을 보장.
  - **(EU 일반정보보호규정(GDPR))** 조직이 개인 데이터를 사용하는 방법을 정의
- **(국내 논의)** 문화체육관광부는 올해 2월부터 9월까지 ‘AI-저작권법

9) 사단법인 유엔미래포럼, ‘5개의 최고 AI 콘텐츠 탐지기’, 2023.5.1., <<http://www.unfuture.org/1591>>

10) 구분권, 앞의 글.

제도개선 워킹그룹’을 발족하고 ▲ AI 학습데이터에 사용되는 저작물의 원활한 이용 방안, ▲ AI 산출물의 법적 지위 문제 및 저작권 제도에서의 인정 여부, ▲ AI 기술 활용 시 발생하는 저작권 침해와 이에 대한 책임 규정 방안 등을 논의한 바 있음<sup>11)</sup>

- (국내 법률개정안 발의 현황) 최근 데이터마이닝 과정에서 사용된 저작물의 공정이용 규정의 내용을 골자로 하는 저작권법 일부개정법률안이 발의되었음

\* 저작권법 일부개정법률안(의안번호 : 2124685), 이인영의원 대표발의, 2023.9.25.)

### 저작권법 일부개정법률안

[신설] 제35조의5(교육·조사·연구 목적의 정보분석을 위한 복제·전송 등)

① 컴퓨터를 이용한 자동화 분석 기술을 통하여 추가적인 정보 또는 가치를 생성하기 위하여 다수의 저작물을 포함한 대량의 정보를 분석(규칙, 구조, 경향 및 상관관계 등의 정보를 추출하는 경우를 말한다. 이하 이 조에서 "정보분석"이라 한다)하는 것으로 다음 각 호의 요건을 모두 충족하는 경우에는 필요한 범위 안에서 저작물을 복제·전송할 수 있다.

1. 대학·연구기관 또는 그 밖에 대통령령으로 정하는 기관이나 단체에서 수행할 것
2. 교육·조사·연구 등 비상업적 목적으로 이용할 것
3. 정보분석의 대상이 되는 해당 저작물에 적법하게 접근할 것

② 제1항에 따라 저작물을 복제하는 자는 정보분석을 위하여 필요한 한도 내에서 복제물을 보관할 수 있다. 이 경우 저작권 및 그 밖에 이 법에 따라 보호되는 권리의 침해를 방지하기 위하여 복제방지 및 보안 등 대통령령으로 정하는 필요한 조치를 하여야 한다.

## ■ 시사점

- AI 학습데이터에 사용되는 저작물의 원활한 이용과 관련 분쟁 예방을 위해, AI 기술 활용 시 학습데이터로 사용되는 저작물의 저작권 침해 문제와 책임 규정 방안 등을 국가 정책 차원에서 논의할 필요가 있음

11) 문화체육관광부 보도자료, 'AI 시대 새로운 저작권 해법 찾기 워킹그룹 첫 회의', 2023.2.24.

