

논문 2024-3-1 <http://dx.doi.org/10.29056/jsav.2024.09.01>

# 생성형 AI와 저작권 침해 이슈의 기술적 고찰

윤영선\*†

## Technical consideration for generative AI and copyright infringement issues

Young-Sun Yun\*†

### 요 약

인공지능 기술의 급속한 발달로 생성형 AI의 생성결과물에 대한 저작권 관련 논의가 활발히 제기되고 있다. 대표적인 논의는 학습에 사용된 데이터의 취득과 생성 결과물의 저작물 인정 및 침해 등이다. 본 논문에서는 이외에도 생성형 AI를 구성하는 기술적 요소들에서 발생할 수 있는 저작권 침해 우려를 기술하였다. 전반부에서는 대표적인 AI 활용 영역인 텍스트, 이미지, 음향/음성 부분에서의 생성형 AI의 구현 방법을 간단하게 살피고, 후반부에서는 기술적 요소에서 발생할 수 있는 저작권 침해 사례를 검토하였다. 기존의 연구가 학습과 결과물의 측면에서 저작권 관련 논의에 접근하였다면, 본 논문에서는 생성형 AI를 구성하는 기술적 요소에서의 저작권 관련 논의를 진행하였다는 데 의의가 있다고 할 수 있다. 본 논의를 통하여 기술적 요소의 저작권 관련 내용이 확립되었으면 한다.

### Abstract

With the rapid development of artificial intelligence technology, the copyright of generative AI outputs is being debated. The main issues are the acquisition of data used for training and the attribution and infringement of the resulting works. This paper describes other copyright infringement concerns that may arise from the technical components of generative AI. In the first half of this paper, I briefly review how generative AI is implemented in text, image, and sound/voice, which are typical application areas, and in the second half, I review the copyright infringement cases that may arise from the technical elements. While previous studies have approached the copyright discussion from the perspective of training and output, this paper is significant in that it discusses copyright in the technical components of generative AI. I hope that this paper has established the copyright content of technical elements.

**한글키워드** : 생성형 AI, 저작권 침해, 모델 구조, 가중치 파일, 프롬프트

**keywords** : Generative AI, copyright infringement, model architecture, weighted files, prompt

\* 한남대학교 정보통신공학과

† 교신저자: 윤영선(email: ysyun@hnu.kr)

접수일자: 2024.08.10. 심사완료: 2024.09.03.

게재확정: 2024.09.20.

## 1. 서론

2016년 알파고의 등장으로 대중의 관심을 끌  
게 된 인공지능 기술은 2023년 발표된 ChatGPT

로 정점을 찍으며, 일상에서 쉽게 접할 수 있는 단어가 되었다. 현재는 산학연 모든 분야에서 적극적으로 관련 기술이 활용되고 있으며, 적용되지 않은 분야를 찾기 어려울 정도이다.

인공지능은 컴퓨터의 개념이 등장했을 때부터 언급된 개념으로, 사람처럼 고차원적인 정보처리를 수행할 수 있는 기계를 만들기 위한 노력에서 출발하였다. 인공지능은 사람과 같이 생각하고 판단하는 지적 능력을 인공적으로 구현하기 위한 시스템 또는 프로그램으로 정의할 수 있다.

인공지능의 발전은 신경회로망 기술의 발전과 밀접한 연관이 있다. 1958년 프랭크 로젠블랫(Frank Rosenblatt, 1928-1971)은 사람 뇌의 동작과 구조를 모방한 뉴런 시스템으로 퍼셉트론을 개념화하였으며[1], 이후 다양한 후속 연구와 대체 연구가 진행되었다. 단층 퍼셉트론의 선형성을 극복하기 위하여 제안된 다층 퍼셉트론 시스템 또한 계산량이나 구현의 제약으로 인해 성능 면에서 한계를 겪기도 하였다. 기존 연구의 약점을 극복하기 위한 다양한 연구가 꾸준히 진행되었으며, 컴퓨팅 기술과 병렬처리 기술의 발전으로 대용량 네트워크를 활용한 딥러닝 기술이 태동하여, 텍스트, 비디오, 음성 등 여러 분야에서 사람의 능력을 초과하는 결과를 가져오게 되었다[2].

1998년 합성곱 신경망 기술의 제안으로 이미지 분류에서의 성능 향상의 발판이 구축되었다[3]. 특히, 2014년 제안된 가변 오토인코더와 생성형 적대 신경망은 이미지 생성 분야에서 중요한 역할을 하였으며, 딥러닝 기술은 이미지 분류와 인식 분야 외에도 이미지 생성 분야의 중요 연구 방법론으로 자리 잡았다[4, 5]. 더불어 2017년 트랜스포머의 등장으로 텍스트와 순차적 시계열 패턴 연구가 가속화되었고, GPT(Generative Pre-trained Transformer)와 같은 모델이 등장하여 대중적인 생성형 AI의 시대를 열었다[6].

생성형 AI는 주어진 데이터나 조건에 따라 텍

스트, 이미지, 오디오, 음향 신호 등 다양한 형태의 콘텐츠와 아이디어를 생성하는 인공지능의 일종이며, 방대한 양의 데이터를 기반으로 사전 훈련된 대규모 모델을 활용하여 생성 작업이 이루어진다. 최근의 생성형 AI는 콘텐츠 제작뿐만 아니라 디지털 이미지 개선, 동영상 편집, 제조용 프로토타입 제작, 데이터 보강 등 다양한 작업에 활용되고 있으며, 그 영향 또한 커지고 있다[7].

