Trends in Intellectual Property Research

Innovative

Article

AI Copyright Infringement: Navigating the Legal Risks of AI-Generated Content

Isha Amjad*

Department of Law, University of Punjab, Jhelum Campus Jhelum, Pakistan
* Correspondence: Ishaamjad340@gmail.com

Abstract: The accelerated growth of generative artificial intelligence (AI) tools that can generate text, images, music, code, and multimodal content has caused a legal and philosophical crisis in the field of copyright law. Current study explores two infringement issues, caused by AI-generated content namely the possibility of an infringement of the existing copyrighted works via the unauthorized need integration and processing of the secure material, and the possibility of the infringement of the individual AI output through reproduction, derivation, or significant imitation of the safeguarded expression. Doctrinal legal analysis, authoritative case law reviewed (2023-2025), the US fair use doctrine and EU text and data mining (TDM) exceptions and the AI Act, indicate that current copyright regimes are under a fundamental challenge by generative AI. The legality of the integration of training data, the use of substantial similarity tests on outputs, the controversial issue of originality when it comes to machine productions, assigning liability along the AI value chain, and the development of defense mechanisms and policy reactions are also discussed. This study describes consistent gaps in the dangers of memorization, the possibility to quantify the damage in the markets, and international harmonization. Although the current legal frameworks (especially the strong fair use scrutiny law in the US and opt-out law in the EU) can cover most of the infringement claims, clarity is still required to stabilize the situation and make sure that transformative innovation is not negated by the rights of creators.

Keywords: copyright infringement; AI-generated content; large language models; intellectual property challenges

Citation: Isha Amjad. 2026. The Ghost in the Machine: Navigating Copyright Infringement in AI-Generated Content. *Trends in Intellectual Property Research* 4(1), 24-28. https://doi.org/10.69971/tipr.4.1.2026.104.

1. Introduction

The generative AI models, including OpenAI GPT series (e.g., GPT-4 and GPT-4o), Stable Diffusion, Midjourney, Claude by Anthropic, Llama by Meta, and Gemini by Google, are trained on huge datasets, which in many cases are trained on statistical patterns (billions or trillions of copyrighted materials scraped off the internet) to create new content. There is a two-fold problem of infringement. On the input side, authors, publishers, musicians, visual artists, and rights owners claim that scraping, copying and use of their works without their permission to train are direct infringement of reproduction right. On the output side, AI systems can produce works which are substantially similar to the works used to train these AI systems, and thus infringe rights of reproduction, adaptation, derived work, or public performance. At the same time, the issue regarding the right to copyright AI products and the respective rights of authors, is a legal issue that is not yet resolved. This is impacting the enforcement, economic incentives, and the position of machine-generated content in the public domain.

Current study explores these two inseparable challenges in detail combining US cases, legal case victories in Europe, like *GEMA v. OpenAI* (Munich District Court I, November 2025)[1], and milestones of implementation of the EU AI Act (transparency obligations begin August 2026)[2]. Although copyright law has got adaptive tools, especially fair use in the US and TDM exceptions/opt out in the EU, their use in AI presents challenges that need to be carefully interpreted by the judiciary and eventually need to be addressed by the legislature, either by not over-regulating innovation or by not undermining the incentives for the creator (IP Watch Dog 2025; Smith 2025).

2. Input Challenges of Copyright and Training Data

1 GEMA v. OpenAI: Munich court rules on AI copyright.2025. https://www.twobirds.com
2 https://artificialintelligenceact.eu/

Every generative AI model is trained on a training dataset, which is often assembled with massive amounts of web scraping, often without permission or license to use by the owners of the rights.

The rights holders believe that reproducing works in training samples is the infringement of a fundamental exclusive right in the US law (17 U.S.C. § 106(1))[3] and in the EU law (Article 2 InfoSoc Directive)[4]. OpenAI, Meta, Anthropic, and Stability AI do not reject the practice of copying but protect it by fair use (US) or TDM (EU) exceptions. In 2025, this was popular in Concord Music Group v. Anthropic, in which music publishers had claimed infringement by use of lyrics training (Hostetler 2026).

