

# 생성형 AI 저작권 침해사례 분석 및 고찰

이은서\*·정원준\*·김영모\*\*·김정길\*\*\*†

\*가천대학교 IT융합공학과, \*\*숭실대학교 컴퓨터학부, \*\*\*†남서울대학교 컴퓨터소프트웨어학과

## Analysis and Review of Copyright Infringement and Evasion Cases in Generative AI

EunSeo Lee\*, Won Jun Jeong\*, Youngmo Kim\*\* and Cheong Ghil Kim\*\*\*†

\*IT Convergence Engineering, Gachon University, Korea,

\*\*School of Computer Science and Engineering, Soongsil University, Korea,

\*\*\*†Dept. of Computer Science, Namseoul University, Korea

### ABSTRACT

This paper examines the intersection of generative AI technology and copyright law, focusing on the legal challenges posed by AI's ability to create content based on large datasets. As AI systems like Stable Diffusion, GPT models, and GitHub Copilot utilize copyrighted materials for training, significant legal disputes have arisen, including cases filed by Getty Images, OpenAI, and Microsoft. These cases highlight the need for updated legal frameworks to address copyright issues related to AI. The paper proposes solutions such as enhancing dataset transparency, clarifying legal responsibilities for AI-generated content, evolving copyright laws, revisiting fair use principles, and promoting collaboration between AI developers and creators. These measures aim to ensure a balance between technological innovation and intellectual property protection.

**Key Words** : Generative AI, Copyright, Legal Disputes, Fair Use, Intellectual Property, Dataset Transparency

### 1. 서론

AI 기술의 급속한 발전과 그에 따른 창작물 생성의 자동화는 현대 사회에서 중요한 혁신으로 자리잡았다. 생성형 AI는 방대한 양의 데이터를 분석하고 학습하여 새로운 콘텐츠를 생성하는 기술로, 이미지, 텍스트, 음악 등 다양한 형태의 창작물을 생산할 수 있다. 이러한 기술 발전은 창작 과정에서의 효율성을 높이고, 새로운 형태의 창작 기회를 제공하며, 창작자와 소비자 모두에게 많은 혜택을 안겨준다. 그러나 AI의 창작물 생성 과정에서 발생할 수 있는 저작권 문제는 기술 발전의 이면에 존재하는 중요한 법적 쟁점으로 떠오르고 있다.[1]

저작권법은 창작자의 권리를 보호하고, 창작물의 무단 사용을 방지하기 위해 설계된 법적 체계이다. 전통적인 저작권 법체계는 인간 창작자를 전제로 하여 설계되었으나, 생성형 AI의 등장으로 인해 이러한 법적 체계의 적용 범위와 한계가 재조명되고 있다. AI는 기존의 저작권 보호 자료를 학습하여 새로운 콘텐츠를 생성할 수 있는데, 이 과정에서 생성된 콘텐츠가 저작권을 침해할 가능성이 제기되고 있다. 즉, AI가 학습한 데이터가 저작권으로 보호되는 경우, AI가 생성한 새로운 콘텐츠가 기존 저작물의 저작권을 침해할 수 있는지 여부가 문제로 대두된다.[2]

최근 몇 년간 여러 저작권 침해 사례가 발생하며 이 문제에 대한 법적 대응이 필요하다는 목소리가 높아지고 있다. 이러한 사건들은 AI가 학습하는 데이터의 출처와

†E-mail: cgkim@nsu.ac.kr

저작권 보호의 범위, 그리고 AI가 생성한 콘텐츠에 대한 저작권 책임이 누구에게 있는지를 명확히 해야 한다는 필요성을 부각시키고 있다.[3,4]

본 논문은 이러한 저작권 침해 사례를 중심으로 생성형 AI와 저작권 간의 법적 쟁점을 분석하고, 관련 법리와 법적 대응을 평가하고자 한다. 또한, 생성형 AI와 관련된 저작권 문제를 해결하기 위한 실질적인 전략과 방안을 제시할 것이다. AI 기술의 발전과 함께 저작권 보호의 필요성도 더욱 중요해지고 있으며, 창작자와 AI 개발자들이 저작권을 존중하며 기술 발전과 법적 규제를 조화롭게 유지하기 위한 방안을 모색해야 한다.