생성형 인공지능은 여러 이점을 살려 다양한 분야에서 활발히 활용되고 있으나, 수반되는 주요한 우려 중 하나는 저작권 침해의 가능성이다. 본 연구는 생성형 AI가 생성하는 주요 결과물인 텍스트, 이미지, 코드, 음악 등에서 발생할 수 있는 저작권 침해 가능성 및 그 유형을 분석하고자 한다. 최근에는 생성형 AI를 구성하는 모델, 코드, 가중치 등의 저작권 인정 여부와 복제, 라이선스 미준수에 따른 저작권 침해 등도 고려되고 있어 기술적 단계에서 AI의 저작권 침해 이슈를 다루하고자 한다.

본 논문은 2장에서 생성형 AI의 활용과 적용되는 기술을 간략하게 서술하고, 3장에서는 생성형 AI에서 발생할 수 있는 저작권 이슈, 4장에서는 생성형 AI를 구성하는 기술적 요소의 저작권 침해 가능성을 검토하고, 마지막으로 결론을 맺는다.

## 2. 생성형 AI의 활용과 적용 기술

생성형 AI 분야는 크게 텍스트, 이미지, 음향/오디오로 구분할 수 있다. 각각의 응용 분야에 따라 다시 텍스트에서도 시·소설 작성과 같은 문학 분야, 문서의 요약, 프로그램 코드 작성 등 다양한 활용 분야로 세분화할 수 있지만 기본적인 이론은 유사하므로 해당 기술을 먼저 살펴보고 저작권 분쟁의 가능성을 검토한다.

## 2.1 텍스트

생성형 AI를 적용하여 텍스트를 생성하는 것은 다음과 같은 활용/기술 분야로 나눌 수 있다.

- 언어 모델(language models): 주어진 텍스트의 문맥을 파악하여 단어나 문장의 빈도 또는 활용 규칙에 따라 정규화된 문장을 생성하는데 사용될 수 있다. GPT-3와 같은 언어 모델은 대규모 텍스트 데이터를 학습하여 문맥과 의미를 이해하고 다음 단어나 문장을 예측할 수 있다.
- 자동 완성(auto-completion): 사용자가 입력한 텍스트에 대해 자동으로 나머지 문장을 완성한다. 사용자가 일부분만 입력한 후 그 의도를 파악하여 나머지를 추가하기 때문에, 이메일, 문서 작성, 코드 구현 등에서 활용될 수 있다.
- 텍스트 생성(text generation): 주어진 주제나 지시에 따라 새로운 텍스트를 생성하는 기술로써, 글 작성, 시나리오 생성, 소셜 쓰기 등에 사용될 수 있으나, 할루시네이션<sup>1)</sup>이 발생할 수 있다.
- 대화 모델(conversational models): AI와 자연스러운 방식으로 대화하는 기술로써, 챗봇이나 가상의 상담사로 사용되어 질문에 답변하거나 대화를 진행할 수 있다.
- 요약(summary): 긴 텍스트를 간결하게 요약하는 기술로, 중요한 내용을 추출하여 요약문을 생성한다. 뉴스 기사나 논문, 긴 문서의 요약에 사용될 수 있다.
- 번역(translation): 한 언어로 작성된 텍스트를 다른 언어로 번역하는 기술이다. 실시간 번역이나 문서 번역에 활용될 수 있다.

1) 할루시네이션(hallucination) : 원래는 환각이라는 뜻이지만, 생성형 AI에서 존재하지 않은 사실을 사실처럼 답변을 생성하는 것, 즉 무의미하거나 충실하지 않게 생성된 콘텐츠를 의미한다.

- 스타일 변환(style transfer): 주어진 텍스트의 스타일을 변경하여 다른 문체로 변환하는 기술이다. 글쓰기 스타일을 고전 양식, 논문식, 서술식, 에세이 등으로 변환할 수 있다.
- 문장 완성 및 수정(sentence completion and editing): 불완전한 문장을 완성하거나 텍스트를 수정하여 더 자연스러운 문장으로 만들 수 있다.

이외에도 다양한 분야에 텍스트 기반 생성형 AI가 활용될 수 있으며, 기본적으로 생성형 AI에 사용되는 대표적인 구현 기술은 Transformer[6] 기반의 BERT(Bidirectional Encoder Representations from Transformers)와 GPT로 나눌 수 있다.

BERT는 2018년에 개발된 모델로, 양방향 Transformer 인코더를 기반으로 동작한다. 이 모델은 문장 안에서의 단어 간 상호 작용을 파악하기 위해 양방향 학습을 수행하는 것이 주요 특징으로, 문맥을 모델링하기 위해 왼쪽과 오른쪽의 문맥을 모두 고려하여 단어의 임베딩을 생성한다. 따라서 BERT는 양방향 학습을 통해서 문맥을 잘 파악할 수 있지만, 새로운 문장을 생성하는데는 한계를 가지고 있으므로, 텍스트 분류, 개체명 인식, 질의/응답 등에 많이 사용되고 있다.

GPT는 BERT와 마찬가지로 Transformer 아키텍처를 기반으로 하지만, 단방향 Transformer 디코더를 사용하여 텍스트를 생성하는데 중점을 둔 방식으로, 문장을 왼쪽에서 오른쪽으로 생성하며, 이전 단어들을 사용하여 다음 단어를 예측한다. GPT는 BERT와 달리 텍스트 생성 작업에 특화되어 있으므로 문장 생성 및 완성에 뛰어난 성능을 보인다. 단방향 학습이 적용되어 일부 생성 문맥이 의도와 다른 경우도 발생할 수 있으므로, BERT에 비교하여 할루시네이션 경향이 강하다고 할 수 있다.

