St. John's Law Review

Volume 98, 2024, Number 2 Article 8

The Law and Ethics of AI Creativity

Haochen Sun

THE LAW AND ETHICS OF AI CREATIVITY

HAOCHEN SUN[†]

INTRODUCTION

The rise of generative artificial intelligence ("AI") systems has triggered a backlash among creatives across the globe. In December 2022, artists initiated the No to AI Art movement on social media,[1] primarily as a response to AI companies exploiting their works "without the slightest concept of ethics."[2] These artists expressed deep concern that AI companies had not only used their copyrighted images without consent or compensation, but had given no credit to their artistic contributions at all.[3] In

[†] Professor of Law, University of Hong Kong Faculty of Law; Affiliated Fellow, Information Society Project, Yale Law School. I presented an earlier draft of this article at the 24th Annual Intellectual Property Scholars Conference, the 9th Columbia-Penn Copyright Scholarship Roundtable, the Technology Policy Research Initiative Workshop at Boston University School of Law, the Ideas Lunch Series at the Yale Information Society Project, the *Reframing Intellectual Property Law in the Age of Artificial Intelligence* conference, and *The First Annual US-Asia Comparative Copyright Law Roundtable*. I am grateful to the participants at these events for their feedback and also to Jack Balkin, Oren Bracha, Robert Brauneis, Shyam Balganesh, Anupam Chander, Stacey Dogan, William Fisher, Gideon Parchomovksy, Frank Pasquale, Christopher Robertson, Matthew Sag, Jessica Silbey, Madhavi Sunder, Xiyin Tang, and Christopher Yoo for their helpful conversations and comments. This Article is an output of a research project funded by the Hong Kong General Research Fund (Project Number: 17613223).

[1] Chloe Xiang, *Artists Are Revolting Against AI Art on ArtStation*, VICE (Dec. 15, 2022, 4:25 PM), https://www.vice.com/en/article/ake9me/artists-are-revolt-against-ai-art-on-artstation [https://perma.cc/3HC6-EXD8].

[2] Verity Babbs, *Digital Artists Are Pushing Back Against AI*, HYPERALLERGIC (Mar. 6, 2023), https://hyperallergic.com/806026/digital-artists-are-pushing-back-against-ai/ [https://perma.cc/PC5N-6Z3A].

[3] Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI's Copyright Crisis*, 110 VA. L. REV. ONLINE 207, 209 (2024) ("To create and improve their [generative] AI models, large technology firms have undermined authors' proprietary control over their works by using these works as training data, without consent and often through opaque processes."); Kyle Chayka, *Is A.I. Art Stealing from Artists?*, NEW YORKER (Feb. 10, 2023), https://www.newyorker.com/culture/infinite-scroll/is-ai-art-stealing-from-artists [https://perma.cc/DNQ7-J4US] ("The artists had not consented to have their copyrighted artwork included in the LAION database; they were not compensated for their involvement, even as companies including Midjourney charged for the use of their tools; and their influence was not credited when A.I. images were produced using

their view, this amounted to "daylight robbery"[4] of their creative outputs and constituted "the biggest act of copyright theft in history."[5] In 2023, Hollywood screenwriters joined the movement by launching an unprecedented 148-day strike.[6] Over 15,000 writers, including prominent novelists such as Dan Brown, Suzanne Collins, and Margaret Atwood, endorsed an open letter demanding ethical practices of fair compensation, credit, and author consent.[7]

Proponents of AI have pushed back, arguing that the technology is "simply outpacing our ability to use it ethically."[8] They believe that AI should be prioritized as a disruptive technology with unparalleled power to enhance and democratize creativity. The emergence of platforms like ChatGPT and Midjourney showcases AI's creative capabilities. In terms of efficiency and accuracy, rapidly advancing AI systems can outperform humans by producing more works[9] accomplishing

---

their work."); Naomi Klein, *AI Machines Aren't 'Hallucinating'. But Their Makers Are*, GUARDIAN (May 8, 2023, 4:02 AM), https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein [https://perma.cc/BF3E-GP4E] ("AI art generators are trained on enormous datasets, containing millions upon millions of copyrighted images, harvested without their creator's knowledge, let alone compensation or consent.").

[4] Klein, *supra* note 3. ("AI art . . . is effectively the greatest art heist in history. Perpetrated by respectable-seeming corporate entities backed by Silicon Valley venture capital. It's daylight robbery.").

[5] Kelly Burke, *'Biggest Act of Copyright Theft in History': Thousands of Australian Books Allegedly Used to Train AI Model*, GUARDIAN (Sept. 28, 2023, 11:00 AM), https://www.theguardian.com/australia-news/2023/sep/28/australian-books-training-ai-books3-stolen-pirated [https://perma.cc/CF3M-QN4X].

[6] Ben Schwartz, *AI and the Hollywood Writers' Strike*, NATION (May 8, 2023), https://www.thenation.com/article/economy/ai-and-the-hollywood-writers-strike/ [https://perma.cc/VN8H-K4SZ]; Jennifer Maas, *The Writers Strike Is Over: WGA Votes To Lift Strike Order After 148 Days*, VARIETY (Sept. 26, 2023, 5:07 PM), https://variety.com/2023/tv/news/writers-strike-over-wga-votes-end-work-stoppage-1235735512/ [https://perma.cc/YQ8A-SQ6W].

[7] Brian Fung, *Thousands of Authors Demand Payment from AI Companies for Use of Copyrighted Works,* CNN, https://edition.cnn.com/2023/07/19/tech/authors-demand-payment-ai/index.html [https://perma.cc/H7MS-EJYY] (July 20, 2023, 1:02 PM); The Authors Guild, *Open Letter to Generative AI Leaders*, ACTION NOW, https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders [https://perma.cc/3P43-KEDJ] (last visited Nov. 10, 2024).

[8] Eric James Beyer, *Fear vs. Ethics: Where AI Art Critics Go Wrong*, NFT NOW (Apr. 7, 2023), https://nftnow.com/features/fear-vs-ethics-where-ai-art-critics-go-wrong/ [https://perma.cc/Z24K-R7LT].

[9] Will Henshall, *4 Charts That Show Why AI Progress Is Unlikely To Slow Down*, TIME (Nov. 6, 2023, 4:13 PM), https://time.com/6300942/ai-progress-charts/ [https://perma.cc/RF23-8V4Q] ("AI has surpassed humans at a number of tasks and the rate at which humans are being surpassed at new tasks is increasing.").

tasks with greater precision.[10] This unprecedented level of creativity positions AI as a driving force behind the Fourth Industrial Revolution.[11]

The ethics of creativity lie at the heart of disputes surrounding the legality of generative AI systems. First initiated by creatives such as Sarah Silverman,[12] and then by organizations such as *The New York Times*,[13] waves of lawsuits against AI companies, including Meta, Microsoft, and OpenAI, will determine the legality of generative AI systems' use of copyrighted works.[14] The resolution of these disputes requires consideration of the ethical values that underpin our legal system.

Landmark copyright cases underscore the critical importance of ethics. When determining the legality of Google's Library Project, which involved making verbatim copies of copyrighted books, the Second Circuit supported its fair use decision by considering the ethical value of copyright law.[15] The court stated that "while authors are undoubtedly important

---

[10] COUNCIL OF ECON. ADVISERS, THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE FUTURE OF WORKFORCES IN THE EUROPEAN UNION AND THE UNITED STATES OF AMERICA (Dec. 5, 2022), https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf [https://perma.cc/6VUT-2BJF] ("AI is important because it has the potential to help humans become better at completing cognitive tasks, and to automate tasks that are currently difficult or impossible for humans to do.").

[11] Interview by Wang Chao with Yang Qiang, *The Fourth Revolution*, UNESCO (Oct. 16, 2023), https://courier.unesco.org/en/articles/fourth-revolution [https://perma.cc/G7SC-8LQ3] ("After the internet and mobile internet triggered the Third Industrial Revolution, artificial intelligence (AI) technologies, driven by big data, are fuelling a Fourth Industrial Revolution.").

[12] Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772, 772 (N.D. Cal. 2024). *See also* Alexandra Alter & Elizabeth A. Harris, *Franzen, Grisham and Other Prominent Authors Sue OpenAI*, N.Y. TIMES (Sept. 20, 2023), https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html [https://perma.cc/CTW5-5HAY].

[13] N.Y. Times Co. v. Microsoft Corp., No. 1:23-cv-11195 (S.D.N.Y. Dec. 6, 2024). *See also* Gerrit De Vynck & Elahe Izadi, *New York Times Sues OpenAI, Microsoft for Using Articles To Train AI*, WASH. POST (Dec. 28, 2023, 3:20 AM), https://www.washingtonpost.com/technology/2023/12/27/new-york-times-sues-openai-chatgpt/ [https://perma.cc/7GQ2-WK97].

[14] *Major Cases: Generative AI—Intellectual Property Cases and Policy Tracker*, MISHCON DE REYA LLP, https://www.mishcon.com/generative-ai-intellectual-property-cases-and-policy-tracker [https://perma.cc/X3W9-TZEE] (Oct. 11, 2024).

[15] Authors Guild v. Google, Inc., 804 F.3d 202, 225 (2d Cir. 2015) ("Accordingly, considering the four fair use factors in light of *the goals of copyright*, we conclude that Google's making of a complete digital copy of Plaintiffs' works for the purpose of providing the public with its search and snippet view functions (at least as snippet view is presently designed) is a fair use and does not infringe Plaintiffs' copyrights in their books.") (emphasis added).

intended beneficiaries of copyright, the ultimate, primary intended beneficiary is the public, whose access to knowledge copyright seeks to advance by providing rewards for authorship."[16] Similarly, in the recent case *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*,[17] the Supreme Court relied on the ethical value of copyright law when addressing the intricate issue of the nature and scope of the first factor of the fair use doctrine.[18]

Generative AI systems have utilized vast amounts of copyrighted works without obtaining consent from rights holders. This has ignited controversy about not only the legality, but also the ethical status, of these systems. Many authors contend that generative AI amounts to outright theft of their creative sparks. On the other hand, a great number of technologists advocate for the legitimacy of generative AI, emphasizing its potential to augment creativity in ways that exceed human capacity.

Is it possible to make generative AI systems ethical through legal regulation? Or should we allow AI to disrupt our ethical landscape without regulatory intervention? As these landmark cases have demonstrated, legal decisions relating to copyright and generative AI systems will need to consider not only the letter of the law but also the ethical values that inform and underpin the legal system. These ethical values have been instrumental in refining our comprehension and implementation of legal norms for resolving major social problems,[19] paving the way towards a just and equitable society.[20]

In this Article, I argue that generative AI systems must be ethical, and copyright law can play a pivotal role in achieving this objective. While maximizing the power of AI creativity in generating wealth and improving aesthetics is important, so is

---

[16] *Id.* at 212.

[17] 598 U.S. 508 (2023).

[18] *Id.* at 531("[T]he first factor [of the fair use doctrine] also relates to the justification for the use. In a broad sense, a use that has a distinct purpose is justified because it furthers the goal of copyright, namely, to promote the progress of science and the arts, without diminishing the incentive to create.").

[19] RONALD DWORKIN, TAKING RIGHTS SERIOUSLY 22 (1997) (arguing that a legal principle is "a standard that is observed . . . because it is a requirement of justice or fairness or some other dimension of morality"); RONALD DWORKIN, LAW'S EMPIRE 1 (1986) ("If [a judicial] judgement is unfair, then the community has inflicted a moral injury on one of its members . . . .").

[20] For instance, in the groundbreaking case of *Brown v. Board of Education*, the Supreme Court scrutinized the ethical and legal aspects of racial segregation in public schools, ultimately determining that "in the field of public education the doctrine of 'separate but equal' has no place. Separate educational facilities are inherently unequal . . . ." 347 U.S. 483, 495 (1954).

fostering the ethics of such creativity. Through thoughtful regulation based on proper application of ethical considerations, it may be possible to ensure that generative AI systems respect creatives' copyrights, promote authenticity of information, and foster innovation and progress in the public interest. In making this argument, I examine three interconnected issues in this Article: what the ethics of human creativity entails; why AI creativity also needs to embrace ethics; and how AI creativity can be made ethical.

I first explore the nature and scope of ethics governing human creativity.[21] Embedded within the complex web of human interactions, creativity driving economic growth and aesthetic progress ought to address the ethics of wealth distribution and cultural equality. The legal principles and rules of copyright law embody the ethics of creativity, helping us to navigate the ethical ramifications of human creation and the utilization of knowledge and information. The legality of an individual's creation of a work or use of another's, for example, raises questions about what is right or wrong, fair or unfair, or even just or unjust.[22] Gaining a deeper understanding of the ethical values underpinning our creative abilities can help determine whether these values should be extended to address generative AI systems.

Drawing primarily upon copyright protection practices, I propose three fundamental principles concerning the ethics of human creativity: originality, attribution, and authenticity. As the first principle, the ethics of originality governs the extent to which creatives could copy or "borrow" from others. Meanwhile, the other two principles concerning ethics of attribution and authenticity dictate that creatives should acknowledge the contributions made by others to their works and strive to incorporate accurate information into their works, respectively.

---

[21] *See infra* Part I.B–D.

[22] *See, e.g.*, William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1756 (1988) (discussing the relationship between fair use and distributive justice); Molly Shaffer Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535, 1537 (2005); James Grimmelmann, *The Ethical Visions of Copyright Law*, 77 FORDHAM L. REV. 2005, 2007 (2009) ("When people buy and sell copies of copyrighted works at fairly negotiated prices, they're respecting each others' needs and autonomy."); Shyamkrishna Balganesh, *The Obligatory Structure of Copyright Law: Unbundling the Wrong of Copying*, 125 HARV. L. REV. 1664, 1677 (2012) (discussing "how the wrong of copying operates as a generative idea around which different values and devices in copyright coalesce"); JESSICA SILBEY, AGAINST PROGRESS INTELLECTUAL PROPERTY AND FUNDAMENTAL VALUES IN THE INTERNET AGE 5 (2022).

The second issue addressed in this Article concerns generative AI systems' opacity and the resulting ethical vacuum in which these systems operate. I examine how and why each major process of AI systems—namely the collection, utilization, and generation of works—is a black box.[23] This *status quo* has insulated generative AI systems from the ethics of human creativity, bringing about a range of serious social problems.[24] In response, I assert the importance of integrating ethics into such systems to ensure the responsible development and deployment of AI technologies.[25]

In the third part of the Article, I explore the integration of the ethics of creativity into generative AI systems by proposing two legal responsibilities for AI companies. First, these companies should be legally required to employ filtering technologies that track and remove copyright-infringing outputs generated by their AI systems.[26] Second, they should assume another legal responsibility for implementing watermarking technologies to distinguish AI-generated outputs from human-created works and to curb the dissemination of harmful content, such as disinformation.[27] By adopting these measures, AI companies will ethically promote originality, attribution, and authenticity in the outputs generated by their systems.

This Article makes two theoretical and policy contributions to the discourse on the legality and ethics of AI technology. First, it proposes three principles of ethical creativity that could be applied to govern the generation of content by AI systems. This approach differs from the contemporary discourse on whether the creative outputs of such systems could be protected by intellectual property law through recognizing them as authors and inventors,[28] a legal status typically reserved for humans

---

[23] *See infra* Part II.A–B.

[24] *See infra* Part II.B. *See also* ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE: WHY COMPUTERS CAN'T THINK THE WAY WE DO 240 (2021) ("Shifting the locus of intelligence from humans to machines is a gambit—a chess term, meaning the sacrifice of material for better position—which unavoidably has consequences for human culture.").

[25] *See infra* Part II.B, III.A.

[26] *See infra* Part III.B.

[27] *See infra* Part III.C.

[28] *See, e.g.*, Ryan Abbott, *I Think, Therefore I Invent: Creative Computers and the Future of Patent Law*, 57 B.C. L. REV. 1076, 1104 (2016); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 408 (2019); Daniel J. Gervais, *The Machine as Author*, 105 IOWA L. REV. 2053, 2072 (2020); Carys Craig & Ian Kerr, *The Death of the AI Author*, 52 OTTAWA L. REV. 31, 67 (2021); Dan L. Burk,

only.[29] The U.S. Copyright Office has consistently denied granting author status to AI systems on the grounds that they do not meet this well-established human authorship requirement in copyright law,[30] as their generation of works is achieved "without any creative input or intervention from a human author,"[31] and a federal district court has recently upheld this position.[32]

In this Article, I argue that scholars and regulators should explore a new theoretical and policy direction by considering the application of the ethics of human creativity to generative AI systems, *irrespective of whether the outputs produced by such systems warrant intellectual property protection*. The ethics of originality, attribution, and authenticity—three principles I propose—aim to guide ethical generation of content by AI systems. The ethics of originality can encourage the development of models that refrain from reproducing identical or substantially similar content to existing works. The ethics of attribution can guide AI developers in creating mechanisms for recognizing and crediting the authors of works used as inputs for AI-generated content. Meanwhile, the ethics of authenticity can inspire efforts to combat disinformation and ensure that AI-generated content maintains a high level of accuracy and integrity.

Thirty-Six Views of Copyright Authorship, by Jackson Pollock, 58 HOUS. L. REV. 263, 265 (2020); Haochen Sun, *Redesigning Copyright Protection in the Era of Artificial Intelligence*, 107 IOWA L. REV. 1213, 1231 (2022); Haochen Sun, *Artificial Intelligence Inventions*, 50 FLA. ST. U. L. REV. 61, 84 (2022).

[29] Naruto v. Slater, 888 F.3d 418, 426 (9th Cir. 2018); Shyamkrishna Balganesh, *Authoring the Law*, 68 J. COPYRIGHT SOC'Y U.S.A. 353, 357–58 (2021) ("Copyright has long embodied a fairly distinctive conception of authorship, one that seeks to identify the source of a work and its creative elements as a precondition to protection. . . . Personalization—through identification—is therefore central to copyright . . . .").

[30] U.S. COPYRIGHT OFF. REV. BD., SR1-7100387071, DECISION AFFIRMING REFUSAL OF REGISTRATION OF A RECENT ENTRANCE TO PARADISE 3 (2022), https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf [https://perma.cc/WZ3G-9JD4] (ruling that that "statutory text, judicial precedent, and longstanding Copyright Office practice" all require human authorship as a condition of copyrightability).

[31] *Id*.

[32] Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023). In November 2023, a Chinese court ruled that an AI-generated image qualified as an artwork. The court attributed copyright ownership to a user of Stable Diffusion, reasoning that the individual's prompts and numerous adjustments ultimately resulted in the image that embodied their aesthetic choice and personalized judgment. *See* Yuqian Wang & Jessie Zhang, *Beijing Internet Court Grants Copyright to AI-Generated Image for the First Time*, KLUWER COPYRIGHT BLOG (Feb. 2, 2024), https://copyrightblog.kluwerlawlaw.com/2024/02/02/beijing-internet-court-grants-copyright-to-ai-generated-image-for-the-first-time [https://perma.cc/3TH5-M6LP].

The second contribution of this Article is to propose new responsibilities for AI companies based on the ethics of human creativity, *regardless of whether their use of copyrighted content to train their models is legal or not*. Up until now, the discourse has focused largely on whether such utilization constitutes fair use,[33] a legal doctrine that exempts legal liabilities arising from copyright infringement. This emphasis has resulted in starkly contrasting positions. Drawing heavily on the fair use ruling in the Google Library Project case, many commentators argue in favor of fair use, asserting that the utilization in question not only transforms the purpose and function of the works involved,[34] but also avoids causing considerable harm to their market value.[35] Others disagree, maintaining that this type of utilization lacks a transformative nature and fails to prevent market substitution for the original works.[36]

In this Article, I shift the focus from fair use to the legal responsibilities of AI companies as dictated by the ethics of human creativity. I agree in principle that using copyrighted works to train AI models constitutes fair use, as it promotes innovation in the public interest.[37] However, the generation of works by trained

---

[33] Will Oremus & Elahe Izadi, *AI's Future Could Hinge on One Thorny Legal Question*, WASH. POST (Jan. 4, 2024, 7:00 AM), https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/ [https://perma.cc/W88F-7JCJ].

[34] Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 748 (2021) (arguing that "a[ ] [machine learning] system's use of the data often *is* transformative as that term has come to be understood in copyright law, because even though it doesn't change the underlying work, it changes the purpose for which the work is used").

[35] Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, UCLA L. REV. (forthcoming 2024) (manuscript at 82), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4631726 [https://perma.cc/FHJ5-DDBJ] ("The class action market harm allegations against Alphabet, Meta, OpenAI, and Stability seem at this stage too speculative to weigh against fair use."); James Vincent, *The Scary Truth About AI Copyright Is Nobody Knows What Will Happen Next*, THE VERGE (Nov. 15, 2022, 10:00 AM), https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data [https://perma.cc/HQ7Y-HHMQ] ("If the model is trained on many millions of images and used to generate novel pictures, it's extremely unlikely that this constitutes copyright infringement. The training data has been transformed in the process, and the output does not threaten the market for the original art.").

[36] Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 67 (2017) ("[E]xpressive machine learning presents a new threat of market substitution that alters the analysis of the fourth fair use factor."); Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295, 330 (2023) (concluding that copyrightable characters may easily provoke copyright infringement by AI systems).

[37] Lemley & Casey, *supra* note 34, at 748 ("Broad access to training sets will . . . ultimately making artificial intelligence systems using ML algorithms better,

models is an entirely different and more complex situation.[38] If AI systems generate works that infringe on copyright or disseminate harmful content such as disinformation, the fair use doctrine cannot protect them from legal liabilities.[39] Consequently, it is crucial to establish the extent of legal responsibility AI companies must assume to prevent and address these illegal and unethical activities.

In light of these challenges, I propose how and why regulators should impose two legal responsibilities on AI companies in line with the ethics of creativity. First, AI companies should be legally required to proactively implement filtering technologies that monitor and remove AI-generated works that appear identical or substantially similar to copyrighted works. Second, they should assume the legal responsibility to utilize watermarking technologies to differentiate between works created by AI and humans, while also preventing the generation of harmful content. The ethics of human creativity would provide stronger justifications for these measures, which regulators are considering adopting in China, the European Union, and the United States.

This Article comprises three parts. In Part I, after demonstrating that creativity is a social process, I explore the ethical principles of originality, attribution, and authenticity of creative activities. In Part II, I scrutinize the opacity of AI systems in the collection, utilization, and generation of works, highlighting the need for a greater focus on the legal and social problems arising from these black box processes. Lastly, in Part III, I explore why the ethical principles of originality, attribution, and authenticity should govern AI creativity, and how they would lead to the adoption by AI companies of filtering and watermarking responsibilities.

