Monthly Report
2026-3호

www.copyright.or.kr

저작권 기술 트렌드

COPYRIGHT TECH TREND

한국저작권위원회
KOREA COPYRIGHT COMMISSION

AI 시대의 웹툰 저작권 보호를 위한 방어 기술과 K-웹툰 산업의 대응 전략

뉴스 브리프

인공지능 기술이 웹툰 산업에 점차 도입되면서 창작 활동의 범위가 넓어지는 한편, 창작자의 화풍을 무단 학습하거나 원작자를 특정하기 어려운 AI 산출물이 유통되는 등 저작권 관련 문제도 함께 제기되고 있다. AI 학습 데이터의 투명성이 충분히 확보되지 않고 산출물의 출처를 객관적으로 검증할 수단이 부족한 상황에서, 창작자의 권리를 기술적으로 보호하려는 논의가 본격화되고 있다. 본 보고서는 이에 대응하는 기술적 접근을 검토하는 것을 목적으로 한다. 생성형 AI의 스타일 모방을 사전에 억제하는 방어 기술과 이미지의 출처를 사후에 확인할 수 있는 워터마킹 기술을 중심으로 두 기술의 보완적 역할을 분석하고, 나아가 표준화·법적 증거 활용 등 산업 전반의 대응 방향과 K-웹툰의 지속 가능한 생태계 구축 가능성을 함께 살펴본다.

뉴스 플러스

Ⅰ. 서론 : AI 시대 웹툰 저작권 보호의 새로운 국면과 기술적 해법의 모색

- **AI 기술 확산과 웹툰 산업의 변화**

국내외 웹툰·만화 산업은 인공지능 기술의 발전과 함께 전략적 활용 단계에 접어들고 있다. AI는 기획 단계에서 시놉시스 구상과 캐릭터 설정을 보조하는 것을 시작으로, 작화 공정에서는 배경 생성·채색·후반 보정 등 단순 반복 작업을 효율화하는 도구로 활용되고 있다. 이는 작가의 창작 부담을 줄이고 제작 시간을 단축함으로써 웹툰 산업 전반의 경쟁력을 높이는 긍정적 요인으로 평가된다. 실제로 만화·웹툰 이용자 대상 설문조사에서 '작가의 창작 보조를 통한 제작 시간 단축 및 생산성 향상'에 대한 기대가 43.7%로 나타난 것은, 이러한 산업적 변화에 대한 인식을 보여준다.[1)]

1) 조윤주, " "창작부터 보안까지 AI가"...AI 빅뱅 맞는 웹툰업계", 파이낸셜뉴스, 2026.02.19, https://www.fnnews.com/news/202602141404331466

AI 기술은 창작 지원을 넘어 산업 생태계의 구조적 문제 해결에도 적용되고 있다. 대표적 사례로, 네이버 웹툰이 2017년부터 운영해온 AI 기반 불법 유통 탐지 시스템인 툰레이더(ToonRadar)는 이미지 워터마킹과 데이터 분석을 결합해 불법 복제물을 추적하고 유포자를 식별함으로써 저작권 보호의 실효성을 높이고 있다.[2] 이처럼 AI는 창작 효율과 권리 보호 양면에서 긍정적 가능성을 보여주며 웹툰 산업의 핵심 기술 인프라로 자리매김하고 있다.

그러나 이러한 기회 요인과 함께 창작자의 권리를 위협하는 문제들도 수면 위로 떠오르고 있다. 특정 작가의 화풍과 스타일을 모방한 AI 산출물이 유통되거나, AI 모델이 학습에 활용한 저작물을 공개하지 않는 불투명한 관행은 저작권 침해 논란의 주요 원인이 되고 있다. 기술 발전이 가져온 편의가 오히려 창작자의 권익을 훼손할 수 있다는 우려가 커지면서, 웹툰 산업은 기술 활용과 권리 보호라는 두 가지 과제를 동시에 안게 되었다.

- **창작자의 딜레마: 기술 수용과 권리 침해의 경계**

기술 변화의 한가운데 놓인 웹툰 작가들은 AI를 둘러싼 복잡한 상황 속에서 쉽지 않은 선택에 직면하고 있다. 한국만화가협회가 창작자의 권리 이해와 기술 변화 대응 기준을 논의하는 포럼을 개최하는 등 산업계 차원의 해법 모색이 이어지고 있으나, 현장에서 체감하는 불확실성은 여전하다. 국내 웹툰 작가의 생성형 AI 활용 경험이 33% 수준에 머물고 있다. AI를 활용하지 않는 주된 이유로 '제3자의 저작권을 침해할 수 있다는 윤리적·법적 부담'(36.5%)이 꼽혔다는 사실은[3] 기술 도입의 필요성과 잠재적 위험 사이에서 창작자들이 겪는 복합적인 현실을 드러낸다.

이 딜레마는 창작자들을 기술 수용과 권리 침해 위험이라는 기로에 서게 한다. 일본에서는 AI 만화 플랫폼이 시장 상위권을 차지하는 등 기술 수용이 이미 상당 부분 현실로 진행되고 있으며, 국내에서도 AI 웹툰에 대한 이용자의 부정적 인식은 12.5%에 불과한 것으로 나타났다.[4] 그러나 AI 도구를 활용하는 과정에서 자신의 작품이 의도치 않게 타인의 저작권을 침해하거나, 반대로 자신의 작품이 AI 학습 데이터로 무단 활용되어 유사한 스타일의 산출물이 유통될 수 있다는 우려는 여전히 해소되지 않고 있다. 결국 창작자들은 생산성 향상이라는 이점에도 불구하고, 자신의 권리를 지킬 명확한 장치가 갖춰지지 않은 상태에서 기술 수용 여부를 결정해야 하는 상황에 놓여 있다.