## 참고자료

- Blake Brittain, ‘Getty Images lawsuit says Stability AI misused photos to train AI’  
<https://www.reuters.com/legal/getty-images-lawsuit-says-stability-ai-misused-photos-train-ai-2023-02-06/>
- Gil Appel, Juliana Neelbauer, and David A. Schweidel, ‘Generative AI Has an Intellectual Property Problem’  
<https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>
- Google LLC v. Oracle America Inc., 593 U.S. \_\_\_\_ (2021)
- Min Chen, ‘Artists and Illustrators Are Suing Three A.I. Art Generators for Scraping and ‘Collaging’ Their Work Without Consent’  
<https://news.artnet.com/art-world/class-action-lawsuit-ai-generators-deviantart-midjourney-stable-diffusion-2246770>
- Thomas H. Davenport and Nitin Mittal, ‘How Generative AI Is Changing Creative Work’  
<https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work>
- AI 리포터, ‘구글, AI 생성 이미지에 워터마크 심는 ‘신스ID’ 공개’  
<https://www.digitaltoday.co.kr/news/articleView.html?idxno=486557>
- 구분권, ‘콘텐츠 바다에 ‘저작권 독물’ 살포…AI는 개 대신 고양이를 그렸다’  
[https://www.hani.co.kr/arti/economy/it/1114108.html?\\_fr=mt2](https://www.hani.co.kr/arti/economy/it/1114108.html?_fr=mt2)
- 문화체육관광부 보도자료, ‘AI 시대 새로운 저작권 해법 찾을 워킹그룹 첫 회의’
- 사단법인 유엔미래포럼, ‘5개의 최고 AI 콘텐츠 탐지기’  
<http://www.unfuture.org/1591>
- 윤정민, ‘“내 작품 도둑질 그만“…네이버웹툰서 ‘AI웹툰 보이콧’ 운동 쇄도’  
[https://www.newsis.com/view/?id=NISX20230604\\_0002327215&cID=13006&pID=13100](https://www.newsis.com/view/?id=NISX20230604_0002327215&cID=13006&pID=13100)



**SPC 'ANGEL' 통계**

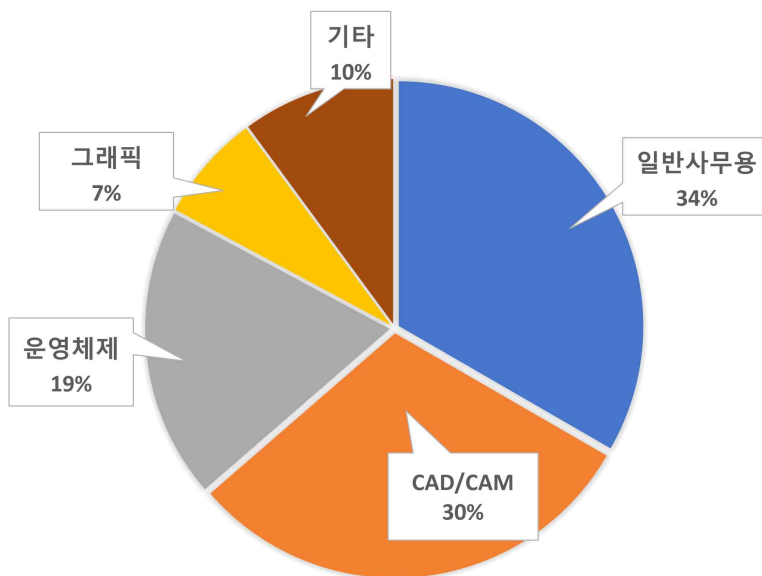
월 1회 제공

**한국소프트웨어저작권협회, 불법복제 SW 제보 'ANGEL' 서비스 9월 통계 현황**

- 한국소프트웨어저작권협회(SPC)가 지난 9월 한 달간(' 23. 9. 1. ~ 9. 30.) 'ANGEL (불법제보)' 서비스를 분석한 결과, 기업 또는 개인의 불법복제 SW 사용 제보는 총 81건으로 나타나 전년도 같은 기간 대비 88% 증가
- SW 용도별로는 일반사무용 14건(33%), 설계(CAD/CAM) 13건(30%), 운영체제 8건 (19%), 그래픽 3건(7%), 기타 5건(10%) 순으로 제보가 접수된 것으로 나타남

**[그림] SPC 'ANGEL(불법제보)' 서비스 2023년 9월 통계 현황**

2023.9. 불법복제 소프트웨어 제보 통계  
-SW 용도별 제보 건수-



\* 한국소프트웨어저작권협회는 2018년 11월부터 제보시스템과 제보 방식의 편의성을 개선한 불법복제 SW 제보시스템 'ANGEL(불법제보)' 서비스를 운영하고 있음

다음 SW·저작권 동향리포트 <제2023-14호> 발간일은 12월 10일입니다.