I. THE ETHICS AND LAWS OF HUMAN CREATIVITY

"Good artists copy; great artists steal." This aphorism, attributed to Pablo Picasso, highlights the importance of deriving inspiration from others, rather than advocating for the theft of ideas. A good artist may merely excel at replication; a great

safer, and fairer."); HAOCHEN SUN, TECHNOLOGY AND THE PUBLIC INTEREST 20 (2022) ("As the Universal Declaration of Human Rights requires, technology should be used in the public interest.").

[38] *See* Pasquale & Sun, *supra* note 3, at 209, 211 (discussing the ways and reasons for holding AI companies legally and ethically accountable for such activities).

[39] *See infra* Sections II.B, III.A.

artist—like Picasso—possesses the ability to learn from others and skillfully combine influences to create something uniquely their own.

Numerous visionaries, including T. S. Eliot[40] and Steve Jobs,[41] have shared similar perspectives. They highlight the ethical dimension intrinsic to human creativity, emphasizing the importance of respecting and acknowledging the sources of inspiration that shape and fuel our innovative pursuits.

In this Part, I show that the creative process is subject to ethical evaluation because it is embedded within a social context. It relies on existing knowledge and information as input, and its engagement with an audience enriches the meaning of what is created.[42] This dynamic process gives rise to three ethical principles that govern the originality, attribution, and authenticity of creative endeavors, ensuring their integrity while acknowledging and respecting their roots in the collective human experience.

A. The Social Nature of Human Creativity

From an individualistic perspective, explorations of human creativity focus on a person's contributions to the generation of new ideas. The emphasis is on individual inspiration and initiative—moments of revelation experienced by the solitary artist or discovery of new knowledge by a researcher working alone.[43] This notion has a long history. Renaissance artists such as Albrecht Dürer, Leonardo da Vinci, and Donatello were portrayed as gifted geniuses possessing "quasi-divine power,"[44]

---

[40] T.S. Eliot declared, "Immature poets imitate; mature poets steal; bad poets deface what they take, and good poets make it into something better, or at least something different." *See* Ben Shoemate, *What Does It Mean—Good Artists Copy, Great Artists Steal*, MEDIUM (Aug. 2, 2012), https://medium.com/ben-shoemate/what-does-it-mean-good-artists-copy-great-artists-steal-ee8fd85317a0 [https://perma.cc/488L-N2VE].

[41] LIBRIS MEDIA, STEVE JOBS: INSANELY GREAT QUOTES 7 (Jesper Bove-Nielsen & Christine Lejre eds., 2011).

[42] Seana Moran, *Introduction: The Crossroads of Creativity and Ethics*, *in* THE ETHICS OF CREATIVITY 1, 12 (Seana Moran et al. eds., 2014) ("Creativity is also social—in collaboration, in interplay of ideas across minds, in judgements of value.").

[43] Ernest Edmonds et al., *Panel: Individual and/versus Social Creativity*, PROC. 3D CONF. ON CREATIVITY & COGNITION 36 (1999). *See also*, Mark A. Lemley, *The Myth of the Sole Inventor*, 110 MICH. L. REV. 709, 710 (2011) ("[T]he very theory of patent law is based on the idea that a lone genius can solve problems that stump the experts, and that the lone genius will do so only if properly incented by the lure of a patent.").

[44] GIANCARLO F. FROSIO, RECONCILING COPYRIGHT WITH CUMULATIVE CREATIVITY 144 (2018).

challenging the perception of creatives as mere professional artisans.

While this individualistic notion of creativity has become less prominent over time, modern forms of creative output continue to exhibit individual creativity. For example, despite questions about the level of creativity involved in photography—a process involving the capture of reality through mechanical and chemical means—we recognize individual artistry in the arrangement and composition of an image.[45]

What challenges traditional conceptions of individual creativity is the reality of human creativity as a social process.[46] While Renaissance artists were advocating for individual genius, there were already those who argued that almost everything there was to say had been said, recognizing "transformative elaboration"[47] as a form of creativity. Essentially, when a moment of individual inspiration occurs, one cannot ignore the numerous conversations and interactions that preceded it.[48] This social nature of human creativity is evident in traditional forms of creative output across domains.

In literature, collaborative and participative practices can be seen in the oral-formulaic tradition that gave rise to Homer's *Iliad* and *Odyssey*, and in the textual art of rewriting, which allowed for the free reuse and remodeling of iconic figures and characters such as King Arthur.[49] Modern works also embody these creative practices. James Joyce's renowned *Ulysses* is an almost point-for-point inversion of the *The Odyssey*,[50] while the works of Shakespeare are constantly reframed and rewritten.[51] *A Song of*

---

[45] Ginsburg & Budiardjo, *supra* note 28, at 354–58; Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 60 (1884).

[46] *See* Haochen Sun, *Patent Responsibility*, 17 STAN. J.C.R. & C.L. 321, 347 (2021) (concluding that "innovation is by nature a social process").

[47] *See* Edmonds et al., *supra* note 43, at 37–38.

[48] *See, e.g.*, RICHARD POSNER, LAW AND LITERATURE 403 (Harvard Univ. Press rev. ed. 1998) ("The literary imagination, as should be apparent from the earlier discussion of the use of living persons as models of fictional characters, is not a volcano of pure inspiration but a weaving of the author's experience of life into an existing literary tradition.").

[49] *See* FROSIO, *supra* note 44, at 70–103.

[50] *Books You Didn't Know Were Based on Other Famous Books*, PENGUIN (Sept. 28, 2020), https://www.penguin.co.uk/articles/2020/09/books-based-on-other-books [https://perma.cc/VG5Q-RKCY].

[51] *See, e.g.*, Jordan Payeur, *10 Popular Movies That Were Actually Shakespeare Adaptations*, CBR (Jan. 11, 2023), https://www.cbr.com/famous-films-shakespeare-adaptations/ [https://perma.cc/AD8R-PVFE]; Andrea Oh, *10 Shakespeare Retellings Adapted for the Modern Era*, ELECTRIC LIT (Apr. 29, 2019),

Ice and Fire, the book series that inspired the TV phenomenon *Game of Thrones*, was partly influenced by the author's curiosity about the tax policy *Lord of the Rings*' Aragorn would implement after becoming king.[52] Musical innovation, spanning diverse genres like jazz, punk, and hip-hop, often emerges from improvisation and risk-taking with existing material.[53] These examples serve to illustrate the inherent interconnectedness and social nature of creativity, as it thrives on building upon and transforming existing ideas and works.

Acknowledging the social nature of the creative process not only refutes the individualistic notion of creativity but also highlights the importance of embracing the interconnectedness and collective influence that shapes it. In contemporary culture, "[c]reativity almost always involves the combination of prior ideas and work."[54] Even the creativity attributed to the archetypal individual author is difficult to separate from its historical and cultural context, as individuals are situated within specific cultural settings. Within their respective creative fields, these individuals either conform to or challenge the prevailing networks of knowledge and domains of expertise.[55]

Another conventional perspective on human creativity views it as a mechanical process, focusing solely on the cognitive aspects of creativity and disregarding the evaluation of its social impact. According to this interpretation, human creativity involves "unfocused thought in which abstract ideas are generated" and then "crystallization of the unfocused thought into concrete perceivable products using task-relevant knowledge and

---

https://electricliterature.com/10-shakespeare-retellings-adapted-for-the-modern-era/ [https://perma.cc/KZ9K-UTZV]; Melissa Baron, *The Best Modern Shakespeare Adaptions*, BOOK RIOT (Sept. 15, 2021), https://bookriot.com/modern-shakespeare-adaptations/ [https://perma.cc/XZC9-SZHT].

[52] Mikal Gilmore, *George R.R. Martin: The Rolling Stone Interview*, ROLLING STONE (Apr. 24, 2014), https://www.rollingstone.com/culture/culture-news/george-r-r-martin-the-rolling-stone-interview-242487/ [https://perma.cc/EQ59-SPU4]; Juliette Harrisson, *How Tolkien and Lord of the Rings Inspired Game of Thrones*, DEN OF GEEK (July 6, 2023), https://www.denofgeek.com/tv/how-tolkien-and-lord-of-the-rings-inspired-game-of-thrones/ [https://perma.cc/L6NX-GJZ7].

[53] Olufunmilayo B. Arewa, *Creativity, Improvisation, and Risk: Copyright and Musical Innovation*, 86 NOTRE DAME L. REV. 1829, 1840–42, 1844 (2011).

[54] Gregory N. Mandel, *To Promote the Creative Process: Intellectual Property Law and the Psychology of Creativity*, 86 NOTRE DAME L. REV. 1999, 2014 (2011).

[55] Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1178–89 (2007).

memory."[56] This notion aligns with the etymology of the verb "to create," which stems from the Latin "*creare*" meaning "to make, to produce in a physical sense."[57] If this is used as a guide for our understanding, "creativity" should refer only to the act of making; it should not describe any mental process of insight or discovery, nor reference the perceived nature of the outputs generated, or be used "when the term *original* would be adequate."[58]

This mechanistic perspective offers a limited view of creativity, neglecting its inherently social and interconnected nature, as well as the importance of recognizing its broader impact on society and culture. In the field of conceptual art, creators often adopt a passive role in the realization of their work, allowing their subjects to shape the direction of the piece rather than asserting direct creative control.[59] This approach highlights the collaborative and social nature of creativity, as well as the importance of acknowledging the contributions of others in the creative process. For instance, in a work titled *Obliteration Room*, an artist gave gallery patrons colored stickers to apply to a white room, allowing the room to become covered over the course of the exhibition.[60] In the creation of *Following Piece*, an artist took a video camera and followed random subjects through public streets until they entered private establishments. Both works involve minimal levels of direct authorial input and offer subjective creative value, but both artists have undeniably engaged in the act of creation.

These examples underscore the fact that creativity is not merely about individualistic genius or mechanistic processes; it is also about fostering an environment that encourages participation, interaction, and exchange of ideas among people. In fostering this environment, artists can tap into the rich and diverse pool of human experiences, which ultimately leads to a more profound and meaningful understanding of creativity and its impact on society and culture.

Human creativity has larger social impacts as a driver of new ideas and technologies. Technological advancements in cloud, big

---

[56] Omri Rachum-Twaig, *Recreating Copyright: The Cognitive Process of Creation and Copyright Law*, 27 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 287, 351 (2017).

[57] Ignacio L. Götz, *On Defining Creativity*, 39 J. AESTHETICS & ART CRITICISM 297, 298 (1981).

[58] *Id.*

[59] Christopher Buccafusco, *How Conceptual Art Challenges Copyright's Notions of Authorial Control and Creativity*, 43 COLUM. J.L. & ARTS 375, 376–77 (2020).

[60] *Id.* at 377.

data, mobile, and social network infrastructure have been made possible through open-source information technology communities, where social collaboration is vital to software development.[61] Remix culture,[62] facilitated by social media and advancements in image, video, and audio editing technology, has enabled amateur and referential creativity to thrive.[63] This allows those "formerly known as the audience"[64] to become active participants in their own culture, further emphasizing the social nature of creativity.[65]

In essence, people find creative inspiration in the work of others, using it as a starting point for original output. Authors routinely recontextualize, remix, substitute, or otherwise mash up existing work to create something new. This phenomenon highlights the importance of recognizing and embracing the inherently social, collaborative, and interconnected nature of human creativity in today's rapidly evolving digital landscape.

B. Three Principles of Ethical Creativity

The social process is crucial in shaping the ethical dimension of human creativity. This is because the ethics of human creativity guide us to deal with "others" in the creation and dissemination of our works. For example, such principles can help mediate on issues such as the extent to which we could "borrow" from others' works, subsequently give credit to others' contributions to our

---

[61] Phil Granof, *Social Creativity: The Engine of Software Development in the Social Era*, WIRED, https://www.wired.com/insights/2013/10/social-creativity-the-engine-of-software-development-in-the-social-era/ [https://web.archive.org/web/20240422044544/https:/www.wired.com/insights/2013/10/social-creativity-the-engine-of-software-development-in-the-social-era/] (last visited Sept. 29, 2023).

[62] *See generally* LAWRENCE LESSIG, REMIX: MAKING ART AND COMMERCE THRIVE IN THE HYBRID ECONOMY (2008).

[63] Rebecca Tushnet, *Legal Fictions: Copyright, Fan Fiction, and a New Common Law*, 17 LOY. L.A. ENT. L. REV. 651, 652 (1997) ("As legends and folktales of Coyote the Trickster or Paul Bunyan previously brought audiences together, modem secondary creativity allows fans to transcend passive reception, using material to which they have easy access."); William W. Fisher III, *The Implications for Law of User Innovation*, 94 MINN. L. REV. 1417, 1418 (2010) ("Digital mashups are created by combining audio, video, graphical, or textual material from preexisting works into new digital works."); Anupam Chander & Madhavi Sunder, *Everyone's a Superhero: A Cultural Theory of "Mary Sue" Fan Fiction as Fair Use*, 95 CALIF. L. REV. 597, 600 (2007).

[64] *See* FROSIO, *supra* note 44, at 280–86.

[65] Christopher S. Yoo, *Rethinking Copyright and Personhood*, 2019 UNIV. ILL. L. REV. 1039, 1069 (2019) (arguing that "authors inevitably internalize their expectations of their readers' likely responses into their work").

works, as well as how we could embody true information into our works as much as we could.

Hence, in this section, I reveal that there are three fundamental ethical principles of originality, attribution, and authenticity that govern the social process of human creativity. Subsequently, I explore how these principles have been integrated into copyright law and other relevant laws that serve as the vehicles for realizing the ethos of ethical creativity.

1. Ethics of Originality

While drawing upon others' intellectual output is essential for human creativity, the ethics of originality guard against improper appropriation of such output.[66] Originality is a widely accepted standard for assessing the quality of a creative work, particularly when determining whether it merits copyright protection.[67] For instance, copyright law adopts originality as a legal standard to establish which works qualify for copyright protection. In the landmark case of *Feist Publications., Inc. v. Rural Telephone Service Co.*, [68] the Supreme Court ruled that originality necessitates an author's independent creation of a work possessing "some minimal degree of creativity." [69] The level of this requirement, according to the Court, is relatively low. As long as an independently created work possesses even the slightest creative spark, it will satisfy the originality criterion. [70] Originality also possesses an ethical dimension, which primarily safeguards against improper appropriation resulting in a work of substantial similarity. [71] As an ethical principle, it helps determine the extent and manner in which an author can draw upon informational resources during the creative process. Although the *Feist* ruling does not precisely define the degree of

[66] Balganesh, *supra* note 22, at 1679 (arguing that "copyright law treats copying as a wrong because copying original expression remains an independent moral wrong").

[67] William W. Fisher III, *Recalibrating Originality*, 54 HOUS. L. REV. 437, 438 (2016) ("In all countries, a work must be 'original' to be entitled to copyright protection.").

[68] 499 U.S. 340 (1991).

[69] *Id.* at 345 ("Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity.").

[70] *Id.* at 345 ("[T]he requisite level of creativity is extremely low; even a slight amount will suffice. The vast majority of works make the grade quite easily, as they possess some creative spark, 'no matter how crude, humble or obvious' it might be.").

[71] Arnstein v. Porter, 154 F.2d 464, 468–69 (2d Cir. 1946).

originality needed for a work to be copyrightable, it does require that an author should not create a work substantially similar to another's work. The ruling implicitly endorses the ethical principle of originality through its interpretation of the concept:

> Originality does not signify novelty; a work may be original even though it closely resembles other works, so long as the similarity is fortuitous, not the result of copying.[72]

This observation establishes an ethical boundary that creators must not cross when incorporating elements from another's work into their own creation. It emphasizes the importance of respecting the rights of original creators while simultaneously allowing for the natural exchange and evolution of ideas that drive human creativity. Possessing "the creative powers of the mind,"[73] creators must take responsibility for ensuring that their work embodies a sufficient amount of "the fruits of [their] intellectual labor."[74]

If an individual deliberately copies a substantial portion of another's work, the resulting work would be deemed ethically inappropriate and, therefore, legally unoriginal.[75] In contrast, if he or she derives ideas from earlier works but infuses those ideas with original expression, that "creative" appropriation would be considered to meet the originality requirement, both ethically and legally. This distinction highlights the delicate balance between drawing inspiration from existing works and maintaining the integrity of the creative process by avoiding improper appropriation.

The ethics of originality also require individuals to refrain from copying the core or essential part of a work. To promote creativity at a societal level,[76] the fair use privilege allows members of the public to make unauthorized use of a work without paying remuneration to its copyright holder. In determining the

---

[72] *Feist*, 499 U.S. at 345.

[73] Trade-Mark Cases, 100 U.S. 82, 94 (1879).

[74] *Id. See also* WILLIAM F. PATRY, PATRY ON COPYRIGHT § 3:28 ("The key is whether original matter in which protection is claimed is the result of plaintiff's ingenuity rather than appropriation of another's material.").

[75] The *Feist* ruling provides an example to illustrate this ethical principle: "assume that two poets, each ignorant of the other, compose identical poems. Neither work is novel, yet both are original and, hence, copyrightable." *Feist*, 499 U.S. at 346. *See also* Sheldon v. Metro-Goldwyn Pictures Corp., 81 F.2d 49, 54 (2d Cir. 1936).

[76] Haochen Sun, *Fair Use as a Collective User Right*, 90 N.C. L. REV. 125, 174 (2011) ("With the embrace of the collective right to fair use, copyright law would be built on the principle that knowledge and information form an essential resource that empowers people to engage in communicative actions in intangible public space.").

legality of such use, one factor is "the amount and substantiality of the portion" of the work used by its appropriator.

Most importantly, taking even a small portion of the "heart," or "most valuable and pertinent portion,"[77] of the work may weigh against a finding of fair use,[78] as it is very likely to negatively affect the commercial value of the work and disrupt market order. In *Harper & Row v. Nation Enterprises*,[79] *The Nation* magazine published a 300–400-word excerpt containing the most significant and newsworthy portions of President Ford's memoir. This prompted its publisher, Harper & Row, to sue *The Nation* for copyright infringement. The Supreme Court ruled in favor of Harper & Row, stating that the third factor of the fair use doctrine weighed against *The Nation* and holding that even though the excerpt was quantitatively small, its qualitative importance and the fact that it was published before the release of the memoir negatively affected the potential market for the copyrighted work.[80]

In *Dr. Seuss Enterprises, L.P. v. ComicMix LLC*,[81] the court found that ComicMix copied the heart of Dr. Seuss's work by replicating its structure, rhyme scheme, and illustrations. The decision to copy these elements was likely driven by ComicMix's intention to create a mash-up that would be recognizable and appealing to fans of both Dr. Seuss and Star Trek. By incorporating the most iconic and distinctive aspects of Dr. Seuss's work, ComicMix aimed to evoke the original work's style and charm while adding a Star Trek-themed twist. The case highlights the importance of considering the qualitative significance of the copyrighted material used and the need for a new work to be sufficiently transformative to qualify for fair use protection.[82]

2. Ethics of Attribution

In the social context of creative activities, the ethics of attribution are integral. These ethical norms encourage creators to give proper credit to the others' original works that they have drawn upon to generate or develop their new works.

---

[77] L.A. News Serv. v. CBS Broad., Inc., 305 F.3d 924, 940 (9th Cir. 2002).

[78] Monge v. Maya Mags, Inc., 688 F.3d 1164, 1178 (9th Cir. 2012) (ruling that taking "the 'heart' of each individual copyrighted picture" tilts the third factor against fair use).

[79] 471 U.S. 539, 544–45 (1985).

[80] *Id*. at 555.

[81] 983 F.3d 443, 454–55 (9th Cir. 2020).

[82] *Id.* at 456.

In various fields of creative and intellectual work, acknowledging the contributions of those who have supported the development of ideas but do not qualify for authorship status is an essential aspect of the ethics of attribution. Words of gratitude and recognition giving credit to these individuals are often offered in a book or article's acknowledgments. In academic writing, a typical biographical footnote first identifies the title and institutional affiliation of an author, and then acknowledges the individuals who have supported the paper or deserve recognition for any reason. While some authors in legal scholarship might use acknowledgments strategically to signal quality, in other fields, they serve more effectively as a means of attribution.[83] For instance, in the hard sciences, research projects often involve a larger network of academic contributors.[84] In this context, acknowledgment data is included in publication records in databases such as the Web of Science.[85]

These examples demonstrate that proper attribution plays a crucial role in promoting ethics.[86] By giving credit to an original author, a researcher, for example, upholds the ethical sense of intellectual honesty and integrity. Intellectual honesty involves being truthful, fair, and transparent regarding the sources and origins of one's ideas and research.[87] Proper attribution signifies an acknowledgment that the presented work is not entirely one's own and relies on the intellectual contributions of others.[88] This practice helps maintain high standards of honesty and accuracy in sharing knowledge and ideas, as it exhibits a commitment to transparency and truthfulness about the work's sources of support and inspiration. Without appropriate attribution, individuals risk being perceived as misleading their audience, potentially damaging their credibility and reputation within their field.[89]

---

[83] Jonathan I. Tietz & W. Nicholson II Price, *Acknowledgments as a Window into Legal Academia*, 98 WASH. U. L. REV. 307, 319 (2020).

[84] Davide Simone Giannoni, *Book Acknowledgements Across Disciplines and Texts*, *in* ACADEMIC DISCOURSE ACROSS DISCIPLINES 165 (Ken Hyland & Marina Bondi eds., 2006).

[85] *See* Tietz & Price, *supra* note 83, at 319.

[86] *See also* Xiyin Tang, *Art After Warhol*, 71 UCLA L. REV. 870, 923–25 (2024) (discussing the ethics values of proper attribution through various examples).

[87] INT'L CTR. FOR ACAD. INTEGRITY, FUNDAMENTAL VALUES OF ACADEMIC INTEGRITY 5 (3d ed., 2021).

[88] *Id.* at 8.

[89] Carol M. Bast & Linda B. Samuels, *Plagiarism and Legal Scholarship in the Age of Information Sharing: The Need for Intellectual Honesty*, 57 CATH. U. L. REV.

Such intellectual honesty, in turn, increases the trustworthiness of the output. To many creators, a sense of "ownership" over output provides significant intrinsic motivation to create,[90] and engaging in active efforts to credits others' authorship can have the effect of strengthening their own creative contribution and claim of authorship. For example, authors who engage transparently with prior works demonstrate to their audience that they have understood pre-existing concepts and practices sufficiently to distinguish from their own, thereby instilling trust in the value of their contribution.[91]

Proper attribution also encourages ethical behavior in others, fostering a dynamic collaborative environment in knowledge production.[92] When individuals consistently practice proper attribution, they set a positive example within their field, helping to establish an ethical culture in which colleagues and peers are more likely to adhere to standards of intellectual honesty and integrity.[93] This approach contributes to the overall ethical climate of a discipline or profession, promoting a shared commitment to honesty, fairness, and transparency.

Given that academic and professional work often builds upon the contributions of others, recognizing authorship acknowledges the collaborative nature of knowledge production and underscores the importance of respecting and valuing the work of others.[94] This acknowledgment cultivates a sense of

---

777, 778 (2008) ("[A] writer who does not cite to the original author risks being unprofessional, giving offense, and being labeled a plagiarist.").