The critical shield in US is the four-factor test of the fair use under 17 U.S.C. § 107)[5] namely purpose and character of use, nature of the copyrighted work, amount and substantiality of use and effect on the market. AI developers argue that training is transformative because it extracts abstract patterns rather than republishing works. Courts in *Bartz v. Anthropic* (N.D. Cal. 2025)[6] characterized training as "extraordinarily" or "quintessentially transformative," drawing analogies to human learning or non-expressive analysis (Skadden 2025). *Kadrey v. Meta* (N.D. Cal. 2025)[7] similarly found "highly transformative" use, though limited by evidentiary gaps (Debevoise & Plimpton LLP, 2025). Nature of the copyrighted work is often neutral, though creative works (e.g., novels, lyrics) weigh slightly in favor of plaintiffs. Regarding amount and substantiality of use, entire works are copied, but courts tolerate this as necessary for effective learning, since only minimal literal portions appear in outputs. The forth element, i.e., effect on the market is the most contested. *Kadrey* warned of "market dilution" from AI flooding markets with competing content (indirect substitution), which with stronger evidence could decisively weigh against fair use (Loeb & Loeb LLP 2025). Bartz rejected broad market harm claims without proof of infringing outputs.

Authors Guild v. Google (2d Cir. 2015)[8] support transformative scanning use for search and indexing purposes, while *Andy Warhol Foundation v. Goldsmith* (US 2023)[9] heightened scrutiny of commercial use. In contrast, Thomson Reuters v. ROSS (3d Cir. 2025)[10] rejected fair use for non-transformative, competitive copying in legal AI tools, distinguishing generative from non-generative applications (Law.com 2025). Bartz held pirated downloads for non-training-related storage to be infringing, leading to a $1.5 billion class settlement (nearly 500,000 works, approximately $3,000 per title) (Copyright Alliance 2026).

Articles 3-4 of the Directive on the Copyright in Digital Single Market (DSM Directive)[11] provide unconditional research exception and commercial exception with opt-out (machine-readable reservations). Germany's § 44b UrhG[12] limits to research and creates stricter compliance requirements. The AI Act (Regulation (EU) 2024/1689)[13] requires generative AI providers to respect opt-outs (Article 53), publish training summaries, and comply with copyright law. Enforcement beginning August 2026 includes fines up to 3% of global revenue or €15 million. *GEMA v. OpenAI* (Munich 2025)[14] rejected TDM for memorization/reproduction of lyrics in models/outputs and confirmed injunction, disclosure, and damages—the first major EU decision on generative AI infringement (Bird & Bird 2025; Two Birds, 2025). Models can "memorize" excerpts. GEMA held that embedding in model parameters constitutes fixation/reproduction (§§ 16, 19a UrhG) [15] and rejected statistical arguments (Complete Music Update 2025).

3. Output Challenges of Substantial Similarity and "Style" Infringement

Even if training is fair-use- or TDM-exempt, outputs can infringe. Copyright protects expression, not ideas and styles. Infringement requires substantial similarity to protectable elements. Proving causation (from training data) is challenging with "black box" models, but verbatim reproduction (e.g., lyrics via prompt) infringes (GEMA ordered remedies). In *In Re: OpenAI Copyright Litigation* (S.D.N.Y. 2025)[16], courts permitted output-based claims—a shift from earlier dismissals (Debevoise & Plimpton LLP, 2025). Styles are unprotected, but distinctive compilations and series can trigger "comprehensive non-literal similarity." Extreme imitation (e.g., Disney/Midjourney lawsuits 2025 for character reproduction) poses infringement risks (Smith 2025; Wilson 2025). Year 2025 showed renewed interest in output infringement, such as in *Warner Bros. Discovery v. Midjourney*[17] (C.D. Cal. 2025), where "strikingly similar" images imitated copyrighted characters (Best Law Firms 2025).