기술의 특성상 텍스트 생성에서 저작권 침해는 GPT 기반의 자동 완성, 텍스트 생성 과정에

서 발생 가능성이 높다. BERT나 GPT 모델의 경우 생성 모델을 대규모 텍스트 데이터로 사전 학습(pre-training)한 후, 후속 작업을 통해 특정 작업에 맞게 세부 조정(fine tuning)한다. 자동 생성/완성의 경우에는 주어진 문장이나 전제(premise)가 높은 신뢰도로 학습에 사용/조정되었거나 반복적으로 학습에 사용된 경우, 동일한 문장이나 코드가 생성될 수 있다. 또한, 텍스트 생성의 경우에도 주어진 주제나 지시(프롬프트<sup>2)</sup>)에 따라 학습에서 사용된 내용이 노출될 수 있다. 이 경우, 생성형 AI가 생성한 결과물은 영 상태(zero state)에서 생성된 것이 아니고, 주어진 전제 조건에 의하여 생성되기 때문에, 저작권 침해가 발생한다면 사용자의 책임을 제외하고 전적으로 생성형 AI를 개발한 개인이나 기관이 책임을 져야 하는지는 논의가 필요하다.

## 2.2 이미지/영상

이미지 분야에서 생성형 AI의 활용은 이미지 생성, 편집 및 변형 등 다양하게 활용되고 있다. 주요 활용 영역은 다음과 같다.

- 이미지 생성(generation): 대표적인 생성형 AI 활용 분야로, 고해상도의 사실적인 이미지를 생성하여 게임이나 만화, 영화(비디오)에 적용할 수 있다.
- 이미지 편집 및 변형(editing/transform): 사진이나 일반적인 이미지를 특정 화가나 유사 이미지로 변환하는 방식이다. 특정 패턴을 모사하거나 과장하여 새로운 패턴을 추가하기도 한다.

2) 프롬프트(prompt)란 생성형 AI 모델이 특정 작업을 수행하거나 특정 종류의 출력을 유도하는 입력 텍스트를 의미한다. 프롬프트에 따라 생성 결과물의 품질이 달라지고 할루시네이션이 적게 발생하기 때문에 프롬프트 자체도 창작물(저작물)로 인정받고자 하는 노력도 존재한다.

- 이미지 복원 및 보정(restoration): 오래된 사진이나 손상된 이미지를 복원하거나, 사진의 피사체를 제외한 객체의 제거, 배경 변경 등에 적용되며, 저화질 영상을 초고해상도로 변환하기도 한다.
- 이미지 캡션 또는 설명 생성(caption): 이미지를 입력받아 이미지 내용을 설명하는 텍스트를 생성하는 분야이다.
- 이미지 합성(synthesis): 여러 이미지를 합성하여 새로운 이미지를 생성한다. 또는 정적 이미지로부터 동영상을 구현하거나 3차원 이미지로 변환한다. 이미지 생성이 텍스트 입력 기반이라면, 이미지 합성은 텍스트와 이미지 모두 입력으로 사용된다는 차이가 있다.

이미지 분야에서 사용되는 생성형 AI는 VAE(가변 오토인코더, Variational Auto Encoder)[4], GAN(생성적 대립/적대 신경망, Generative Adversarial Networks)[5], Flow기반 생성 모델[8], Stable Diffusion[9] 방식이 대표적인 모델로 현실적이고 다양한 이미지를 생성할 수 있다.

VAE는 입력과 출력을 같게 만들기 위하여 평균과 분산으로 구성되는 잠재 공간(latent space)을 구축하고, 이 잠재 공간으로부터 원하는 이미지를 생성(또는 decoding)하는 생성 모델이다. 기존의 오토인코더(AE)는 입력된 데이터를 압축한 후 다시 복원하는 구조로 압축된 데이터를 생성하는데 사용(인코더가 목적임)되지만, VAE는 확률 분포로부터 이미지를 생성하는 데 주로 사용된다. 즉, VAE는 데이터의 특징을 확률적으로 파악하고 그 분포를 기반으로 새로운 이미지를 생성하는 모델이다. 기존의 오토인코더가 원본 이미지를 잘 복원할 수 있는 잠재 변수를 찾는 게 목적이라면, VAE는 원본 이미지를 잘 생성할 수 있는 잠재 공간을 구성하는 것이 목적이며, 인코더를 학습하기 위하여 디코더를 추가하거나

(AE), 디코더를 학습하기 위하여 인코더를 추가 (VAE)하는 방법론의 차이가 발생한다. VAE는 흐릿하거나 품질이 낮은 이미지를 생성하는 경향이 있다고 알려졌으며, 데이터의 구조를 포착하기 위한 저차원 공간인 잠재 공간이 복잡하고 작업하기 어렵다는 약점이 있어, 고품질 이미지나 잠재 공간에 대한 명확한 이해가 중요한 애플리케이션에서 VAE의 효과가 제한될 수 있다고 한다.