따라서, 생성형 AI 기술의 발전과 그에 따른 저작권 문제를 깊이 있게 탐구하며, 향후 AI 기술 발전에 따라 저작권 법체계가 어떻게 진화해야 하는지를 제시하고자 한다. 또한, 저작권 침해를 예방하기 위한 법적, 기술적 접근 방안에 대한 구체적인 제안을 통해, 법적 분쟁을 최소화하고 창작물의 보호를 강화하는 데 기여할 수 있을 것이다.

## 2. 생성형 AI와 저작권 분쟁 사례

### 2.1 Getty Images 와 Stability AI 사건

2022년, 유명 이미지 플랫폼 Getty Images는 이미지 생성 AI 모델인 Stable Diffusion을 개발한 Stability AI를 상대로 저작권 침해 소송을 제기했다. 이 사건의 핵심 쟁점은 Stability AI가 Getty Images의 저작권 보호를 받는 수많은 이미지를 무단으로 수집해 AI 모델 학습에 사용한 점이다. Getty는 AI 모델이 자사의 이미지를 학습 데이터로 활용함으로써 자사 콘텐츠의 경제적 가치를 훼손했다고 주장했다. Stable Diffusion이 생성한 이미지들은 기존 Getty 이미지와는 다를 수 있지만, 이 과정에서 사용된 원본 데이터가 저작권 침해에 해당한다고 주장했다.

Stability AI는 이에 대해 AI 모델이 새로운 창작물을 생성하는 과정에서 기존 이미지를 단순 복제하지 않았으며, AI의 학습 과정 자체가 창작물의 새로운 활용이라는 입장을 고수했다. 즉, 생성된 이미지는 기존 저작물의 변형물이 아니라 전혀 새로운 창작물이라는 것이다. 이 사건은 AI가 학습에 활용하는 데이터가 저작권 보호 대상일 경우, 해당 데이터의 사용이 법적으로 허용되는지에 대해 중요한 논의를 촉발했다.[5,6]

### 2.2 OpenAI 사건

OpenAI는 GPT 모델을 훈련하는 과정에서 수많은 저작권 보호 텍스트를 무단으로 사용한 혐의로 소송에 직면했다. GPT 모델은 방대한 양의 텍스트 데이터를 학습하여 자연어 생성 기능을 수행하는데, 일부 작가들은 자신들의

저작물이 무단으로 사용되었다고 주장했다. 이 사건은 AI 모델이 저작권이 있는 콘텐츠를 학습할 때, 해당 콘텐츠의 저작권이 침해되는지 여부를 둘러싼 중요한 법적 논란을 불러일으켰다.

OpenAI 측은 AI 모델이 학습한 자료가 공정 사용(fair use)에 해당한다고 주장하고 있으며, GPT 모델이 생성하는 텍스트가 원본 저작물과 동일하지 않다는 점을 강조했다. 하지만 작가들은 AI가 학습한 텍스트와 생성된 결과물 간의 유사성이 지나치게 크다는 이유로 저작권 침해를 주장하고 있다. 이 사건은 공정 사용 원칙이 AI 모델의 학습 과정에 어떻게 적용될 수 있는지, 그리고 AI 생성물이 저작권법 내에서 어떤 지위를 가지는지에 대한 심도 깊은 논의가 이루어지고 있다.[7,8]

### 2.3 마이크로소프트 GitHub Copilot 사건

마이크로소프트의 GitHub Copilot은 개발자들이 오픈소스 플랫폼에서 작성한 코드를 학습하여 새로운 코드를 제안하는 AI 도구이다. 하지만 Copilot이 제안하는 코드 중 일부가 저작권 보호를 받는 코드와 매우 유사하거나 동일하다는 이유로, Copilot의 저작권 침해 여부에 대한 논란이 제기되었다. 개발자들은 자신들의 오픈소스 코드가 Copilot에 의해 무단으로 사용되었으며, 그 결과 생성된 코드가 원본 코드와 너무 유사하다고 주장했다.