4. Copyright in AI-Generated Outputs

US and EU systems require human authorship. The US Copyright Office (USCO) and courts reject purely AI works (*Thaler v. Perlmutter* 2023/2025 affirmed)[18]. EU "own intellectual creation" excludes machine outputs without personal decisions. In China,

---

3 https://www.law.cornell.edu/uscode/text/17/106
4 https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:en:HTML
5 https://www.law.cornell.edu/uscode/text/17/107
6 https://authorsguild.org/advocacy/artificial-intelligence/what-authors-need-to-know-about-the-anthropic-settlement/
7 https://www.robinskaplan.com/newsroom/insights/no-harm-no-win-a-cautionary-tale-of-kadrey-v-meta-platforms-inc
8 https://www.copyright.gov/fair-use/summaries/authorsguild-google-2dcir2015.pdf
9 https://www.supremecourt.gov/opinions/22pdf/21-869_87ad.pdf
10 https://www.courtlistener.com/docket/70622297/thomson-reuters-enterprise-centre-gmbh-v-ross-intelligence-inc/
11 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L0790
12 https://itmedialaw.com/en/%C2%A7section-44b-urhg-in-the-context-of-data-mining-of-ai/
13 https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
14 Ibid,1
15 https://itmedialaw.com/en/%C2%A7section-44b-urhg-in-the-context-of-data-mining-of-ai/
16 https://www.courtlistener.com/docket/69879510/in-re-openai-inc-copyright-infringement-litigation/
17 https://www.courtlistener.com/docket/71271014/warner-bros-entertainment-inc-v-midjourney-inc/
18 https://media.cadc.uscourts.gov/opinions/docs/2025/03/23-5233.pdf

courts like the Beijing Internet Court (2025) require proof of creative effort for AI images (Emanuel 2025). Detailed prompts/iterations/edits may be protectable as derivative works/compilations (Zarya of the Dawn partial registration 2023/2025). The USCO 2025 report confirms protection for human-AI hybrids when creativity is human (U.S. Copyright Office 2025). Public domain status discourages investment and complicates enforcement. In Zhangjiajie (China 2025), lack of creative effort led to denial of protection (China IP Law Update 2025).

5. Liability Challenges and Emerging Models

Users are liable for infringing prompts; developers for enabling infringement (knowledge/inducement; Grokster 2005). The AI Act imposes risk management obligations (Article 53). Platforms enjoy safe harbor provisions (DMCA, § 512; EU DSA)[19] with notice and takedown duties[20]. Cases of class actions are on the rise: Bartz settlement highlights piracy punishments. Music cases like *UMG v. Udio* (2025 settled)[21] to licensing solutions (Smith 2025). *OverDrive v. OpenAI*[22] (2025) introduces the element of trademarks (Internet Lawyer Blog 2025). Publishers escalated it in 2025 (e.g., *Advance Local Media v. Cohere*)[23] and studios (*Disney/Universal/Warner v. Midjourney*)[24] where the input and output infringement are considered (Hostetler 2026).

Transformative training is often safeguarded, yet piracy/market substitution bust in. 2025 decisions split: pro-AI (Bartz, Kadrey) vs. restrictions (*Thomson Reuters v. ROSS* 2025[25], no fair use for competitive, non-transformative copying) (Neudata 2025). The AI Act requires summaries and respect for opt-outs; code of conduct (2025) sets compliance benchmarks. Mandatory provenance (C2PA), collective funds, sui generis rights. International harmonization needed given US flexibility vs. EU protection orientation (Brown 2025).US favors training fair use (absent piracy) while EU strict on opt-outs/memorization. Settlements increasing; transparency becoming mandatory (Reuters 2026).