[90] Diane Leenheer Zimmerman, *Copyrights as Incentives: Did We Just Imagine That?*, 12 THEORETICAL INQUIRIES L. 29, 44 (2011) (quoting Eric von Hippel & Georg von Krogh, *Open Source Software and the "Private-Collective" Innovation Model*, 14 ORG. SCI. 209, 216 (2003)).

[91] ROBERT A. HARRIS, USING SOURCES EFFECTIVELY: STRENGTHENING YOUR WRITING AND AVOIDING PLAGIARISM 8–9 (5th ed. 2017).

[92] For example, a study examined the difficulties encountered by amateur animators participating in collaborative online projects, some of which involved up to fifty contributors. These animators faced the challenge of attributing authorship to a maximum of ten individuals. Project leaders often determine attribution based on the quality of work, sometimes even before the project commences. In these situations, an animator's desire to distinguish themselves undoubtedly shapes their creative approach. However, they must also adhere to a pre-established project direction and consider the contributions of other participants. *See* Kurth Luther, Nicholas Diakopoulos & Amy Bruckman, *Edits & Credits: Exploring Integration and Attribution in Online Creative Collaboration*, 2010 ASSOC. COMPUTING MACHINERY CONF. ON HUM. FACTORS COMPUTING SYS., 2823, 2828 (2010), https://dl.acm.org/doi/abs/10.1145/1753846.1753869 [https://perma.cc/EB9M-RF6N].

[93] *Id.*

[94] *Id.*

community and shared responsibility within a field, where individuals understand that their success and the contributions of colleagues and peers are interconnected.[95]

The ethical norms governing proper attribution protect the reputational and pecuniary interests of creators whose intellectual outputs are used by others. First and foremost, proper attribution safeguards a creator's reputation as the author of a work.[96] Authorship is a "source of honor and pride," and "authors have a 'deep interest' in having their works correctly attributed to them."[97] While U.S. copyright law emphasizes the importance of economic incentives for authors to create, many are driven by "more complex reputation-based interests" rather than "the straightforward pecuniary interests of publishers."[98]

In academia, a field in which professional success is driven by research output, the importance of appropriate authorship credit is particularly apparent. For instance, the recent outcry in response to the publication of articles listing AI models like ChatGPT as contributing authors highlights the significance of proper attribution within the academic community.[99]

The incentive to create new works indeed often rests on recognition, particularly for academic authors and creators of user-generated content on the internet. Many academics, such as professors, devote a significant amount of time to writing papers, books, and blogs to gain recognition and enhance their reputation within their field.[100] This can lead to career advancement, increased collaboration opportunities, and a sense of personal fulfillment.

Modern forms of internet-based creativity, such as user-generated content, are significantly influenced by the pursuit of

---

[95] *Id.*

[96] Jeanne C. Fromer, *Expressive Incentives in Intellectual Property*, 98 VA. L. REV. 1745, 1790 (2012) ("Attribution makes it easy to broadcast a creator's involvement, enabling the public to give kudos to the creator.").

[97] U.S. COPYRIGHT OFF., AUTHORS, ATTRIBUTION, AND INTEGRITY: EXAMINING MORAL RIGHTS IN THE UNITED STATES 34 (2019), https://www.copyright.gov/policy/moralrights/full-report.pdf [https://perma.cc/Q2WW-P8S8].

[98] Greg Lastowka, *Digital Attribution: Copyright and the Right to Credit*, 87 B.U. L. REV. 41, 42 (2007).

[99] Chris Stokel-Walker, *ChatGPT Listed as Author on Research Papers: Many Scientists Disapprove*, NATURE (Jan. 18, 2023), https://www.nature.com/articles/d41586-023-00107-z [https://perma.cc/U7PF-6GXJ]; Ian Sample, *Science Journals Ban Listing of ChatGPT as Co-Author on Papers*, GUARDIAN (Jan. 26, 2023), https://www.theguardian.com/science/2023/jan/26/science-journals-ban-listing-of-chatgpt-as-co-author-on-papers [https://perma.cc/Q3JM-ZHVQ].

[100] Lastowka, *supra* note 98, at 60.

attribution. This drives participants to create distinctive contributions that warrant recognition. Content creators, including bloggers, vloggers, and social media influencers, devote substantial time and effort to crafting content that captivates, educates, or entertains. Their ultimate goal is to amass a loyal following, expanding their online presence and solidifying their status as thought leaders or influencers within their respective niches.[101] This phenomenon not only highlights the competitive nature of the digital landscape but also emphasizes the importance of producing high-quality content that resonates with the target audience. Further, by participating in open platforms, creators can gain recognition and respect within their communities and foster the exchange of ideas and democratization of knowledge.

In the open-source software movement, where ownership is often discouraged, attribution is important for many participants.[102] Removing a person's name from a project credit file without consent is generally considered unacceptable. For those involved, recognition serves as a guiding motivation behind their creative process, as the credit file acts as a reference for future work opportunities.[103] Engaged in a collaborative practice that involves modifying and distributing software,[104] open-source software creators are driven to ensure their contributions are distinctive enough to warrant recognition, all while building upon previous work.

The expansion of the open copyright production market—which includes initiatives such as open-source software, Creative Commons licensing, and open-access publishing—indicates that many producers and authors value personal satisfaction, reputation, and knowledge sharing over monetary rewards. In fact, Creative Commons, an organization dealing with open resource licenses, discovered that ninety-seven percent of authors and copyright holders chose some form of license requiring attribution of their authorship.[105] This finding suggests that numerous creators, while promoting widespread dissemination for non-commercial purposes, maintain a strong

---

[101] Luther et al., *supra* note 92, at 2823–25.

[102] Catherine L. Fisk, *Credit Where It's Due: The Law and Norms of Attribution*, 95 GEO. L.J. 49, 89 (2006).

[103] *Id.* at 89–90.

[104] *Id.* at 88–89.

[105] Lastowka, *supra* note 98, at 80.

interest in receiving acknowledgment and credit for their creative efforts.

Beyond recognition and integrity, attribution also has a direct impact on the economic value of creative works. Gaining recognition for one's work contributes to a positive reputation, which, in turn, can lead to an increase in the value of existing works.[106] Some cultural theorists argue that market demand arises from social expectations and cultural pressures, implying that recognition plays a crucial role in determining market exchanges and the pricing of creative works.[107]

In situations where information providers receive little to no market benefit other than widespread attention, proper attribution of the information to its source becomes crucial.[108] As authors may lack monetary incentives to produce work, proper recognition serves as the primary means to encourage content creation. The right of attribution is especially vital in the rapidly evolving digital era, characterized by a free flow of information and a growing preference among authors and researchers for open-market licensing over traditional copyright rights. Protection for reliable authorial attribution should be "one cornerstone of a well-balanced copyright edifice."[109] Ensuring proper attribution not only promotes trust and reliability within the information-sharing ecosystem but also incentivizes the continuous production of valuable content.

Attribution indeed plays a significant role in providing exposure and potentially financial gains for artists and authors whose work is reproduced. "For instance, an independent musician whose work is sampled or featured in a Billboard Hot 100 song (and credited accordingly) can gain instant name recognition and reap associated financial" benefits.[110] This aligns with the public interest in expanding knowledge, as it enables consumers and users to access the original work, enriching their understanding and appreciation of the content. This is a primary

---

[106] Fromer, *supra* note 96, at 1790 ("A strongly positive reputation can provide the creator with financial rewards, such as increased professional opportunities and a higher salary. In this sense, providing attribution to creators is nothing more than a traditional pecuniary incentive."); U.S. COPYRIGHT OFF., *supra* note 97, at 35.

[107] Lastowka, *supra* note 98, at 62.

[108] *Id.* at 68.

[109] *Id.* (citing Julie E. Cohen, *The Place of the User in Copyright Law*, 74 FORDHAM L. REV. 347, 371, 348 (2005)).

[110] John S. Ehrett, *Fair Use and an Attribution-Oriented Approach to Music Sampling*, 33 YALE J. REGUL. 655, 662 (2016).

reason why most open-market licensing platforms, such as Creative Commons, demand proper attribution for any secondary uses of resources.[111] By ensuring appropriate credit is given, these platforms promote a healthy ecosystem where creators are recognized and encouraged, while users benefit from an ever-growing pool of accessible knowledge.

3. Ethics of Authenticity

Authenticity is another crucial aspect of ethical creativity, promoting truthfulness in one's claim of creativity and narration of creative ideas. Individuals must strive to be genuine, presenting their thoughts, ideas, and creations without pretense or deception and reflecting the true origins and sources of inspiration for their work.[112] Plagiarism and disinformation are prime examples of violations of the ethics of authenticity, despite both involving the creation of new works of authorship.

a. Plagiarism

Plagiarism involves the misappropriation of someone else's creative work and presenting it as one's own without giving due credit. This dishonest act leads to undeserved benefits derived from the copied material while contributing no original value.[113] Plagiarism and copyright infringement both involve wrongful copying, but they are distinct concepts with different implications. While both practices are considered unethical, not every instance of appropriation will necessarily constitute a violation of copyright law.[114] Applicable to both copyrighted and non-copyrighted works, plagiarism primarily focuses on the moral and ethical ramifications of unauthorized copying.

Individuals who engage in plagiarism may not necessarily face legal consequences, but they often face social repercussions such as public disapproval. For example, chefs caught plagiarizing

---

[111] CREATIVE COMMONS, *Attribution 4.0 International*, https://creativecommons.org/licenses/by/4.0/legalcode.en#:~:text=Moral%20rights%2C%20such%20as%20the,limited%20extent%20necessary%20to%20allow [https://perma.cc/DLR5-R6M9] (last visited Nov. 10, 2024) ("Moral rights, such as the right of integrity, are not licensed under this Public License . . . .").

[112] CHARLES TAYLOR, THE ETHICS OF AUTHENTICITY 47 (1991) (arguing that "the importance of recognition has been modified and intensified by the understanding of identity emerging with the ideal of authenticity").

[113] Laurie Stearns, *Copy Wrong: Plagiarism, Process, Property, and the Law*, 80 CALIF. L. REV. 513, 516–17 (1992).

[114] *Id.* at 522 (pointing out plagiarism "as the generic equivalent of copying").

recipes—which typically cannot be protected under copyright law—have been called out by those affected, mocked by the wider online community, and experienced significant damage to their professional reputation.[115] In 1987, President Joseph Biden withdrew from the presidential race due to public scrutiny surrounding his use of phrases and mannerisms borrowed from other politicians, as well as the subsequent emergence of plagiarism accusations from his law school days.[116]

In academia, sensitivity to plagiarism is especially high, further emphasizing the importance of ethical conduct in such environments. Plagiarism not only constitutes an ethical breach but also jeopardizes the perceived meritocracy within academia, the reputation of an institution, and the value of its diplomas in the job market.[117] Students who engage in plagiarism may face serious consequences, including failing grades for assignments or courses, and a negative impact on their academic and employment prospects.[118] Academic dishonesty in research staff and professors can have an even more significant effect on a university's reputation, potentially leading to career-ending outcomes. For instance, in 2023, an art professor in China was dismissed after social media posts revealed he had plagiarized works from international artists.[119] Similarly, a professor in Singapore was

---

[115] Priya Krishna, *Who Owns a Recipe? A Plagiarism Claim Has Cookbook Authors Asking*, N.Y. TIMES, https://www.nytimes.com/2021/11/29/dining/recipe-theft-cookbook-plagiarism.html [https://perma.cc/5QC6-M4LQ] (June 22, 2023); *see also* James Hansen, *London Chef Elizabeth Haigh's Cookbook Withdrawn After Plagiarism Allegations*, EATER LONDON (Oct. 12, 2021, 8:57 AM), https://london.eater.com/22720370/makan-cookbook-plagiarism-elizabeth-haigh-sharon-wee-nonya-kitchen [https://perma.cc/WT4V-5HFB]; Alison Flood, *Former MasterChef contestant's book pulled amid plagiarism accusations*, GUARDIAN (Oct. 11, 2021, 11:30 AM), https://www.theguardian.com/books/2021/oct/11/former-masterchef-contestants-book-pulled-amid-plagiarism-accusations [https://perma.cc/9VHC-94VH].

[116] Neena Satija, *Echoes of Biden's 1987 Plagiarism Scandal Continue to Reverberate*, WASH. POST (June 5, 2019, 5:17 PM), https://www.washingtonpost.com/investigations/echoes-of-bidens-1987-plagiarism-scandal-continue-to-reverberate/2019/06/05/dbaf3716-7292-11e9-9eb4-0828f5389013_story.html [https://perma.cc/MPD6-HJ3Q].

[117] Sean Zwagerman, *The Scarlet* P*: Plagiarism, Panopticism, and the Rhetoric of Academic Integrity*, 59 COLL. COMPOSITION & COMMC'N 676, 677 (2008).

[118] *Id.* at 686.

[119] Yang Caini, *China Academy of Art Fires Professor over Plagiarism Allegations*, SIXTH TONE (Apr. 27, 2023), https://www.sixthtone.com/news/1012807 [https://perma.cc/7XT7-A7DX].

fired after being accused of plagiarizing a student's thesis.[120] These examples highlight the severe implications of plagiarism within the academic community and the importance of maintaining ethical standards in research and education.

In light of the heightened sensitivity to plagiarism, considerable measures are taken to promote awareness and education on the subject within academia. Academic publishing companies often provide prospective authors with editorial policies on ethical publishing, which include comprehensive definitions and examples of plagiarism, information on citation practices, and warnings about submissions being screened using plagiarism detection software.[121] Universities implement similar policies for their research staff and students.[122] While their effectiveness in reaching students is debatable, ongoing research in the field of education consistently aims to understand and improve student awareness of academic plagiarism.[123] However, these policies have limited impact on non-academic creatives and the broader public's understanding of plagiarism as an unethical practice that does not necessarily infringe on copyright. Consequently, it is crucial to continue raising awareness about the importance of ethical conduct and originality across creative fields and among the public.

---

[120] *NTU Confirms Termination of Employment for Tianjin-Born Associate Professor Implicated in Plagiarism Allegations*, GUTZY ASIA (Oct. 12, 2023), https://gutzy.asia/2023/10/12/ntu-confirms-termination-of-employment-for-tianjin-born-associate-professor-implicated-in-plagiarism-allegations/ [https://perma.cc/7VWD-JAJA].

[121] *See, e.g.*, *Editorial Policies: Plagiarism and Duplicate Publication*, NATURE, https://www.nature.com/nature-portfolio/editorial-policies/plagiarism [https://perma.cc/KGG6-HLLZ] (last visited Nov. 10, 2024); *Journal Information: Publishing Ethics*, CAMBRIDGE UNIV. PRESS, https://www.cambridge.org/core/journals/flow/information/journal-policies/publishing-ethics [https://perma.cc/5N5E-KGPY] (last visited Nov. 10, 2024).

[122] *Harvard University Plagiarism Policy*, HARVARD UNIV., https://usingsources.fas.harvard.edu/harvard-plagiarism-policy [https://perma.cc/MBJ2-V7NE] (last visited Nov. 10, 2024); *Academic Integrity: Faculty Statement on Academic Integrity*, COLUMBIA COLL., https://www.college.columbia.edu/academics/academicintegrity [https://perma.cc/MWS3-GUK4] (last visited Dec. 15, 2024); *University Policies: Plagiarism*, UNIV. OF OXFORD, https://www.ox.ac.uk/admissions/graduate/applying-to-oxford/university-policies/plagiarism [https://perma.cc/9JY8-MDTA] (last visited Dec. 15, 2024).

[123] *See, e.g.*, J.M. Gullifer & G.A. Tyson, *Who has Read the Policy on Plagiarism? Unpacking Students' Understanding of Plagiarism*, 39 STUD. IN HIGHER EDUC. 1202, 1203–04 (2014); Wan Zah Wan Ali, Habsah Ismail & Tan Tien Cheat, *Plagiarism: To What Extent it Is Understood?*, 59 PROCEDIA SOC. & BEHAV. SCIS. 604, 605 (2012).

Plagiarism on the Internet is particularly problematic. Social media has provided opportunities for individuals to gain attention or notoriety for personal gratification and economic benefits. Consequently, various online platforms, including the service formerly known as Twitter, have become inundated with viral posts and jokes taken from other users without credit.[124] Popular TikTok videos demonstrate how to exploit the platform's algorithms by actively copying the ideas and content of other users.[125] A new genre of podcasts has emerged, repackaging previously published work from investigative journalists and historians, often with little to no acknowledgement of the original creators.[126] In terms of mainstream Internet awareness, the most-followed individual creator on YouTube has shared the story of his struggle with a Russian creator who managed to amass forty million followers and become the most subscribed channel in Russia by plagiarizing video concepts, thumbnails, and content.[127] This widespread recognition of plagiarism highlights the importance of promoting ethical practices and discouraging content theft across various online platforms and creative fields.

b. Disinformation

Plagiarism and disinformation both undermine authenticity, but they do so in different ways. Plagiarism involves falsely presenting oneself as the creator of a piece of work, thereby disrupting the authenticity of one's authorship. On the other hand, disinformation aims to distort the authenticity of facts. For instance, fake news involves the deliberate spread of false, misleading, or inaccurate information, often with the intention of

---

[124] Jacob Shamsian, *Meet the man who has dedicated himself to righting the biggest wrong on the internet*, BUS. INSIDER (Aug. 31, 2017, 5:38 PM), https://www.businessinsider.com/twitter-joke-stealing-fat-jew-dory-common-white-girl-kale-salad-2017-8 [https://perma.cc/KKK6-JK43].

[125] Rebecca Jennings, *How the internet gets people to plagiarize each other*, VOX (May 24, 2022, 9:00 AM), https://www.vox.com/the-goods/23137820/plagiarism-growth-hacks-tiktok-instagram [https://perma.cc/AD7E-HB99].

[126] *Id.*

[127] Rituraj Halder, *Infamous Story of MrBeast Copycat Who Gained over 40 Million Subscribers Just Ripping off YouTube King's Videos*, ESSENTIALLY SPORTS (May 25, 2023, 8:00 AM), https://www.essentiallysports.com/esports-news-mrbeast-the-copycat-who-gained-over-million-subscribers-just-ripping-off-mrbeast/ [https://perma.cc/E6PQ-6CBG].

manipulating public opinion, promoting specific ideologies, or causing confusion and chaos.[128]

Disinformation indeed has far-reaching consequences for individuals, organizations, and society as a whole.[129] Although not a new social problem, our digital age reliance on online platforms for information has made us more vulnerable to its spread. Social media, particularly, facilitates the rapid and widespread dissemination of false news stories, targeting users and distorting their perceptions of reality.[130] Studies have shown that many social media users are exposed to at least one fake news article daily.[131] Such exposure can erode trust in institutions, the media, and even our own judgment. It can also foster polarization, divisiveness, and misinformation, ultimately undermining the democratic process and informed decision-making.

In 2016, a fabricated conspiracy theory emerged, claiming that a Washington, D.C. pizzeria called Comet Ping Pong was the center of a child trafficking ring run by high-ranking members of the Democratic Party. This fake news story spread rapidly on social media, leading to harassment of the pizzeria's staff and ultimately prompting a man to enter the restaurant with a gun, attempting to "self-investigate" the allegations. No one was injured, but the incident highlighted the potential for fake news to incite dangerous actions.[132] In 2018, a viral hoax called the "Momo Challenge" claimed that a sinister character named Momo was encouraging children to perform dangerous tasks and self-harm through online videos and social media platforms.[133] The challenge was widely reported by the media, leading to panic

---

[128] Daniela C. Manzi, *Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News*, 87 FORDAM L. REV. 2623, 2625, 2630, 2632 (2019).

[129] MARTHA MINOW, SAVING THE NEWS: WHY THE CONSTITUTION CALLS FOR GOVERNMENT ACTION TO PRESERVE FREEDOM OF SPEECH 24–25 (2021).

[130] Manzi, *supra* note 128, at 2628.

[131] *See, e.g.*, Amy Watson, *Frequency of seeing false or misleading information online among adults in the United States as of April 2023, by local area*, STATISTA (Apr. 17, 2024), https://www.statista.com/statistics/1462059/false-news-consumption-frequency-us-by-area/ [https://perma.cc/D36E-AZDU].

[132] Cecilia Kang & Adam Goldman, *In Washington Pizzeria Attack, Fake News Brought Real Guns*, N.Y. TIMES (Dec. 5, 2016), https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html [https://perma.cc/JFR5-VNHJ].

[133] Allyson Chiu, *The 'Momo Challenge': A Sinister Threat to Young People or an Urban Myth?*, WASH. POST (Sept. 5, 2018, 6:40 AM), https://www.washingtonpost.com/news/morning-mix/wp/2018/09/05/the-momo-challenge-a-sinister-threat-to-young-people-or-an-urban-myth/ [https://perma.cc/EHE5-AKQW].

among parents and educators.[134] However, investigations found no evidence to support its existence, and it was ultimately debunked as a fake news phenomenon.[135]

Due to the spread of fake news concerning the coronavirus and vaccines, an "infodemic" [136] became part of the COVID-19 pandemic, causing mental distress, breeding mistrust in health authorities, and undermining pandemic relief measures.[137] For instance, information circulated on social media that consuming alcohol could serve as a preventative measure or even a cure for COVID-19; a myth firmly debunked by the World Health Organization ("WHO"), emphasizing that alcohol not only fails to prevent or treat the virus but can also cause harm if consumed in excessive amounts.[138] Numerous fake news stories propagated false claims about COVID-19 vaccines, including allegations that they contain microchips for tracking individuals, alter human DNA, or induce severe side effects. Although COVID-19 vaccines are subject to rigorous testing and have been proven safe and effective in preventing severe illness and fatalities caused by the virus, unfounded claims fueled vaccine hesitancy and hindered vaccination efforts in certain regions.[139]

As the these examples illustrate, fake news presents a considerable threat to the authenticity of facts, as it distorts, manipulates, and fabricates information, resulting in far-reaching consequences for individuals, societies, and democratic institutions. Intentionally obscuring the boundary between fact and fiction, fake news frequently combines elements of truth with falsehoods.[140] Confusion among information consumers in discerning reality from fabrication can breed skepticism about the

---

[134] *Id.*

[135] *Id.*

[136] *See Infodemic*, WORLD HEALTH ORG., https://www.who.int/health-topics/infodemic#tab=tab_1 [https://perma.cc/AQ9Q-A2TJ] (last visited Nov. 10, 2024).

[137] *See* Michael A. Gisondi et al., *A Deadly Infodemic: Social Media and the Power of COVID-19 Misinformation*, 24 J. MED. INTERNET RSCH. 1 (2022).

