COPYRIGHT REGENERATED: HARNESSING GENAI TO MEASURE ORIGINALITY AND COPYRIGHT SCOPE

Uri Y. Hacohen* & Niva Elkin-Koren**

ABSTRACT

The rise of Generative Artificial Intelligence ("GenAI") models is revolutionizing the creative domain. By using models like GitHub Copilot, OpenAI's GPTs, or Stability AI's Stable Diffusion, non-professional users can generate high-quality content such as text, images, music, or code. These powerful tools facilitate new, unimaginable ways of human creativity on a large scale, disrupting the professional creative sectors. This Article proposes a novel approach that leverages the capacity of GenAI to assist in copyright legal disputes.

GenAI models are trained on examples, generalizing expressive patterns and applying those learnings to perform different tasks, such as autocompleting sentences or generating visual outputs in response to a textual prompt. These models are designed to grasp complex probability distributions from training samples by identifying recurring relationships between input and output data.

Similarly, humans learn from a corpus of preexisting materials, memorize impressions, learn styles, extract themes from text, generalize principles from new materials, and engage in deconstructing and reconstructing processes. Unlike human learning, which occurs within the confines of the human mind, GenAI learning involves digital replication. Consequently, GenAI has sparked numerous class actions alleging copyright infringement. These claims assert that the models infringe copyright, either because they are trained on copyrighted materials without authorization, generate derivative works of those materials, or both.

While copyright law prohibits the unauthorized copying of protected expressions, it permits the extrapolation and learning of ideas.

For a work to be copyrighted, it must be original, meaning the author must originate it. As a result, the law does not protect expressions that are generic and, therefore, cannot be attributed to any particular author, such as ideas, scènes à faire, or conventional programming standards.

For centuries, courts have struggled to consistently differentiate between original expressions and generic ones, resulting in systematic overprotection of copyrighted works. GenAI presents an unprecedented opportunity to inform and improve this legal analysis. By learning from data at various levels of granularity, GenAI systems are revealing the shared patterns in preexisting works that were previously difficult to measure accurately.

In this Article, we propose a novel approach for measuring originality to assist in copyright legal disputes. We harness the powerful learning capacity of GenAI to gain more nuanced insights into the genericity of expressions on a significantly larger scale. Based on interdisciplinary research in computer science and law, we propose employing data-driven bias — a fundamental aspect of inductive machine learning — to assess the genericity of expressive compositions in preexisting works.

During learning, GenAI models distill and rank expressive compositions based on their prevalence in the models' datasets. The more frequently these expressive compositions appear in the GenAI models' datasets (indicating their "generic" nature), the more likely GenAI models are to utilize them when generating new works. Conversely, the rarer expressive compositions appear in the GenAI models' datasets (indicating their "original" nature), the less likely GenAI models are to utilize them.

Leveraging the capacity of GenAI to learn with greater nuance and on a much grander scale could have groundbreaking implications for copyright law. It could assist courts in determining copyright scope, potentially leading to more efficient and equitable resolutions. Moreover, it has the potential to inform the Copyright Office's registration practices and provide a valuable signal to facilitate market licensing transactions. Finally, by harnessing GenAI to measure originality at scale, our approach offers valuable insights for policymakers seeking to adapt copyright law to meet new challenges imposed by an era of "cheap creativity" enabled by GenAI.

TABLE OF CONTENTS

I. INTRODUCTION

I WANT TO BE A MACHINE.

— ANDY WARHOL[1]

Recently, in *Warhol v. Goldsmith*,[2] the Supreme Court faced a difficult dilemma. The Court had to decide whether Andy Warhol's depiction of singer-songwriter Prince (Figure 1, left side) was sufficiently transformative to break free from Lynn Goldsmith's exclusive right to control the uses of her copyrighted photograph (Figure 1, right side), or whether Warhol's creation was an unlawful "derivative" work, subservient to Goldsmith's exclusivity.[3] The Court sided with Goldsmith, but

---

1. Keith Hartley, *Andy Warhol and Eduardo Paolozzi: I Want to Be a Machine*, NAT'L GALLERIES SCOT. (Nov. 14, 2019), https://www.nationalgalleries.org/art-and-artists/ features/andy-warhol-and-eduardo-paolozzi-i-want-be-machine [https://perma.cc/9TS3-W37X]. In reality, Warhol's famous quote was a bit different. According to Gene Swenson, who conducted the interview that produced Warhol's quote, Warhol actually said "everybody should be a machine," in the context of being nonjudgmental about gender expression. *See* Jennifer Sichel, *"Do You Think Pop Art's Queer?" Gene Swenson and Andy Warhol*, 41 OXFORD ART J. 59, 62 (2018).

2. Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023).

3. Christopher Sprigman & Kal Raustiala, *Why Andy Warhol's 'Prince Series,' the Subject of a Long-Term Copyright Dispute, Should Be Considered Fair Use After All*, ARTNET NEWS (Apr. 27, 2021), https://news.artnet.com/art-world-archives/andy-warhol-prince-series-op-

not before revealing strong disagreement among the Justices,[4] the lower courts,[5] and the numerous advising amici.[6]

Figure 1: Andy Warhol's Depiction of Prince (left); Lynn Goldsmith's Photograph of Prince (right)

Recently, several pending class action lawsuits regarding generative AI ("GenAI") models brought similar issues before lower courts.[7] For example, in *Getty Images v. Stability AI*, the Delaware District court

---

ed-1962050 [https://perma.cc/LM3K-HEZH] ("The real challenge in the Warhol case is articulating exactly what the 'new and different' elements are. And this illustrates an age-old problem in copyright law that has never been solved . . . ."). The Court conducted this analysis under the first prong of the fair use doctrine, which looks at the extent to which the challenged use added "new expression, meaning, [and] message" to the original work. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

4. The dissent accused the majority of "hamper[ing] creative progress and undermin[ing] creative freedom." *Warhol*, 598 U.S. at 560 (Kagan, J., dissenting). On the other hand, the majority accused the dissent of reciting "theme[s] . . . familiar to any student of copyright law" and "offer[ing] no theory of the relationship between transformative uses of original works and derivative works that transform originals." *Id.* at 548–49.