GAN은 위조지폐범과 경찰의 예시로 잘 알려져 있으며, 데이터를 생성하는 생성자(generator)와 생성된 데이터가 실제 데이터와 얼마나 유사한지를 판별하는 판별자(discriminator)로 구성되어 서로 교차하면서 학습하는 방식이다. 학습이 종료된 후에는 생성자를 이용하여 원하는 이미지를 생성하게 되며, 목시적으로 확률 분포를 모델링하기 때문에 모델 구성에 제한이 없으며 생성된 결과물의 품질이 뛰어나다고 알려져 있다. 다만, 생성자와 판별자를 같이 학습해야 한다는 점, GAN에 BCE(Binary Cross-Entropy) loss를 사용할 때 생성자가 다양한 이미지를 생성하지 못하고 비슷한 이미지만 계속해서 생성하는 모드 붕괴(Mode Collapse) 등과 같이 학습에 어려움이 있을 수 있다.

VAE나 GAN은 주로 인코딩(encoding)과 디코딩(decoding)을 수행하는 두 개의 독립적인 신경망을 학습하는 방식으로 구성된다. 이와는 대조적으로, Flow 생성 모델은 입력 영상을 잠재 공간 또는 임베딩 벡터로 변환하는 역변환 가능한 함수를 찾는 방식으로 작동하며, 이 함수를 역변환하여 원래의 입력 영상을 생성한다. 요약하면, GAN이나 VAE에서는 독립적인 인코딩 및 디코딩 모델을 사용하는 반면, Flow 생성 모델은 잠재 공간으로의 변환 함수를 학습하고, 이 함수를 역변환하여 이미지를 생성하는 방식이다. Flow 생성 모델은 데이터의 특성을 보존하면서 저차원에서 고차원 공간으로의 변환을 가능하게 하며,

생성된 변환 함수를 통해 새로운 데이터를 생성하므로, 고품질의 이미지를 생성할 수 있게 된다.

최근에 이미지 생성 모델로 널리 사용되는 모델은 안정 확산(stable diffusion) 모델이다. 안정 확산모델은 흐름 기반 생성 모델과 유사하게 직접 변환 방식을 채택하고 있으며, 기존의 GAN이나 VAE와는 다른 인코딩/디코딩 방식을 적용하고 있다. 확산모델은 데이터로부터 잡음을 조금씩 더해가면서 완전한 잡음 모델을 만드는 전방향 과정(forward process)과 반대로 잡음으로부터 조금씩 원 영상을 복원해 가는 역방향 과정(reverse process)으로 구성된다. 확산모델은 비평형 열역학(non-equilibrium thermodynamics)에서 영감을 받았으며[10], 영상에서 잡음으로 변환하는 과정을 마르코프 연쇄(Markov chain)로 정의하고, 그 역변환을 통하여 잡음으로부터 영상을 얻는 방식이다.

이미지 생성 모델을 학습시키기 위해서는 텍스트 응용과 비슷하게 원본 입력 데이터가 필요하다. 따라서 이미지 생성 모델은 원본 입력 영상을 어떻게 수집하였느냐가 저작권의 침해 여부와 관련성이 있다고 할 수 있다. 텍스트 분야와 다르게 사용자가 생성 결과물이 학습데이터의 저작권 침해 여부를 쉽게 확인할 수 있으므로, 허락되거나 승인되지 않은 학습데이터의 사용은 바로 모델에 반영될 수 있으므로 이미지 생성 모델의 저작권 이슈는 학습데이터의 수집, 공정한 이용 또는 이용 허락에 있다고 할 수 있다.

### 2.3 음향/음성

음향/음성 생성 분야에서도 텍스트/프롬프트 기반의 생성 모델을 제안하는 연구가 활발히 진행되고 있다. 오디오/음성 분야에 적용된 기술이나 활용 분야는 데이터를 처리하는 과정에서 많은 차이가 있지만, 핵심 부분인 생성 모델은 이미지 분야와 많은 유사점이 존재한다.

- 음성 합성(synthesis): 음성인식 분야와 더불어 연구의 역사가 매우 길며, 활용 분야 또한 다양하다. 텍스트를 입력받아 음성을 생성한다.
- 음성 변환(transform): 입력받은 사람의 목소리를 다른 사람의 목소리로 변환하거나, 특정 스타일로 변환한다. 음성 합성과 결합하여 소실된 목소리를 복원하는 데 사용하기도 한다.
- 오디오 생성 및 변환(audio generation) : 새로운 음악을 작곡하거나 특정 스타일로 음악을 생성한다.
- 소리 합성(sound synthesis): 이미지 합성과 유사하게 소리를 입력받아 다양한 소리를 생성하는 기술로 게임, 영화, 가상현실과 같이 가상 세계의 소리를 생성한다.
- 음성 강화 및 잡음 제거(enhancement): 음성 신호에서 배경 소음을 제거하거나 억제하고, 특정인의 목소리만 강조하기도 한다.

음향/음성 생성 모델과 이미지 생성 모델의 공통점은 텍스트와 비텍스트로 구성된 데이터를 이용하여 학습하고, 텍스트를 입력하면 비텍스트를 생성한다는 것이다. 음향/음성 처리분야에서는 음향을 효과적으로 나타내기 위하여 시간과 주파수 특성을 표현한 스펙트로그램을 이용하여 음향학적 특징을 표현하고, 생성된 스펙트로그램으로부터 음향/음성/음악 등을 생성하고 있다. 영상 등의 경우에는 시각적으로 모델의 변이 등을 쉽게 파악할 수 있으나, 음향/음성/음악 등은 음파의 전압 차에 의하여 출력을 확인하므로 모델과의 상관성이나 그 효과를 명확하게 확인할 수 없다는 어려움이 따른다. 음향/음성 신호는 시간의 변화에 따라 특징이 변화하며, 공간·시간적으로 지역성(locality)을 가지기 때문에, 초기에는 음향/음성에 특화된 통계적인 방법이 널리 사용되었다. 점차 음성/음향에 특화된 고유의 신경회로망 및 딥러닝 모델로 발전된 후 최근에는 VAE,

GAN, Diffusion, Transformer 방식 등 다양한 생성 모델이나 그 기술이 적용되면서 생성물의 결과물 또한 향상되고 있다.