마이크로소프트는 이에 대해 Copilot이 학습한 코드는 공개된 오픈소스 라이브러리에서 가져온 것이며, 이를 기반으로 완전히 새로운 코드를 생성하는 과정에서 원본 코드를 단순히 복제하지 않았다고 주장했다. 하지만 이 사건은 AI가 학습한 코드의 출처와 AI가 제안하는 코드 간의 저작권 침해 여부를 논의하는 중요한 판례로 자리잡을 가능성이 높다. AI가 오픈소스 코드를 학습하는 과정에서 어떤 저작권 보호가 적용될 수 있는지, 그리고 그 결과물이 원본 코드와 동일하거나 유사한 경우 법적으로 어떤 책임을 질 것인가가 핵심 쟁점으로 떠오르고 있다.[9,10]

## 3. 저작권 침해사례 분석

### 3.1 Stability AI 사건 분석

Stability AI 사건에서 가장 주목해야 할 점은 AI가 저작권 보호를 받는 이미지를 학습 데이터로 사용함으로써 발생하는 법적 책임이다. Getty Images는 AI가 자사의 이미지를 무단으로 학습하여 이익을 취했고, 그 결과 생성된 이미지가 기존 저작물의 가치를 훼손한다고 주장한다. 이 사건은 AI가 학습에 사용하는 데이터의 저작권 침해 여부를 명확히 규정할 필요성을 부각시킨다.

특히 AI가 기존 저작물을 직접적으로 복제하는 것이 아닌, 이를 학습 데이터로 사용해 새로운 창작물을 생성하는 경우에도 저작권 침해로 간주될 수 있는지가 핵심 쟁점이다. Getty Images 측은 AI가 생성한 이미지가 직접적인 복제가 아니더라도, 학습 과정에서 저작권 침해가 발생했다고 주장하고 있다. 이는 AI 기술 발전에 따른 법적, 윤리적 논의를 더욱 복잡하게 만들며, 법적으로 AI 학습 데이터에 대한 더 엄격한 규제가 도입될 가능성을 제기하고 있다.

### 3.2 OpenAI 사건 분석

OpenAI 사건은 AI 모델이 학습한 데이터와 생성된 텍스트 간의 유사성 문제를 다루고 있다. AI 모델이 방대한 텍스트 데이터를 학습하는 과정에서 해당 데이터가 저작권 보호를 받는 경우, 이를 기반으로 생성된 텍스트가 저작권 침해로 간주될 수 있는지에 대한 논의가 중심이다. OpenAI는 GPT 모델이 학습한 텍스트를 바탕으로 새로운 형태의 텍스트를 생성했다고 주장하며, 이는 공정 사용 원칙(fair use)에 부합한다고 주장하고 있다.

그러나 저작권자들은 AI가 학습한 원본 텍스트와 생성된 텍스트 간의 유사성이 지나치게 높아 공정 사용의 범위를 넘어서고 있다고 주장한다. 이 사건은 AI 모델의 학습 과정에서 공정 사용 원칙이 어떻게 적용될 수 있는지에 대한 중요한 판례를 남길 가능성이 크다. 만약 법원이 AI 모델의 학습 과정에서 공정 사용이 인정되지 않는다고 판결할 경우, AI 개발자들은 학습 데이터에 대해 더 엄격한 라이선스 절차를 밟아야 할 가능성이 커진다.