The *GEMA v. OpenAI* [26](Munich District Court I, Case No. 42 O 14139/24, judgment of November 11, 2025) is one of the first major European judicial decisions on copyright infringement in the context of generative AI. GEMA (Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte), Germany's central collecting society for musical performance and mechanical reproduction rights, sued two OpenAI entities: OpenAI LLC (US parent company) and OpenAI Ireland Ltd. (European operator of ChatGPT). GEMA brought claims on behalf of songwriters and publishers whose lyrics were allegedly trained in ChatGPT models (GPT-4 and GPT-4o) and reproduced in outputs without permission. The copyright-specialized Munich District Court I (42nd Civil Chamber) ruled largely in GEMA's favor: injunction, disclosure, and damages were awarded, while moral rights claims were dismissed. The judgment, issued just two months before full enforcement of the EU AI Act's transparency obligations in August 2026, has far-reaching implications for AI developers, rights holders, and the debate over innovation and intellectual property (Bird & Bird 2025; Two Birds 2025).

GEMA represents thousands of composers, lyricists, and music publishers and initiated the lawsuit in late 2024 as part of a broader campaign for fair compensation for works powering AI systems. The case concerned nine well-known German song lyrics, including "Atemlos durch die Nacht" by Helene Fischer and "Mensch" by Herbert Grönemeyer. GEMA alleged these lyrics were scraped from the internet, incorporated into OpenAI's training datasets without license, and "memorized" in model parameters such that simple prompts produced near-verbatim reproductions. OpenAI defended by arguing that models do not store literal copies but learn statistical patterns, infringement is user-initiated, and the process falls under the TDM exception (§ 44b UrhG[27], Article 4 DSM Directive[28]) (Complete Music Update 2025).

The proceeding took place before the copyright-specialized Munich District Court I. An oral hearing in October 2025 foreshadowed the outcome: the court showed skepticism toward OpenAI's user liability argument. The complete 65-page judgment appeared shortly after November 11, 2025. It is not yet final (appeal to the Munich Higher Regional Court expected) but marks a turning point. The case is part of GEMA's strategy, including an ongoing lawsuit against Suno AI (also Munich I) and earlier proceedings against other platforms (Smith 2025).

The court rejected OpenAI's technical and legal defenses, creating precedent. Embedding lyrics in model parameters constitutes fixation and thus reproduction (§ 16 UrhG)[29]. Even without literal storage, the ability to regenerate via prompt proves causal connection. The court rejected the "statistical pattern" argument: "When a model outputs recognizable passages of protected works in response to simple prompts, memorization exists, constituting reproduction." The model is viewed as a "carrier" of protected content (Bird & Bird 2025). TDM exceptions are inapplicable as § 44b UrhG [30](Article 4 DSM Directive) permits analysis, not reproduction or making available to the public. Memorization and regurgitation exceed "extraction for analysis" since outputs serve commercial-generative purposes. Opt-outs were irrelevant as no exception applied (Two Birds 2025). ChatGPT outputs infringe § 19a UrhG. Liability lies primarily with OpenAI due to control over training, architecture, and deployment. User liability was rejected;

---

19 https://www.law.cornell.edu/uscode/text/17/512 ; https://en.wikipedia.org/wiki/Digital_Millennium_Copyright_Act

20 https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

21 https://www.universalmusic.com/universal-music-group-and-udio-announce-udios-first-strategic-agreements-for-new-licensed-ai-music-creation-platform/

22 https://www.courtlistener.com/docket/71932115/overdrive-inc-v-openai-opco-llc/

23 https://www.bakerlaw.com/advance-local-media-v-cohere/

24 https://ipwatchdog.com/2025/09/08/warner-bros-complaint-alleges-midjourneys-copyright-infringement-systematic-willful/

25 https://www.courtlistener.com/docket/70622297/thomson-reuters-enterprise-centre-gmbh-v-ross-intelligence-inc/

26 Ibid,1

27 Ibid,12

28 https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng

29 Ibid, 12

30 Supra

OpenAI negligently failed to implement filters (Complete Music Update 2025). Lyrics of approximately 15 words or more are protectable if characteristic. Moral rights (e.g., distorted lyrics) were denied for lack of proof of reputational harm. OpenAI's research origins do not shield commercial use (Debevoise & Plimpton LLP 2025). The court ordered OpenAI to cease and desist, disclose training data, and pay damages (amount still to be determined). No worldwide injunction, but enforceable in Germany/EU.