### 3. 생성형 AI와 저작권 이슈

앞 장에서 살펴본 바와 같이 생성형 AI는 다양한 분야에서 사람의 능력에 버금가거나 능력을 웃도는 결과물을 생성하고 있다. 최근 들어 생성형 AI의 결과물이 저작물인지 또는 저작권으로 인정받을 수 있는지 많은 논의가 진행되고 있다. 생성형 AI의 결과물이 창작물(저작권을) 인정받기 위해서는 생성물의 출처나 창작자가 분명한 작품으로부터 변형 또는 창작적 요소가 가미된 것인지, 또는 출처가 불분명하더라도 기존의 저작물과 차별화된 창작물로 인정받아야 한다. 그러나 생성형 AI가 결과물을 생성하기 위해서는, 학습이 필요하며, 이 과정에서 학습을 위한 많은 양의 데이터가 필요하다는 제한이 존재한다. 따라서 생성형 AI 기술 분야에서 저작권 분쟁 또는 침해가 발생할 수 있는 경우는 학습에 사용된 데이터의 이용과 생성 결과물의 창작성에 따른 두 가지 경우를 고려할 수 있으며, 이외에도 생성형 AI를 구현하는 기술적 결과물에 대한 검토가 필요하다. 즉, 학습에 사용되는 데이터, 생성형 AI의 결과물, 그 모델을 구축하는 데 필요한 기술 또는 학습 결과물의 저작권 이슈에 대해 논의가 필요하다. 데이터와 생성 결과물에 대한 논의는 지금까지 학계나 산업계에서 지속해서 논의되고 있는 부분이지만, 기술적 요소는 논의가 미흡한 실정이다.

생성형 AI는 학습 과정에서 많은 양의 데이터가 필요하며, 이는 수집하는 과정에서 저작권 보호를 받는 저작물의 복제 및 전송 행위가 수반될 수 있다. 저작물 수집 과정에서 저작자의 이용 허락을 받지 않을 때는 저작권 분쟁이 일어날 수

있는데, 공정이용이나 TDM 면책 규정<sup>3)</sup>, 또는 사적 계약을 도입하여 해결할 수 있다[11].

다음으로 생성형 AI의 결과물이 저작권을 침해할 가능성을 살펴보자. 생성형 AI가 학습데이터를 그대로 복제(데이터 복제, data replication)하는 경우는 매우 적은 확률로 가능하다고 알려져 있다. 만약 학습에 사용된 데이터가 저작권 분쟁 소지를 가지고 있는 경우에는 당연히 생성물 또한 저작권 분쟁 대상이 될 수 있다. 데이터 복제의 경우에는 학습된 데이터가 그대로 출력되는 데 비하여, 학습된 이미지와는 다른 원저작물을 변형하여 새로운 저작물을 생성하였다면 원저작자의 2차 저작물 작성권 침해가 될 수 있다 [12]. 또한 공공 데이터<sup>4)</sup>만을 사용하였으나 특정 저작권자의 저작물과 유사하다면, 생성물과 저작물의 유사성 결과만을 가지고 저작권 위반 여부를 판단하기는 쉽지 않아 보인다. 저작권법 위반을 다루기 위해서는 접근성(의거성)과 실질적 유사성 모두를 만족하여야 하기 때문이다. 따라서 저작권 위반 여부를 판단하기 위해서는 학습에 사용된 데이터의 분석으로부터 의거성 여부를 확인할 필요가 있다.<sup>5)</sup> 또한 결과물의 저작권 침해를 다루기 위해서는 결과물을 생성하는 프롬프트의 역할 또한 무시할 수 없다. 프롬프트는 생성형 AI의 결과물의 품질을 결정하는 중요한 요소

로써, 기술적 측면에서 살펴본다면 결과물을 생성하는 사용자의 창작성이 표현된 입력이라 할 수 있을 것이다.

마지막으로 고려할 사항은 생성형 AI를 구현하는 기술적 요소들에서 발생가능한 저작권 위반 요인이다. 학습데이터나 결과물이 아닌 소프트웨어 소스 코드<sup>6)</sup>의 경우에는 원 저작자와 관련된 저작권 위반 여부가 아니라 동일 또는 유사 기능을 구현하는 두 개의 생성형 AI 시스템 또는 프로그램의 저작권 분쟁 가능성을 고려한 것이다. 따라서 생성형 AI를 구성하는 기본 구성, 학습된 모델(모델 가중치 파일), 생성형 AI를 구현하는 소스 코드 간의 실질적 유사성을 판단할 수 있다면 저작권 침해로 인정할 수 있을 것이다. 최근의 연구 경향은 연구 논문 발표 시, 연구에 사용된 프로그램을 같이 공개하거나 코드의 재현을 요구하고 있다. 반면에 상업적 목적으로 개발된 시스템(프로그램)의 구성 요소는 공개하지 않거나 특정 조건(또는 라이선스)에서만 이용하기를 허용하고 있다. 따라서 공개되지 않은 두 시스템이 동일/유사 기능을 제공하고 있으며 구성 요소 간의 실질적 유사성이 발견된다면 저작권 분쟁의 이슈가 발생할 가능성이 높다. 이에 따라 4장에서 기존 연구에서 언급되지 않은 저작권 분쟁 요소를 살펴본다.