[138] *Busting myths on alcohol and COVID-19*, WORLD HEALTH ORG. (Apr. 14, 2020), https://www.who.int/news-room/feature-stories/detail/global-news-busting-myths-on-alcohol-and-covid-19 [https://perma.cc/2Z54-3XCG].

[139] *Id.*

[140] Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCIENCE 1146, 1146 (2018) ("Here we investigate the differential diffusion of true, false, and mixed (partially true, partially false) news stories using a comprehensive data set of all of the fact-checked rumor cascades that spread on Twitter from its inception in 2006 to 2017. The data include ~126,000 rumor cascades spread by ~3 million people more than 4.5 million times.").

authenticity of information, ultimately eroding the credibility of legitimate news sources and the accuracy of factual data.

Fake news often takes advantage of cognitive biases like confirmation bias and availability heuristic to lend a veneer of credibility to false information. By targeting emotions, pre-existing beliefs, and sensationalist narratives, fake news fosters an illusory sense of authenticity, increasing the likelihood of people accepting and disseminating it.[141] This phenomenon contributes to the formation of echo chambers, where individuals are predominantly exposed to information that aligns with their pre-existing beliefs, irrespective of factual authenticity. Consequently, it becomes increasingly difficult for them to engage in rational discussions, assess the veracity of information, and reach a consensus based on objective evidence.[142]

C. Transforming Ethics of Creativity into Law

Serving as guiding principles that underpin a society's values and norms, ethics play a significant role in shaping and informing the content of laws.[143] Through legislative, judicial, or customary processes, they can be transformed into law and legal rules that govern what is considered right and wrong, fair and unfair, or just and unjust.[144] Legal institutions subsequently implement and enforce these legal rules.

Ethical principles arise from sources such as religious teachings, philosophical theories, cultural norms, and social conventions, and are identified and widely accepted through social activities. For example, the principle of not causing harm to others can be found in numerous religious and philosophical

---

[141] Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 388, 395–96 (2020) ("Some falsehoods can hurt or even ruin individual lives. For all these reasons, it is sensible to hope that social norms and even laws will chill them.").

[142] Haochen Sun, *Regulating Algorithmic Disinformation*, 46 COLUM. J.L. & ARTS 367, 377 (2023) ("Recommendation algorithms can also amplify disinformation to influence user beliefs by creating echo chambers and filter bubbles. Echo chambers refer to homogenous information environments resulting from users' own choices to follow like-minded individuals on social media, whereas filter bubbles are similar information environments created surreptitiously by a platform's automated efforts to understand individual user preferences and reflect them in its recommendations.").

[143] *See* LON L. FULLER, THE MORALITY OF LAW 4 (1964) (arguing that "a proper respect for the internal morality of law limits the kinds of substantive aims that may be achieved through legal rules"); RONALD DWORKIN, LAW'S EMPIRE 1 (1986) ("There is inevitably a moral dimension to an action at law, and so a standing risk of a distinct form of public injustice.").

[144] *See id.* at 6–8 (discussing how the morality of duty is transformed into legal rules).

traditions, serving as a foundational ethical guideline for many societies.[145]

Once ethical principles are established, they can be transformed into legal norms through legislative processes, judicial decisions, or customary practices.[146] Lawmakers and judges often draw upon them when drafting legislation or interpreting legal rules. For instance, the principle of not causing harm to others has been incorporated into legal norms in various forms, such as laws prohibiting assault, theft, or defamation.

Subsequently, legal rules are implemented and enforced through legal institutions, such as courts, the police, and regulatory agencies. These institutions ensure that individuals and organizations adhere to the legal norms derived from moral principles and impose sanctions or remedies when violations occur. For example, a person who causes harm to another by committing assault may be arrested, prosecuted, and sentenced to imprisonment or required to pay compensation to the victim.

The three principles of ethical creativity—originality, attribution, and authenticity—are fundamental to the fair use doctrine. In upholding the ethics of originality, it denies the legality of a substantial amount of copying where a copier fails to provide persuasive justifications for doing that.[147] The fair use doctrine supports the ethics of attribution by dismissing unauthorized uses of copyrighted works that fail to properly credit the original creators.[148] Additionally, the doctrine safeguards the ethics of authenticity by addressing bad-faith uses of copyrighted works, such as instances involving illegal

---

[145] *See id.* at 168–69.

[146] *See id.* at 6–8.

[147] *See, e.g.*, Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 550 (2023) ("Goldsmith's original photograph of Prince, and AWF's copying use of that photograph in an image licensed to a special edition magazine devoted to Prince, share substantially the same purpose, and the use is of a commercial nature. AWF has offered no other persuasive justification for its unauthorized use of the photograph.").

[148] *See, e.g.*, Williamson v. Pearson Educ., Inc., No. 00 CIV. 8240(AGS), 2001 WL 1262964, at *5 (S.D.N.Y. Oct. 19, 2001) ("Another relevant consideration within the first of the four fair use factors is the propriety of the defendant's conduct."); Rogers v. Koons, 960 F.2d 301, 309 (2d Cir. 1992) (tearing off of copyright notice); Weissmann v. Freeman, 868 F.2d 1313, 1324 (2d Cir. 1989) (removal of the original author's name and substitution of the copier's).

access[149] or false claims of use.[150] Hence, the fair use doctrine embodies all three principles in its legal framework.

Broadly, these three ethical principles have significantly influenced the development and enforcement of copyright law and other related laws, such as anti-fake news statutes. They help protect against improper appropriation and encourage responsible use of creative works. First, the ethics of originality has been adopted in legal norms to determine the existence of copyright infringement.[151] For example, to establish a case for infringement of the right of reproduction, the plaintiff must prove that the defendant had access to their work and that there is substantial similarity between both parties' works.[152] This "substantial similarity" standard deems the unauthorized use of copyrighted material as "improper or wrongful" and legally actionable.[153] In *Arnstein v. Porter*, Judge Jerome Frank captured the nature of such ethical wrongdoing as follows:

> The plaintiff's legally protected interest is not, as such, his reputation as a musician but his interest in the potential financial returns from his compositions which derive from the lay public's appreciation of his efforts. The question, therefore, is

---

[149] *See, e.g.*, Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 562 (1985) (ruling that the defendants' "use had not merely the incidental effect but the *intended purpose* of supplanting the copyright holder's commercially valuable right of first publication"); NXIVM Corp. v. Ross Inst., 364 F.3d 471, 478 (2d. Cir. 2004) (pointing out that the defendant "could have acquired the copyrighted manuscript legitimately," by "pa[ying] the requisite fee to enroll in NXIVM's seminars").

[150] *See, e.g.*, Atari Games Corp. v. Nintendo of Am. Inc., 975 F.2d 832 (Fed. Cir. 1992).

[151] An alternative explanation for the originality requirement is that it functions as an evidentiary rule, simplifying the process for courts to handle independent creation claims. This idea stems from worries about infringement accusations involving simple or fact-heavy works, where a high level of similarity is intrinsic to the subject matter. Mandating a minimum degree of creativity could offer a judicial instrument that allows courts to identify content that is improbable to have been independently generated. *See* WILLIAM F. PATRY, PATRY ON COPYRIGHT § 3:33 (2024).

[152] *See*, *e.g.*, Arnstein v. Porter, 154 F.2d 464, 468–69 (2d Cir. 1946) ("In some cases, the similarities between the plaintiff's and defendant's work are so extensive and striking as, without more, both to justify an inference of copying and to prove improper appropriation."); North Coast Indus. v. Jason Maxwell, Inc., 972 F.2d 1031, 1033–34 (9th Cir.1992) (ruling that "to establish that the plaintiff copied a preexisting work, a defendant must show that plaintiff had access to the prior work and that plaintiff's work is substantially similar to the prior work in both ideas and expression").

[153] Shyamkrishna Balganesh, *The Normativity of Copying in Copyright Law*, 62 DUKE L.J. 203, 206 (2012) ("The substantial-similarity analysis has courts focus entirely on the *significance* of the similarity between the plaintiff's and the defendant's works for their assessment of actionability.") (emphasis added).

> whether defendant took from plaintiff's works so much of what is pleasing to the ears of lay listeners, who comprise the audience for whom such popular music is composed, that defendant *wrongfully appropriated* something which belongs to the plaintiff.[154]

The act of creating a substantially similar copy of someone else's work is considered unethical because it can potentially harm the commercial value of the original work by diverting its consumers, such as "lay listeners," to the appropriator's work. Additionally, it may disrupt the expressive value of the original work by forcing the original author to be unwillingly associated with the appropriator or by making the appropriator's work appear as if it were the author's expression.[155] The ethics of originality, therefore, impose a responsibility on individuals not to create works that bear substantial similarity to those of others.[156] By adhering to these ethical principles, creators can contribute to a diverse and respectful environment that values originality of works. This not only protects the rights and interests of original creators but also fosters a culture of innovation and fair competition in the creative industries.

Second, copyright law acknowledges the ethical importance of proper attribution by safeguarding authors' honor and reputation through the recognition of their right of attribution.[157] The Berne Convention for the Protection of Literary and Artistic Works,[158] an international agreement governing copyright, protects the right of attribution as a significant aspect of an author's moral rights. Known as "the right to claim authorship," this right ensures that the author should be identified as the creator of their work whenever it is published, displayed, or otherwise used.[159] This is

---

[154] 154 F.2d at 473 (emphasis added).

[155] ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYING? 115–16 (2015).

[156] Balganesh, *supra* note 22, at 1682–83 ("Copyright thus imposes on individuals an obligation not to produce a substantially similar copy of the original work through the act of reproduction (that is, copying).").

[157] ROBERTA ROSENTHAL KWALL, THE SOUL OF CREATIVITY: FORGING A MORAL RIGHTS LAW FOR THE UNITED STATES 37–52 (2010).

[158] Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, 828 U.N.T.S. 221.

[159] *Id.* at 235 ("Independently of the author's economic rights, and even after the transfer of the said rights, the author shall have the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation.").

important because proper attribution helps to establish the author's reputation and can be crucial for their career and professional standing. By recognizing and protecting the right of attribution, copyright law promotes ethical behavior in the creative industries and encourages creators to respect one another's contributions. It also enables authors to benefit from the recognition and appreciation of their work, which can lead to further opportunities and advancements in their field.

In the United States, moral rights protection is not as comprehensive as under the Berne Convention, though the Visual Artists Rights Act ("VARA") of 1990[160] and the Lanham Act offers some safeguards for authors' rights of attribution. These legal measures help promote ethical behavior within the creative industries and foster respect for original works and their creators.

The VARA specifically protects the right of attribution for authors of works of visual art, such as paintings, sculptures, drawings, and photographs.[161] They have the right to claim authorship and prevent the use of their name in association with any distorted, mutilated, or modified versions of their work that would be prejudicial to their honor or reputation.[162] The Lanham Act,[163] which primarily deals with trademark law, indirectly offers some protection for the right of attribution. The Act prohibits the use of false or misleading representations that are likely to cause confusion, mistake, or deception as to the origin, sponsorship, or approval of goods or services.[164] In cases where an author's work is falsely attributed, the Lanham Act can help protect their reputation and commercial interests by preventing others from exploiting their name or creative output.[165]

The rise of disinformation and fake news has further complicated the ethics of authenticity. In response, several countries have enacted laws to regulate the creation and dissemination of fake news. These laws aim to maintain the integrity of information, protect citizens from misinformation, and uphold the ethical principles of authenticity and truthfulness. In 2019, the Parliament of Singapore adopted the Protection from Online Falsehoods and Manipulation Act

---

[160] Visual Artists Rights Act of 1990, 17 U.S.C. § 106A.
[161] *See id.*
[162] *See id.*
[163] Lanham Act, 15 U.S.C. §§ 1051–1129.
[164] 15 U.S.C. § 1125(a)(1)(A).
[165] *Id.*; *see also* Fromer, *supra* note 96, at 1796–97.

("POFMA"), which enables the government to take countermeasures against the creation and circulation of fake news.[166] The Act targets "false statements of fact"[167] and enables a minister to issue a Correction Direction and a Stop Communication Direction to the publisher of the falsehood.[168] Germany has introduced one of the broadest existing Internet regulatory regimes. The German Network Enforcement Law designates hate speech or fake news as illegal if it falls under a provision of the German criminal code. This law allows for the removal of such content and imposes fines on social media platforms that fail to remove illegal content within a specified time frame.[169]

In the United States, several legislative proposals have been put forward to tackle the issue of fake news and misinformation. While certain proposals concentrate on specific areas, such as political advertising or social media platforms, others focus on enhancing media literacy and promoting accurate information.[170] For example, the Algorithmic Accountability Act, introduced in 2019, mandates that large tech companies evaluate their algorithms and automated decision-making systems for potential

---

[166] David Tan & Jessica Teng Sijie, *Fake News, Free Speech and Finding Constitutional Congruence*, 32 SING. ACAD. L.J. 207, 208 (2020).

[167] *Singapore Fake News Laws: Guide to POFMA (Protection from Online Falsehoods and Manipulation Act)*, SING. LEGAL ADVICE (Apr. 25, 2022), https://singaporelegaladvice.com/law-articles/singapore-fake-news-protection-online-falsehoods-manipulation/ [https://perma.cc/6Q5J-FLH2].

[168] Under section 7(1) of POFMA, a person commits an offense if he communicates a statement which he knows or has reason to believe is false and the communication is likely to affect social and political stability under certain circumstances. Protection from Online Falsehoods and Manipulation Act 2019 § 7(1). Although section 7(4) excludes intermediaries from this liability, under section 20 a minister can issue a Part 4 direction if a false statement has been communicated and he believes such issuance to be in the public interest. *Id.* § 20. Part 4 directions include the Targeted Correction Direction, which requires an intermediary to communicate a correction notice to any end user who accesses the statement, a General Correction Direction, which involves sharing the notice with all users, and a Disabling Direction, which disables end-user access to the falsehood in Singapore. *Id.* §§ 21–23.

[169] The law further requires platforms to establish an effective complaints management infrastructure to delete or block such content within a specified time frame. Facebook has already received a €2 million fine for reporting the deletion of only 1,704 pieces of problematic content in the first half of 2018, a very low figure compared to the 260,000 and 215,000 deletions made by Twitter and YouTube, respectively. *See* Martin Gerecke, *Facebook fined EUR 2 m for infringement of Germany's Network Enforcement Act*, CMS LAW-NOW (Aug. 28, 2019), https://cms-lawnow.com/en/ealerts/2019/08/facebook-fined-eur-2-m-for-infringement-of-germany-s-network-enforcement-act [https://perma.cc/PXG7-KZ2P].

[170] Sun, *supra* note 142, at 385.

biases, discrimination, and other adverse consequences. This legislation is designed to hold technology companies responsible for the content and information disseminated by their platforms, including the propagation of fake news, thus promoting transparency and accuracy in the digital information landscape.[171] These examples demonstrate the efforts made by various countries to address the ethical issues surrounding fake news and disinformation. By implementing legal measures to regulate the creation and dissemination of false information, they aim to protect the authenticity of information and foster an environment of trust and credibility in the digital age.

D. Summary

Creativity is an ethical idea. In this Part, I have demonstrated that creative activities occur in a social setting, in which individuals borrow ideas and draw inspiration from others' works, while societal networks give credit to creatives' original contributions to the knowledge ecosystem and maintain authenticity of creative activities. The social context of creativity gives rise to three ethical principles, concerning originality, attribution, and authenticity. The interplay between these ethical principles and their corresponding laws seeks to ensure that creative activities foster a culture of innovation while maintaining the integrity of artistic and intellectual pursuits.

II. THE ETHICAL VACUUM OF AI CREATIVITY

If he were alive today, how might Picasso assess AI-generated art? Could it ever be comparable to that of the most gifted artists? Many are presently grappling with these questions, which measure AI creativity in terms of intelligence or artistic ability against our own.[172] However, as demonstrated in the preceding part, human creativity also has an ethical dimension that is highly valued by creatives such as Picasso.

Can AI systems genuinely "think" and "act" as ethically as humans do? If so, how can we decipher the ethical components of AI creativity? It has now become crucial for us to explore these

---

[171] *Id.*

[172] Bernard Marr, *The Intersection Of AI And Human Creativity: Can Machines Really Be Creative?*, FORBES (Mar. 27, 2023, 2:48 AM), https://www.forbes.com/sites/bernardmarr/2023/03/27/the-intersection-of-ai-and-human-creativity-can-machines-really-be-creative/?sh=40f6896a3dbc [https://perma.cc/A823-YGY6].

questions, scrutinizing the ethical implications of AI-generated art and creativity.

In this Part, I will explore how and why the generation of outputs by AI systems, through their collection and utilization of data, involves black-box processes. The resulting opacity has made it exceedingly challenging for us to assess the ethics of AI creativity. Hence, this *status quo* has led to an ethical vacuum in the domain of generative AI, causing a myriad of adverse effects on originality, attribution, and authenticity, three ethical values and principles deeply ingrained in human creativity.

A. The Opacity of AI Models and Datasets

1. Opacity of the Data Collection Process

Generative AI systems require access to massive amounts of data, and tech firms have become increasingly secretive about how they collect datasets. Factors such as escalating competition in the field, along with efforts to avoid potential legal or regulatory scrutiny, have incentivized AI companies to withhold crucial information about their methods of data collection, as well as the origins and contents of their datasets. This growing lack of transparency poses challenges in ensuring ethical and responsible AI development while maintaining competitiveness in the rapidly evolving AI landscape.

Our extremely limited knowledge about the data used to train prominent generative AI services is of particular concern in these services' earlier stages. For example, Stability AI, which faces ongoing lawsuits from copyright owners, is known to have utilized the LAION-5B dataset to train its Stable Diffusion system.[173] The LAION-5B dataset consists of 5.85 billion hyperlinks pairing images with text descriptions collected from the Internet by the German research organization Large-Scale Artificial Intelligence Open Network ("LAION").[174] OpenAI, the company responsible for breakthrough ChatGPT service launched in late 2022, initially seemed to live up to its name by offering relative transparency regarding the data used to train its GPT-3 model. An article by OpenAI contributors revealed that GPT-3 was trained using a

---

[173] *See* Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCIENCE 158, 159 (2023).

[174] Romain Beaumont, *LAION-5B: A New Era Of Open Large-Scale Multi-Modal Datasets*, LAION (Mar. 31, 2022), https://laion.ai/blog/laion-5b [https://perma.cc/3RKU-QM54].

combination of a specially filtered version of the publicly available CommonCrawl dataset derived from web scraping (60%), the Reddit-sourced WebText2 dataset (22%), the more obscure Books1 and Books2 datasets (16% combined), and Wikipedia (3%).[175]

Following ChatGPT's initial launch, OpenAI introduced GPT-4, an upgraded version of its service.[176] This time, in the face of growing competition, the company adopted a more secretive, "black box" approach.[177] Although OpenAI released a ninety-eight-page technical report about the new model, AI researchers criticized the published information for its lack of meaningful transparency.[178] The company cited "both the competitive landscape and the safety implications of large-scale models like GPT-4"[179] as means of justifying their approach.

Even when companies like Stability AI and OpenAI are willing to share the datasets they use, this action may fall short of providing meaningful transparency. For example, OpenAI's disclosure that Books1 and Books2 datasets were used in training GPT-3 offers limited insight.[180] Similarly, while we know that Stable Diffusion was trained using the LAION-5B dataset, LAION cautioned potential users that the collection of "5,85B CLIP-filtered image-text pairs" lacks data curation.[181] Consequently, they "cannot entirely exclude the possibility for harmful content being still present in safe mode."[182] This highlights the need for a more comprehensive approach to transparency in AI development,

---

[175] Tom B. Brown et al., Language Models are Few-Shot Learners 9 tbl. 2.2 (July 22, 2020) (arXiv preprint), arXiv:2005.14165v4. Concerned online bloggers and commentators have decried the lack of publicly available information concerning the Books1 and Books2 datasets. *See, e.g.*, *AI Training Datasets: The Books1+Books2 that Big AI eats for Breakfast*, VISIONS OF FREEDOM (Dec. 14, 2022), https://gregoreite.com/drilling-down-details-on-the-ai-training-datasets/ [https://perma.cc/9PRW-56N7].

[176] Lisa Lacy, *The Free Version of ChatGPT Just Got a Big Upgrade*, CNET (May 17, 2024, 3:09 PM), https://www.cnet.com/tech/services-and-software/the-free-version-of-chatgpt-just-got-a-big-upgrade/ [https://perma.cc/S6TT-CC8Y].

[177] *Unveiling the Flaws of Revolutionary GPT: Black-Box Nature and Hallucinations*, FLUID AI (June 25, 2024), https://www.fluid.ai/blog/unveiling-the-flaws-of-revolutionary-gpt-black-box-nature-and-hallucinations/ [https://perma.cc/YT7J-8KV5].

[178] Chloe Xiang, *OpenAI's GPT-4 Is Closed Source and Shrouded in Secrecy*, VICE (Mar. 16, 2023, 9:21 AM), https://www.vice.com/en/article/openais-gpt-4-is-closed-source-and-shrouded-in-secrecy/ [https://perma.cc/KP5B-5AB8].

[179] OpenAI, GPT-4 Technical Report 2 (Mar. 4, 2024) (arXiv preprint), arXiv:2303.08774v6.

[180] *See* Beaumont, *supra* note 174.

[181] *Id.*

[182] *Id.*

encompassing not only the disclosure of datasets but also the methods of data collection, curation, and filtering.

In pursuit of offering a diverse range of generative services, AI companies often rely on multiple datasets to train their models. OpenAI, for instance, has been secretive about GPT-4's training data while actively promoting the new functions ChatGPT can provide. Among these new features are prompts combining text and images that can now be used to generate responses.[183] OpenAI President Greg Brockman has explicitly mentioned that this new capability required the model to be trained using both text and images.[184] However, apart from vaguely stating that GPT-4 "was trained using publicly available data (such as Internet data) as well as data we've licensed,"[185] specific details about the additional data required for this development remain undisclosed.

The web scraping practices employed to gather data from the Internet present several significant concerns. The pervasive role of social media in today's society ensures that a vast amount of personal information is readily available online. As evidenced by the Cambridge Analytica scandal, insufficient platform protections can enable external actors to harvest and exploit this data.[186] Web scraping technology has greatly improved Internet data mining. Typically, this is an automated process that starts with acquiring access to raw, target, or pre-processed data. The data is then extracted or copied to facilitate knowledge discovery through a series of steps, including data cleaning and pre-processing, data transformation, and pattern evaluation.[187] The use of web-scraped data in AI training raises ethical questions of user privacy and consent, as well as the potential for biases to be

---

[183] OpenAI, *GPT-4*, OPENAI (Mar. 14, 2023), https://openai.com/index/gpt-4-research/ [https://perma.cc/9348-FRAX].

[184] Kyle Barr, *GPT-4 Is a Giant Black Box and Its Training Data Remains a Mystery*, GIZMODO (Mar. 16, 2023), https://gizmodo.com/chatbot-gpt4-open-ai-ai-bing-microsoft-1850229989 [https://perma.cc/82FU-R68A].