5. The district court's opinion was that Warhol's Prince Series works are transformative and protected by fair use. *See* Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 382 F. Supp. 3d 312, 324–26 (S.D.N.Y. 2019). The Second Circuit disagreed, finding that the district court "stretched [precedent] too far" and calling the district court's transformativeness standard "overly liberal . . . [which] risks crowding out statutory protections for derivative works." Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 11 F.4th 26, 38–39 (2d Cir. 2021).

6. *See* Martin Adams, *A "Brief" Update on the Warhol Case — Amicus Briefs and the Solicitor General's View*, AUTHORS ALL. (Aug. 25, 2022), https://www.authorsalliance.org/2022/08/25/a-brief-update-on-the-warhol-case-amicus-briefs-and-the-solicitor-generals-view/ [https://perma.cc/UK36-7DV7] ("Last week, the final set of amicus curiae . . . briefs were filed. In total, there were 38 amicus briefs filed: 8 in support of the AWF, 20 in support of Goldsmith, and 9 in support of neither party.").

7. Doe 1 v. GitHub, Inc., No. 22-CV-06823, 2023 WL 3449131 (N.D. Cal. May 11, 2023); Andersen v. Stability AI, Ltd., No. 23-CV-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023).

must consider whether works created with GenAI models such as Stable Diffusion (e.g., Figure 2, left side) constitute lawful independent creations, or whether they are infringing derivatives of underlying copyrighted works used to train the models (e.g., Figure 2, right side).[8] Such lawsuits put courts in an even more challenging position than cases like *Warhol*. Plaintiffs argue that GenAI models are "21st-century collage tools" that violate the rights of millions of authors.[9] Defendants argue that GenAI models are "expanding the boundaries of human creativity."[10] Both sides have persuasive arguments.

Figure 2: A Photograph Generated by Stable Diffusion (left); A Picture Owned by Getty Images (right)

Copyright scholars are only beginning to grapple with the ways in which GenAI exacerbates existing challenges in copyright law.[11] Copyright law's ultimate goal is to foster the creation and dissemination of expressive works by granting authors limited exclusive rights to their

8. Complaint & Demand for Jury Trial at 18, Getty Images, Inc. v. Stability AI, Inc., No. 1:23-cv-00135 (D. Del. Feb. 3, 2023).

9. Complaint & Demand for Jury Trial at 1, Andersen v. Stability AI, Ltd., No. 23-CV-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023).

10. Amended Notice of Motion, Motion to Dismiss & Memorandum of Points and Authorities in Support of Motion to Dismiss at 1, Andersen v. Stability AI, Ltd., No. 23-CV-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023).

11. *See, e.g.*, Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCIENCE 158 (2023) (mapping GenAI challenges for copyright doctrines); Oren Bracha, The Work of Copyright in the Age of Machine Production (Sept. 24, 2023) (unpublished manuscript) (on file with the University of Texas at Austin School of Law), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4581738 [https://perma.cc/VUW8-NQ59] (same); Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley & Percy Liang, *Foundation Models and Fair Use* (Stan. L. & Econ. Olin Working Paper, Paper No. 584, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4404340 [https://perma.cc/5RTM-NVBT] (mapping GenAI challenges for the fair use doctrine); Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295 (2023).

respective expressions.[12] Thus, copyright law's greatest challenge is allocating legal entitlements to expressive works by consecutive authors.[13] This task has always been a challenge because authorship derives from cultural contexts.[14] Authors routinely engage with preexisting materials to convey a meaningful message.[15] This task grows more complicated when GenAI augments human creativity. Generative models draw upon preexisting works created by an often unidentifiable multitude of authors. Consequently, allocating legal entitlements among all possible claimants becomes an insurmountable objective.

In this Article, we argue that GenAI may also be able to address the same challenges that it creates. To further its goal, copyright law strives to allocate legal entitlements only to original works of authorship, meaning those expressions which originated from the author.[16] Therefore, when deciding copyright disputes, courts are often tasked with assessing the originality of given works to delineate the scope of their legal protections. So far, courts have performed this task by applying numerous legal doctrines such as the idea/expression dichotomy,

---

12. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1107–09 (1990).

13. *See* JAMES BOYLE, SHAMANS, SOFTWARE AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY 53 (1996) (tackling the problem of information ownership).

14. *See* Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1163–65 (2007) (explaining that creativity is a social construct).

15. *See* Paul Edward Geller, *Copyright History and the Future: What's Culture Got To Do With It?*, 47 J. COPYRIGHT SOC'Y U.S.A. 209, 231 (2000) (emphasizing the cumulative nature of copyright expression).

16. 17 U.S.C. § 102(a) ("Copyright protection subsists, in accordance with this title, in original works of authorship . . . .").

merger, scènes à faire, substantial similarity, and fair use.[17] These doctrines are notoriously vague and unpredictable.[18] In practice, they lead to the overprotection of preexisting works.[19]

GenAI introduces new opportunities to enhance these methods by informing legal analysis with quantitative measures. Since GenAI documents the output of human creativity on an unprecedented scale, it facilitates a systematic study of the concealed interconnections among elements of expressive works.[20] Consequently, GenAI may facilitate

17. See discussion *infra* Section II.B.

18. *See* SAS Inst., Inc. v. World Programming, Ltd., 64 F.4th 1319, 1326 (Fed. Cir. 2023) ("The scope of protection is 'not constant' across all literary works . . . . Nor is it necessarily constant across all elements in a single work." (citing Eng'g Dynamics, Inc. v. Structural Software, Inc., 26 F.3d 1335, 1348 (5th Cir. 1994))); Walker v. Time Life Films, Inc., 784 F.2d 44, 48 (2d Cir. 1986) (noting that the idea/expression dichotomy is "a distinction easier to state than to apply"); Jessica Litman, *The Public Domain*, 39 EMORY L.J. 965, 975 (1990) ("[T]he boundaries of copyright are inevitably indeterminate."); *see also* Mark A. Lemley, *Our Bizarre System for Proving Copyright Infringement*, 57 J. COPYRIGHT SOC'Y U.S.A. 719, 741 (2010) ("The line between idea and expression . . . [is] extraordinarily difficult . . . to draw."); Richard H. Jones, *The Myth of the Idea/Expression Dichotomy in Copyright Law*, 10 PACE L. REV. 551, 553 (1990) ("No 'expressionless idea' exists and . . . it makes no sense to speak of an 'idealess expression' . . . . [T]h[is] scheme of differentiating idea from expression does not aid courts in their task . . . ."); Robert Yale Libott, *Round the Prickly Pear: The Idea-Expression Fallacy in a Mass Communications World*, 14 UCLA L. REV. 735, 740 (1967) (arguing that the idea/expression distinction turns not on concrete definitions but rather "demonstrat[es] the court's ad hoc decision that *this* plot, *this* theme, or *these* characters should not be granted the sanctuary of the copyright laws" (emphasis in original)) ; Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (discussing levels of abstraction at which copyrightability may be tested but noting that "[n]obody has ever been able to fix that boundary, and nobody ever can"); Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) ("The test for infringement of a copyright is of necessity vague . . . no principle can be stated as to when an imitator has gone beyond copying the 'idea,' and has borrowed its 'expression.'").