3) 공정이용(fair use)은 저작권자의 허락없이도 저작물을 제한적으로 허용할 수 있도록 허용하는 원칙이며, TDM(Text and Data Mining) 면책 규정은 웹 크롤링 등 불특정하게 수집되는 저작물에 대하여 저작권 침해의 예외를 인정하는 규정이다.

4) 저작권 보호기간이 종료되거나 특정 라이선스 조건 유/무에 따라 누구나 사용할 수 있는 데이터

5) 본 논문에서는 생성형 AI의 결과물을 저작물로 인정할 것인지 여부는 판단하지 않고, 결과물의 저작권 위반 위험 가능성만을 다룬다.

#### 4. 생성형 AI 구성 요소의 저작권 침해 가능성 검토

인공지능 모델이 적법한 방법으로 학습된 후 그 학습 모델이나 모델 구조가 공개 또는 비공개

6) 생성형 AI에 의하여 작성된 소스 코드의 저작권은 학습데이터의 공개 여부에 따라 창작성의 인정 여부가 중요 쟁점이 될 것이며, 여기에서는 생성형 AI를 구축하는 시스템의 소스 코드를 말한다.

될 때 저작권이 침해되는 가능성을 살펴볼 수 있다. 모델 구조는 논문 등에서 공개하는 경우가 많아, 인공지능 모델의 신규성이나 창작성을 인정받기 어려울 수 있으므로, 공개된 모델 구조의 유사성은 저작권 침해로 보기 어려울 수 있다. 그러나 학습된 결과물인 가중치나 소스 코드의 복제, 공개되지 않은 기존 모델 구조가 배포된 가중치 파일에서 복원되어 사용된다면 저작권 침해 가능성을 생각할 수 있다.

**가) 미공개 모델 구조의 복제·사용**

AI 모델의 내부 신경망 구조는 사용 목적에 따라 다양하게 구성될 수 있다. 시계열 패턴이나 이미지, 텍스트, 관심 영역 등에 따라 적용 구조가 다르며 공개된 모델의 부분 사용, 강조, 변형 등에 따라 성능의 차이가 발생한다. 이런 점을 고려할 때 앞으로 최신 연구나 매우 중요한 영역의 연구 결과는 미발표되거나 지연 발표될 가능성이 존재한다[13]. 따라서 앞으로는 미공개 모델에 대한 취득 위협이 증가할 것이며, 이 경우 AI 시스템에 사용된 모델 간의 저작권 분쟁이 발생할 수 있다. 모델 구조를 저작물로 볼 것인지의 논란은 존재할 수 있지만, 특허권, 영업비밀 보호법을 통해서도 모델 구조는 보호받을 수 있기 때문에 개발자/회사의 저작물(창작물)로 판단하고 저작권의 보호를 받는 것으로 가정한다.

모델 간의 유사도를 구하기 위해서는 전체 모델 구조, 일부 구조, 또는 노드의 구조를 비교하여야 한다. AI 시스템의 모델 구조는 동적 그래프 또는 정적 그래프로써 표현되기 때문에 모델 간의 비교는 그래프 유사도를 통하여 계산될 수 있으며, 의거성이 반영된다면 저작권 침해 여부도 확인할 수 있을 것으로 보인다.

**나) 학습 가중치 결과를 그대로 사용하는 경우**  
가중치 결과가 공개되는 경우와 공개되지 않

은 경우로 구분할 수 있다. 먼저 가중치 결과물이 특정 조건이나 라이선스 하에서 사용을 허용할 때는 관련 조건이나 라이선스를 준수하여 사용 시에는 문제가 없으나, 그 조건을 위반하거나 라이선스를 미준수할 때는 저작권 문제가 아닌 계약이나 라이선스 위반 등이 발생할 수 있다. 반대로 공개되지 않은 학습 가중치 결과를 임의로 사용할 때는 저작권법 이외에도 영업비밀 보호법, 계약법, 특허법, 데이터 보호법 등 여러 가지 문제가 발생할 수 있다<sup>7)</sup>.

**다) 원본 가중치 파일을 복사하여 다른 모델의 초깃값으로 사용하거나 변형하여 사용하는 경우**

나) 항에서 기술한 특정 조건이나 라이선스의 허용 범위를 벗어난 경우를 고려한다. 가중치 파일은 모델 구조와 밀접한 관계가 있으므로 모델의 배포 당시 정한 조건이나 라이선스를 위반하고 변형하여 저작권 침해나 다른 법률의 적용을 회피하고자 할 수 있다. 가중치 파일은 AI 시스템/프로그램의 창작자(개발자)가 특정 조건에서 창작한 고유한 결과물로 간주할 수 있으므로 무단 사용 또는 변형의 형태로 사용하려고 할 수 있기 때문이다. 동일한 데이터를 이용하더라도 학습 환경에서 적용하는 다양한 변수를 개발자가 조절하여야 하며, 동일한 매개변수 환경에서도 완전히 동일한 결과물을 생성할 수 없으므로 학습 가중치 파일 또한 생성형 AI의 고유 결과물로 볼 수 있으며, 이는 생성형 AI의 저작물이 아닌 개발자의 저작물로 판단하여야 한다고 생각한다.<sup>8)</sup>

7) 관련 법 적용 여부는 저자의 개인적 견해이다.

8) AI 프로그램의 개발자를 창작자로 간주하고 가중치 파일을 개발자의 저작물로 인정할지는 논의의 대상이다.