### 3.3 GitHub Copilot 사건 분석

GitHub Copilot 사건은 AI가 학습한 코드 데이터와 AI가 생성한 코드 간의 저작권 문제를 중심으로 논의되고 있다. Copilot이 학습한 코드가 공개된 오픈소스 코드라고 해도, AI가 제안하는 코드가 원본 코드와 유사하거나 동일할 경우 저작권 침해로 간주될 수 있다. 마이크로소프트는 Copilot이 학습한 코드는 완전히 새로운 형태의 코드라고 주장하고 있지만, 개발자들은 자신들의 코드를 무단으로 사용한 것은 저작권 침해라고 주장하고 있다.

이 사건은 AI가 학습한 코드 데이터에 대한 저작권 보호와 AI가 생성한 코드의 저작권 귀속 문제를 다루는 중요한 판례가 될 수 있다. AI가 학습하는 과정에서 기존 코드가 어느 정도로 사용되는지, 그리고 생성된 코드가 원본 코드와 유사한 경우 책임이 누구에게 있는지에 대한 명확한 법적 기준이 필요하다. 이 사건은 AI와 저작권법의 경계를 설정하는 중요한 사례로 자리 잡을 것이다.

## 4. 저작권 침해 예방에 대한 고찰

### 4.1 데이터셋의 투명성 확보

AI 개발자들은 학습 데이터의 출처를 명확히 하고, 저작권 보호 자료를 사용할 경우 적절한 라이선스를 확보해야 한다. 데이터셋의 투명성 확보는 저작권 침해 문제를 예방하는 핵심 전략이다. 데이터 출처를 명확히 하고, 저작권 보호 자료를 사용할 때는 해당 저작물의 권리를 보유한 자로부터 허가를 받아야 한다. 이러한 조치는 AI 개발자들이 법적 문제를 사전에 예방하고, 생성형 AI의 법적 책임을 명확히 할 수 있는 기초가 된다.

### 4.2 법적 책임 명확화

AI가 생성한 콘텐츠의 법적 책임을 명확히 규정하는 것도 중요하다. AI 개발자와 사용자는 생성된 콘텐츠에 대한 저작권 문제를 명확히 인식하고, 법적 대응을 미리 준비해야 한다. AI가 생성한 콘텐츠가 저작권을 침해했을 경우, 책임의 귀속 문제를 명확히 하고, 이에 대한 법적 조치를 취할 수 있는 체계를 마련해야 한다. AI 개발자들은 AI가 생성하는 콘텐츠에 대한 법적 책임을 사전에 검토하고, 필요한 법적 절차를 준수해야 한다.

## 5. 저작권 법체계의 진화 방향

AI 기술의 발전은 저작권 법체계의 진화를 요구하고 있다. 기존 저작권 법체계는 주로 인간 창작자를 염두에 두고 설계되었으나, AI 기술의 발전에 따라 새로운 법적 도전과제가 제기되고 있다. 향후 저작권 법체계가 AI 기술 발전에 따라 어떻게 진화해야 하는지에 대한 논의는 다음과 같은 방향으로 진행될 수 있다.[11-15]

### 5.1 AI 창작물의 법적 지위 명확화

AI가 생성한 콘텐츠에 대한 저작권의 법적 지위를 명확히 해야 한다. 현재의 저작권 법체계는 인간 창작자에게만 저작권을 부여하고 있으나, AI가 창작한 콘텐츠에 대해서도 법적 보호가 필요하다. 이와 관련하여, AI가 생성한 콘텐츠에 대한 저작권을 인정하는 법적 기준을 마련할 필요가 있다.

### 5.2 AI 학습 데이터의 저작권 보호 강화

AI가 학습하는 데이터에 대한 저작권 보호를 강화할 필요가 있다. AI 개발자는 학습 데이터의 출처를 명확히 하고, 저작권 보호 자료를 사용할 때는 적절한 라이선스를 확보해야 한다. 이는 AI 학습 데이터에 대한 저작권 보

호를 강화하고, 저작권 침해를 예방하는 데 중요한 역할을 할 것이다.