- **학습의 불투명성: 앤트로픽 판결과 공정 이용 쟁점**

웹툰 작가들이 느끼는 불안의 주요 원인 중 하나는 AI 모델의 학습 과정이 충분히 공개되지 않는다는 점이다. AI 기업들이 학습에 사용한 데이터셋을 공개하지 않는 관행이 이어지면서, 창작자들은 자신의

2) 조윤주, " "창작부터 보안까지 AI가"...AI 빅뱅 맞는 웹툰업계", 파이낸셜뉴스, 2026.02.19, https://www.fnnews.com/news/202602141404331466
3) 김민수, "웹툰 작가들 'AI와 저작권' 논의…한국만화가협회 포럼 개최", 노컷뉴스, 2026.03.05, https://www.nocutnews.co.kr/news/6479784
4) 김민수, "웹툰 작가들 'AI와 저작권' 논의…한국만화가협회 포럼 개최", 노컷뉴스, 2026.03.05, https://www.nocutnews.co.kr/news/6479784

저작물이 동의 없이 학습에 활용되었을 가능성을 배제하기 어려운 상황이다. 이러한 가운데 최근 미국 법원의 판결은 관련 논의에 적지 않은 영향을 미쳤다. 미국의 AI 기업 앤트로픽(Anthropic)을 상대로 한 저작권 침해 소송에서 법원은 AI 기업이 수백만 권의 책을 스캔해 모델을 학습시킨 행위를 공정 이용(fair use)에 해당한다고 판결했다.[5)]

재판부가 이를 공정 이용에 해당한다고 판결한 이유는 AI의 학습 과정을 인간의 학습과 유사한 것으로 판단했기 때문이다. 이 판결은 저작권자의 명시적 동의 없이 이루어진 AI 학습 행위의 법적 해석에 있어 하나의 선례가 될 수 있다는 점에서 콘텐츠 산업 전반에 작지 않은 파장을 불러일으킨 것으로 평가된다. 미국의 판례가 곧바로 국제적 기준으로 작용하는 것은 아니지만, 공정 이용의 범위를 저작권자보다 기술 기업의 관점에서 넓게 해석하는 경향에 대한 우려를 불러일으키기에 충분했다는 평가가 나온다.

이러한 법적 해석은 기술 기업의 학습 편의를 위해 창작자의 권리가 제약될 수 있다는 우려를 확산시켰다. 특히 학습 데이터의 출처와 활용 범위가 공개되지 않는 불투명한 관행과 공정 이용 인정이 맞물릴 경우, 창작자는 자신의 저작물이 어떻게 활용되는지조차 알지 못한 채 권리 침해 가능성을 감수해야 하는 처지에 놓일 수 있다.

• 객관적 검증의 부재와 다층적 기술 해법의 필요성

현재 웹툰 산업이 직면한 저작권 문제의 핵심 중 하나는 객관적 검증 수단의 부족이다. 특정 AI 산출물이 자신의 저작물을 학습하여 만들어졌다는 의심이 있더라도, 이를 명확히 입증할 기술적·법적 방법론은 아직 확립되지 않았다. 저작권 침해 소송과 같은 전통적인 사후 대응 방식은 침해 사실의 입증 책임을 원작자가 부담하는 구조인데, AI 모델 내부의 복잡한 메커니즘을 분석해 인과관계를 증명하는 것은 현실적으로 쉽지 않다. 이러한 어려움은 기존 법 제도의 적용을 가로막고, 창작자의 권리 구제를 제한하는 요인으로 작용한다.

이에 법적·제도적 논의와 병행하여 실효성 있는 기술적 해법을 모색하는 움직임이 이어지고 있다. AI의 학습 방식에 개입해 사전에 스타일 모방을 방지하는 선제적 방어 기술과, 변형되거나 복제된 산출물의 출처를 사후에 증명하는 추적 기술을 결합한 '다층적 방어 체계'가 유력한 대안으로 거론된다. 본 보고서는 이러한 기술적 가능성을 구체적으로 살펴보기 위해 이어지는 본론에서 글레이즈(Glaze)와 SiGRRW 두 가지 기술을 심층적으로 분석한다.

5) 김지원, “책 수백만권 스캔한 앤스로픽 충격…AI가 뒤흔든 ‘저작권’ ”, 주간경향, 2026.02.23., https://weekly.khan.co.kr/article/2026022306000031#ENT

Ⅱ. 본론 1: 사전 예방을 위한 방어 기술, 글레이즈 심층 분석

• 글레이즈의 개념: AI의 스타일 모방을 막는 클로킹(cloaking) 기술

AI 기술의 발전은 창작의 가능성을 넓힌 동시에 창작자의 예술적 정체성을 위협하는 새로운 쟁점을 낳았다. 그 대표적인 문제가 바로 스타일 모방(style mimicry)이다. AI가 특정 작가의 화풍을 학습해 이를 복제하거나 변형한 산출물을 만들어 내는 이 현상은, 작가가 오랜 시간 쌓아온 독창성의 가치를 희석할 수 있어 창작 생태계의 주요 위협 요인으로 인식되고 있다. 글레이즈는 이러한 문제에 대응하기 위해 시카고 대학(University of Chicago) 연구팀이 개발한 기술로, 창작자가 작품을 온라인에 공개하기 전 저작물을 보호할 수 있도록 설계된 선제적 방어 수단이다.

글레이즈의 핵심 원리는 클로킹(cloaking)*이라는 개념에 기반한다. 클로킹은 AI 모델이 학습 데이터로부터 정확한 정보를 습득하지 못하도록 데이터를 의도적으로 변형하는 기법으로, 원본의 핵심 정보를 숨기거나 다른 정보를 제공해 오학습을 유도하는 방식으로 작동한다. 글레이즈는 이 원리를 적용해 창작자가 자신의 작품 이미지에 육안으로는 거의 감지하기 어려운 미세한 섭동(perturbations)**를 추가함으로써 AI 모델이 해당 작품의 스타일을 전혀 다른 것으로 인식하게 만든다. 이 섭동은 원본의 시각적 품질에 미치는 영향을 최소화하도록 설계되었다.[6]

* 클로킹(cloaking): 원본 데이터에 미세한 노이즈나 변형을 추가하여 AI 모델이 데이터의 진짜 특징을 학습하지 못하도록 방해하는 기술. '은폐' 또는 '위장' 기술로도 불림

** 섭동(perturbation): 원본 데이터에 의도적으로 추가하는 미세한 변화값. 인간의 눈으로는 거의 감지되지 않으나 AI 학습 과정에는 유의미한 영향을 미침

• 작동 원리: 특징 공간 교란의 메커니즘

글레이즈가 AI의 스타일 모방을 방지하는 원리를 이해하려면 먼저 AI가 예술 스타일을 인식하는 방식을 살펴볼 필요가 있다. AI 모델은 대량의 이미지를 학습하는 과정에서 각 화풍의 특징을 추출하고, 이를 다차원의 가상 공간에 벡터 형태로 기록한다. 이 공간을 특징 공간(feature space)*이라 한다. AI는 특정 스타일의 특징 벡터 좌표를 토대로, 사용자의 요청에 맞는 스타일을 재현하는 방식으로 산출물을 생성한다.