그러나 가중치 파일이 개발자의 저작물로 인정받지 못할 때는 저작권 법외에 영업비밀이나 특허법(특허로 인정받는 경우), 데이터 보호 등을 고려하여야 할 것이다. 다만, 가중치 파일은 이진 파일이기 때문에 조그마한 환경변화에도 동일성을 판단하기 어려울 것이다. 따라서 학습 환경에 적용된 매개변수나 데이터 조합 등을 함께 한 경우라면 비슷한 성능의 다른 가중치 파일을 생성할 수 있으며 이는 저작권법 대신 영업비밀 등의 방법으로 보호해야 할 것으로 보인다. 다만 유사한 환경에서 학습을 진행하여 유사한 가중치 파일을 생성한 것이 아니라, 타 시스템의 학습 가중치 파일을 직접 변형 또는 조작한 것은 저작권 침해로 다룰 수 있을 것이므로, 가중치 파일의 벡터 특성 등을 반영한 유사성 비교 연구가 필요할 것으로 보인다.

최근 학습 과정에서 학습 결과물을 평균하거나 가중 결합하여 사용하여 성능을 높이려는 시도가 많아, 특정 가중치 파일이 사용되었는지 단순 비교만으로는 확인하기 어렵다. 따라서 저장된 대응 모델 구조의 가중치를 비교하여 그 경향을 비교하거나, 생성 결과물 또는 인식 결과를 분석하여 의거성에 따라 저작권 침해 여부를 고려해야 할 것이다.

#### 라) 적정하지 않은 방법으로 취득한 학습데이터를 데이터 정제에 이용

적정한 계약이나 동의 없이 무단으로 데이터를 취득하여 AI 학습에 사용하는 경우는 데이터의 부족 현상을 해결하기 위한 가장 쉬운 방법이다. 생성형 AI의 약점 중의 하나는 데이터 복제로 특정 조건이 맞으면 낮은 확률이라도 학습에 사용된 데이터와 유사한 결과를 얻을 수 있다는 것이다. 이런 문제를 해결하기 위하여 데이터 정제(filtering) 작업을 진행할 수 있는데 가장 쉬운 방법은 저작권으로 보호받는 데이터를 이용하여

학습에 사용한 후, 관련 데이터를 정제하는 경우를 고려할 수 있다. 적절한 방법으로 저작권 보호 데이터를 사용하더라도, 라이선스에 따라 AI 학습에 사용할 수 없는 경우가 있을 수 있다. 이를 위하여 전체 학습데이터에서 포함된 저작권 데이터를 정제하기 위하여 학습에 사용하는 경우를 저작권 침해로 볼 수 있는지는 논란이 될 수 있다. 또는 직접적으로 저작물을 사용하지 않고, 적법한 방법으로 사용 허가를 취득한 생성형 AI의 생성 결과물로 학습에 사용하는 경우와 저작물이 포함된 기존 학습된 가중치 모델에서 저작물 전체 또는 핵심 영역을 모델에서 제외(subtract)하는 효과를 가지도록 조율(tuning)하는 경우를 생각할 수 있다. 이들 경우에는 자신의 저작권 침해 사실을 은폐하기 위하여 2차 저작물을 사용하거나, 사후 작업을 통하여 모델을 변경하는 경우라고 할 수 있는데 현실적으로 생성된 가중치 파일만 가지고는 저작권 침해 사실을 입증하기는 매우 어려울 것으로 보인다.

#### 마) 프롬프트의 저작물 인정

프롬프트는 생성형 AI의 결과물을 유도하는 입력 텍스트로서 특정 생성물을 얻거나 결과물의 품질을 높이기 위하여 매우 중요한 역할을 수행한다. 생성형 AI의 결과가 사람의 고유한 창작물로 인정받기 위해서는 생성형 AI의 입력인 프롬프트가 사람의 창작적 기여인지를 살펴봐야 할 것이다[14]. 만약 프롬프트가 사람의 창작적 기여라면, 생성형 AI의 결과물은 프롬프트를 작성한 사람의 창작물로 인정받아야 할 것이며, 프롬프트가 일반적인 단순 요구조건이라면 생성 결과물은 논란의 여지가 많지만, 사람의 창작물이 아닌 AI의 창작물로 보아야 할 것이다. 위 논의에서 프롬프트를 사람의 창작 기여물로 인정한다면 문장의 단순, 복잡함을 넘어서 프롬프트 또한 사용자의 창작물로 보고 저작권으로 보호해야 할 것

이다. 따라서 소량의 텍스트로 구성된 프롬프트의 창작성을 판단하기는 쉽지 않을 수 있지만, AI의 결과물을 창작물로 인정하기 위해서는 프롬프트의 창작성을 반드시 확인하여야 할 것으로 보인다.

본 장에서는 생성형 AI를 구현하기 위한 모델 구조, 학습 결과물인 가중치, 결과물을 생성하기 위한 입력인 프롬프트에서 발생할 수 있는 저작권 분쟁 가능한 이슈를 살펴보았다. 여기서 언급되지 않은 상황도 존재할 수 있으나 기술적 관점에서 고려해야 할 사항을 정리하였다.