19. *See* James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 884–85 (2007) (arguing how the unpredictability of liability leads risk-averse users to seek a license where none is needed); *see also* NEIL WEINSTOCK NETANEL, COPYRIGHT'S PARADOX 6–7 (2008) (discussing how copyright's exponential growth is in "ungainly distension" with free speech); Jessica Litman, *Billowing White Goo*, 31 COLUM. J.L. & ARTS 587, 587 (2008) (noting how fair use has failed to keep up with huge expansions in copyright protections); John Tehranian, *Infringement Nation: Copyright Reform and the Law/Norm Gap*, 2007 UTAH L. REV. 537, 543–48 (explaining how ordinary behaviors, such as unauthorized video recording, could be criminal infringement under current copyright law); Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 357 (1999) (weighing the public interest costs of broadening copyright law over information production and exchange processes).

20. For example, Google's T5 and Facebook's LLaMA datasets were trained on content from over fifteen million websites. Similarly, OpenAI's GPT-3 features a model size of 175 billion parameters and was trained with forty-five terabytes of data. This means hundreds of billions of words, or essentially, the entire Internet. *See* Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang et al., A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT (May 2023) (unpublished manuscript), https://arxiv.org/pdf/2302.09419.pdf [https://perma.cc/5LR2-MR4T]; RISHI BOMMASANI, DREW A. HUDSON, EHSAN ADELI, RUSS ALTMAN, SIMRAN ARORA, SYDNEY VON ARX ET AL., ON THE OPPORTUNITIES AND RISKS OF FOUNDATION MODELS (2021), https://crfm. stanford.edu/assets/report.pdf [https://perma.cc/TBG8-AAZK].

the development of new and more accurate measures to assess the originality of these works.

Copyrighted works demonstrate various compositions of basic elements (e.g., lines, shapes, colors, textures, words, musical sounds). These elements are the building blocks of creative expression and therefore not protected by copyright law.[21] However, compositions of these elements ("expressive compositions") can be protected. The scope of protection that copyright law affords expressive compositions varies with their originality, which, in turn, depends on their prevalence and cultural embedment.[22] The more ubiquitous the compositions of elements are and the more they are absorbed in preexisting works, the less likely they are to be considered original under copyright law.[23] This Article refers to this implicit dynamic in copyright law as the *genericity principle*.[24]

In addition, the more "generic" expressive compositions become, the more likely GenAI models are to capture their patterns and deploy them when generating new expressive works.[25] Indeed, genericity is a fundamental principle that GenAI operates on. During training, models learn recurrent patterns among basic elements in works included in the models' training set.[26] During deployment, the models apply these patterns to generate new content in response to users' prompts.[27]

Building on these insights, we demonstrate how new computational procedures can utilize GenAI models to produce originality scores for copyrighted works.[28] Specifically, we propose to rank the originality of works based on the genericity of their expressive compo-

---

21. *Cf.* Boisson v. Banian, Ltd., 273 F.3d 262, 272 (2d Cir. 2001) ("[O]riginal works broken down into their composite parts would usually be little more than basic unprotectible elements like letters, colors, and symbols."). *See generally* Justin Hughes, *Size Matters (or Should) in Copyright Law*, 74 FORDHAM L. REV. 575 (2005) (suggesting that copyright protection over small pieces of creative expressions threatens "recombinant culture" in software programming, collage art, and more).

22. *See* discussion *infra* Section II.B. Furthermore, the rate of genericity is linked to the level of the expressive compositions' complexity. The more complex expressive compositions are, the longer it takes them to become generic when used. Highly complex works may still be protected even if they achieved "fully" generic status, but their scope of protection is likely to be very narrow and limited to verbatim copying. *See* Justin Hughes, *Fair Use Across Time*, 50 UCLA L. REV. 775, 778 (2003).

23. *See* Jamie Lund, *Copyright Genericide*, 42 CREIGHTON L. REV. 132, 140 (2009).

24. Scholars and practitioners of intellectual property law often associate genericity with trademark law. In this Article, we argue that genericity is also implicitly considered in copyright law. *See* discussion *infra* Section II.B.

25. *See* Uri Hacohen, Niva Elkin-Koren, Amit Beremano, Roi Livni & Adi Haviv, Measurable Copyright Similarity for Generative Models (2023) (unpublished manuscript) (on file with authors).

26. *See* Bracha, *supra* note 11, at 6.

27. *See* Bracha, *supra* note 11, at 10 ("GenAI['s] distinctive feature is generating new information goods.").

28. *See* discussion *infra* Section III.C.; Hacohen et al., *supra* note 25.

sitions. When GenAI models treat such patterns as generic, their originality score will be low.[29] This approach takes advantage of GenAI's capacity to learn from data at different granularity levels, thereby revealing the underlying shared patterns in preexisting works.

This capacity to measure genericity and quantify originality against the knowledge captured by a GenAI model could enable copyright law to distinguish more accurately between original works and those in the public domain. It may accordingly introduce more nuance into copyright analysis in deciding copyright scope.

These abilities are particularly crucial today as numerous copyright owners are suing the makers of GenAI systems, claiming that the outputs of these systems are substantially similar to their original works.[30] To the extent that such similarity arises from genericity, it may not result in copyright liability.