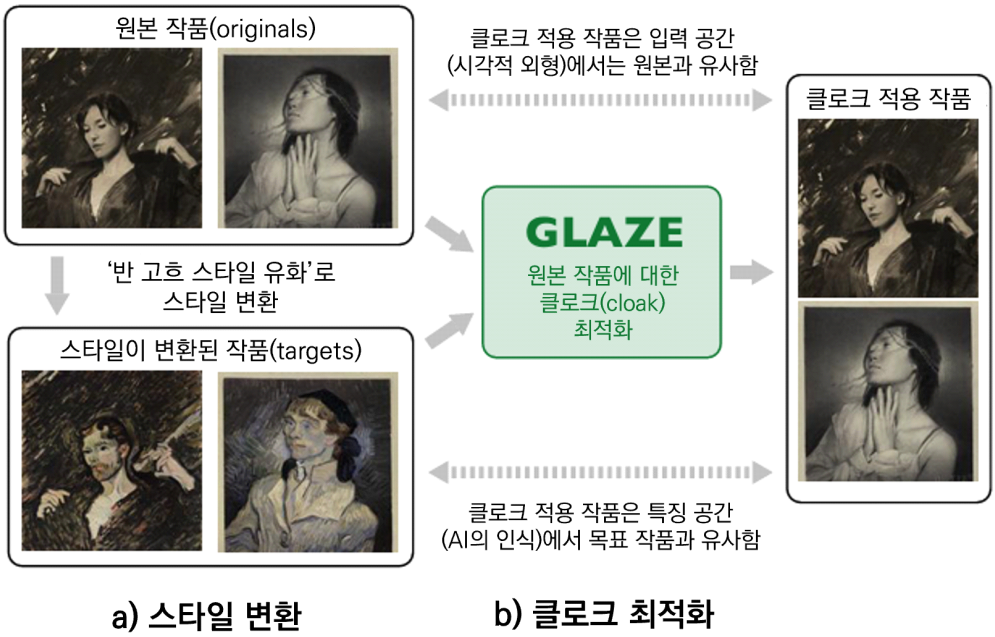
글레이즈는 단순히 노이즈를 추가하는 것을 넘어, 아래 [그림 1]과 같이 두 단계의 과정을 통해 특징 공간의 좌표 체계를 의도적으로 교란한다. 첫 번째 단계에서 글레이즈는 원본 작품과 전혀 다른 화풍을 가짜 목표 스타일(target style)로 설정한다. [그림 1]에서는 반 고흐의 유화 스타일이 가짜 목표 스타일로 사용되었다. 이어서 원본 작품의 구도와 내용은 유지한 채, 이 목표 스타일을 입힌 변환 이미지(style-transferred art)를 가상으로 생성한다(a단계). 이 변환 이미지는 AI가 원본의 화풍 대신 목표 스타일의 특징을 학습하도록 유도하는 역할을 한다.

두 번째 단계에서 글레이즈는 원본 이미지에 육안으로는 감지하기 어려운 미세한 섭동, 즉 클로크

6) Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222

(cloak)를 계산하여 추가한다(b단계). 클로크는 AI가 해당 이미지를 학습할 때 원본의 스타일 특징 대신 첫 번째 단계에서 생성한 변환 이미지의 특징을 인식하도록 유도한다.

[그림 1] 글레이즈의 핵심 작동 원리 개요도



출처: Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222 (Figure 6)

앞서 설명한 것처럼 AI는 각 화풍의 특징을 특징 공간 안의 좌표로 기록한다. 클로크가 적용된 이미지를 학습한 AI는 작가 A의 스타일 좌표를 원래 위치가 아닌 가짜 목표 스타일, 즉 반 고흐 유화 스타일의 좌표 근처에 잘못 기록하게 된다. 이후 사용자가 A 스타일을 요청하더라도 AI는 오학습된 반 고흐 스타일을 기반으로 전혀 다른 결과물을 산출한다. 이처럼 글레이즈는 AI의 인식 체계를 교란함으로써 스타일 도용의 가능성을 낮추는 방어 메커니즘이다.

* 특징 공간(feature space): AI가 데이터(이미지, 텍스트 등)의 복잡한 특징들을 수학적으로 표현하고 분류하기 위해 사용하는 다차원의 가상 공간. 유사한 특징을 가진 데이터는 이 공간에서 서로 가까운 위치에 배치됨

• 예술가를 위한 방어막: 글레이즈의 실제 적용 효과와 사례

글레이즈의 기술적 효과는 창작자를 대상으로 한 사용자 연구를 통해 실증적으로 검증되었다. 논문에 따르면 글레이즈 효과 평가 설문에 참여한 예술가 중 93%가 AI 모델의 스타일 모방을 성공적으로 방어했다고 응답했다.[7] 아울러 이미지에 추가되는 미세한 섭동에 대해서도 참여자의 92%가 작품의 가치를 훼손하지 않는 수준이라고 평가해 기술의 실용성 측면에서도 긍정적인 반응을 얻었다. 이는 해당 기술이 이론적 모델에 그치지 않고 실제 현장의 창작자에게 실질적인 보호 효과를 제공할 수 있음을 보여주는 결과다.

7) Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222

예를 들어, 한 작가가 신작을 소셜 미디어에 게시하기 전 이 기술을 적용하는 상황을 생각해 볼 수 있다. 해당 이미지가 AI 모델 학습에 활용되더라도 AI는 작가의 스타일을 정확히 습득하기 어려워지고, 제3자가 유사한 화풍의 산출물을 얻으려는 시도 역시 성공 가능성이 낮아진다. 이처럼 글레이즈는 온라인 환경에서 작품을 공개하는 창작자들이 스타일 도용의 위험을 줄이는 데 활용할 수 있는 기술적 선택지 중 하나다.

[그림 2] 글레이즈의 실제 방어 효과 비교

출처: Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222 (Figure 8)

- **방어 기술의 진화: 글레이즈에 대한 우회 시도와 대응**

모든 방어 기술이 그렇듯 글레이즈 역시 이를 무력화하려는 기술적 시도, 즉 우회 공격(bypass attacks)의 가능성에서 자유롭지 않다. 우회 공격은 주로 이미지에 적용된 미세한 섭동을 제거하거나 약화시키는 방향으로 이루어진다. 별도의 노이즈 추가, 이미지 압축 및 복원 과정 반복, 이미지 전체를 부드럽게 만드는 스무딩(smoothing) 필터 적용 등이 대표적인 방식이다.