## 5. 결론

인공지능 기술, 특히 생성형 AI의 빠른 확산과 깊은 관심으로 생성 결과물에 대한 보호 및 저작권 침해에 대한 논의가 활발하게 진행되고 있다. 최근 논의의 대부분은 생성형 AI의 모델을 학습하기 위하여 사용된 데이터와 생성 결과물에만 집중된 경향이 있다. 이에 따라 본 논문에서는 기존의 논의 외에 기술적 측면에서 생성형 AI 시스템에서 발생할 수 있는 저작권 분쟁 요소를 살펴보기 위하여, 대표적인 적용 분야인 텍스트, 이미지, 음성/음향 부분에 적용된 기술을 간략하게 정리하였다. 이를 바탕으로 생성형 AI의 기술적 구현 요소에서 발생할 수 있는 저작권 분쟁 사례를 검토하였다.

생성형 AI와 관련된 저작권 분쟁은 개발과 사용의 양 측면에서 살펴봐야 할 것이다. 개발자의 생성물과 학습 과정의 생성물, 사용자의 생성물은 적용단계에서 명확히 구분되지는 않을 수도 있지만, 서로 다른 관점에서 창작물(저작물) 논의가 필요할 것으로 보인다.

생성형 AI의 성능을 결정하는 중요한 요소는

모델 구조, 학습 결과인 가중치 파일의 기술적 요소와 그리고 추론(또는 결과 생성)을 위한 프롬프트 입력으로 구분할 수 있다. 이 중 모델 구조와 가중치 파일은 개발 측면에서 프롬프트는 사용 측면에서 접근할 수 있다. 모델 구조는 적용하고자 하는 분야와 밀접하게 관련되어 있으며 공개되지 않은 모델 구조는 개발자의 창작물로 간주할 수 있으므로 보호받아야 할 것이다. 또한 동일한 구조의 AI 시스템이라 할 지라도 학습 환경이나 사용된 데이터에 따라 동일한 이진파일을 생성하지 않기 때문에 이 또한 시스템 개발자의 저작물로 인정받아야 할 것이다. 마지막으로 프롬프트는 생성형 AI 결과물의 품질이나 원하는 생성물을 위해 중요한 입력이 되며, 사용자의 창작성이 표현될 수 있으므로 AI 결과물이 창작물로 인정받을 수 있다면, 그 입력 또한 창작물로 인정하여야 할 것이다.

본 논문에서 논의한 내용은 아직 산업계나 학계 등에서 정리되지 않았기 때문에 저자의 의견과 다른 해석이 충분히 존재할 수 있을 것으로 보이며, 검토된 내용이 기술적 관점에서 인공지능과 저작권 관련 논의의 출발이 되기를 희망한다.

이 논문은 2023학년도 한남대학교  
학술연구비 지원에 의하여 연구되었음

## 참고 문헌

- [1] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 65(6), 386-408, 1958, DOI: <https://doi.org/10.1037/h0042519>
- [2] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning. *Nature*", 521(7553), 436-444, 2015,

- DOI: <https://doi.org/10.1038/nature14539>
- [3] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, 86(11), 2278-2324.1998, DOI: <https://doi.org/10.1109/5.726791>
- [4] D. P. Kingma, M. Welling, "Auto-Encoding Variational Bayes", arXiv preprint arXiv:1312.6114, 2013, DOI: <https://doi.org/10.48550/arXiv.1312.6114>
- [5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, et al. "Generative adversarial nets", In Advances in neural information processing systems(pp. 2672-2680), 2014, DOI: <https://dl.acm.org/doi/10.5555/2969033.2969125>
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al., "Attention is all you need", In Advances in neural information processing systems(pp. 5998-6008). 2017, DOI: <https://dl.acm.org/doi/10.5555/3295222.3295349>
- [7] Wikipedia, Generative Artificial Intelligence, [https://en.wikipedia.org/wiki/Generative\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Generative_artificial_intelligence), Accessed 2023/08/09
- [8] Lilian Weng, Flow-based Deep Generative Models, <https://lilianweng.github.io/posts/2018-10-13-flow-models/>, Accessed 2023/08/09
- [9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models", Proceedings of 2022 IEEE/CVF CVPR, pp. 10674-10685, 2022, DOI: <https://doi.org/10.1109/CVPR52688.2022.01042>
- [10] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics", Proceedings of 2015 ICML, vol 37. pp. 2256-2265, 2015, DOI: <https://dl.acm.org/doi/10.5555/3045118.3045358>
- [11] C.Koo, "Generative AI and copyright issues", Broadcasting Culture 2023 Winter, p.30-43, 2023, <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11635571>
- [12] Y.Shin, "Copyright issues and responses to generative AI", Cultural information issue report, vol. 2023-2(no. 42), 2023, KCISA, <https://webzine.kcisa.kr/vol019/>
- [13] N. Tiku, G. De. Vynck, Google shared AI knowledge with the world - until ChatGPT caught up, The Washington Post, May 5, 2023, <https://www.washingtonpost.com/technology/2023/05/04/google-ai-stop-sharing-research/>, Accessed 2024/08/09
- [14] Y.Kim, "Is Prompt Generation Creative under Copyright Law? -Focusing on the Legal Consideration of Creative Use of Generative AI-", Justice no.200, pp. 261-297, 2023, DOI: <https://doi.org/10.29305/tj.2024.2.200.261>

저 자 소 개



윤영선(Young-Sun Yun)

2001.2 KAIST 전산학과 박사  
 2006.4-2007.2 한국전자통신연구원 초빙 연구원  
 2012.8-2013.7 University of Washington 방문학자  
 2001.3-현재 한남대학교 교수  
 <주관심분야> 음성인식, 음성변환, 화자인식, 인공지능, 저작권침해, 유사도, 완성도 감정, 오픈소스 등