[185] *See* OpenAI, *supra* note 183.

[186] Natasha Lomas, *Social Media Giants Urged to Tackle Data-Scraping Privacy Risks*, TECHCRUNCH (Aug. 25, 2023, 10:20 AM), https://techcrunch.com/2023/08/24/data-scraping-privacy-risks-joint-statement/ [https://perma.cc/8WDN-5GRR].

[187] Artha Dermawan, *Text and Data Mining Exceptions in the Development of Generative AI Models: What the EU Member States Could Learn from the Japanese "Nonenjoyment" Purposes?*, 27 J. WORLD INTELL. PROP. 44, 47 (2023) ("[T]o increase the dependability of the data and its effectiveness, data cleaning and preprocessing will look for missing data and delete noisy, redundant, and low-quality data from the data collection.").

embedded in the AI systems. The legal implications surrounding copyright and data protection laws further complicate the matter.

Currently, there are no truly effective methods to prevent text and data mining tools from accessing Internet platforms. Websites can opt to implement robots.txt, a robot exclusion protocol that allows them to block specific web scraping tools.[188] However, companies can easily bypass these exclusions by changing their tool's name, refusing to disclose its name, or simply ignoring the voluntary protocol and extracting information regardless.[189] In the past, web scraping primarily served the purpose of indexing information, making it more accessible through search engines. However, with the rise of generative AI services, owners and creators of online content are now inadvertently contributing to the growth of wealthy tech companies rather than promoting the overall public benefit of the Internet.[190]

As a result of automated data collection methods and the lack of data transparency offered by AI companies, there is much we do not know about the contents of large-scale datasets used in training generative models. However, some external parties have started addressing this issue. For example, a coalition of machine learning and legal experts from various U.S. universities and organizations have developed a tool called the Data Provenance Explorer.[191] This tool enables users "to identify, track, and learn about the legal status of training datasets for generative AI."[192] The Data Provenance Explorer has revealed that many large language models, particularly open-source ones trained using datasets crowdsourced by aggregators like GitHub and Papers with Code, "have an extremely high proportion of missing data licenses . . . ranging from 72% to 83%."[193]

---

[188] Kali Hays & Alistair Barr, *AI is Killing the Bargain at the Heart of the Web. 'We're in a Different World'*, BUS. INSIDER (Aug. 30, 2023), https://www.businessinsider.com/ai-killing-web-grand-bargain-2023-8 [https://perma.cc/ELC3-AT7S].

[189] *Id.* (arguing "web crawlers can also simply ignore the blocking instructions and siphon the information from a site anyway").

[190] *Id.*

[191] Jon Gold, *Generative AI training data sets are now trackable—and often legally complicated*, COMPUTERWORLD (Oct. 26, 2023), https://www.computerworld.com/article/3709490/generative-ai-training-data-sets-are-now-trackable-and-often-legally-complicated.html [https://perma.cc/J6LT-352G].

[192] *Id.*

[193] *Id.*

Another complicating factor is the presence of copyright-protected material in the data. While some large datasets contain purely factual information that is not protectable under copyright law, "the majority of training datasets consist of copyrighted [contents]."[194] For AI services aiming to develop models capable of generating text, images, or recognizing faces, the corpus of works used for data training often includes those protected by copyright law.[195] This reality is evident in the rising number of lawsuits brought by copyright owners against text generators like ChatGPT[196] and image generators like Stable Diffusion.[197] AI companies rarely seek permission in advance, meaning copyright owners often only become aware that their works have been used when they recognize their works in generated outputs or investigative journalistic efforts reveal that their works were included within the datasets utilized by the companies.[198]

As the legal landscape surrounding AI-generated content continues to evolve, several uncertainties remain. One significant area of concern is the use of data with debatable copyright status. While traditional copyrighted works like books, paintings, and photographs are clearly protected, the status of online content such as user-generated posts is less clear. These short pieces of content may not reach the level of creativity required for copyright protection or may be predominantly fact-based.[199] Additionally, much of the user-generated content posted online incorporates or directly copies pre-existing copyrighted material, raising questions about fair use principles.[200] Should authors succeed in

---

[194] Nicola Lucchi, *ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems*, 15 EUR. J. RISK REGUL. 602, 613 (2023).

[195] *Id.*

[196] *See, e.g.*, Alter & Harris, *supra* note 12.

[197] Emilia David, *Getty Lawsuit Against Stability AI To Go To Trial in the UK*, THE VERGE (Dec. 4, 2023, 5:46 PM), https://www.theverge.com/2023/12/4/23988403/getty-lawsuit-stability-ai-copyright-infringement [https://perma.cc/Z4YY-KTUN].

[198] *See, e.g.*, Alex Reisner, *What I Found in a Database Meta Uses To Train Generative AI*, ATLANTIC (Sept. 25, 2023), https://www.theatlantic.com/technology/archive/2023/09/books3-ai-training-meta-copyright-infringement-lawsuit/675411/ [https://perma.cc/J6UL-U2W5]; *You Just Found Out Your Book Was Used to Train AI. Now What?*, THE AUTHOR'S GUILD (Sept. 27, 2023), https://authorsguild.org/news/you-just-found-out-your-book-was-used-to-train-ai-now-what/ [https://perma.cc/7J76-3YAZ].

[199] Consuelo Reinberg, *Are Tweets Copyright-Protected?*, WIPO MAG. (July 14, 2009), https://www.wipo.int/wipo_magazine/en/2009/04/article_0005.html [https://perma.cc/D8ME-8TZL].

[200] Daniel Gervais, *The Tangled Web of UGC: Making Copyright Sense of User-Generated Content*, 11 VAND. J. ENT. & TECH. L. 841, 861, 865 (2009).

their ongoing lawsuits, it remains to be seen whether ordinary Internet users with sufficiently original content could also make similar claims.

2. Opacity of the Training Process

One of the primary challenges in understanding the role of datasets in generative AI development is determining how specific data pieces or datasets influence the parameters[201] of generative AI models during the training process. As noted above, prior to GPT-4, OpenAI was relatively transparent about the training processes used in ChatGPT's development. GPT-3 is a Large Language Model ("LLM"), a complex neural network or deep learning model that processes vast amounts of textual data. It learns to infer relationships between words by predicting the next word, or "token," in a sequence.[202] To create the widely used ChatGPT service, OpenAI fine-tuned the GPT-3 LLM using the Reinforcement Learning from Human Feedback ("RLHF") machine learning model, enabling it to generate outputs in response to user prompts.[203]

The RLHF training process starts with the creation of a Supervised Fine Tuning ("SFT") model. In this stage, human "labelers" craft appropriate responses to actual user prompts sampled from an OpenAI dataset, which are then used for supervised learning by pairing inputs and outputs.[204] Next, labelers rank a set of responses generated by the SFT model from best to worst, and this data is utilized to train a reward model.[205] In the final stage, reinforcement learning takes place using the

---

[201] In AI, parameters refer to the variables that a model acquires during its training phase. These internal variables enable the model to generate predictions or make decisions. Within a neural network, parameters encompass the neurons' weights and biases. *What Are Parameters in AI?*, TED AI, https://www.ai-event.ted.com/glossary/parameters [https://perma.cc/ZG95-TJFV] (last visited Nov. 10, 2024).

[202] Alberto Romero, *A Complete Overview of GPT-3—The Largest Neural Network Ever Created*, TOWARDS DATA SCI. (May 24, 2021), https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd [https://perma.cc/V6LZ-YT8R]; Molly Ruby, *How ChatGPT Works: The Model Behind the Bot*, TOWARDS DATA SCI. (Jan. 31, 2023), https://towardsdatascience.com/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286 [https://perma.cc/47Z2-QYXR].

[203] *Introducing ChatGPT*, OPENAI (Nov. 30, 2022), https://openai.com/blog/chatgpt [https://perma.cc/MP4C-QQSG]; Ruby, *supra* note 202.

[204] Ruby, *supra* note 202.

[205] *Id.*

Proximal Policy Optimization ("PPO") algorithm.[206] The SFT model's established policy guides the generation of new responses to prompts. The reward model then calculates a reward for each output, and the PPO algorithm uses this reward to update the policy. This process is iteratively repeated to refine the model's performance.[207]

Since introducing GPT-4 into the ChatGPT service, OpenAI has revealed little about the updated training process.[208] GPT-4, as a multimodal model capable of processing both text and images, differs from the previous LLM-based GPT-3 and the initial ChatGPT model, necessitating changes in its training approach.[209] However, OpenAI has not publicized any useful information about these changes. OpenAI does identify slight differences in the RLHF post-training processes used for generating responses to prompts, stating that GPT-4 incorporates an additional safety reward signal during RLHF training to reduce harmful outputs, as defined by their usage guidelines, by training the model to refuse requests for such content.[210] Despite this, even the most comprehensive publications about ChatGPT's earlier iteration leave us with limited understanding of how individual pieces of data are utilized during an AI system's training process.

In generative AI systems designed for specific purposes and relying on much smaller datasets, it might seem feasible to quantify the contribution of data to a model's training. For example, The Next Rembrandt project aimed to create a piece of art in the style of the renowned Dutch painter using a dataset comprising his complete works,[211] a relatively limited dataset. Generating a convincing Rembrandt-like painting required training that employed deep learning algorithms and facial recognition techniques. This training enabled the final model to analyze more than 400 faces and over 6,000 facial landmarks

---

[206] *Id.*

[207] *Id.*

[208] OPENAI, GPT-4 TECHNICAL REPORT 2 (2023), https://cdn.openai.com/papers/gpt-4.pdf [https://perma.cc/7QYL-Y5X7] (in addition to declaring their intent to increase dataset secrecy in response to concerns relating to competition, in the GPT-4 Technical Report, OpenAI declared an intent to provide "no further details about the architecture (including model size), hardware, training compute, . . . training method, or similar").

[209] *Id.* at 1.

[210] *Id.* at 12–13.

[211] *Our Work: The Next Rembrandt*, VML, https://www.wundermanthompson.com/work/next-rembrandt [https://perma.cc/Q2BF-55XG] (last visited Nov. 10, 2024).

painted by the artist, thereby learning how to replicate his unique style.[212]

To 3D print the painting, the AI model needed to replicate Rembrandt's brushstrokes. This involved recognizing similarities between the new work and his existing pieces, analyzing the brushstrokes he employed, and aligning over 165,000 painting fragments to the generated artwork.[213] Some observers might be inclined to argue that the final output closely resembles a particular Rembrandt painting. However, given that the system has learned to consider numerous minute details from Rembrandt's entire body of work, the task of quantifying the actual contribution of a specific painting to the model's training is extremely challenging.

Understanding how individual pieces of data are utilized during training of an AI system becomes significantly more challenging when considering the breadth of functions generative models like ChatGPT intend to serve. While its capacity in each field vastly differs, ChatGPT has been trained to respond to prompts concerning healthcare information, education, industry, marketing, finance, translation, mathematics, society, and matters relating to AI and machine learning.[214] To generate such diverse outputs, LLMs like GPT-3 and GPT-4 are trained using billions of parameters.[215] Throughout this process, each system iteratively updates the weight assigned to each parameter to improve functionality, meaning weights do not reflect any single source and are not the result of any single round of training.[216] When this extensive training is combined with the reinforcement learning processes OpenAI employs to train the ChatGPT prompt response function,[217] any attempt to understand the contribution of individual pieces of data to the development of this service becomes futile.

---

[212] Dutch Digital Design, *The Next Rembrandt: bringing the Old Master back to life*, MEDIUM (Jan. 24, 2018), https://medium.com/@DutchDigital/the-next-rembrandt-bringing-the-old-master-back-to-life-35dfb1653597 [https://perma.cc/9URC-XVFL].

[213] *Id.*

[214] Konstantinos I. Roumeliotis & Nikolaos D. Tselikas, *ChatGPT and Open-AI Models: A Preliminary Review*, 15 FUTURE INTERNET 1, 9–17 (2023).

[215] Funmi Looi Somoye, *How is Chat GPT Trained*, PC GUIDE (Apr. 8, 2024), https://www.pcguide.com/apps/chat-gpt-trained/ [https://perma.cc/EF46-XD8N].

[216] Steve Hook, *What is an LLM?—Large Language Model Explained,* PC GUIDE (Dec. 5, 2023), https://www.pcguide.com/ai/llm-large-language-model-ai/ [https://perma.cc/J7VM-P5E4].

[217] *See supra* notes 198–202.

This lack of clarity around the exact contribution of individual data pieces or datasets to the final AI model's behavior complicates efforts to determine the legal and ethical implications of using specific datasets in AI training, as well as assessing the potential biases and limitations introduced during the training process. As AI technology continues to advance, it is crucial for developers and researchers to prioritize transparency and explore methods that allow for better understanding and tracking of how data influences AI model outcomes.

3. Opacity of the Generation Process

While the previous section explored the challenges of understanding how data contributes to the training of AI services offered to the public, this section deals with the actual generation of creative works. The black box nature of generative AI systems limits our understanding of how user prompts produce particular outputs in the form of text, an image, or a piece of music,[218] and the efficacy of measures designed to prevent the generation of problematic content. Such opacity compels ongoing research and development in AI safety and ethics.

In LLMs like GPT-3 and GPT-4, when users input a prompt, the system produces a response based on extensive training data.[219] Tracing the precise steps leading to a specific output is difficult, as the model's decision-making hinges on complex interactions among billions of parameters. To reduce the likelihood of generating harmful or undesirable content, AI developers adopt safety measures such as content filters, reinforcement learning using human feedback, and guidelines for human reviewers.[220] Nonetheless, problematic outputs may still arise due to the inherent intricacies of generative models and their dependence on large datasets, which can harbor biases and contentious information.

Among all black box problems, particularly prominent is how generative AI systems use copyrighted content in datasets to produce outputs based on user prompts. It is hard to pinpoint the extent to which these systems, as opposed to human creators, adhere to copyright laws or their underlying principles. The

---

[218] *See* Luke Tredinnick & Claire Laybats, *Black-Box Creativity and Generative Artificial Intelligence*, 40 BUS. INFO. REV. 98, 100 (2023).

[219] Sag, *supra* note 36, at 299.

[220] *See, e.g.*, *Our Approach to AI Safety*, OPENAI (Apr. 5, 2023), https://openai.com/blog/our-approach-to-ai-safety [https://perma.cc/B3EG-F3WC].

disconnect between the training data and the fully trained model has been discussed earlier[221] has implications for analyzing potential copyright infringements in AI-generated outputs. For example, when instructing an image generator to produce images of a coffee cup, our current technological understanding allows us to view the generated images as "a combination of vectors that encode a latent idea of a coffee cup as represented in the training data," rather than a replication of "any individual coffee cup."[222] Since it is currently impossible to quantify the number of sources that influence the generation of a potentially copyright-infringing output, or the contribution and copyright status of individual sources, we must examine the copyright policies of the AI companies developing these systems.

The opacity surrounding the artificial generation of works creates some of the most difficult implications for copyright law. The phenomenon of dataset "memorization" is currently being examined by U.S. courts, as evidenced by the lawsuit Getty Images filed against Stability AI in February 2023. One of the arguments presented by Getty Images was that the Stable Diffusion system occasionally "produces images that are highly similar to and derivative of the Getty Images proprietary content."[223]

To support this claim, Getty Images showcased a picture of two soccer players generated by Stable Diffusion that closely resembled an image from their library. In both images, players wear the same team kits, strike similar poses, and are captured from comparable angles.[224] Most notably, the generated photograph, along with numerous others under dispute, includes a slightly distorted version of the Getty Images watermark.[225] While the similarities between the images are not definitive enough to guarantee a finding of substantial similarity, the generated output did contain several distortions in the players' bodies, faces, and the ball's position.[226] Regardless of whether the

---

[221] *See supra* Section II.B.

[222] Sag, *supra* note 36, at 124.

[223] Complaint at 17–18, Getty Images (US), Inc. v. Stability AI Inc., No. 1:23-cv00135-UNA, (D. Del. Feb. 3, 2023). *See* Sag, *supra* note 36, at 115–16.

[224] Sag, *supra* note 36, at 116.

[225] James Vincent, *Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its content*, THE VERGE (Jan. 17, 2023, 5:30 AM), https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit [https://perma.cc/2RJH-8UTH].

[226] *See* Sag, *supra* note 36, at 311–12 ("The original photo is compelling because of the specific angle of the shot and the way it captures Henderson's attempts to tackle the ball from Eriksen and the way Eriksen uses his body to block the tackle. It is also

images are ultimately deemed substantially similar, the black box issues discussed earlier make it challenging to determine the exact process and reasons behind AI-generated images.

In contrast to their efforts to counter inaccurate information, AI companies have been quiet regarding their policies to prevent copyright infringement. As outlined above, generative AI services are often reticent about revealing the origins and contents of their datasets.[227] In their efforts to promote the safety of their services, companies such as OpenAI have been relatively transparent about their content moderation efforts to ensure user safety. While not publicly disclosing its internal content moderation practices, OpenAI provides a Content Moderation endpoint for developers working with their models.[228] This helps developers create applications that adhere to OpenAI's usage policies, enabling the detection and filtering of harmful inputs and outputs, such as hate speech, threats, harassment, self-harm references, sexual content, and violent content.[229] The company's usage policies and publications relating to the Content Moderation endpoint do not, however, refer to the moderation of copyright-infringing outputs.[230]

Microsoft has pledged to offer legal defense for users of its Copilot system who face copyright infringement lawsuits, highlighting the company's confidence in "incorporated filters and other technologies . . . designed to reduce the likelihood that Copilots return infringing content."[231] Other AI companies seem to exhibit indifference towards preventing copyright infringement, placing the responsibility on end users instead. OpenAI and Stability AI, for example, have usage policies that prohibit users from employing their services for intellectual property

compelling because in silhouette the two players form a windmill—making the photo artistic as well as communicative. In the Stable Diffusion photo, the unique perspective is lost, no one is in control of the ball, there is no tackle, no windmill silhouette, just two bizarre, disfigured football golems haunting the field.").

[227] *See supra* Section II.A.

[228] Todor Markov et al., *New and improved content moderation tooling*, OPENAI (Aug. 10, 2022), https://openai.com/blog/new-and-improved-content-moderation-tooling.

[229] *Id.*; *Moderation*, OPENAI, https://platform.openai.com/docs/guides/moderation/overview [https://perma.cc/877J-WHYK] (last visited Nov. 10, 2024).

[230] *See Moderation*, *supra* note 229; *Usage Policies*, OPENAI (Jan. 10, 2024), https://openai.com/policies/usage-policies.

[231] Brad Smith, *Microsoft Announces New Copilot Copyright Commitment For Customers*, MICROSOFT (Sept. 7, 2023), https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/.

infringement, yet disclose little about their own efforts to prevent copyright-infringing content generation.[232] Furthermore, some companies argue that existing copyright laws are sufficient to address the challenges posed by generative AI. In a statement to the U.S. Senate, Stability AI expressed their belief that "existing legal frameworks effectively govern AI outputs."[233]

B. Ethical Hazards in AI Black Boxes

In this section, I will reveal that the black box nature of AI creativity has brought about negative consequences for creatives and society at large. This opacity has resulted in an ethical vacuum concerning AI-generated works, given that it has severely deviated from the three fundamental principles of human creativity as discussed in Part I.

1. Disrupting Originality

The ethical principle of originality, as I have shown, is intended to ensure that creative activities do not lead to copyright infringing outcomes.[234] Current AI systems, however, have disrupted the ethics of originality by generating outputs that are identical or substantially similar to copyrighted works.

One may excuse this disruption of the ethics of originality on the basis that the unauthorized use of copyrighted content for training AI models constitutes legitimate text and data mining[235] or fair use,[236] which is deemed both ethical and legal in many jurisdictions. Indeed, a text and data mining exemption has been

---

[232] *Terms of Use*, OPENAI https://openai.com/policies/row-terms-of-use/ [https://perma.cc/8C6A-GKAT] (Oct. 23, 2024), ("You may not use our Services . . . in a way that infringes, misappropriates or violates anyone's rights."); *Acceptable Use Policy*, STABILITY AI, https://stability.ai/use-policy [https://perma.cc/YB5Q-9P7K] (Mar. 1, 2024) ("You agree you will not use, or allow others to use, Stability Technology: To violate the law or others' rights (including . . . intellectual property . . . rights . . .).").

[233] Ben Brooks, *Statement to the U.S. Senate AI Insight Forum on Transparency, Explainability, and Copyright*, STABILITY AI (Nov. 29, 2023), https://stability.ai/news/copyright-us-senate-open-ai-transparency [https://perma.cc/269D-X74D].

[234] *See supra* Section I.B.

[235] *See, e.g.*, Eleonora Rosati, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity*, 27 ASIA PAC. L. REV. 198, 199–200 (2019); Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC'Y U.S.A. 291, 343–44 (2019); Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893, 938 (2019).

[236] *See, e.g.*, Lemley & Casey, *supra* note 34, at 748.

adopted in many jurisdictions.[237] For example, E.U. copyright law provides a copyright infringement exemption "for reproductions and extractions made by research organizations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter."[238] It is this exemption that may have allowed LAION to compile the LAION-5B dataset used in the training of Stable Diffusion, and which may provide a defense for parent company, Stability AI, in its ongoing lawsuits.[239] However, the exemption only applies when the mining is carried out for *research* purposes. Any text and data mining for a *commercial* purpose, pursuant to E.U. copyright law, requires that the party have *legitimate* access to the content concerned and that the copyright owner has not expressly opted out from text and data mining.[240] Most generative AI systems are commercial platforms and should therefore, under E.U. copyright law, acquire permission from copyright owners before using their content for text and data mining.

Under U.S. copyright law, there may be a fair use defense available for AI companies relying on text and data mining following the decision of the court in the Google Library case,[241] where Google's digitization of published books was held to be permissible.[242] In reaching this decision, the court considered the fact that the project aimed to improve discoverability of existing works, rather than directly compete with them by making verbatim and substantially similar copies of their copyrighted works. In contrast, through text and data mining, generative AI

---

[237] *See* Sean M. Flynn et al., *Legal Reform to Enhance Global Text and Data Mining Research*, 378 SCIENCE 951, 952 (2022).

[238] Directive 2019/790, of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, PE/51/2019/REV/1, art. 3 (EU) [hereinafter Directive (EU)].

[239] *See* Samuelson, *supra* note 173 ("LAION's creation of this dataset was very likely lawful because the European Union (EU) adopted an exemption allowing nonprofit research organizations to make copies of in copyright works for text and data mining (TDM) purposes. . . . This exemption cannot be overridden by contract.").

[240] Directive (EU), *supra* note 238, at art. 4(1) ("Member States shall provide for an exception or limitation . . . for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining. . . . The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.").