이러한 이미지 변형 기법들은 글레이즈의 방어 효과를 실질적으로 약화시킬 수 있다. 변형 과정에서 섭동 신호가 손상되면 AI 모델이 교란된 정보 대신 원본 이미지에 가까운 스타일 특징을 학습할 가능성이 생기기 때문이다. 이는 해당 기술의 보호 기능이 특정 조건 아래에서는 제한될 수 있음을 보여주는 한계로, 창작자들이 효과를 과신하지 않고 관련 기술 동향을 지속적으로 살펴볼 필요가 있는 이유이기도 하다.

• **사전 예방 기술의 의의와 저작권 생태계에 미치는 영향**

글레이즈와 같은 사전 예방 기술은 AI 시대의 저작권 논의에 새로운 시각을 제공한다. 지금까지의 저작권 보호는 침해가 발생한 후 법적·행정적 조치를 취하는 사후 대응 방식에 주로 의존해왔다. 그러나 이 방식은 시간과 비용이 상당히 소요되고, AI 산출물의 경우 침해 사실을 입증하기가 쉽지 않다는 구조적 한계를 안고 있다. 이에 비해 사전 예방 기술은 침해의 원인이 될 수 있는 AI의 학습 단계에 직접 개입함으로써 문제를 사전에 차단한다는 점에서 결이 다르다. 창작자가 자신의 저작물 학습에 대한 일정한 통제 가능성을 갖는다는 점에서 능동적 보호 수단으로서의 의의를 지닌다.

나아가 이러한 사전 예방 기술의 확산은 저작권 생태계 전반에도 영향을 미칠 수 있다. 섭동이 적용된 데이터를 학습에 활용할 경우 모델 성능이 저하될 수 있다는 점에서, AI 기술 기업들이 온라인 데이터를 무분별하게 수집하던 방식을 재검토하는 계기가 될 수 있다. 이는 기업들이 창작자에게 정당한 이용 허락을 구하는 등 투명한 방식으로 데이터셋을 구축하도록 유도하는 시장 기제로 작용할 수 있다. 궁극적으로 이러한 기술은 창작자의 권리를 기술적으로 보호하는 데 그치지 않고, 기술 기업과 창작자 간의 신뢰를 바탕으로 한 AI 창작 생태계 조성에도 기여할 것으로 기대된다.

Ⅱ. 본론 2: 사후 증명을 위한 추적 기술, SiGRRW 심층 분석

• **SiGRRW의 개요: 강인하고 복원 가능한 워터마킹 프레임워크**

글레이즈가 AI의 학습 단계에 개입하는 사전 예방 기술이라면, SiGRRW(Single-Watermark Robust Reversible Watermarking Framework)는 이미 유통 중인 디지털 이미지의 출처를 증명하는 사후 추적 기술에 해당한다. 이미지에 저작권 정보를 육안으로는 감지하기 어렵게 삽입하는 워터마킹 기법의 일종으로, 강인성(robustness)과 가역성(reversibility)이라는 두 가지 핵심 특성을 동시에 구현하도록 설계되었다.

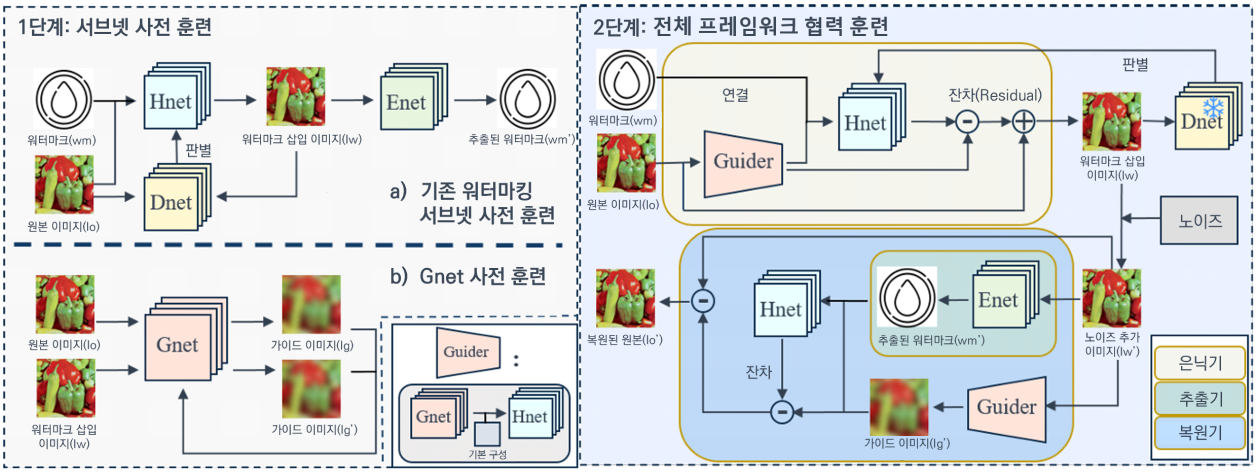
첫 번째 특성인 강인성은 이미지에 삽입된 워터마크가 공격을 받더라도 훼손되지 않고 유지되는 성질을 말한다. JPEG 압축이나 필터 적용과 같은 일반적인 이미지 처리는 물론, AI를 이용해 이미지를 미세하게 변형하여 워터마크를 제거하려는 재생성 공격에도 저항할 수 있도록 설계되었다. 불법 복제된 웹툰이 다양한 형태로 변형·유통되더라도 그 안에 숨겨진 원작자 정보를 추출해 권리를 확인할 수 있다는 의미다.

두 번째 특성인 가역성은 이미지에 삽입된 워터마크를 필요에 따라 완전히 제거하고, 원본 이미지를 픽셀 단위의 손상 없이 복원할 수 있는 성질을 말한다. 기존의 많은 워터마킹 기술이 원본에 미세한 손상을 남기는 비가역적(irreversible) 방식이었던 것과 달리, SiGRRW는 원본의 가치를 온전히 보존할 수 있다. 원작자가 작품을 판매하거나 전시하는 등 상업적으로 활용할 때 저작권 정보가 포함되지 않은 원본을 언제든 확보할 수 있다는 점에서 실용적 의의를 갖는다.