**5.3 공정 사용 원칙의 재정립**

AI가 학습하는 과정에서의 공정 사용(fair use) 원칙의 적용 범위를 재정립해야 한다. 공정 사용 원칙은 기존의 저작권 법체계에서 인간 창작자에게 적용되던 원칙으로, AI가 학습할 때의 공정 사용 범위를 명확히 할 필요가 있다. AI 학습 과정에서 공정 사용이 인정되는 경우, 그 범위를 명확히 정의하고, AI 기술 발전에 따른 새로운 법적 요구를 충족해야 한다.

**5.4 AI 개발자와 창작자 간의 협력**

AI 개발자와 창작자 간의 협력 모델을 구축하여, 저작권 문제를 사전에 예방하고 해결할 수 있는 시스템을 마련해야 한다. AI 개발자와 창작자 간의 원활한 협력을 통해, AI 학습 데이터에 대한 권리 문제를 해결하고, 창작물의 저작권을 보호할 수 있는 공동의 노력 필요하다

**5.5 법적 대응 체계의 정비**

AI 기술 발전에 따른 저작권 침해 문제를 효과적으로 대응하기 위해 법적 대응 체계를 정비해야 한다. 법원과 규제 기관은 AI와 관련된 저작권 분쟁을 신속하게 처리할 수 있는 체계를 마련하고, AI 기술 발전에 따른 새로운 법적 쟁점을 해결하기 위한 정책을 마련해야 한다

**6. 결 론**

생성형 AI는 새로운 창작의 가능성을 열어주지만, 그 과정에서 저작권 침해의 문제가 발생할 수 있다. AI와 저작권 간의 균형을 유지하기 위해서는 AI 학습 데이터의 출처를 명확히 하고, 저작권 침해를 방지하기 위한 법적 규제와 기술적 전략을 마련해야 한다. 앞으로 AI 개발자와 창작자들은 저작권을 존중하고, AI 기술 발전과 저작권 보호 간의 균형을 이루기 위한 노력이 필요하다.

AI 기술 발전에 따라 저작권 법체계도 진화해야 한다. 첫째, AI가 생성한 콘텐츠에 대한 법적 지위를 명확히 하여 저작권 보호를 강화해야 한다. 둘째, AI 학습 데이터에 대한 저작권 보호를 강화하고, 적절한 라이선스 확보를 통해 저작권 침해를 예방해야 한다. 셋째, 공정 사용 원칙을 재정립하여 AI 학습 과정에서의 법적 기준을 명확히 해야 한다. 넷째, AI 개발자와 창작자 간의 협력 모델을 구축하여 저작권 문제를 사전에 예방하고 해결할 수 있는 시스템을 마련해야 한다. 마지막으로, 법적 대응 체계를 정비하여 AI 기술 발전에 따른 저작권 분쟁을 효과적으

로 처리할 수 있는 체계를 구축해야 한다

이러한 법적, 정책적 조치를 통해 저작권 침해를 예방하고, 창작물의 보호를 강화하며, AI 기술 발전과 저작권 보호 간의 균형을 맞추는 노력이 필요하다. 창작자와 AI 개발자들은 AI가 학습하는 데이터가 저작권으로 보호될 수 있음을 인식하고, 저작권을 존중하는 자세를 가져야 한다. AI 기술의 발전에 따라 저작권 법체계도 진화하여, 법적 문제를 사전에 예방하고 해결할 수 있는 체계를 마련하는 것이 중요하다.

**7. 제 언**

나아가 생성형 AI가 반도체와 디스플레이 설계에 활용되면서 저작권 문제에 대한 고려가 필요합니다.[16] 반도체 설계 데이터는 각 기업의 지적 재산으로 보호되며, 학습 데이터로서의 사용은 원작자의 권리에 직접적인 영향을 미칠 수 있습니다. 따라서 AI 모델을 훈련시키기 위해 기존 설계 데이터를 사용할 경우, 학습 데이터의 출처와 라이선스 상태를 명확히 해야 하며, 이러한 절차를 통해 저작권 침해를 방지할 수 있습니다.