[241] Authors Guild v. Google, Inc., 804 F.3d 202 (2d. Cir. 2015).

[242] Lucchi, *supra* note 194, at 12–13.

services "have the potential to empower users to easily produce content that may directly compete with the original ingested material."[243]

Some commentators, however, have made the case that text and data mining is transformative. For instance, it has been argued that following the initial copying, the processing of mined information to produce usable datasets transforms the information into factual data lacking too much of the original expression to be considered a real copy.[244] It is technological factors such as this that give rise to the most difficult questions and create challenges reconciling generative AI with copyright law at the data collection stage. Should AI company practices be out of the scope of copyright law, or would that enable unprecedented theft of copyright infringing content?

But generative AI systems' potential fair use of copyrighted works by no means makes it legal for them to generate outputs that are identical or substantially similar to any of such works. For instance, AI image generators are able to "memorize" images contained within the datasets they are trained on and have often been found to produce outputs which duplicate the original works, or are near identical copies.[245] Commentators have concluded that the chance that the output of generative AI systems would infringe on copyrighted content is high.[246] Many of them believe that text-to-image AI generators have a "[v]isual [p]lagiarism [p]roblem."[247]

Copying by image generators sometimes occurs only in part but the implications are, nevertheless, significant. One study of Stable Diffusion suggested that the model copied from its training data roughly 1.88% of the time.[248] With AI estimated to have

---

[243] *Id.* at 13.

[244] Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. DAVIS L. REV. 893, 954 (2019).

[245] Kyle Barr, *Researchers Prove AI Art Generators Can Simply Copy Existing Images*, GIZMODO (Feb. 1, 2023), https://gizmodo.com/ai-art-generators-ai-copyright-stable-diffusion-1850060656 [https://perma.cc/ER83-HU2F].

[246] Sag, *supra* note 36, at 330 (concluding that "copyrightable characters, provoking copyright infringement [through AI systems] is easy").

[247] Gary Marcus & Reid Southen, *Generative AI Has a Visual Plagiarism Problem*, IEEE SPECTRUM (Jan. 6, 2024), https://spectrum.ieee.org/midjourney-copyright [https://perma.cc/8H3U-BVLJ].

[248] Kyle Wiggers, *Image-generating AI Can Copy and Paste from Training Data, Raising IP Concerns*, TECHCRUNCH (Dec. 13, 2022, 4:30 AM), https://techcrunch.com/2022/12/13/image-generating-ai-can-copy-and-paste-from-training-data-raising-ip-concerns/ [https://perma.cc/2A37-KPQW].

generated over 150 billion images in a single year,[249] even such a small percentile represents an unprecedented number of copyrighted infringing images.

2. Silencing Attribution

The opacity surrounding generative AI systems has a "silencing effect" in terms of proper attribution, and a major question concerning AI training is what is owed to copyright owners whose works are utilized. The technical features of the training process make it difficult to claim that the process itself infringes copyright, as when an AI model learns from data, this process does not include "copying in any legally cognizable sense," as "it is not copying any given text" but simply updating "probabilities to reflect statistical patterns from the training data."[250] The black box of AI training creates further difficulties when grappling with this issue, as there is so little we currently know about how data informs the training process. While a lack of meaningful company transparency efforts is a contributing factor,[251] ultimately it is the complexities of machine and deep learning processes that prevent us from quantifying the exact contribution of data to the functionality of a fully trained generative AI model.[252]

Without efforts to reconcile these issues, the opacity of generative AI systems could effectively avoid giving credit and paying royalties to copyright owners. Copyright licensing is at the core of the copyright law, as it ensures society benefits from access to a work when an author cannot effectively market it or enables those who wish to build upon an existing work to do so.[253] The goal of copyright royalty contracts is to guarantee the economic incentive for creation filters back to the creator of a work and to determine how the market surplus is divided among all parties involved.[254] Although we cannot currently comprehend the exact

---

[249] *AI has generated 150 years worth of images in less than 12 months, study shows*, DESIGNBOOM (Aug. 21, 2023), https://www.designboom.com/technology/ai-has-generated-150-years-worth-of-photographs-in-less-than-12-months-study-shows-08-21-2023/ [https://perma.cc/8LYX-CZJW].

[250] *See* Matthew Sag, *Fairness and Fair Use in Generative AI*, FORDHAM L. REV. 1887, 1907–08 (2024).

[251] *See, e.g.*, *supra* note 208.

[252] *See supra* Section II.A.

[253] Yafit Lev-Aretz, *The Subtle Incentive Theory of Copyright Licensing*, 80 BROOK. L. REV. 1357, 1358 (2015).

[254] Richard Watt, *Licensing and Royalty Contracts for Copyright*, 3 REV. ECON. RSCH. ON COPYRIGHT ISSUES 1, 1 (2006).

contributions made by works of human authorship to the development of generative AI models, we can identify the features of the services that make them so marketable and ask whether they could have been achieved without reliance on pre-existing copyrighted material.

The first, and most obvious, feature of generative AI systems that makes them so marketable is the sheer scale of their capacity for generation. In their efforts to attract customers, AI companies will themselves promote this feature of their systems.[255] Customers also frequently discover and share new ways to capitalize on the potential of various generative AI services.[256] However, ensuring a high volume of potential outputs necessitates a high volume of valuable inputs. Without contributions from copyright owners, the generative function of AI models would be diminished.

A second feature which has been hugely popular with consumers is the ability of AI systems to generate works in the style of specific human authors and artists. When publicly available generative AI services first emerged early in 2023, a great deal of popular interest and excitement revolved around this new imitative ability;[257] however, a system cannot be trained to fulfill such a function unless the work of authors and artists is included in its training dataset.

As human creativity is increasingly social in nature, copyright law includes protections, most notably through fair use and fair dealing exceptions, which ensure opportunities to draw upon

---

[255] *See, e.g.*, OpenAI, *supra* note 183 ("We've created GPT-4, the latest milestone in OpenAI's effort in scaling up deep learning. GPT-4 is a large multimodal model (accepting image and text inputs, emitting text outputs) that, while less capable than humans in many real-world scenarios, exhibits human-level performance on various professional and academic benchmarks. For example, it passes a simulated bar exam with a score around the top 10% of test takers; in contrast, GPT-3.5's score was around the bottom 10%.").

[256] *See, e.g.*, Anna Cooban, *ChatGPT can pick stocks better than your fund manager*, CNN (May 5, 2023, 8:04 AM), https://edition.cnn.com/2023/05/05/investing/chatgpt-outperforms-investment-funds/index.html [https://perma.cc/SU2D-6TFE]; Jeffrey M. Perkel, *Six Tips for Better Coding with ChatGPT*, NATURE (June 5, 2023), https://www.nature.com/articles/d41586-023-01833-0 [https://perma.cc/8TPF-PS5H].

[257] Joe Coscarelli, *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*, N.Y. TIMES (Apr. 24, 2023), https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html [https://perma.cc/5L5F-TNCV]; Jaelani Turner Williams, *An AI Eminem Rap About Cats Was Taken Down by Major Record Label*, OKAYPLAYER (Apr. 7, 2023), https://www.okayplayer.com/music/universal-copyright-eminem-ai.html [https://perma.cc/AR7H-5Z6Z].

existing works in the creative process.[258] Humans creating in good faith can often be observed acknowledging the contributions made by existing works, justifying these protections by shaping their creative process around compliance with the substance of copyright law, such as through active efforts to ensure works are original or lack substantial similarities with those that came before.[259] Those with less substantive knowledge of copyright law can still be observed reciprocating, though more indirectly, through their adherence to the ethical norms which underpin human creativity, such as avoiding plagiarism and accurately attributing authorship of ideas, concepts, and styles.[260] While the opacity around AI company datasets certainly colors this conclusion, it does not presently appear that the developers of generative AI models are providing similar acknowledgement or reciprocation.

Due to the current lack of AI company compensation efforts, copyright owners and authors are engaged in ongoing lawsuits to compel remuneration.[261] Regarding works generated in the style of specific authors and artists, some AI companies have stressed that copyright makes "experimentation with style" permissible.[262] This may be true for human creators, whose creativity exhibits efforts to engage with both legal and ethical norms. For instance, when musicians or their legal advisors notice potential similarities between the songs they are writing and existing ones, the artists will often endeavor to substantially distinguish their own. Moreover, despite the fact that copyright law does not protect styles, when an original artist believes their style is evident in a new song, the secondary artists will sometimes choose to offer royalties and acknowledge the influence through songwriting credits.[263] Therefore, AI companies should not feel confident in

---

[258] *See supra* Section I.B.

[259] *See supra* Section II.A

[260] *See supra* Section II.B

[261] *See, e.g.*, Alter & Harris, *supra* note 12; David, *supra* note 197.

[262] Brooks, *supra* note 233.

[263] *See* Edward Lee, *Fair Use Avoidance in Music Cases*, 59 B.C. L. REV. 1873, 1902–03, 1910 (2018) ("It is not uncommon for song writers accused of copyright infringement by other song writers, especially ones who are established musicians, to agree to share—sometimes begrudgingly—songwriting credit and royalties."); Jem Aswad, *Olivia Rodrigo Gives Taylor Swift Songwriting Credit on Second 'Sour' Song, 'Deja Vu'*, VARIETY (July 9, 2021, 7:54 AM), https://variety.com/2021/music/news/olivia-rodrigo-taylor-swift-songwriting-credit-deja-vu-1235015769/ [https://perma.cc/UN9A-RWJ3] ("Olivia Rodrigo and her main collaborator, songwriter-producer Daniel Nigro, have given Taylor Swift, Jack Antonoff and St.

their entitlement to permissions granted for human creativity when they have yet to make clear their own efforts to acknowledge the contribution of pre-existing works to AI creativity.

3. Harming Authenticity

Another problem with the opacity of generative AI systems arises from their potential to produce content that lacks authenticity. The rise of fake news and deepfakes clearly demonstrates the ability of such systems to produce legally and ethically problematic content. For instance, despite the progress it offers over the GPT-3 LLM, the GPT-4-powered ChatGPT model still " 'hallucinates' facts and makes reasoning errors." [264] Moreover, many users of the service have reported that outputs often include false or made-up sources,[265] with one pre-GPT-4 study of ChatGPT responses to a small number of physical geography and geography education prompts finding that every response included false or inaccurate citations.[266]

If generative AI services continue to be perceived as an increasingly important knowledge resource, they risk becoming substantial purveyors of misinformation. Public health commentators have already expressed concern that public perception of AI-generated texts as "equally or more credible than human-written texts" could "magnify the already existing problem

---

Vincent a songwriting credit on "Deja Vu"—the second song from Rodrigo's blockbuster debut album "Sour" to receive such a non-collaborative credit . . . . The influence of "Cruel Summer" on "Deja Vu" is less tangible, however—amounting basically to yelling on the bridge, which is more of an arrangement touch than a songwriting one—even though Rodrigo has acknowledged the influence in interviews.").

[264] *See* OPENAI, *supra* note 183.

[265] Chris Glorioso, *Fake News? ChatGPT Has a Knack for Making Up Phony Anonymous Sources*, NBC N.Y. (Feb. 23, 2023), https://www.nbcnewyork.com/investigations/fake-news-chatgpt-has-a-knack-for-making-up-phony-anonymous-sources/4120307/ [https://perma.cc/5G2V-WXUH].

[266] Terence Day, *A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT*, 75 PRO. GEOGRAPHER 1024, 1025–26 (2023) ("Although the reference and citations are in English, in highly regarded journals, and look legitimate on first examination, they are all fake and hard to detect on initial examination. The citations are in a consistent style and the page numbers fall within the volume and issue of the real journal, but in all cases the relevant journal pages are occupied by papers unrelated to the listed reference. Many of the authors listed by ChatGPT are highly regarded in their field, but they did not write the papers they have been credited with.").

of misinformation . . . and . . . threaten public health globally."[267] There is further concern that the systems could be abused by bad actors seeking to propagate misinformation by enabling convincing fake news stories and deepfakes to be produced easily at scale,[268] or facilitating the dissemination of political deepfake videos to undermine election campaigns.[269] Though this Article focuses primarily on copyright, the artificial generation of misinformation can still be contrasted with the examples of social creativity displayed by human authors in Section II; those producing new works, at least in good faith, will strive to establish veracity and accurately attribute authorship in order to emphasize the value of their contribution to the state of the art.[270]

Due to the black box problem, it is difficult to understand why generative AI systems produce inaccurate information, so when searching for adherence to legal and ethical norms in the artificial creation of work we must examine the efforts of programmers during the development stage. As the marketability of generative AI systems relies on consumer belief in the accuracy of the information they provide, companies like OpenAI do take observable steps to "be as transparent as possible that ChatGPT may not always be accurate" and publicize the measures they are taking to "further reduce the likelihood of hallucinations and to educate the public on the current limitations of these AI tools."[271] Such measures include "leveraging user feedback on ChatGPT outputs . . . flagged as incorrect" on the GPT-3 model during the

---

[267] Dan W. Meyrowitsch, Andreas K. Jensen, Jane B. Sørensen & Tibor V. Varga, *AI Chatbots and (Mis)information in Public Health: Impact on Vulnerable Communities*, 11 FRONTIERS PUB. HEALTH 1, 2 (2023).

[268] *See, e.g.*, Tiffany Hsu & Stuart A. Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots*, N.Y. TIMES (Feb. 8, 2023), https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html [https://perma.cc/T85K-AG5C]; Pranshu Verma, *The Rise of AI Fake News is Creating a 'Misinformation Superspreader'*, WASH. POST (Dec. 17, 2023), https://www.washingtonpost.com/technology/2023/12/17/ai-fake-news-misinformation/ [https://perma.cc/XWY4-FC7Q].

[269] *See, e.g.*, Adam Edelman, *States Are Lagging in Tackling Political Deepfakes, Leaving Potential Threats Unchecked Heading into 2024*, NBC NEWS (Dec. 16, 2023), https://www.nbcnews.com/politics/artificial-intelligence-deepfakes-2024-election-states-rcna129525 [https://perma.cc/WA52-8HKY] ("They have become significantly more common online in recent months—an increase that has prompted some experts to warn that the 2024 race could be the first 'deepfake election' because voters could see political disinformation videos online and not be able to determine what's real and what's not.").

[270] *See supra* Section II.B.

[271] OpenAI, *supra* note 183.

training of GPT-4[272] and developing external benchmarks like TruthfulQA to monitor the progress being made in efforts to ensure accuracy.[273] However, OpenAI admits that these efforts have only led to minimal improvements regarding hallucinations [274] and, due to the difficulties we face in understanding how any piece of data contributes during training to the creation of the final model,[275] we remain in the dark about the weight ChatGPT affords to accuracy as a parameter.

Another uncertainty with authenticity again lies in AI systems' generation of works without properly identifying true authorship. The opacity of AI creates difficulties when attempting to assign legal responsibility for the generation of AI outputs bearing substantial similarity to copyrighted works. Within the field, there are a variety of perspectives on where practical responsibility lies. As noted above, some companies have shared little about their efforts to prevent the generation of such works but have implemented policies emphasizing that users should not use the systems for the purpose of infringing intellectual property.[276] For its DALL-E 3 image generation service, OpenAI has gone further by teaching the system to decline requests for art in the style of particular artists but, contrary to the provisions of copyright law, the safeguard only applies to artists who are still alive.[277] Microsoft has expressed extreme confidence in its "broad range of guardrails such as classifiers, metaprompts, content filtering, and operational monitoring and abuse detection, including that which potentially infringes third-party content," and promised to fund the legal defense of any user sued for infringing copyright despite using these safeguards.[278] However, study of these safeguards and their effectiveness is necessary.

It remains difficult to conclude where practical responsibility lies and, therefore, how it should inform the development of copyright law in the AI era. In the creative processes of humans,

---

[272] *Id.*

[273] *See* Parinay Chaturverdi, *GPT4 Explained in 5 Minutes!*, MEDIUM (Mar. 17, 2023), https://parinayachaturvedi.medium.com/gpt4-explained-in-10-minutes-4b955fc2ae96 ("TruthfulQA . . . tests the model's ability to separate fact from an adversarially-selected set of incorrect statements. These questions are paired with factually incorrect answers that are statistically appealing.").

[274] *Id.*

[275] *See* Sag, *supra* note 36, at 343.

[276] *See* OpenAI, *supra* note 232.

[277] *DALL-E 3*, OPENAI, https://openai.com/dall-e-3 [https://perma.cc/VCR8-QSSA] (last visited Nov. 10, 2024).

[278] *See* Smith, *supra* note 231.

we can observe engagement with legal and ethical norms in acknowledgement, either direct or indirect, of the contributions of what has come before.[279] Are infringing AI outputs a consequence of misuses which contravene AI platform policies or escape internal safeguards, or are platforms being disingenuous or, at the very least, reckless concerning the capacity of their systems to be used for the purpose of infringing copyright? Only greater transparency regarding training processes and copyright policies can provide a credible answer to these questions. The existence of policies and safeguards is promising, but far more information is necessary to understand whether these safeguards sufficiently engage with legal and ethical norms and reciprocate the social nature of human creativity.

III. FOSTERING ETHICS OF AI CREATIVITY

How can we solve the ethical vacuum problems arising from the opacity of generative AI systems? Proposals to enhance respect for existing works include improved dataset recordkeeping and use of smaller datasets, which induce AI systems to learn abstractions rather than memorize.[280] However, these measures only provide minor adjustments to the current functioning of generative AI systems. They do not alter their black box nature in any significant way, nor do they necessarily contain the generation of outputs too closely resembling copyrighted works. They also do not guarantee proper attribution or filtering out of fake news and other malicious content.

In this Part, I explore the rationale for applying the ethics of human creativity to AI creativity as a means of addressing these issues. To better foster ethics of originality, regulators should legally require AI companies to implement filtering technologies that minimize occurrences of copyright infringement. Furthermore, AI companies should be encouraged to uphold the ethics of attribution and authenticity. Consequently, regulators should mandate the adoption of watermarking technologies that can differentiate between human-generated outputs and those created by AI systems, thereby enhancing the authenticity of the AI systems' outputs. By aligning the ethical principles of human and AI creativity, we can work towards establishing a more

---

[279] *See supra* Sections II.A–B.

[280] *See* Sag, *supra* note 36, at 339.

transparent landscape for responsible, trustworthy AI-generated content.

A. Why Ethics for AI Creativity

The ethical vacuum stemming from the opacity of AI systems necessitate the extension of ethics of human creativity to govern AI creativity. In this section, I suggest two additional reasons for this: first, AI systems rely on human works to enable their creativity; and second, the dissemination of works generated by such systems takes place within human society.

Recognizing the interdependence between human and AI creativity, it is essential to ensure that legal and ethical frameworks are in place to support the responsible development and use of generative AI technologies. As this Article has shown, the ethics of creativity require human authors to comply with legal norms in the creative process. Artificial creativity undeniably relies on the human creativity that precedes it. Even generative AI companies, such as OpenAI, acknowledge that it would be impossible to create tools like ChatGPT without access to copyrighted material.[281] There is limited potential for growth in AI systems without ongoing human creativity. Current AI systems often struggle with emotion, context, hyperbole, and prose,[282] relying on pre-existing data and patterns to produce text, "which limits their ability to think outside the box and offer fresh perspectives."[283]

Even as advancements diminish the limitations of technology, the training sets involved require continued creation and utilization of human works. Artificial systems possess, at best, moderate levels of incremental creativity and can only generate new and effective variations of existing ideas, solutions, systems,

---

[281] Dan Milmo, *'Impossible' to create AI tools like ChatGPT without copyrighted Material, OpenAI says*, GUARDIAN (Jan. 8, 2024, 8:40 AM), https://www.theguardian.com/technology/2024/jan/08/ai-tools-chatgpt-copyrighted-material-openai [https://perma.cc/UZH5-HDQF] ("[I]t would be impossible to create tools like its groundbreaking chatbot ChatGPT without access to copyrighted material.").

[282] Sam Johnson, *Why AI Can't Replace a Client-focused Team of Dedicated Human Writers*, SPICEWORKS (Apr. 3, 2023), https://www.spiceworks.com/tech/artificial-intelligence/guest-article/why-ai-cant-replace-human-writers/ [https://perma.cc/VA4L-U3LP].

[283] *The Future of Writing: Are AI Tools Replacing Human Writers?*, AICONTENTFY (Nov. 6, 2023), https://aicontentfy.com/en/blog/future-of-writing-are-ai-tools-replacing-human-writers [https://perma.cc/3JHV-KPAJ].

and artifacts.[284] AI systems trained solely on works generated by other AI systems create an "autophageous loop," [285] or self-consuming process, akin to an animal not just chasing its tail but eating it. If this process continues generation after generation, the artifacts will be amplified, and synthetic data will start to drift away from reality. AI system exposure to human creativity is crucial to prevent the degradation of generated content and maintain a connection with real-world ideas and concepts. This reinforces the importance of AI systems' adhering to ethical principles that protect human creativity. Through encouraging responsible use of copyrighted materials, regulators would promote a mutually beneficial coexistence of human and AI creativity.

The generative AI field also relies on larger cultural aspects of human creativity. One of the most novel and marketable aspects of generative models is their ability to replicate the style of human creatives, which is not possible without the shared cultural understandings that original authors create. Human parody practices, for example, ironize and criticize specific works or prevailing aesthetic practices, traditions, and styles, conducting a "double-coded dialogue between the present and the cultural past."[286] The art of parody lies in the tension between a known original and its parodic twin, relying on a recognizable allusion to the original through distorted imitation. Similarly, artificially-generated works that imitate particular styles, even if for non-parodic but fair purposes, will lack social value in the absence of links to specific human-generated reference points. These must be respected through adherence to legal and ethical frameworks protecting intellectual property and original human creative works.

Moreover, AI-generated works are predominantly disseminated to and utilized by humans, making it likely to give rise to legal and ethical disputes. This necessitates the application

---

[284] David H. Cropley, Kelsey E. Medeiros & Adam Damadzic, *The Intersection of Human and Artificial Creativity*, *in* CREATIVE PROVOCATIONS: SPECULATIONS ON THE FUTURE OF CREATIVITY, TECHNOLOGY & LEARNING, CREATIVITY THEORY AND ACTION IN EDUCATION 19, 32 (Danah Henriksen & Punya Mishra eds., 2022).

[285] Maggie Harrison Dupré, *When AI Is Trained on AI-Generated Data, Strange Things Start to Happen*, FUTURISM (Aug. 2, 2023, 10:09 AM), https://futurism.com/ai-trained-ai-generated-data-interview [https://perma.cc/W7BF-9P4V].