Our approach also opens new opportunities for increasing the fairness and efficiency of copyright registration and licensing practices. For example, quantified originality measures could assist the U.S. Copyright Office in distinguishing "cheap creativity" from sufficiently original GenAI-augmented works worthy of legal protection.[31] Our approach could also support objective measures for remunerating authors of original works generated through interaction with GenAI systems.[32]

This Article proceeds in three parts. Part II describes the objectives of copyright law and presents genericity as an implicit organizing principle that underlies copyright law's originality and scope delineating doctrines. Part III proposes leveraging GenAI models to measure and quantify copyright originality. We explore the rise of GenAI technology, its disruptive effects on copyright law, and the shortcomings of contemporary approaches to remedy these effects. Building on these shortcomings, we propose using GenAI models to produce originality scores for expressive works of authorship.

---

29. *See* discussion *infra* Part III. Our procedure essentially distills the works' protected "expressions" from its unprotected "ideas," or, more accurately, it ascertains how original and therefore protected the expressive elements of each copyrighted works are. *Cf.* Jones, *supra* note 18, at 598 (arguing that "[t]he only relevant criteria for distinguishing unprotectible from protectible expressions are originality and creativity of the expressions in a work").

30. *See* Amended Notice of Motion, *supra* note 10; *see also* Michael M. Grynbaum & Ryan Mac, *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*, N.Y. TIMES (Dec. 27, 2023), https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html [https://perma.cc/RXR2-SWEG] (alleging that millions of the newspaper's articles were used to train chatbots to produce competitive content).

31. *See* discussion *infra* Section IV.B. Dan Burk coined the term "cheap creativity" to imply that as GenAI reduces the costs associated with creativity, the economic justification for intellectual property protection is reduced accordingly. Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669, 1673 (2023).

32. *See* discussion *infra* Section IV.B and Section IV.C.

Lastly, Part IV explores the policy implications of our quantification approach. By delineating copyright scope more precisely and predictably, our proposed methodology may benefit not only the copyright system throughout its lifecycle, but also form registration, licensing practices, and infringement litigation procedures.[33]

II. DIVIDING ENTITLEMENTS AMONG AUTHORS

THE THOUGHTS OF EVERY MAN ARE, MORE OR LESS, A COMBINATION OF WHAT OTHER MEN HAVE THOUGHT AND EXPRESSED . . . .

— JOSEPH STORY[34]

Copyright law protects original works created by authors, such as original novels, articles, musical compositions, images, and software.[35] It assigns authors a set of exclusive rights to their respective works of authorship. However, new ideas, impressions, and learning are inherently linked to one another. They are interconnected by shared features and inseparably tied in a culturally expressive ecosystem.

Copyright law is called upon to draw the line between intangible aspects of works which originate with a particular author and those which do not.[36] In a creative ecosystem of interactive exchange where works and authors intermingle, these boundaries are often diffused. The organizing principle of copyright law, which protects authors' claims to entitlements to their expressions, is originality.[37] To be copyrighted, a work must have originated with the person claiming the entitlement.

Yet, for a work to convey meaning, it needs to draw from shared expressive elements that are common within a particular culture. Without such familiar elements, people may struggle to grasp its significance.

This Part examines the tension between originality and the genericity of common expression in copyright law. Section II.A discusses copyright law's objectives and originality's role in serving copyright goals. Section II.B unfolds the inherent tension between originality and ubiquitous expression and underscores the role of genericity in limiting the scope of copyright protections.

---

33. *Cf.* Peter S. Menell & Michael J. Meurer, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1, 12 (2013) (explaining how various social harms flow from copyright scope's vagueness).

34. Emerson v. Davies, 8 F. Cas. 615, 619 (C.C.D. Mass. 1845) (No. 4,436).

35. 17 U.S.C. § 102(a).

36. *See* BOYLE, *supra* note 13.

37. *See* Gideon Parchomovsky & Alex Stein, *Originality*, 95 VA. L. REV. 1505, 1505 (2009).

A. Originality and Copyright Goals

As the Constitution mandates, copyright law seeks "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[38] Accordingly, copyright law incentivizes creation and dissemination of original works by granting authors exclusive rights to their respective works.[39] These rights ensure that authors can commercially exploit their works and sustain incentives to invest in creating future works. However, the "promot[ion]" of "Progress" is inconsistent with granting unlimited rights to control copyrighted materials.[40] Instead, it often requires setting limits on the rights granted to authors.[41] That is because creative processes are situated in cultural contexts which involve interaction with preexisting materials.

The process of generating original works is often nonlinear, opaque, and unintelligible, even to the human author herself. We listen, watch, observe and absorb facts, symbols, images, narratives, ideas,

---

38. U.S. CONST. art. I, § 8, cl. 8.

39. Leval, *supra* note 12, at 1107; *see also* Robert D. Cooter & Uri Y. Hacohen, *Progress in the Useful Arts: Foundations of Patent Law in Growth Economics*, 22 YALE J.L. & TECH. 191, 197 (2020) (exploring how patent law provides economic incentives for cumulative innovation and creativity); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 993 (1997) (analyzing how the primary justification for intellectual property is incentivizing inventions and creations); Menell & Meurer, *supra* note 33, at 26 ("[C]opyrights create temporary, artificial scarcity relating to the use of knowledge for the purpose of promoting progress in technology and expressive creativity."). The Supreme Court has affirmed this goal on many occasions. *See, e.g.*, Mazer v. Stein, 347 U.S. 201, 219 (1954) (explaining how the primary consideration of copyright law is to promote the "production of literary [or artistic] works of lasting benefit to the world" (quoting Washingtonian Publ'g Co. v. Pearson, 306 U.S. 30, 36 (1939))); Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 546 (1984) ("[Copyright] is intended to motivate creativity . . . and to allow the public access to the products of their genius after the limited period of exclusive control has expired." (quoting Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984))).