Critics accuse the court of not comprehending AI: memorization is not specific to overfitting, but a deliberate storage; parameters as reproduction obscures the fact that it is probabilistic (IP Watch Dog 2025). The ruling would put a hold on doctrine prior to AI Act consideration and slack EU progress. Adherents glorify the reinforcement of the rights to the creators in opposition to the uncompensated use of data (Wilson 2025). As AI Act is implemented in 2026, additional reviews of TDM edge cases will take place. The approach by GEMA illustrates the gathering of roles by societies on the licensing of AI (Complete Music Update 2025).

6. AI Challenges in Academics

Detecting AI-generated plagiarism is not easy. Tools like Turnitin can flag similarities, but AI detectors are unreliable, often misidentifying non-native speakers' work as AI-produced. The plagiarism checking tool cannot read AI and paraphrase it, which requires human intervention and prevention training. Institutions suggest not to be too dependent on detectors but on assignment development that encourages original thinking. Although it is the essence of plagiarism to say that the work produced by the AI was created by one as their own, original work, the specialization of AI as a non-human, algorithmic author introduces some new complications (Cotton, Cotton, and Shipway2023).

Submitting such a product as own work is an inquiry of an ultimate level: Can one plagiarize a machine? Who owns the copyright of AI text, and who is that? The question is how honesty can be redefined in creation where the tool can make up the information without the human consciousness? The conventional plagiarism is a man author whose intellectual contribution is stolen. In the case of AI, the designer is not a human author, as it traditionally is. The AI is a statistical algorithm that is trained on a large body of text created by humans and produces results based on the patterns and prompts.

The fundamental sin is not stealing off a person, but to lie to the process. In case a student presents AI generated work, they deceive claiming that it is a cognitive work of research, synthesis, and articulation. They instill automated collection as individual intellectual attainment hence overstepping the main aim of education, which is to achieve the ability and knowledge. The source (AI) does not matter in this opinion. Critics say that since there is no human writer to whom one can be wrong, then it is just a breach of the use policy and not plagiarism. The AI does not own any rights, and its results are frequently a mashup of the information that is publicly free. The crime, however, is not plagiarism in itself, but on an institutional level against rules of unauthorized assistance i.e. in the manner of contract cheating (Weber-Wulff et al. 2023).

Artificial intelligence is non-deterministic and it is trained on millions of resources. They do not copy and paste, but they interpolate, and therefore traditional text-matching software (e.g., Turnitin) is not as effective. The boundary between AI-assisted writing (e.g. writing with Grammarly, or an AI to brainstorm) and AI-generated writing is unclear. Paraphrasing tools, so-called humanizers, and automatic prompting to evade detection have found quick responses to AI detectors (such as GPTZero, Originality.ai) (Perkins 2023). False positives also affect these detectors especially when the author is not a native writer of the English language or the author has a consistent and simple writing style; this poses a serious equity issue. In the majority of jurisdictions, such as the United States, AI-generated works which lack substantial human authorship have been held to be ineligible to receive copyright. This brings about a paradox, in the event that no one is the owner of the content, who is the plagiarist victimizing? The moral violation is totally transferred to the situation of the submission (academic, professional) and not to the property protection. The academic institutions must be very careful about the adoption of AI generated work with strict rules of disclosure (Khalil and Er 2023).

Good policies should be clear, fair and process oriented. They are supposed to make the difference between its permission (e.g. AI as an editor or brainstorming partner, with documentation) and its unauthorized submission (offering raw AI output under their own name). There should as well be policies that concern the responsible application of AI detectors considering their limitations. The latter is not a punitive but rather a pedagogical solution. Evaluations should change and be more process rather than product. As AI continues to be implemented in writing and creativity software (e.g., Microsoft Copilot, the AI-based Workspace of Google) the idea of pure human authorship will grow more and more of an anachronism.