[286] Ansgar Nünning, *The Creative Role of Parody in Transforming Literature and Culture: An Outline of a Functionalist Approach to Postmodern Parody*, 3 EUR. J. ENG. STUD. 123, 128–30 (1999).

of ethical principles that govern human creativity to AI systems. As of November 2023, it has been estimated that 64% of global workers have claimed AI-generated work as their own.[287] Furthermore, predictions suggest that up to 90% of online content could be produced by AI by 2025.[288] Considering the disregard that AI companies have shown towards copyrighted content, the magnitude of potential online infringement is immeasurable. The growing potential for illegal and unethical content generated by AI models has also made it imperative to address the extent of responsibility that AI companies should bear. Given these concerns, it is crucial to establish legal and ethical frameworks that extend to AI creativity, ensuring responsible development and use of generative AI technologies. By doing so, the challenges posed by the opacity of AI systems can be effectively addressed, protecting both human creators and the broader society from the negative impacts of AI-generated content.

Technologically, it is feasible for AI companies to make greater efforts to authentically protect the creative process of humans and adhere to legal and ethical norms. There are strong arguments for requiring generative AI companies to develop and implement new technologies that can make them achieve those goals properly. For example, it is now possible for them to adopt such technologies to monitor and eliminate copyright-infringing outputs of their systems.[289] By integrating these technological measures into their AI platforms, companies can fulfill their ethical responsibilities for originality, attribution, and authenticity and thereby foster a culture of trust. This not only helps safeguard copyrighted material but also encourages responsible practices within the AI industry.

From a historical perspective, advancements in digital technologies have prompted tech companies to follow ethics of human creativity through modernization of copyright law. Initially designed for the bricks-and-mortar world, copyright was made applicable to the peculiarities of electronic distribution in cyberspace through the safe harbor mechanism, which shields an

---

[287] Vala Afshar, *64% of workers have passed off generative AI work as their own*, ZDNET (Nov. 22, 2023, 9:51 AM), https://www.zdnet.com/article/64-of-workers-have-passed-off-generative-ai-work-as-their-own/ [https://perma.cc/4ZW2-XEEY].

[288] Bruna Ferreira, *Generative AI: To Infinity and Beyond*, QUIDGEST (Aug. 2023), https://quidgest.com/en/blog-en/generative-ai-by-2025/ [https://perma.cc/BXW4-Q7VY].

[289] *Id.*

internet service provider from copyright infringement liabilities caused by its users.[290] But that mechanism also requires internet service providers to remove copyrighted works if they know or should have known the unauthorized making available of such works on their platforms.[291] It is crucial for lawmakers and the AI industry to adapt to rapidly evolving technologies and address the potential consequences of widespread copyright infringement and other illegal and unethical content. This may involve updating existing laws, implementing more robust technologies, and promoting responsible practices within the AI industry to ensure the protection of intellectual property and original content.

While the case for increased legal and ethical responsibility has been established, it is crucial to outline why this should be considered at the stage of the generation of works by AI systems. As mentioned earlier, the expansive services offered by leading AI companies rely on vast datasets.[292] Introducing strict copyright duties at the data collection stage could potentially hinder advancement of the technology. This might "impede innovation from start-ups and entrants who don't have the resources to obtain licenses,"[293] or it may lower the quality of datasets, leading to subpar systems or even shutting down projects entirely.[294] Furthermore, asking AI companies to retroactively alter their datasets after discovering their contribution to the generation of infringing outputs is often challenging, as it typically requires a complete reset of the model, forfeiting the significant resources invested in its training.[295]

AI companies have asserted that their collection and utilization of copyrighted content for training their models constitute fair use. However, the fair use doctrine does not grant permission to generate new works that infringe on another's

---

[290] Mike Scott, *Safe Harbors Under the Digital Millennium Copyright Act*, 9 N.Y.U. J. LEGIS. & PUB. POL'Y 99, 116 (2005).

[291] *Id.* at 120.

[292] Ferreira, *supra* note 288.

[293] Wes Davis, *AI Companies Have All Kinds of Arguments Against Paying for Copyrighted Content*, THE VERGE (Nov. 4, 2023, 6:17 PM), https://www.theverge.com/2023/11/4/23946353/generative-ai-copyright-training-data-openai-microsoft-google-meta-stabilityai [https://perma.cc/CW26-RH8K].

[294] Jane Hillman, *Data Quality and AI Safety: 4 ways bad data affects AI and how to avoid it,* PROLIFIC (Oct. 24, 2023), https://www.prolific.com/resources/data-quality-and-ai-safety [https://perma.cc/F56L-E3G4].

[295] *Challenges in removing data from AI Models*, DIGWATCH (Sept. 4, 2023), https://dig.watch/updates/challenges-in-removing-data-from-ai-models [https://perma.cc/7DA6-4U4X].

copyright, for example, by making a verbatim copy without transformative use.[296] Targeting AI's companies' legal and ethical responsibility at the generation stage only allows authors to take legal actions when they identify AI-generated works that have infringed their copyrights. Therefore, the mechanism does not entitle authors to prevent AI companies from using their works for data training processes without a finding of copyright infringement. By and large, AI companies can still maintain their freedom to use works for model training purposes within the bounds of fair use.

Given those concerns, focusing on legal and ethical responsibility at the generation stage can be a more practical approach. As the following two sections will show, implementing filtering and watermarking technologies to monitor and remove copyright-infringing outputs and harmful content generated by AI can help strike a balance between fostering innovation and protecting intellectual property. By addressing potential copyright issues at this stage, the AI industry can continue to advance while respecting legal and ethical frameworks, ultimately benefiting creators, consumers, and the technology sector as a whole.

B. Using Filters to Safeguards Originality

1. Protecting Originality Through the Filtering Responsibility

Copyright filtering technology has immense potential for application in generative AI systems, promoting the ethics of originality. By integrating this technology into their generative AI systems, companies can ensure that the creative outputs generated by their AI do not infringe upon existing copyrighted material. This combination of technologies could lead to more original, innovative, and legally compliant content, benefiting both the creators and users of generative AI services.

As leaders in AI technology, companies that create and provide generative services are well-positioned to implement copyright filtering systems. While copyright filtering responsibility at the training stage is possible, it comes with limitations. For example, RLHF[297] has been suggested as a means

[296] Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 532–33 (2023).

[297] *See supra* note 203.

of introducing copyright awareness into AI models.[298] However, its challenges include acquiring good quality human feedback for the systems to learn from, providing proper oversight of the learning process, and potential issues with reward misgeneralization or hacking.[299] While flaws in content filtering technology should not form the basis for an argument against a specific legal copyright responsibility at the training stage,[300] there is significant weight in the argument that excessive regulation at system development stages may hinder AI's full potential and interfere with creative destruction, whereby long-standing norms and practices must be dismantled for innovation to thrive.[301] Given that the technology already exists and appears to have the potential for application in this specific context, a post-development copyright filter on AI-generated outputs is the most logical option for the time being.

The ethics of AI creativity should address the complete reproduction of works as a consequence of dataset memorization. This category of generated work is both the most copyright-incompatible and the most easily detectable form of output. Memorization, in technical terms, occurs when a "string 's' is extractable with 'k' tokens of context from a model 'f' if there exists a (length-k) string 'p', such that the concatenation [p || s] is contained in the training data for 'f', and 'f' produces 's' when prompted with 'p' using greedy decoding."[302] Any legal definition

---

[298] *See* Sag, *supra* note 36, at 339 ("[A] similar approach to interactive reinforcement learning could help reduce the probability of copyright infringement. Moreover, this strategy should also extend to closely related concerns raised in relation to trademark, right of publicity, and privacy.").

[299] Stephen Casper & Xande Davies, Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback 9 (Sept. 11, 2023) (arXiv preprint), arXiv:2307.15217v2 ("[R]eward models are trained to reflect human approval instead of human benefit which can result in actions that would be approved of by humans while nevertheless being undesirable. . . . This type of problem is known as 'reward hacking', and has been observed in AI systems, including those trained with RLHF.").

[300] *See* Sag, *supra* note 36, at 340 ("It is unclear whether a universal filter that compared model output to the training data is feasible.").

[301] Stan Karanasios, Olga Kokshagina & Pauline C. Reinecke, *Calls to regulate AI are growing louder. But how exactly do you regulate a technology like this?*, THE CONVERSATION (Apr. 4, 2023, 8:48 PM), https://theconversation.com/calls-to-regulate-ai-are-growing-louder-but-how-exactly-do-you-regulate-a-technology-like-this-203050 [https://perma.cc/W498-R523].

[302] Nicholas Carlini, Quantifying Memorization Across Neural Language Models 1, 3 (Mar. 6, 2023) (arXiv preprint), arXiv:2202.07646v3; *Greedy Algorithm*, NAT'L INST. OF STANDARDS & TECH. (Feb. 2, 2005), https://xlinux.nist.gov/dads//HTML/greedyalgo.html [https://perma.cc/MKV8-G482] ("An algorithm that always takes the

must be capable of covering instances where outputs duplicate training data simply because it is the simplest way for the system to respond to a prompt. The current copyright law understanding of "reproduction" is likely to be sufficient for this purpose. Under present law, "a copyrighted work would be infringed by reproducing it in whole or in any substantial part, and by duplicating it exactly or by imitation or simulation."[303] By incorporating this understanding of reproduction into AI-generated content filtering mechanisms, AI companies can effectively address instances of dataset memorization and better adhere to copyright obligations, protecting original works and their creators.

As long as generative AI companies have a system in place to filter out such reproductions, there should be a case for compliance with the most basic part of the filtering responsibility. As filtering is unlikely to have a perfect success rate, the implementation of a system could be sufficient, or additional requirements could be put in place to mitigate this imperfection. Taking inspiration from the Digital Millennium Copyright Act ("DMCA"), some form of requirement could be established to ensure that, where a user has repeatedly taken advantage of the imperfect filtering capability to generate and make publicly available reproductions of copyright-infringing works, a system is in place that "provides for the termination in appropriate circumstances of subscribers and account holders of the service provider's system or network who are repeat infringers."[304]

More complex is the issue of reproduction of copyrighted material in part, through inclusion of extracts in a generated work. While there is reason to believe that a filtering system can detect such reproductions, determining whether they should be blocked is a more intricate challenge. The EU Directive on Copyright and Related Rights in the Digital Single Market requires that platforms make their best efforts to ensure the unavailability of copyright-infringing works and prevent their future uploads.[305] This mandate effectively calls for the implementation of copyright

---

best immediate, or local, solution while finding an answer. Greedy algorithms find the overall, or globally, *optimal solution* for some *optimization problems*, but may find less-than-optimal solutions for some instances of other problems.").

[303] 17 U.S.C. § 106.

[304] *Id.* § 512(i)(1)(A).

[305] Directive (EU), *supra* note 238, at art. 17(4).

filtering technology.[306] In Germany, it was decided that user-generated content involving minor uses, such as short extracts of audio-visual works, soundtracks, texts, and visual arts files, would remain online until the end of the complaint and redress mechanism.[307] Introducing a similar provision in the generative AI context would not be a decision that minor reproductions are fair use, but would empower companies to refrain from immediately blocking them.

However, this approach would necessitate the establishment of a complaint and redress system that facilitates a more qualitative evaluation of generated content. Several potential formats exist for such a system. For example, works that fall under a predefined reproduction threshold could be freely generated for users. If copyright owners deem these works to be infringing, they could report them to the respective generative AI companies, who would then determine if they concur and whether a strike should be issued to the responsible user under their repeat infringer policy. Alternatively, the system could proactively block partial reproductions exceeding a certain threshold, while still allowing users of the generative AI service to contest the copyright-infringing status of the work and opt to remove the pre-emptive block. Inspired by the DMCA's counter-notification system, this may entail requirements like a "statement under penalty of perjury that the subscriber has a good faith belief that the material was removed or disabled as a result of a mistake."[308]

2. Implementation of Copyright Filtering Responsibility

While it is indeed critically important to impose a copyright filtering responsibility, is it technologically feasible for AI companies to adopt and implement measures to meet this responsibility? Generative AI companies have demonstrated their capacity to implement copyright filtering technology. OpenAI's ChatGPT, for example, refuses to output entire copyrighted works when asked directly. Its AI image generator DALL-E refrains from producing images in the style of living artists and "allows

---

[306] Matthias Leistner, *The Implementation of Art. 17 DSM Directive in Germany—A Primer with Some Comparative Remarks*, 71 GRUR INT'L 909, 915 (2022).

[307] In Germany, it was decided that for "limited user-generated contents covering so-called minor uses," those which "are limited to 15 seconds of audiovisual works, 15 seconds of sound tracks, 160 characters of a text and up to 125 kbyte of visual arts files," would remain "online until the end of the complaint and redress mechanism." *Id.*

[308] 17 U.S.C. § 512(g).

artists to opt out of having their work used to train the bot."[309] However, it is not clear whether these copyright precautions involve the type of filtering discussed earlier or if they are more general filters introduced during the training stage of the generative models. Microsoft has provided more direct indications of output filtering technology in place, outlining that their AI system, Copilot, uses "a broad range of guardrails such as classifiers, metaprompts, content filtering, and operational monitoring and abuse detection, including that which potentially infringes third-party content."[310]

Copyright filtering, in the context of content hosting, emerged as a natural technological extension of the § 512(c) DMCA notice-takedown system.[311] Under this system, platforms are expected to act swiftly to remove content upon receiving information from copyright owners about its infringing nature.[312] In the context of generative AI, certain limitations typically associated with copyright filtering technology may not be as concerning. For example, content editing tactics used by internet users to avoid detection and upload existing copyrighted works are less likely to be effective, because implementing a filter at the generation stage means users would not have access to the material to apply such changes.[313] However, alternative bypass techniques may emerge. One example is the Do Anything Now ("DAN") prompt, which some internet users have employed to circumvent ChatGPT restrictions.[314] This approach involves asking ChatGPT for two responses: one as it would normally provide, labeled as "ChatGPT," "Classic," or something similar, and a second response in "Developer Mode" or "Boss" mode, which has fewer restrictions.[315] To counteract such strategies, companies must continuously update their generation services and filters, ensuring

---

[309] Isaiah Portiz, *OpenAI Faces Existential Threat in New York Times Copyright Suit*, BLOOMBERG L. (Dec. 30, 2023), https://news.bloomberglaw.com/ip-law/openai-faces-existential-threat-in-new-york-times-copyright-suit [https://perma.cc/D6H8-4R7X].

[310] *See* Smith, *supra* note 231.

[311] 17 U.S.C. § 512(c)(3)(A)(i)–(iii).

[312] *Id.*

[313] *See YouTube's Content ID Easily Fooled*, VOX INDIE (Jan. 5, 2019), https://voxindie.org/youtubes-content-id-easily-fooled/ [https://perma.cc/VM7Q-NNVW].

[314] Jon Martindale, *What is a DAN Prompt for ChatGPT?*, DIGIT. TRENDS (July 19, 2023), https://www.digitaltrends.com/computing/what-is-dan-prompt-chatgpt/ [https://perma.cc/YXU9-SPNG].

[315] *Id.*

that bypass methods like the DAN prompt do not remain effective for long.[316]

Advancements in AI have been leveraged to improve content fingerprinting algorithms. For instance, YouTube's engineers have integrated Google Brain's deep learning system into Content ID, allowing it to more quickly ingest new imagery and find matches and thereby ensuring the system remains both discerning and speedy.[317] However, there is a fundamental difference between creating technology that can match fingerprints effectively and developing systems with the potential capacity of today's generative AI systems. Generative AI can provide users with plausible analyses of complex questions based on just a few prompts, such as defining the literary style of Salman Rushdie or explaining the facts and significance of *Marbury v. Madison*. Moreover, generative AI's facility with style transfer allows it to translate the same content into various styles, like a hip-hop rendition or a fifth-grade reading level.[318]

The implementation of fingerprinting technology has proven to be effective in detecting copyrighted material, whether in whole or in part, within generated outputs. However, there are both theoretical and practical challenges when it comes to distinguishing fair uses from infringing uses of copyrighted material and deciding which extracts to block.[319] Furthermore, technical and theoretical challenges may also arise when detecting and making decisions regarding works that imitate the style of existing human creatives.

First, as the creation of a fingerprint pattern relies on the analysis of the actual content within a reference file, a hypothetical output that has learned from a collection of reference files,[320] without memorizing or partially reproducing any of them, should not match any specific reference file.

Using AI-generated paintings as an example, deep neural text-to-image generators are trained on a dataset through a process known as Stable Diffusion. This iterative denoising procedure involves gradually adding noise to an image until it

---

[316] *Id.*

[317] John Paul Titlow, *How YouTube Is Fixing Its Most Controversial Feature*, FAST CO. (Sept. 13, 2016), https://www.fastcompany.com/3062494/how-youtube-is-fixing-its-most-controversial-feature [https://perma.cc/NG7M-7DJV].

[318] *See* Sag, *supra* note 36, at 299.

[319] Titlow, *supra* note 317.

[320] *Id.*

becomes white noise. The process is then recorded and reversed for the AI to learn from.[321] The consumption of training images in this manner allows the AI to understand what humans truly desire from a given image description, rather than simply assembling materials.[322] Once the training process is complete, only the algorithm is relevant at the generation stage, with no reliance on the original training data from which the AI learned the artistic style. This implies that the final AI-generated artwork contains no original work or fragments of original work, but rather a mathematical expression abstracted from the original work.[323] When an AI-generated output successfully imitates an artist's style, without accidentally memorizing or reproducing a part of their work, it is unclear how the two fingerprints could overlap. This presents a challenge in detecting potential infringements in cases where AI-generated content imitates the style or essence of copyrighted works without directly copying their content.

Second, even if AI systems can detect outputs that imitate an artist's style, there are several arguments suggesting that these outputs should not be blocked. Copyright law clearly states that style is not copyrightable, as "only specific expressions of an idea may be copyrighted."[324] For instance, Picasso may have copyright on his cubist portrait of three women, but another artist could paint a similar subject in the cubist style without violating Picasso's copyright, as long as their work is not substantially similar to Picasso's specific expression.[325]

However, when it comes to generative AI systems, a notice serves as a warning against future content generations that bear varying resemblances to the copyrighted work.[326] Filtering for such inexact matches is technically challenging. Despite these challenges, major social networks have made significant progress in matching material against catalogs of copyrighted works. Generative AI companies can adopt similar approaches, although

---

[321] Zihang Lan, Shuhan Yang, Rui Fan, Bo Zhao & Yanru Yan, *Innovation or Piracy? Empirically Demarcating AI Painting Copyright Infringement Boundary*, PROC. 3D INT'L CONF. ON PUB. MGMT & INTELLIGENT SOC'Y 1328, 1334 (2023).

[322] *Id.*

[323] *Id.*

[324] Dave Grossman Designs, Inc. v. Bortin, 347 F. Supp. 1150, 1156 (N.D. Ill. 1972).

[325] *Id.* at 1156–57.

[326] Katherine Lee, A. Feder Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Train*, J. COPYRIGHT SOC'Y (forthcoming 2024) (manuscript at 143).

it may not provide an absolute solution. Based on our understanding of data collection processes,[327] it appears that AI companies have access to a potential catalog of reference files. However, due to the vast scale of data on which generative AI systems are trained, it remains "unclear whether a universal filter that compare[s] model output to the training data is feasible."[328] Implementing such a filter would demand sophisticated technical solutions and continuous adaptation to the ever-evolving nature of generative AI content.

Nevertheless, the large scale of current training datasets should not be used as an excuse for avoiding filtering obligations. In fact, there are potential advantages to implementing such measures. First, concerns have been raised about generative AI service providers taking a reckless approach to business growth. The United Nations General Secretary has criticized AI companies for prioritizing profit over the prevention of "serious unintended consequences."[329] Recklessness has been particularly evident in data collection, with companies facing backlash for unlicensed use of copyrighted content[330] and the inclusion of personal information in training data, potentially leading to privacy breaches.[331] Second, deduplication of training data has been suggested as a solution to mitigate dataset "memorization."[332] Although it may not address all instances of dataset replication, research indicates that "increased duplication in the training data tends to yield increased replication during inference."[333] If companies were legally required to filter copyright-infringing outputs, the process

---

[327] *See supra* Section III.A.

[328] *See* Sag, *supra* note 36, at 340.

[329] Sebastian Klovig Skelton, *UN Chief Blasts AI Companies for Reckless Pursuit of Profit*, COMPUT. WKLY. (Jan. 23, 2024), https://www.computerweekly.com/news/366567272/UN-chief-blasts-AI-companies-for-reckless-pursuit-of-profit [https://perma.cc/T7X4-PNR5].

[330] *See supra* Section III.A.

[331] Cat Zakrzewski, *FTC Investigates OpenAI over Data Leak and ChatGPT's Inaccuracy*, WASH. POST (July 13, 2013, 7:26 PM), https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan/ [https://perma.cc/29MY-24PR] ("The FTC also asked the company to provide records related to a security incident that the company disclosed in March when a bug in its systems allowed some users to see payment-related information, as well as some data from other users' chat history. The FTC is probing whether the company's data security practices violate consumer protection laws.").

[332] Gowthami Somepalli, Vasu Singla, Jonas Geiping & Tom Goldstein, Understanding and Mitigating Copying in Diffusion Models 1–2 (May 31, 2023) (arXiv preprint), arXiv:2305.20086v1.

[333] *Id.* at 4.

of creating and maintaining reference file databases could promote more responsible data management and deduplication.

Moreover, the law must allow for some degree of reference and re-interpretation of copyrighted material for essential social and cultural purposes, such as commentary, criticism, and parody.[334] Users of generative AI systems already utilize these services to create such works. One notable example occurred in the early months of 2023 when a YouTube user received a copyright strike for uploading a parody rap song about cats in the style of Eminem, following a complaint from his record label.[335] While efforts to remove future generated works of this nature are expected from copyright industries, attempting to filter out style imitation could introduce additional obstacles to the dissemination of such works.

However, it is important not to overstate the extent to which works imitating style are non-infringing and undetectable. Many AI-generated works that claim to only replicate style do, in fact, exhibit substantial similarities with the expressions of works they were trained on or even directly copy parts of them. For example, a recent New York Times article asked several AI systems, including Midjourney, ChatGPT, and Microsoft Bing, to generate images from movies and television shows, either directly or indirectly.[336] In several instances, the AI systems produced outputs that were close to or nearly identical to screenshots from the referenced works.[337] Even when copyright-related restrictions are imposed on user prompts, users of AI systems like DALL-E 3 have reported being able to generate images of characters that closely resemble copyrighted images by simply describing the

---

[334] *See* Samantha Cole, *Record Label Wipes AI-Generated Eminem Rapping About Cats from the Internet*, VICE (Apr. 4, 2023, 11:38 AM), https://www.vice.com/en/article/88xadz/ai-generated-eminem-rap-youtube-chatgpt [https://perma.cc/FA74-YTR8]; Erin Williamson, *Is My Parody Fair Use?*, COPYRIGHT ALL., (Mar. 16, 2021), https://copyrightalliance.org/is-my-parody-fair-use/ [https://perma.cc/2U5T-3RQU].

[335] Cole, *supra* note 334.