40. The notion of absolute property rights is conceptually incoherent in general, given property is inherently a social construct. *See* Anna di Robilant & Talha Syed, *Property's Building Blocks: Hohfeld in Europe and Beyond, in* THE LEGACY OF WESLEY HOHFELD: EDITED MAJOR WORKS, SELECT PERSONAL PAPERS, AND ORIGINAL COMMENTARIES 223, 229 (Shyamkrishna Balganesh, Ted M. Sichelman & Henry E. Smith eds., 2022). However, while some private ownership provides a public good by counteracting the "tragedy of the commons," absolute rights are an extremely unattractive goal from a social utilitarian perspective, especially with respect to information goods which benefit from a free flow of content. *See* Julie E. Cohen, Lochner *in Cyberspace: The New Economic Orthodoxy of "Rights Management,"* 97 MICH. L. REV. 462, 498–502 (1998); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1037–38 (2005); Oren Bracha, *Give Us Back Our Tragedy: Nonrivalry in Intellectual Property Law and Policy*, 19 THEORETICAL INQUIRIES L. 633, 669 (2018).

41. di Robilant & Syed, *supra* note 40; *see also* Peter S. Menell, *Intellectual Property and the Property Rights Movement*, 30 REGULATION 36, 40 (2007) ("Congress has limited the duration and the rights of intellectual property owners in significant respects to achieve an appropriate balance between incentives to create and opportunities for later improvers.").

contexts, connections, and links. We often do this subconsciously.[42] Once we have processed such new intakes, we make them "our own." We often feel that ideas are "ours" simply because we comprehend them. As authors, we tend to believe that "our" works are original, new, self-made, and crafted out of thin air. Therefore, we often view those who copy us as exploiters who misappropriate or steal what is rightfully ours.[43] However, when we are the copiers, we often describe our work with different terms. We might describe our own copying as: "I was inspired"; "I had already thought of that idea myself"; "This is so trivial, everyone knows it"; or "Something so basic cannot possibly belong to anyone."

The creative process often involves ongoing interactions with multiple stakeholders rather than siloed authorship. Creating new works involves human capital trained on existing knowledge, engagement with preexisting materials, and sharing a cultural language.[44] A cultural language consists of common words, phrases, slogans, symbols, melodies, stanzas, and images, enabling effective communication within a culture.[45]

Consequently, the use of copyrighted materials is an essential input in any creative process. As Jessica Litman points out:

> [T]he very act of authorship in *any* medium is more akin to translation and recombination than it is to creating Aphrodite from the foam of the sea. Composers recombine sounds they have heard before; playwrights base their characters on bits and pieces drawn from real human beings and other playwrights' characters; novelists draw their plots from lives and other

---

42. Litman, *supra* note 18, at 996–97.

43. *See* Niva Elkin-Koren, *Does Discursive Authorship Justify User Rights?*, 9 JURISPRUDENCE 174, 174 (2017).

44. *See generally* JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE (2012) (introducing a decentered model of creativity, which perceives authors as individuals situated in a cultural context and the creative process as an ongoing process of negotiating meaning); ROSEMARY J. COOMBE, THE CULTURAL LIFE OF INTELLECTUAL PROPERTY: AUTHORSHIP, APPROPRIATION AND THE LAW (1998) (describing the semiotic dimension of culture); Niva Elkin-Koren, *Cyberlaw and Social Change: A Democratic Approach to Copyright Law in Cyberspace*, 14 CARDOZO ARTS & ENT. L.J. 215 (1996) (highlighting the discursive dimension of cultural production, where various social agents engage in meaning-making processes).

45. William Fisher grounds this intuition in what he calls "the cultural theory of copyright law." William W. Fisher III, *Recalibrating Originality*, 54 HOUS. L. REV. 437, 452 (2016); *see also* William W. Fisher III, *When Should We Permit Differential Pricing of Information?*, 55 UCLA L. REV. 1, 33–35 (2007) (advocating for liberalization of copyright law for things essential to the development of culture and individual personhood); William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1752–53 (1988) (explaining how "shared language" opens more opportunities for creative communication among members of a culture).

> plots within their experience; software writers use the logic they find in other software; lawyers transform old arguments to fit new facts; cinematographers, actors, choreographers, architects, and sculptors all engage in the process of adapting, transforming, and recombining what is already 'out there' in some other form. This is not parasitism: it is the essence of authorship.[46]

Copyright law is thus designed to foster the creation of *original works of authorship* by securing incentives to authors and, at the same time, ensuring authors' freedom to use preexisting works in furtherance of the cultural creative cycle.[47]

From an economic perspective, copyright law achieves this balance by weighing the social benefits of restricting access to copyrighted expression (i.e., incentivizing initial creation) against the social costs incurred by such restriction (i.e., frustrating subsequent creation).[48] The greater the ratio of the social benefits relative to the social costs, the more vigorously copyright law will protect expressive compositions of elements, and vice versa.[49]

The social costs associated with protecting some expressive compositions are always prohibitive.[50] For this reason, copyright law never

---

46. Litman, *supra* note 18, at 966–67 (emphasis in original).

47. *See* 44 MARTIN SENFTLEBEN, THE COPYRIGHT/TRADEMARK INTERFACE: HOW THE EXPANSION OF TRADEMARK PROTECTION IS STIFLING CULTURAL CREATIVITY 26–27 (2020). At the same time, however, the challenge of delineating the line between conflicting claims of different authors endures in copyright systems which are rights-based and focused on the fundamental rights of the individual author. *See* ABRAHAM DRASSINOWER, WHAT'S WRONG WITH COPYING? 187–88 (2015).

48. *See* William M. Landes, *Copyright*, *in* A HANDBOOK OF CULTURAL ECONOMICS 132, 132–34 (Ruth Towse ed., 2003) ("[T]he question of how extensive copyright protection should be . . . depends on the costs as well as the benefits of protection."); Jones, *supra* note 18, at 561 ("The justification for protecting expressions but not ideas rests in balancing the interests of society in the free flow of information against the property interests of authors."); Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 580 (1984) ("Congress thus seeks to define the rights included in copyright so as to serve the public welfare and not necessarily so as to maximize an author's control over his or her product."); Sony Corp. of Am. v. Universal Studios, Inc., 464 U.S. 417, 429 (1984) (describing as a "difficult balance between the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce, on the other hand . . .").

49. *See* Landes, *supra* note 48; Lund, *supra* note 49 at 139 ("[L]imiting doctrines such as fair use, statutory licensing, independent creation, and the right of adaptation help exclude copyright protection when the cost to society exceeds the incentive to create . . . .").