• SiGRRW의 전체 구조와 작동 원리

SiGRRW 프레임워크는 아래 [그림 3]와 같이 세 가지 핵심 모듈로 구성되며, 각 모듈이 서로 다른 역할을 수행하면서 강인성과 가역성을 동시에 구현한다. 세 모듈은 은닉기(hider), 추출기(extractor), 복원기(restorer)로, 창작자가 저작권을 보호하고 원본을 관리하는 전 과정을 단계적으로 지원한다.

[그림 3] SiGRRW 프레임워크 전체 구조도



출처: Zikai Xu 외 4인, "SiGRRW: A Single-Watermark Robust Reversible Watermarking Framework with Guiding Strategy", arXiv, 2026.02.22, https://arxiv.org/html/2602.19097v1 (Figure 2)

각 모듈의 역할은 다음과 같다. 은닉기는 창작자가 원본 이미지와 저작권 정보가 담긴 메시지를 입력하면, 잠재 공간을 활용해 최적화된 방식으로 해당 메시지를 이미지의 픽셀 값에 미세하게 반영하여 워터마크가 적용된 이미지를 출력한다. 이렇게 생성된 이미지는 원본과 시각적으로 거의 구별되지 않으며 온라인 플랫폼 등을 통해 배포할 수 있다. 저작권 분쟁이 발생하거나 유통 경로를 추적해야 할 상황에서는 추출기가 불법 유통 중인 이미지에서 숨겨진 메시지를 읽어내어 법적 증거로 활용될 수 있도록 한다. 마지막으로 복원기는 원작자가 상업적 이용 등 다른 목적으로 원본이 필요할 때 워터마크가 삽입된 이미지에서 삽입 정보를 제거하여 원본을 복원한다.

• 핵심 메커니즘 1: 워터마크의 은닉과 잠재 공간 활용

SiGRRW가 기존 워터마킹 기술과 구별되는 핵심은 워터마크를 숨기는 방식에 있다. 기존 기술들이 이미지의 픽셀값을 단순하게 변경하는 방식에 주로 의존했다면, SiGRRW는 AI 모델이 이미지를 인식하는 근본 원리인 잠재 공간(latent space)을 계산의 기반으로 삼는다. 워터마크를 어디에 어떻게 삽입할지를 먼저 결정한 뒤, 그 결과를 픽셀 값에 반영하는 방식이다. 잠재 공간이란 AI가 이미지의 색감·구도·질감 등 수많은 시각적 특징을 압축하여 표현하는 고차원의 데이터 공간으로, 워터마크가 저장되는 장소가 아니라 삽입 방식을 최적화하는 처리 공간에 해당한다.

은닉기는 저작권 정보를 픽셀에 무작위로 새기는 대신, 잠재 공간 안에서 이미지의 특징 표현을 분석하여 어느 영역에 어떤 방식으로 정보를 심을지를 정밀하게 계산한 뒤 그 결과를 픽셀 값에 미세하게 반영한다. 이는 두 가지 측면에서 강점을 갖는다. 첫째, 인간의 눈으로는 변화를 거의 감지할 수 없어 원본의 시각적 품질을 높은 수준으로 유지할 수 있다. 둘째, 이미지의 근원적인 특징값에 정보가 결합되어 있어 이미지를 자르거나 압축하는 등 외형적 변형이 가해지더라도 픽셀에 반영된 정보는 비교적 안정적으로 보존된다. 이것이 SiGRRW의 강인성을 뒷받침하는 기술적 토대다.

* 잠재 공간(latent space): AI가 이미지나 텍스트 같은 고차원 데이터의 핵심적인 특징들을 저차원의 벡터 형태로 압축하여 표현하는 가상 공간. 데이터의 본질적 의미를 담고 있는 영역임

• 핵심 메커니즘 2: 정보의 추출과 원본의 무손실 복원

저작권 증명이 필요한 상황에서는 추출기가 그 역할을 맡는다. 추출기는 워터마크가 삽입된 이미지를 입력받아 은닉기의 삽입 과정을 역으로 추적함으로써 이미지 픽셀에 담긴 저작권 메시지를 해독하고 추출한다. JPEG 압축·노이즈 추가 등 이미지에 어느 정도의 왜곡이 가해진 상태에서도 원래의 메시지를 높은 정확도로 복원할 수 있도록 훈련되어, 불법 복제 과정에서 발생하는 다양한 형태의 이미지 훼손을 극복하고 저작권 정보를 식별해 낸다.

원작자가 소유권 증명과 무관하게 원본 이미지를 활용해야 할 때는 복원기가 사용된다. 복원기는 워터마크가 삽입된 이미지에서 은닉기가 가한 픽셀 수준의 변화를 역산하여 제거함으로써 원본 이미지를 복원한다. 이것이 SiGRRW의 가역성이 구현되는 방식이다. 추출기와 복원기는 각각 저작권 보호라는 대외적 목적과 원본 보존이라는 대내적 목적을 담당하는 상호 보완적 도구로, 두 모듈이 함께 작동함으로써 창작자는 자신의 저작물을 유연하게 관리할 수 있다.

• 기술적 성능 분석: AI 기반 재생성 공격 방어 효과

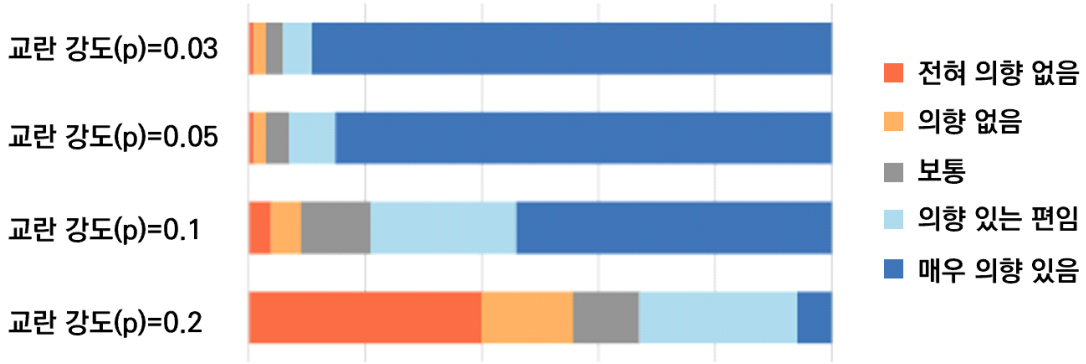
SiGRRW의 주목할 만한 성능 중 하나는 재생성 공격(regeneration attacks)에 대한 저항력이다. 재생성 공격이란 워터마크가 삽입된 이미지를 AI에 입력해 미세하게 변형된 새로운 이미지를 생성함으로써 기존 워터마크를 무력화하려는 시도를 말한다. AI를 이용해 저작권 추적을 회피하려는 침해 수법의 하나로 볼 수 있다.