또한, 반도체 설계 데이터의 사용 기준을 설정하여 AI가 생성하는 새로운 설계가 기존 데이터와 유사성을 가질 경우에도 저작권 침해가 발생하지 않도록 법적 기준을 마련할 필요가 있습니다. [17] AI가 특정 설계 데이터를 학습하거나 이를 기반으로 생성된 설계에 대한 권리와 책임을 명확히 규정하는 표준이 저작권 관련 분쟁을 예방하고, 산업 내 AI 활용을 활성화하기 위한 발판이 될 것입니다.

**감사의 글**

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업으로 수행되었음(과제명: 시니어의 콘텐츠 제작 접근성 향상을 위한 생성형 AI 기반 콘텐츠 창·저작 플랫폼 기술 개발, 과제번호: RS-2024-00340342, 기여율: 100%).

**참고문헌**

1. Koo, C., "Generative AI and Copyright Issues," *Broadcasting Culture*, Vol. 427, pp. 30-43, 2023.
2. Park, Y., "A Study on Copyright Infringement of Style Imitation by Generative Artificial Intelligence (AI)," *Industrial Property Rights*, Vol. 76, pp. 431-454, 2023.
3. Lee, K., Cooper, A. F., & Grimmelmann, J., "Talkin' 'Bout AI Generation: Copyright and the Generative-

- AI Supply Chain,” arXiv preprint arXiv:2309.08133, 2023.
4. Lemley, M. A., “How Generative AI Turns Copyright Law on Its Head,” Available at SSRN 4517702, 2023.
  5. Coulter, M., “Aiming for Fairness: An Exploration into Getty Images v. Stability AI and Its Importance in the Landscape of Modern Copyright Law,” *DePaul Journal of Art, Technology & Intellectual Property Law*, Vol. 34, No. 1, pp. 124, 2024.
  6. Chesterman, S., “Good Models Borrow, Great Models Steal: Intellectual Property Rights and Generative AI,” *Policy and Society*, 2024, puae006.
  7. Kavia, A., & Sharma, K. S., “Chat GPT and Copyright: Legal and Ethical Challenges,” Available at SSRN 4848807, 2023.
  8. Lucchi, N., “ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems,” *European Journal of Risk Regulation*, pp. 1-23, 2023.
  9. Basanagoudar, V., & Srekanth, A., “Copyright Conundrums in Generative AI: Github Copilot’s Not-So-Fair Use of Open-Source Licensed Code,” *J. Intell. Prot. Stud.*, Vol. 7, pp. 58, 2023.
  10. Shin, S., “Legal Issues of Image-Generating AI and Training Data,” *The Journal of Law & IP*, Vol. 13, No. 1, pp. 1-38, 2023.
  11. Son, Y., “Copyright and Works Created by Generative AI,” *Law and Policy Studies*, Vol. 23, No. 3, pp. 357-389, 2023.
  12. Kim, Y., “Generative AI and Copyright Issues,” *KISDI AI Outlook*, Vol. 2023.13, pp. 1-35, 2023.
  13. Zhong, H., et al., “Copyright Protection and Accountability of Generative AI: Attack, Watermarking, and Attribution,” In: *Companion Proceedings of the ACM Web Conference 2023*, pp. 94-98, 2023.
  14. Lee, W., “Consideration of Copyright Law Amendments for Future Society with Generative AI,” 2024.
  15. Kim, Y., “Legal Considerations on Prompt Creation in Generative AI,” *Justice*, Vol. 200, pp. 261-297, 2024.
  16. Aramoon, O. *Intellectual Property Protection: From Integrated Circuits to Machine Learning Models*, 2022.
  17. Thongmeensuk, S. Rethinking copyright exceptions in the era of generative AI: Balancing innovation and intellectual property protection. *The Journal of World Intellectual Property*, 27(2), 278-295, 2024.

---

접수일: 2024년 10월 10일, 심사일: 2024년 11월 8일,  
 게재확정일: 2024년 11월 28일