50. Copyright law usually does not call expressions that are never protected "expressions." Instead, they are called "ideas." However, this binary distinction is illusory. *See* Jones, *supra* note 18, at 598 ("All expressions in a writing can be placed on a continuum . . . ."); Christopher Buccafusco, *Authorship and the Boundaries of Copyright: Ideas, Expressions, and Functions in Yoga, Choreography, and Other Works*, 39 COLUM. J.L. & ARTS 421, 425

protects "idea[s], procedure[s], process[es], system[s], method[s] of operation, concept[s], principle[s], or discover[ies]."[51] The same is generally true of facts.[52] However, because creative processes are cumulative and situated in a semantic environment (cultural, artistic, linguistic), many expressive compositions' social cost/benefit ratio is dynamic rather than fixed.[53] Thus, expressive compositions worthy of strong legal protection at one point in time may warrant weaker or no legal protection at a later point, and vice versa.[54] This notion is reflected in the limited duration of copyright, which eventually expires, unlike other

---

(2016) ("[T]he idea/expression dichotomy is better understood as a spectrum rather than as a dichotomy."); Lund, *supra* note 49 at 137 (noting that "courts interpreting *Feist*'s application of the merger doctrine have imputed a spectrum of copyright protection"); Landes, *supra* note 48, at 350–51.

51. 17 U.S.C. § 102(b). This principle, known as the idea/expression dichotomy, traces back to the seminal Supreme Court case of *Baker v. Selden*, 101 U.S. 99 (1879). On the importance of this principle, see *Harper & Row*, 471 U.S. at 589; Narell v. Freeman, 872 F.2d 907, 910 (9th Cir. 1989); Warner Bros. v. Am. Broad. Cos., 654 F.2d 204, 208 (2d Cir. 1981); Ashton-Tate Corp. v. Ross, 728 F. Supp. 597, 601 (N.D. Cal. 1989); Edward Samuels, *The Idea-Expression Dichotomy in Copyright Law*, 56 TENN. L. REV. 321, 322 (1989); Andrew F. Christie, *Copyright Protection for Ideas: An Appraisal of the Traditional View*, 10 MONASH U. L. REV. 175, 176 (1984); WILLIAM F. PATRY, LATMAN'S THE COPYRIGHT LAW 30 (6th ed. 1986); MELVILLE NIMMER & DAVID NIMMER, 1 NIMMER ON COPYRIGHT § 2.03[D][1] (1989); David E. Shipley, *Conflicts Between Copyright and the First Amendment After* Harper & Row Publishers, Inc. v. Nation Enterprises, 1986 BYU L. REV. 983, 987 (1986). The dichotomy of idea and expression is also accepted in English law. *See* E.P. SKONE JAMES, JOHN F. MUMMERY, J.E. RAYNER JAMES, ALAN LATMAN & STEPHEN SILMAN, COPINGER AND SKONE JAMES ON COPYRIGHT 175–76 (12th ed. 1980).

52. *See Harper & Row*, 471 U.S. at 547 ("[N]o author may copyright facts or ideas."); NIMMER & NIMMER, *supra* note 51, § 2.11[A]. However, unlike "undisputed facts" as "concept[s], principle[s], and discover[ies]," the so called "created facts" are not excluded from protection ab initio. These types of "facts," like other expressive compositions, adhere to the genericity principle. *Cf.* Justin Hughes, *Created Facts and the Flawed Ontology of Copyright Law*, 83 NOTRE DAME L. REV. 43, 101 (2007) (advocating for an "essential facilities" type doctrine which would render created facts unprotectable when they become widely used and thus essential for subsequent authors); Wendy J. Gordon, *Reality as Artifact: From* Feist *to Fair Use*, LAW & CONTEMP. PROBS., Autumn 1992, at 93, 97 (similar). Genericity also explains why small linguistic features, such as slogans and neologisms, are unprotected ab initio. 37 C.F.R. § 202.1(a) (1992). It's because they immediately become essential for communication, even if they originated with the author. *See* Litman, *supra* note 1819, at 1013. For examples of neologisms, see GYLES BRANDRETH, THE JOY OF LEX: HOW TO HAVE FUN WITH 860,341,500 WORDS 9 (1983) ("aerosol," "automation"); WILLIAM SAFIRE, ON LANGUAGE 9, 282, 289 (2d ed. 1981) ("deplane," "uptight," "pseudoevent"). Consider also "palimony," "yuppie," "significant other," and "Ms. DD."

53. But there are also other views. *See, e.g.*, Brief for Arthur R. Miller as Amicus Curiae Supporting Respondent at 2, Google, LLC, v. Oracle Am., Inc., 593 U.S. 1 (2021) (No. 18-956) ("Petitioner would have this Court declare that use of a copyrighted work is 'necessary' whenever it is popular, and an infringer wants to take advantage of that popularity. That is not the law.").

54. Lund, *supra* note 2349, at 139; *see also* William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 326, 347 (1989).

property rights.[55] However, given the extraordinary length of copy-rights, a composition's cost/benefit ratio may change during the lifespan of individual copyrighted works.[56]

In practice, copyright law tailors the scope of legal protection it affords expressive compositions by their originality, applying numer-ous scope-delineating doctrines. These doctrines are sensitive to the fact that the impact of expressive compositions on culture, language, and function is enhanced as these compositions become more preva-lent. We call this dynamic the genericity principle, which we explore in the next Section.

B. Genericity in Copyright Law

Scholars and practitioners of intellectual property law often asso-ciate genericity with trademark law. Trademark law protects the dis-tinctive power of a symbol to identify the signified product or service. Therefore, in trademark law, the genericity principle dismisses legal protection from marks so commonly used as to become "generic," namely synonymous with a general class of products or services.[57]

In this Section, we explain that the notion of originality, in addition to other scope delineating doctrines in copyright law, also adhere to similar genericity dynamics.[58] The more widely that expressive com-

---

55. 17 U.S.C. § 302. Copyright protection extends for a period of the life of the author plus seventy years. *Id.* § 302(a). The term for anonymous works and works made for hire is 120 years from the date of creation, or ninety-five years from the date of first publication, which-ever expires first. *Id.* § 302(c). Prior to 1998, when the Copyright Act was amended, the term of copyright was the life of the author plus fifty years, or, for anonymous works and works made for hire, 100 years from the date of creation or seventy-five years from the date of first publication. *Id.* § 302(a), (c) (1976).