대부분의 전통적인 워터마킹은 픽셀의 특정 패턴에 의존하기 때문에 이미지 전반에 미세한 변화를 가하는 재생성 공격에 취약한 편이다. 반면 SiGRRW는 잠재 공간에 정보를 숨기기 때문에 AI가 이미지의 겉모습을 변형하더라도 그 근본적인 특징 안에 내재된 워터마크까지 제거하기는 어렵다. 논문에 따르면 SiGRRW는 다양한 재생성 공격 시나리오에서 높은 정확도로 워터마크를 추출하는 데 성공하여 기존 기술 대비 우수한 방어 성능을 보인 것으로 나타났다. AI를 이용한 저작권 침해 시도가 앞으로 더욱 고도화될 것을 고려할 때, 이러한 방향의 기술이 실효성 있는 방어 수단으로 기능할 수 있음을 시사하는 결과다.

• 기술의 한계점과 향후 발전 방향

물론 SiGRRW 역시 모든 상황에 완벽히 대응할 수 있는 기술은 아니며 한계와 취약점을 지니고 있다. 첫째, 워터마크의 강인성과 비가시성 사이에는 본질적인 상충 관계(trade-off)가 존재한다. 강인성을 높이려면 이미지에 더 많은 정보를 더 깊이 삽입해야 하는데, 이는 원본 이미지에 미세한 시각적 변형을 유발할 가능성을 높인다. 원본의 품질을 중시하는 창작자에게는 민감하게 받아들여질 수 있는 지점이다. 실제로 [그림 4]에서 볼 수 있듯이, 교란 강도(p)가 0.03·0.05 수준일 때는 대다수 창작자가 게시 의향을 보이지만, 교란 강도가 0.2로 커질수록 거부 응답 비율이 크게 증가한다. 이는 방어 강도와 창작자의 수용성 사이의 균형이 기술 설계에서 중요한 고려 요소임을 보여준다.

[그림 4] 창작자의 클로킹 이미지 게시 의향



출처: Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222 (Figure 10)

둘째, 새로운 유형의 공격에 대한 취약성이다. 현재 SiGRRW는 알려진 재생성 공격 등에는 효과적인 방어력을 보이지만, 향후 등장할 새로운 방식의 AI 이미지 처리 기술이나 워터마크 제거에 특화된 AI 모델에 대해서는 성능을 보장하기 어렵다. 특정 워터마킹 알고리즘의 패턴을 학습해 이를 전문적으로 찾아내고 제거하는 표적 제거 공격(targeted removal attack) AI가 개발될 경우 현재의 방어 체계는 무력화될 수 있다. 글레이즈와 마찬가지로 SiGRRW 역시 창과 방패의 경쟁에서 자유롭지 않다는 점을 보여주는 사례다.

셋째, 기술의 보편성 문제다. SiGRRW 프레임워크는 현재 특정 구조의 AI 모델을 기반으로 설계되어 있어, 다른 구조의 AI 모델이나 비(非) AI 기반의 정교한 이미지 편집 소프트웨어에는 적용이 어렵거나 효과가 제한될 수 있다. 이러한 한계를 극복하고 더 다양한 디지털 환경과 공격 시나리오에 범용적으로 대응할 수 있는 유연성과 확장성을 확보하는 것이 향후 기술 발전의 과제로 남는다.

Ⅱ. 본론 3: 기술의 융합과 사회·산업적 파급 효과

• 이중 방어 체계: 글레이즈와 SiGRRW의 상호 보완적 역할

지금까지 분석한 글레이즈와 SiGRRW는 개별적으로도 저작권 보호 수단으로서 일정한 의의를 갖지만, 결합될 때 보완적 효과가 더욱 온전히 발휘될 수 있다. 두 기술은 AI로 인한 저작권 침해 과정의 서로 다른 단계에 각각 대응하도록 설계되어, 어느 한 기술만으로는 다루기 어려운 영역을 보완하는 관계에 있다. 이는 창작자의 권리를 보다 다층적으로 보호하는 구조로 기능할 수 있다.

[표 1] 글레이즈와 SiGRRW 비교

	Glaze	SiGRRW
목적	AI의 스타일 모방 사전 방지	저작물 출처의 사후 증명
개입 시점	작품 공개 이전, AI 학습 단계	침해 발생 이후,유통 추적
핵심 원리	특징 공간 교란을 통한 오학습 유도	잠재 공간에 저작권 정보 삽입
주요 특성	비가시성, 선제적 방어	강인성, 가역성
검증 결과	설문 예술가 93% 방어 성공 평가	재생성 공격 시나리오에서 기존 기술 대비 우수한 방어 성능
주요 한계	스무딩·압축 등 우회 공격에 취약 가능성	강인성·비가시성 간 상충 관계, 보편성 한계
법적 활용	침해 예방,증거 기능 없음	워터마크 추출을 통한 법적 증거 제공 가능

출처: Zikai Xu 외 4인, "SiGRRW: A Single-Watermark Robust Reversible Watermarking Framework with Guiding Strategy", arXiv, 2026.02.22, https://arxiv.org/html/2602.19097v1
Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222 (논문 정보 기반 재구성)

먼저 글레이즈는 사전 예방의 역할을 담당한다. 저작물이 온라인에 공개되어 AI 학습 데이터로 수집되는 단계에서 작동하여 AI 모델이 작가의 스타일을 정확히 학습하기 어렵도록 이미지에 미세한 섭동을 가한다. 이를 통해 특정 작가의 스타일을 모방한 AI 산출물이 생성될 가능성을 낮추는 효과를 기대할 수 있다. 다만 이 방식이 우회되거나 섭동이 적용되지 않은 작품이 무단으로 학습·복제되는 상황까지 차단하기는 어렵다.

이러한 한계를 보완하는 것이 SiGRRW의 역할이다. 불법으로 유통되는 AI 산출물이나 복제물이 발견되었을 때 이미지에 삽입된 워터마크를 추출함으로써 원작자를 확인하고 저작권 침해 여부를 판단하는 데 활용할 수 있다. 원본이 어떻게 변형되었는지 추적하고 유통 경로를 파악하는 단서로도 기능한다. 두 기술이 함께 활용될 때 창작자는 사전 예방과 사후 대응을 아우르는 보호 체계를 갖출 수 있다.