[336] Stuart A. Thompson, *We Asked A.I. to Create the Joker. It Generated a Copyrighted Image.*, N.Y. TIMES (Jan. 25, 2024), https://www.nytimes.com/interactive/2024/01/25/business/ai-image-generators-openai-microsoft-midjourney-copyright.html [https://perma.cc/XY6N-RCBM].

[337] *Id.* ("When Times journalists asked ChatGPT to create an image of SpongeBob SquarePants, the children's animated television character, it produced an image remarkably similar to the cartoon. The chatbot said the image only resembled the copyrighted work. The differences were subtle—the character's tie was yellow instead of red, and it had eyebrows instead of eyelashes. . . . When Times journalists omitted SpongeBob's name from another request, OpenAI created a character that was even closer to the copyrighted work.").

characters' appearance rather than using their names.[338] This highlights the challenge in ensuring that AI-generated content does not infringe on copyright even when attempting to focus on imitating style rather than directly copying content.

In order to effectively enforce copyright law in the era of generative AI, it may be necessary to exert some effort in filtering generated works that imitate the style of existing works. While identifying fair use can be challenging, there are proposals suggesting that AI could be employed to aid in the analysis of substantial similarity, thus offering potential for filtering in this specific context.

One approach to this issue is using supervised learning, where experts create a training dataset of works, like songs, using previously litigated cases and labeling them as "similar" or "not similar." The algorithm would then learn from this training dataset, enabling it to predict an outcome when presented with a new set of inputs.[339] By implementing such a system, it would be possible to analyze AI-generated content for similarities with existing copyrighted works more efficiently and accurately. This would help in distinguishing between imitations of style and instances of copyright infringement, allowing for better enforcement of copyright law while preserving the creative potential of generative AI systems.

By employing AI algorithms in the enforcement of copyright law, particularly in the context of generative AI, it could become possible to better differentiate between imitations of style and actual copyright infringements. As mentioned, supervised learning can be used to classify works as "similar" or "not similar" or to determine the degree of similarity between two works.[340] Unsupervised learning, on the other hand, can help identify correlations and trends in unlabelled datasets, which could further contribute to the development of more accurate filtering systems.[341]

---

[338] Wei Mao, *How to Get Around DALL-E 3's Copyright Restrictions: Hints and Tricks for Creating Copyrighted Characters in DALL-E 3*, PLAIN ENG. (Oct. 21, 2023), https://plainenglish.io/community/how-to-get-around-dall-e-3-s-copyright-restrictions-0fd7ec [https://perma.cc/9TZL-REVG].

[339] Shine Sean Tu, *Use of Artificial Intelligence to Determine Copyright Liability for Musical Works*, 123 W. VA. L. REV. 835, 855 (2021) (pointing out that the algorithm "would learn from this training dataset so that it would then be able to forecast an outcome from a new set of inputs").

[340] *Id.* at 855–56.

[341] *Id.* at 856.

While these AI-based filtering systems may not be perfect initially, they can be refined and improved over time, in much the same way that AI companies have been releasing generative AI systems that are not yet fully prepared to tackle potential copyright infringement issues. By gradually enhancing these filtering systems, it would be possible to strike a better balance between preserving the creative potential of generative AI and enforcing copyright law effectively.

In conclusion, the development and implementation of AI-based filtering systems could prove to be a valuable tool in addressing copyright concerns in the era of generative AI. Although challenges may arise in distinguishing between fair use, imitations of style, and actual infringements, the ongoing refinement of such systems can contribute to a more effective and balanced approach to copyright enforcement.

C. Applying Watermarking to Promote Attribution and Authenticity

In addition to the filtering responsibility, regulators should require companies to incorporate watermarking technology into AI-generated content. This technology can contribute to increased attribution and authenticity, as it allows for the tracking and identification of AI-generated content in cases of potential copyright infringement, content manipulation, or misuse.

1. The Nature of AI Watermarking

AI companies can apply watermarking technologies to embed unique and detectable signals into the outputs of their generative systems.[342] Closely related to AI watermarking is content labelling, which "involves applying visible content warnings to alert users to the presence of AI-generated media online."[343] Labelling has helped reduce sharing of debunked content, and could yield similar results in the context of AI-generated

---

[342] Tambiama Madiega, *Generative AI and Watermarking*, at 2 (Dec. 2023), https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf [https://perma.cc/C4TW-68EX].

[343] Chloe Wittenberg, Ziv Epstein, Adam J, Berinsky & David G. Rand*, Labelling AI-Generated Content: Promises, Perils, and Future Directions*, MASS. INST. OF TECH. (Mar. 27, 2024), https://mit-genai.pubpub.org/pub/hu71se89/release/1 [https://perma.cc/84GR-KVZ7].

content.[344] AI watermarking, though, is a subtle but robust method for tracking and identifying AI-generated media, while labelling depends on visible indicators that can be easily ignored or removed.[345]

AI watermarking typically involves encoding and detection. During the encoding step, the AI model is trained to embed specific signals into its output. This encoding process can involve modifying the AI algorithm to include watermark patterns or integrating the watermarking process into the model's architecture, ensuring that the watermark is consistently and effectively incorporated into the generated content.[346] At the detection step, specialized algorithms are employed to detect the embedded signals within the AI-generated content. Specifically, these detectors are designed to identify the unique watermarks incorporated during the encoding process. Once detected, the presence of the watermark allows the detector to determine whether the output was generated by an AI system and, in some cases, trace its origin back to a specific AI model or company.[347]

So far, AI watermarking techniques typically take three forms. First are visible watermarks, discernible to the naked eye. These watermarks are a pattern or image superimposed on the AI-generated image, allowing humans to immediately identify the AI origin of the content.[348] Second are hidden watermarks, physically present but not immediately perceptible to the naked eye, requiring a detector for identification.[349] Watermarks can be hidden by manipulating the pixels of an image or adding syntax patterns, such as specific word choices and punctuation, to AI-generated text.[350] Third, cryptographic watermarks, detectable using certain types of detectors, exist in the metadata attached to

---

[344] *Id*.

[345] *Id*.

[346] Madiega, *supra* note 342, at 2.

[347] *Id*.

[348] *See* Gustaf Björksten & Daniel Leufer, *Identifying Generative AI Content: When and How Watermarking Can Help Uphold Human Rights* 5 (Access Now, Discussion Paper, 2023), https://www.accessnow.org/wp-content/uploads/2023/09/Identifying-generative-AI-content-when-and-how-watermarking-can-help-uphold-human-rights.pdf; *see also* Instasafe Mktg., *Digital Watermarking and its Types*, ZERO TR. BLOG (Feb. 13, 2024), https://instasafe.com/blog/digital-watermarking-and-its-types/ [https://perma.cc/5285-EU43].

[349] Björksten & Leufer, *supra* note 348, at 5.

[350] *Id.*

the output, identifying the AI signee and potentially additional provenance data.[351]

Each of these three watermarking techniques serves a different purpose and offers varying levels of protection and traceability. Visible watermarks are more suitable when immediate recognition of AI-generated content is needed, whereas hidden watermarks provide a more subtle method for tracking content without altering its appearance.[352] Cryptographic watermarks, on the other hand, offer a secure way to store provenance data without directly affecting the content's visual or textual elements.[353] AI companies should choose the most appropriate watermarking technique based on their specific requirements and the desired level of transparency, security, and user experience.[354]

Watermarking is already a well-established technology, and numerous AI companies have begun applying it to their generative systems. For instance, Google has introduced a watermarking method called SynthID to identify images,[355] videos, and even text[356] generated by its AI systems. Similarly, since early 2024, OpenAI has embedded C2PA metadata into images generated by ChatGPT and DALL-E 3.[357] These watermarks are designed to assist in verifying the origin and history of AI-generated images. In 2024, to bolster trust and transparency, Meta also unveiled its plan to incorporate invisible watermarks into its text-to-image generation systems.[358]

Given the various functions of watermarking technology, it is essential for legislators to legally obligate AI companies with generative systems to apply such technology to their outputs. Legally, enforcing watermarking requirements could encourage

---

[351] *Id*.

[352] *Id.*

[353] *Id.* at 8.

[354] *Id*. at 6.

[355] Sven Gowal & Pushmeet Kohli, *Identifying AI-generated images with SynthID,* GOOGLE DEEPMIND (Aug. 29, 2023), https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/ [https://perma.cc/T95D-5ZYN].

[356] *Watermarking AI-generated Text and Video with SynthID*, GOOGLE DEEPMIND (May 14, 2024), https://deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid/ [https://perma.cc/3U2F-JQ58].

[357] *C2PA in DALL·E 3*, OPENAI, https://help.openai.com/en/articles/8912793-c2pa-in-dall-e-3 [https://perma.cc/YLT9-FKK6] (last visited Nov. 10, 2024).

[358] Nick Clegg, *Labeling AI-Generated Images on Facebook, Instagram and Threads*, META (Feb. 6, 2024), https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/ [https://perma.cc/D37H-L9GB].

ethical practices of attribution and authenticity in AI-generated content. The subsequent two sections will examine the potential positive effects that this legal mandate could bring about.

2. Achieving Attribution

Watermarking, as previously mentioned, can help distinguish between AI-generated and human-generated content by embedding unique, identifiable patterns or signals into the AI-generated content during its creation.[359] This promotes transparency, accountability, and responsible use of AI systems.

This approach can serve as an initial step towards recognizing the contributions of human-authored works within AI-generated content. To achieve this, AI systems must clearly identify the machine-generated status of their outputs, enabling people to differentiate between human and AI-generated creations. As generative AI outputs become increasingly sophisticated and nearly indistinguishable from human-generated content,[360] it is essential to provide consumers with the ability to identify the creator. This empowers consumers to make informed decisions based on their preferences and expectations.[361] For instance, some readers may favor human-authored books for the emotional depth, nuance, and authenticity of personal experience they provide. Conversely, others may choose AI-generated books, which potentially offer a more objective perspective, reduced human biases, and a lower likelihood of errors.

AI watermarking techniques can also facilitate the identification process by providing clear, detectable signals about content's origin.[362] This allows customers to make educated choices about the content they consume, whether it is AI-generated or human-created. Additionally, this transparency enables customers to support creators and companies whose values, ethical considerations, and preferences align with their own. As AI-generated content continues to expand, promoting transparency and accountability in the creation process will be

---

[359] *Id.*

[360] Björksten & Leufer, *supra* note 348, at 4 ("[G]enerative AI is, or will soon be, at a point where its content is often indistinguishable from human-generated content to the perception of most customers.").

[361] *See* Madiega, *supra* note 342, at 4.

[362] *Guide to Watermarking AI-Generated Content for Trust*, QUICKCREATOR (July 18, 2024), https://quickcreator.io/blog/guide-watermarking-ai-generated-content-trust/ [https://perma.cc/LX7L-7CCU].

vital for fostering trust and encouraging responsible use of AI technologies.

AI watermarks can also play a crucial role in alerting users that AI-generated content is a mechanical response rather than a creation with moral responsibility or human-like intentions.[363] This can help prevent incidents where users become emotionally attached, for example, to an AI chatbot to the extent that they are unable to distinguish it from a sentient being.[364] Watermarks can help users to recognize AI creativity and remain cautious about fully immersing themselves in it. This approach can be particularly effective when combined with "self-disclosing AI machines" that adopt language and patterns intuitively associated with AI sources while avoiding language that might be mistaken for human-generated content.[365]

AI-watermarking techniques can also help deter various forms of malicious exploitation of copyrighted works, such as plagiarism, infringement, license breaches, and derogatory treatment of copyrighted materials, all of which violate the ethics of attribution.[366] By attaching metadata with comprehensive provenance data to the content, cryptographic watermarks allow any traces of malicious exploitation to be recorded and made available for the author to examine, acting as a powerful tool against numerous types of malicious activities.[367] For instance, if an AI-altered version of an author's work is circulated on the Internet, the author can leverage the cryptographic watermarks to capture the adaptation history and trace it back to the individual who utilized the AI.

---

[363] *See generally* Alexei Grinbaum & Laurynas Adomaitis, The Ethical Need For Watermarks in Machine-Generated Language (Sept. 7, 2022) (arXiv preprint), arXiv:2209.03118.

[364] Imane El Atillah, *Man Ends His Life After an AI Chatbot 'Encouraged' Him To Sacrifice Himself To Stop Climate Change*, EURONEWS (Mar. 31, 2023), https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate- [https://perma.cc/2VU7-7XQQ].

[365] Maurice Jakesch, Jeffrey T. Hancock & Mor Naaman, *Human Heuristics for AI-Generated Language Are Flawed*, 120 PROC. NAT'L ACAD. SCIS., no. 11, 2023, at 1, 5.

[366] Isha Marathe, *Combating Deepfakes: 4 Types of AI Watermarks And Their Unique Challenges*, LEGALTECH NEWS (June 25, 2024, 5:25 PM), https://www.law.com/legaltechnews/2024/06/25/combating-deepfakes-4-types-of-ai-watermarks-and-their-unique-challenges/ [https://perma.cc/U7BZ-WHAM].

[367] Björksten & Leufer, *supra* note 348, at 9. *See also, e.g.*, Dominique Guinard, *How Digital Watermarks Strengthen C2PA Content Credentials*, MEDIUM (Nov. 2, 2023), https://domguinard.medium.com/how-digital-watermarks-strengthen-c2pa-content-credentials-764be8b15a38 [https://perma.cc/93MR-U6DR].

In instances of exploitation through adaptation, such as derogatory treatment or unauthorized modification of a work, visible and hidden watermarks can serve as effective tools to discourage malicious activities involving AI models. These watermarks, embedded within the adapted content, can be traced by copyright owners, making it easier to identify unauthorized adaptations. Furthermore, original authors can use these watermarks to establish that the watermarked content represents an adapted version, rather than their own original creation. This can safeguard authors' moral rights against false attribution, preserving the integrity of their work.

3. Establishing Authenticity

AI watermarks play a crucial role in establishing the authenticity of content, particularly as generative AI outputs become increasingly indistinguishable from human-created content. The ease of creating deepfake content that can mislead the public highlights the importance of watermarking for verifying content origins. A notable example of an indistinguishable deepfake is the image generated by Midjourney in March 2023 that depicted the Pope wearing a puffer jacket.[368] AI-generated content poses several threats to society, such as a proliferation of fake news and fraudulent documents. Such threats are not merely theoretical; on May 22, 2023, a falsified photograph of an explosion near the Pentagon, believed to be AI-generated, went viral on social media. Despite the swift response from public authorities clarifying that there had been no explosion, the photo led to a 0.3% decline in the S&P 500, resulting in a $500 billion loss in the stock market.[369] This incident highlights the importance of timely identification and distinction of deepfake content, especially in time-sensitive contexts such as those affecting the stock market or upcoming elections.

Establishing authenticity is particularly crucial when fundamental societal values are at stake and integrity is deemed

---

[368] Chris Stokel-Walker, *We Spoke To the Guy Who Created the Viral AI Image of the Pope That Fooled the World*, BUZZFEED NEWS (Mar. 28, 2023, 2:15 PM), https://www.buzzfeednews.com/article/chrisstokelwalker/pope-puffy-jacket-ai-midjourney-image-creator-interview [https://perma.cc/2PXR-Y935].

[369] Titus Wu, *California Lawmakers Push for Watermarks on AI-Made Photo, Video*, BLOOMBERG L. (Jan. 26, 2024, 5:02 AM), https://news.bloomberglaw.com/artificial-intelligence/california-lawmakers-push-for-watermarks-on-ai-made-photo-video [https://perma.cc/7GNH-VBQ4].

indispensable, as in scientific research and legal proceedings.[370] Basing technological advancements on fake scientific research or judgments on fraudulent evidence could have far-reaching and damaging implications. Establishing authenticity is also crucial in circumstances such as the Liar's Dividend, [371] where an individual or group falsely claims that genuine information is fake in an attempt to evade accountability. Experts have revealed that these tactics have been employed in contexts such as the U.S. Presidential Election and the Israel-Hamas War.[372]

The solution to the challenges posed by deepfakes is the application of AI watermarks. Mandating their use, particularly in industries that are time-sensitive or require extra attention to authenticity, for example, media, education, and legal industries, can help create a norm where content bearing watermarks is presumed to be fake until proven otherwise. Visible watermarks are suitable in the case of time-sensitive matters because they can mitigate harm caused by delays in watermark detection. Hidden and cryptographic watermarks are more suitable for specialized industries where larger organizations can employ costly detectors. For example, courts, publishers, universities, and public authorities can deploy watermark detection systems before accepting a piece of work. Upon detecting watermarked content, relevant authorities can issue detection reports, which serve to falsify the content in question.

In the media, where deepfakes proliferate, watermark detection technology can be incorporated into standard Digital Rights Management ("DRM") systems to limit content that bears watermarks associated with generative AI. [373] Regarding the Liar's Dividend, once watermarks become mandatory for certain content, individuals will be discouraged from claiming genuine content as deepfakes, as true content would unlikely bear any watermarks. This approach can help preserve the integrity of

---

[370] *See* Lev Craig, *AI Watermarking*, TECHTARGET, https://www.techtarget.com/searchenterpriseai/definition/AI-watermarking [https://perma.cc/A9EN-5MGB] (Oct. 2023).

[371] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1785 (2019).

[372] *Id.* at 1768 n.78*;* Tiffany Hsu & Stuart A. Thompson, *A.I. Muddies Israel-Hamas War in Unexpected Way*, N.Y. TIMES (Oct. 30, 2023), https://www.nytimes.com/2023/10/28/business/media/ai-muddies-israel-hamas-war-in-unexpected-way.html [https://perma.cc/4XAZ-3XYT].

[373] *See* Björksten & Leufer, *supra* note 348, at 6.

authentic content and contribute to a more reliable and trustworthy digital media landscape.

The incorporation of watermarking technology in the E.U. Artificial Intelligence Act ("AI Act")[374] represents a significant step towards promoting transparency and accountability in the use of generative AI systems.[375] By mandating that AI-generated outputs are marked in a machine-readable format, the AI Act aims to ensure that users can easily identify content that has been artificially generated or manipulated.[376] This adaptability ensures that the technology remains effective while accommodating the diverse range of AI systems and their respective outputs. Moreover, the exemption provided for AI systems with assistive functions or those that do not substantially alter input data or semantics demonstrates the EU's approach towards regulation. By focusing on AI systems that have a more significant impact on generated content, the AI Act ensures that the transparency obligation is targeted and proportionate.

While the U.S. federal government has not yet enacted comprehensive legislation specifically targeting generative AI, the executive order signed by President Biden in October 2023 demonstrates a commitment to addressing the challenges posed by these technologies, including issues related to content authenticity and provenance.[377] The focus on developing standards, tools, and methods for content authentication, labelling, and detection underscores the increasing awareness of the potential risks and benefits associated with AI-generated content.[378]

At the state level, California has been at the forefront of addressing AI-related concerns. For example, in 2019, it enacted a law requiring the disclosure of AI-generated content in the context of political campaigns and advertisements.[379] This

---

[374] Council Regulation 2024/1689, 2024 O.J. (L) (EU).

[375] In its preamble, the AI Act sets out the need for AI transparency requirements:

> A variety of AI systems can generate large quantities of synthetic content that becomes increasingly hard for humans to distinguish from human-generated and authentic content. The wide availability and increasing capabilities of those systems have a significant impact on the integrity and trust in the information ecosystem, raising new risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception.

Id. pmbl. 133.

[376] *Id.* art. 50(1)–(2).

[377] Madiega, *supra* note 342, at 4.

[378] Exec. Order No. 14110, 88 Fed. Reg. 75191, 75202–03 (Nov. 1, 2023).

[379] A.B. 730, 2019–2020 Reg. Sess. (Cal. 2019) (enacted at CAL. ELEC. CODE § 20010 (West 2023)).

legislation aims to mitigate the potential impacts of AI-generated "deepfake" content on elections and democratic processes.[380] Additionally, California passed a law in 2021, targeting the use of AI-generated bots in commercial transactions and requiring businesses to disclose when they use AI-based communication systems for sales and customer service purposes.[381]

These state-level actions in California represent significant steps towards addressing some of the challenges posed by generative AI. As this Article advocates, a comprehensive federal approach to regulating service providers and users of generative AI systems remains necessary. This would help establish a uniform legal framework across the United States and ensure that AI-generated content is used responsibly and ethically, while minimizing potential risks and negative impacts.

CONCLUSION

This Article advocates for the application of human creativity ethics to govern AI-generated content, thus urging regulators to adopt new legal responsibilities for technology companies concerning their generative AI systems. Drawing on copyright law and professional practices as a source of insight, the Article first establishes three fundamental principles of the ethics of human creativity: originality, attribution, and authenticity. While individuals may borrow ideas and draw inspiration from others, they may not unfairly reduce the value of past creative works by engaging in copyright-infringing activities. Social practices and norms routinely reinforce these ethical values by demanding credit for creatives' intellectual contributions to the knowledge ecosystem and upholding the authenticity of creative endeavors.

AI companies have, however, shielded their generative systems from these ethical principles by keeping each major process of those systems—namely the collection, utilization, and generation of works—opaque. This *status quo*, as the Article shows, has created an ethical vacuum in the domain of generative AI, with serious social consequences. In response, the Article explores how and why AI companies should abide by the ethical principles of originality, attribution, and authenticity. It proposes that they should be first legally required to implement filtering technologies that monitor and remove AI-generated works that

[380] Wu, *supra* note 369.
[381] CAL. BUS. & PROF. CODE § 17941 (West 2019).

appear identical or substantially similar to copyrighted works. Furthermore, they should assume another legal responsibility to employ watermarking technologies to differentiate between works created by AI and humans, while also preventing the generation of harmful content, such as disinformation. By embracing these new legal responsibilities, AI companies can help foster a more responsible approach to AI creativity, addressing concerns related to copyright protection and mitigating the potential negative impacts of AI-generated content on society.

By proposing new approaches based on the ethics of creativity, this Article seeks to move beyond the fair use debate and provide a more comprehensive framework for addressing the legal and ethical challenges posed by AI-generated content. Recognizing that creativity is a social process, I have established three ethical principles—originality, attribution, and authenticity—that guide human engagement in creative activities. The opacity of AI systems in their collection, utilization, and generation of works should not exempt them from these ethical principles. Instead, we should require AI companies to assume ethical responsibilities by integrating filtering and watermarking technologies into their generative systems, thereby promoting a more responsible and transparent approach to AI creativity.

Making generative AI systems ethical is not only possible but necessary, and copyright law can play a significant role in achieving this goal. By understanding the ethical foundations of human creativity and addressing the unique challenges posed by AI-generated content, we can develop legal frameworks that protect the rights of creators, promote authenticity, and foster innovation in the public interest. This approach will ensure that generative AI systems are used legally and ethically, increasing public trust in such systems and benefiting both individuals and society at large.