56. *Cf.* Joseph S. Miller, *Hoisting Originality*, 31 CARDOZO L. REV. 451, 466–67 (2009) (suggesting that the expansion of copyright protections over the years justifies raising the minimum threshold of originality); Cooter & Hacohen, *supra* note 39, at 214 (arguing that for inventions used to innovate, patents should be short to accommodate changes over time).

57. *See, e.g.*, Miller Brewing Co. v. Falstaff Brewing Corp., 655 F.2d 5, 8–9 (1st Cir. 1981) ("Lite" beer); King-Seely Thermos Co. v. Aladdin Indus., Inc., 321 F.2d 577, 581 (2d Cir. 1963) ("Thermos" bottle). For further analysis of the difficulties general names present for trademark protection, see William M. Landes & Richard A. Posner, *Trademark Law: An Eco-nomic Perspective*, 30 J.L. & ECON. 265, 292 (1987) (explaining that protecting generic marks amounts to "language monopoly"). *See generally* J. THOMAS MCCARTHY, TRADEMARKS AND UNFAIR COMPETITION § 12 (2d ed. 1984).

58. Others have observed that there is a common thread underlying seemingly disparate copyright doctrines. *Cf.* Samson Vermont, *The* Sine Qua Non *of Copyright is Uniqueness, Not Originality*, 20 TEXAS INTELL. PROP. L.J. 327, 334–35 (2012) (asserting that uniqueness is copyright's main theory); Robert A. Gorman, *Fact or Fancy? The Implications for Copy-right*, 29 J. COPYRIGHT SOC'Y U.S.A. 560, 560–61 (1982) (asserting a common policy goal of "enriching the public domain" underlying copyrightability, infringement, and defenses of privilege in copyright law); John Shepard Wiley, Jr., *Copyright at the School of Patent*, 58 U. CHI. L. REV. 119, 120 (1991) (suggest that copyright should import lessons from patent law

positions of copyrighted works are embedded in a culture and incorporated into existing expressions, the less vigorously copyright law will likely protect them.[59] Thus, copyright law less zealously protects works that incorporate many "generic" expressive compositions than works that contain few or no such compositions, and vice versa.[60]

In either case, genericity reduces the justification for protecting expressive compositions by changing the ratio of the social costs to benefits attributed to such protection.[61] When expressive compositions turn generic before copyright protection is sought, the justification for denying copyrightability is intuitive. Copyright's *raison d'être* is to enrich the domain of expressive works and thereby improve society's well-

---

on shaping incentives to creation). Peter Menell was the first to invoke the concept of copyright genericity in the context of software applications. *See* Peter S. Menell, *An Analysis of the Scope of Copyright Protection for Application Programs*, 41 STAN. L. REV. 1045, 1066–67, 1101–02 (1989) (analyzing the tension between promoting standardization of and granting legal protections for software interfaces). Others have since applied this genericity dynamic more broadly. *See, e.g.*, Stephen Preonas, *Mergercide, When Good Copyrights Go Bad: A Recommendation for a Market-Based, Defendant-Centric Approach to the Merger Doctrine in the Context of Compilations*, 11 INTELL. PROP. L. BULL. 89, 105–07 (2006) (advocating "mergercide" to invalidate copyright protections for compilations that become industry standard); Timothy S. Teter, *Merger and the Machines: An Analysis of the Pro-Compatibility Trend in Computer Software Copyright Cases*, 45 STAN. L. REV. 1061, 1088–97 (1993) (advocating for a "dynamic merger" to invalidate copyright protections for user interface elements that become de facto standards); *see also* 1 PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § 2.3.2.1, at 2:40.4–41 (3d ed. 2005) (discussing application of merger to de facto standards); Lund, *supra* note 23, at 153 (noting that the "merger doctrine limits copyright over limited means of expression to thin or no protection"); Vermont, *supra* note 58, at 328, 357 (noting that the "dominance principle" limits protection for works whose "social value is due largely to inputs from parties other than the work's author"). But genericity is not the only logic that guides the tailoring of copyright protection. As explained below, ideas, facts, and methods of operation are excluded ab initio regardless of genericity. For a more detailed overview, see Hacohen et al., *supra* note 25.

59. Lund, *supra* note 23, at 153 ("Copyright genericide occurs when diminishing means of expression cause a thin copyright to get thinner, possibly to the point of nonexistence.").

60. The rate of genericity is linked to the level of the expressive compositions' complexity. The more complex expressive compositions are, the longer it takes them to become generic when used. For example, short excerpts from Victor Fleming's famous film *The Wizard of Oz*, such as "Toto, I've a feeling we're not in Kansas anymore," will grow generic faster than Fleming's entire masterpiece. THE WIZARD OF OZ (Metro Goldwyn Mayer 1939). The reason for this is intuitive. Simple expressive patterns — what Justin Hughes calls "microworks" — are likely to be integrated into language or culture more quickly than complex expressive compositions. *See* Hughes, *supra* note 21, at 605 (finding that courts withhold copyright protection from short phrases that "lack the modicum of creativity"); GOLDSTEIN, *supra* note 58, § 2.7.3, at 2:96–96.1 (noting that "the shorter a phrase is, the less likely it is to be original and the more likely it is to constitute an idea rather than an expression"); *see also* sources cited *supra* note 52 and accompanying text on neologisms. Nevertheless, over time the scope of protection for Fleming's entire work will also reduce. *Cf.* Hughes, *supra* note 22, at 778 (suggesting narrowing copyright scope for complex works over time).

61. A cost-benefit analysis is implicit in the antitrust flavor of some of the conceptual configurations of the genericide doctrine. *See, e.g.*, Fisher, *Recalibrating Originality*, *supra* note 45 at 449 (describing originality as "the degree to which the work reflects the exercise of choice by the author"); Vermont, *supra* note 58, at 357 (conceptualizing originality as limited by market dominance). *See generally* Lund, *supra* note 23.

being.[62] Society has no interest in protecting generic expressions that do not affect advancements in art, literature, education, or other creative endeavors.[63] Thus, at least after expressions become generic and before copyright protection is sought, the social benefits associated with protecting those expressions are nil.[64]

Legal protection for nongeneric expressive compositions is justified but only if and to the extent that their nongenericity is preserved. The more widely that expressive compositions are used over time, the more likely they are to become functional or otherwise influence the development of language, culture, and artistic tradition. Zealous protection of expressive compositions in such an environment imposes an increasing toll on future users' ability to pursue interoperability[65] or participate in established expressive discourses.[66] In other words, copyrighted expressions generate network externalities; they become more valuable to users the more they are used.[67]

---

62. *See* sources cited *supra* note 39 and accompanying text.