• **콘텐츠 식별 기술의 표준화 필요성과 로드맵**

개별 보호 기술이 산업 전반에서 실효성을 갖추려면 기술 표준화가 중요한 과제로 떠오른다. 현재 연구기관과 기업마다 서로 다른 방식의 보호 기술을 개발하고 있어, 플랫폼별로 기술이 달리 적용될 경우 상호 호환성 문제가 발생할 수 있다. 예를 들어 한 플랫폼에서 삽입한 워터마크를 다른 플랫폼이나 사법기관이 인식하지 못한다면 기술의 사회적 효용이 제한될 수밖에 없다.

이에 웹툰 플랫폼, 기술 개발사, 창작자 단체, 정부 기관이 함께 참여하는 협의체를 구성하여 콘텐츠 식별·보호 기술의 사회적 표준 수립 논의를 본격화할 수 있다. 표준화 로드맵은 크게 세 단계로 구성할 수 있다. 1단계에서는 워터마크 추출 정확도나 섭동 기반 방어 성공률 등 기술의 최소 성능 요건을 정의한다. 2단계에서는 이를 충족하는 기술들이 상호 운용될 수 있는 공통 기술 프로토콜을 개발한다. 3단계에서는 해당 표준을 준수하는 기술에 공식 인증을 부여하고, 웹툰 플랫폼들이 인증 기술을 도입하도록 정책적 지원을 병행한다.

이러한 표준화 과정은 기술의 신뢰도를 높여 법적 증거로 채택될 가능성을 넓힌다. 창작자가 어떤 플랫폼을 이용하더라도 일관된 수준의 보호를 받을 수 있는 기반이 마련된다는 점에서도 의미가 있다. 아울러 기술의 투명성을 확보함으로써 특정 기업의 기술 독점을 방지하고 건전한 기술 개발 생태계를 뒷받침한다. 결국 표준화는 개별 기술을 사회적 인프라로 전환하는 핵심 과정으로 볼 수 있다.

• **저작권 침해 소송에서 기술적 증거의 가능성과 과제**

이러한 기술들이 사회적으로 공인되고 표준화될 경우, 저작권 침해 소송의 양상에도 변화가 생길 수 있다. 현재 AI 관련 저작권 소송에서 가장 큰 난관은 원고인 창작자가 '피고의 AI 산출물이 자신의 저작물에 의존하여 만들어졌다'는 인과관계를 직접 입증해야 한다는 점이다. 이는 고도의 기술적 분석이 필요하며 창작자에게 상당한 부담으로 작용한다.

SiGRRW와 같은 워터마킹 기술은 이 지점에서 객관적인 기술적 증거를 제공할 잠재력을 지닌다. 침해물로 의심되는 이미지에서 원작자의 정보가 담긴 워터마크가 명확하게 추출된다면, '실질적 유사성'과 같은 주관적 판단에 의존하지 않고도 저작물의 불법 복제·변형 사실을 직접 확인하는 자료가 될 수 있다. 소송 과정을 단축하고 창작자의 입증 부담을 덜어주는 효과로도 이어질 수 있다.

물론 기술적 증거가 법정에서 실제로 인정받기까지는 여러 과제가 남아 있다. 해당 기술의 신뢰성과 보안성에 대한 법원의 공식적인 인정이 선행되어야 하고, 데이터가 위·변조되지 않았음을 확인하는 디지털 포렌식 절차도 필요하다. 아울러 전문가의 증언을 통해 재판부가 기술의 원리를 이해하고 증거의 의미를 올바르게 판단할 수 있도록 지원하는 제도적 체계의 정비도 함께 논의될 필요가 있다. 이러한 기반이 갖춰진다면 기술은 가능성에 머무르지 않고 창작자의 권리를 실질적으로 구제하는 수단으로 자리할 수 있다.

• **기술이 만드는 신뢰: 창작자·플랫폼·기업의 역할 변화**

이러한 기술의 확산은 개별 저작권 보호를 넘어 산업 참여자 각각의 역할과 행동 방식에 변화를 가져올 수 있다. 창작자 입장에서는 자신의 작품이 무단으로 학습되거나 도용될 수 있다는 불안에서 벗어나 작품을 공개하고 AI 기술을 창작의 동반자로 받아들이기가 수월해진다. 이는 창작 의욕을 높이고 더 다양하고 풍부한 콘텐츠가 생산되는 선순환으로 이어질 수 있다.

웹툰 플랫폼은 이러한 기술을 도입함으로써 창작 친화적 플랫폼으로서의 신뢰를 높이고, 소속 작가들을 보호하는 동시에 장기적인 경쟁력을 확보하는 기반을 마련할 수 있다. AI 기술 기업 입장에서도 무분별한 데이터 스크래핑 관행을 재검토하고, 창작자에게 정당한 대가를 지불하며 합법적인 데이터를 확보하는 투명한 사업 모델로 전환하는 계기가 될 수 있다.

Ⅲ. 결론 및 전망

• **기술적 증명과 예방: 창작자 권익 보호의 새로운 패러다임**

AI 기술 확산에 따라 웹툰 산업이 직면한 저작권 위협은 세 가지 문제로 요약된다. AI 모델의 불투명한 학습 관행, 앤트로픽 판결로 상징되는 공정 이용 해석의 확대, 그리고 침해 사실을 객관적으로 입증할 수단의 부족이다. 본 보고서는 이 세 가지 문제에 대해 사전 예방 기술과 사후 추적 기술로 대표되는 기술적 접근이 각각 어떤 가능성을 갖는지를 핵심 검토 과제로 삼았다.

두 기술은 이러한 문제에 각각 다른 방향에서 접근한다. 사전 예방 기술(글레이즈)은 AI가 스타일을 학습하는 최초의 단계를 교란함으로써 침해 자체의 가능성을 낮추고, 사후 추적 기술(SiGRRW)은 이미지의 픽셀에 저작권 정보를 삽입해 침해 발생 시 이를 확인하고 대응하는 수단을 창작자 스스로 갖출 수 있도록 한다. 두 기술이 공통적으로 지향하는 방향은 창작자의 '기술적 자율성' 확보다. 창작자가 플랫폼이나 기술 기업의 결정을 기다리는 대신 자신의 저작물을 보호하기 위한 능동적 조치를 직접 취할 수 있게 된다는 의미다.