7. Conclusions

AI infringement is perceived to have adaptation tensions. US fair use is soft, (Bartz, Kadrey favor transformative training is absent/piracy/harming the market), EU is concerned with opt-outs/transparency/memorization liability (GEMA). Although AI-generated content can be plagiarized to disrupt established ideas about authorship, it supports the long-term academic principles of honesty, hard work, and genuine learning. Institutions have to react with subtle policies, reformed evaluation and active learning. It is not possible to stop the wave of technologies, and it is necessary to prepare students to move in that direction without resorting to AI as a consumable good, but as an instrument of enhancing the real human knowledge and imagination. The problem of AI plagiarism is, after all, a chance to reestablish and rebrand our commitment to the things that education is, in fact, meant to develop.

References

Baker, Hostetler. 2026. *Concord Music Group v. Anthropic* case analysis. https://www.bakerlaw.com
Best Law Firms. 2025. Entertainment litigation and AI copyright. https://www.bestlawfirms.com
Brown, Mayer. 2025. EU AI Act implementation and copyright compliance. https://www.mayerbrown.com
China IP Law Update. 2025. AI-generated content and authorship in China. https://www.chinalawupdate.com
Complete Music Update. 2025. GEMA wins landmark case against OpenAI. https://www.completemusicupdate.com

Congress.gov. 2025. AI copyright legislation developments. https://www.congress.gov
Copyright Alliance. 2026. The *Bartz v. Anthropic* settlement: Implications for AI training. https://copyrightalliance.org
Cotton, Dabby R. E., Peter A. Cotton, and J. Reuben Shipway. 2023. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International* 61: 228-239. https://www.tandfonline.com/doi/full /10.1080/14703297.2023.2190148
Debevoise and Plimpton LLP. 2025. Generative AI and copyright: 2025 litigation review. https://www.debevoise.com
Emanuel, Quinn. 2025. AI authorship in Chinese courts. https://www.quinnemanuel.com
Internet Lawyer Blog. 2025. OverDrive v. OpenAI and trademark issues. https://www.internetlawyerblog.com
IP Watch Dog. 2025. AI training and fair use: Current state of litigation. https://www.ipwatchdog.com
Khalil, Mohammad, and Erkan Er. 2023. Will ChatGPT get you caught? Rethinking of Plagiarism Detection. Proceedings of the *arXiv*: 1-13. https://arxiv.org/abs/2302.04335
Law.com. 2025. Thomson Reuters v. ROSS: Third Circuit rejects fair use defense. https://www.law.com
Loeb & Loeb LLP. 2025. Market harm and AI-generated content. https://www.loeb.com
Neudata. 2025. Legal AI and copyright challenges. Retrieved from https://www.neudata.com
OpenAI. 2023. GPT-4 System Card.
Perkins, Mike. 2023. Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice* 20:7-24. https://doi.org/10.53761/1.20.02.07
Reuters. 2026. AI copyright settlements and licensing trends. https://www.reuters.com
Skadden. 2025. Fair use in AI training: The Bartz decision. https://www.skadden.com
Smith, McKool. 2025. Copyright litigation trends in generative AI. https://www.mckoolsmith.com
U.S. Copyright Office. 2023. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence.
U.S. Copyright Office. 2025. Report on copyright and artificial intelligence. https://www.copyright.gov
Weber-Wulff Debora, Alla Anohina-Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero-ib, Olumide Popoola, Petr Šigut and Lorna Waddington. 2023. Can AI-generated text be reliably detected? A review of the current state of detection tools. *International Journal for Educational Integrity* 19:1-21. https://doi.org/10.1007/s40979-023-00146-z
Wilson, Munck. 2025. Character rights and AI image generation. Retrieved from https://www.munckwilson.com