• **지속 가능한 기술 발전을 위한 제언**

우회 공격 시도, 강인성과 비가시성 사이의 상충 관계, 새로운 AI 모델의 등장에 따른 적용 한계는 방어 기술이 끊임없는 창과 방패의 경쟁 속에 놓여 있음을 보여준다. 현재의 보호 효과가 미래에도 유지되리라는 보장은 없다.

이에 대응하기 위해서는 학계와 산업계가 긴밀히 협력하여 우회 공격 시도를 상시 모니터링하고 방어 알고리즘을 신속히 업데이트하는 공동 연구개발 체계가 필요하다. 동시에 기술의 접근성을 높이는 노력도

병행될 필요가 있다. 아무리 뛰어난 기술이라도 소수의 전문가에게만 닿는다면 실질적 보호 효과는 제한적이다. 웹툰 작가들이 일상적으로 사용하는 클립 스튜디오, 포토샵 등의 창작 소프트웨어에 플러그인(plug-in) 형태로 통합하는 방안도 검토해볼 만하다. 기술의 정교함과 현장에서의 접근성, 이 두 조건이 함께 갖춰질 때 보호 체계는 실질적으로 기능할 수 있다.

• 제도적 흐름과 기술의 접점

기술이 창작자 보호의 실질적 수단으로 기능하기 위해서는 제도적 신뢰가 함께 형성될 필요가 있다. 이 점에서 국내외의 최근 움직임은 주목할 만하다. 미국, 유럽연합(EU), 일본 등 주요국에서 AI와 저작권의 관계를 정립하기 위한 법적·제도적 논의가 활발히 전개되고 있으며, 각국의 접근 방식은 다양하지만 창작자 보호를 강화하는 방향으로 논의가 수렴되는 경향이 관찰된다.

국내에서도 의미 있는 진전이 이루어지고 있다. 문화체육관광부와 한국저작권위원회는 2025년 'AI 활용 저작물의 저작권 등록 안내서'를 공개하며 AI 활용 저작물의 법적 지위에 관한 기준을 처음으로 제시했고,[8] 저작권 보호 미래 포럼을 통해 법·기술·산업계가 함께 저작권 거버넌스를 논의하는 장을 이어가고 있다. 본론에서 제안한 3단계 표준화 로드맵이 실현되기 위한 제도적 토대는 이러한 흐름 속에서 점차 형성되어 가고 있다. 워터마크가 법적 증거로 인정받기 위한 요건, 디지털 포렌식 절차, 기술 전문가 감정 체계에 관한 구체적 기준이 마련될수록 기술의 실효성도 높아질 것으로 기대된다.

• 기술과 제도가 함께 만드는 상생의 AI 창작 생태계

본 보고서에서 검토한 기술들은 AI의 무단 학습을 사전 방지하고, 침해 발생 시 출처를 사후에 증명한다는 두 축에서 창작자 보호에 접근한 기술적 사례다. 사전 예방 기술은 창작자의 스타일이 무단으로 학습되지 않도록 기술적 장벽을 만들고, 사후 추적 기술은 저작물에 추적 가능한 흔적을 남겨 침해 발생 시 책임 소재를 확인하는 데 활용될 수 있다. 이와 같은 방향의 기술이 확산될수록 AI 기업이 무분별한 데이터 수집보다 정당한 이용 허락을 구하는 방식을 선택할 유인이 높아질 수 있다. 기술이 시장 행동에 영향을 미치고, 변화된 시장 관행이 제도화의 근거로 이어지는 흐름이 형성될 수 있다는 점에서 주목할 만하다.

2005년 PC 기반의 무료 만화 코너로 출발한 K-웹툰은 20년 만에 연간 수조 원대 시장으로 성장하며 K-콘텐츠 산업의 핵심 축이 되었다.[9] 이 성장의 토대가 창작자의 독창성이었다면, 다음 20년의 경쟁력은 그 독창성을 AI 시대에도 지켜낼 수 있는 기술적·제도적 인프라와 맞닿아 있다. 두 기술은 그러한 인프라를 향한 구체적인 기술적 시도로서 의미를 갖는다. 이를 뒷받침하는 사회적 대화와 제도적 노력이 함께 이어질 때, K-웹툰 산업이 AI와 함께 지속 가능한 발전의 방향을 모색해 나갈 수 있을 것으로 기대된다.

8) 권혜미, ""AI 활용 창작물도 등록 가능"…정부 첫 가이드라인 제시", 전자신문, 2025.07.01, https://www.etnews.com/20250701000303
9) 백서현, "스무살 맞은 K웹툰, 다음 20년 주역은 AI" 아주경제, 2025.12.08., https://www.ajunews.com/view/20251208143903034

참고문헌

- Shawn Shan 외 5인, "Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models", arXiv, 2025.04.05., https://arxiv.org/pdf/2302.04222
- Zikai Xu 외 4인, "SiGRRW: A Single-Watermark Robust Reversible Watermarking Framework with Guiding Strategy", arXiv, 2026.02.22. https://arxiv.org/html/2602.19097v1
- 조윤주, ""창작부터 보안까지 AI가"...AI 빅뱅 맞는 웹툰업계", 파이낸셜뉴스, 2026.02.19., https://www.fnnews.com/news/202602141404331466
- 김민수, “웹툰 작가들 'AI와 저작권' 논의…한국만화가협회 포럼 개최”, 노컷뉴스, 2026.03.05., https://www.nocutnews.co.kr/news/6479784
- 김지원, “책 수백만권 스캔한 앤스로픽 충격…AI가 뒤흔든 ‘저작권’ ”, 주간경향, 2026.02.23., https://weekly.khan.co.kr/article/202602230600031#ENT
- 안희정, "이용자 거부감보다 기대감 컸다…웹툰에 AI 도입, 업계 인식과 ‘온도차’", ZDNet Korea, 2026.01.07., https://zdnet.co.kr/view/?no=20260106163340
- 권혜미, ""AI 활용 창작물도 등록 가능”…정부 첫 가이드라인 제시", 전자신문, 2025.07.01, https://www.etnews.com/20250701000303
- 백서현, “스무살 맞은 K웹툰, 다음 20년 주역은 AI” 아주경제, 2025.12.08., https://www.ajunews.com/view/20251208143903034
- Emma Roth, “Anthropic wins a major fair use victory for AI — but it’s still in trouble for stealing books”, The Verge, 2025.06.25., https://www.theverge.com/news/692015/anthropic-wins-a-major-fair-use-victory-for-ai-but-its-still-in-trouble-for-stealing-books