63. *See* Miller, *supra* note 56, at 464; Parchomovsky & Stein, *supra* note 37, at 1517; Fisher, *Recalibrating Originality*, *supra* note 45 at 454.

64. *See infra* notes 80–81 and accompanying text. Although the threshold for originality is misleadingly low, the actual scope of protection afforded to generic works is thin. *See infra* Section II.B.2.

65. In the seminal case of *Baker v. Selden*, the Supreme Court recognized that once Selden designed his successful bookkeeping system (as explained in his copyrighted book), defendant Baker's design choices for arranging columns and headings to implement that same system were constrained by the choices that Selden had made. *See* 101 U.S. 99, 101 (1879); *see also* Brief of 72 Intellectual Property Scholars as Amici Curiae in Support of Petitioner at 26, Google, LLC, v. Oracle Am., Inc., 141 S. Ct. 1183 (2023) (No. 18-956); LIBR. OF CONG., FINAL REPORT OF THE NATIONAL COMMISSION ON TECHNOLOGICAL USES OF COPYRIGHTED WORKS 20 (1978); Apple Comput., Inc. v. Microsoft Corp., 799 F. Supp. 1006, 1032–35 (N.D. Cal. 1992), *aff'd*, 335 F.3d 1435 (9th Cir. 1994) (reuse of graphical user interface elements excused because they had become industry standards); Lotus Dev. Corp. v. Borland Int'l, Inc., 49 F.3d 807, 816–18 (1st Cir. 1995); GOLDSTEIN, *supra* note 58, § 2.3.2.1 (analogizing copyright's merger doctrine to trademark genericide); Menell, *supra* note 58, at 1101; Teter, *supra* note 58, at 1088.

66. Litman, *supra* note 18, at 999 ("Some authors have suggested that ideas are not protected because of the strictures imposed on copyright by the first amendment.") (citation omitted); *see* David E. Shipley & Jeffrey S. Hay, *Protecting Research: Copyright, Common-Law Alternatives, and Federal Preemption*, 63 N.C. L. REV. 125, 129–51 (1984); Shipley, *supra* note 51, at 987–91; Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 413–14 (1990). For more discussion of the intersection of copyright protection and free speech, see Robert C. Denicola, *Copyright and Free Speech: Constitutional Limitations on the Protection of Expression*, 67 CALIF. L. REV. 283 (1979); Celia Goldwag, *Copyright Infringement and the First Amendment*, 79 COLUM. L. REV. 320 (1979); Melville B. Nimmer, *Does Copyright Abridge the First Amendment Guarantees of Free Speech and Press?*, 17 UCLA L. REV. 1180 (1970); Lionel S. Sobel, *Copyright and the First Amendment: A Gathering Storm?*, 19 COPYRIGHT L. SYMP. 43 (1969); James L. Swanson, *Copyright Versus the First Amendment: Forecasting an End to the Storm*, 7 LOY. ENT. L.J. 263 (1987); Leila N. Sadat-Keeling, Comment, *Copyright and the First Amendment: Where Lies the Public Interest?*, 59 TUL. L. REV. 135 (1984).

67. Glynn S. Lunney, Jr., *Fair Use and Market Failure:* Sony *Revisited*, 82 B.U. L. REV. 975, 1024 (2002) (noting that because "enjoyment of copyrighted works exhibits network externalities . . . private copying is likely to occur disproportionately with respect to the most popular works") (citations omitted)); Vermont, *supra* note 5858, at 357.

Generic expressions are also costlier to administer compared to nongeneric expressions.[68] Because many authors can claim to have excludability rights in generic expressions, the copyright system would need to untangle complex thickets of overlapping claims to assign these rights to their proper owners.[69]

While courts rarely invoke the genericity principle explicitly, this principle is nevertheless baked into copyright law's originality standard and related scope-delineating doctrines.[70]

Below, we divide these doctrines into two categories: (1) eligibility doctrines, which include the originality standard and its limiting doctrines such as idea/expression, merger, useful article, and scènes à faire; and (2) infringement doctrines, which include substantial similarity and fair use. Eligibility doctrines exclude generic expressive compositions from legal protection altogether. Conversely, infringement doctrines prescribe narrower legal protection to copyrighted works incorporating generic or partially generic expressive compositions by excusing a downstream use.[71]

---

68. *See, e.g.*, Lund, *supra* note 2349, at 132 (discussing administrability costs); Parchomovsky & Stein, *supra* note 37, at 1520 (discussing negotiation costs).

69. *See* Litman, *supra* note 1819, at 997. Nevertheless, the genericity principle is only a rough proxy to the network value of expressive patterns. The social benefits from preventing access to some expressive patterns (namely, the need to incentivize their creation) may outweigh the social costs associated with such prevention (namely, the toll imposed on subsequent users), even if these expressive patterns were used expansively. *Id.*; *cf.* Brief for Arthur R. Miller, *supra* note 5356, at 2 (arguing against a genericity logic in copyright law). Moreover, the genericity principle is agnostic as to how copyrighted expressions become generic. *Cf.* Deven R. Desai & Sandra L. Rierson, *Confronting the Genericism Conundrum*, 28 CARDOZO L. REV. 1789, 1826–29 (2007) (invoking such criticism in the trademark context). *See generally* Lund, *supra* note 23 (discussing genericity in the copyright context).49 For example, copyrighted expressions may become generic as a result of permissible licensing schemes that benefit the copyright owners, as well as by a widespread infringing activity that damages them. *See id.* at 147–48, 151–52; *cf.* Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 562–81 (2004) (arguing that unlicensed copying of common cultural expressions is essential for freedom of speech and for promoting copyright objectives). It is unclear whether or not the agnosticism of the genericity principle is socially desirable. On the one hand, it might be sensible to hold copyright owners accountable for their lack of enforcement, which made the genericity of the protected expressions possible in the first place. *See* Lund, *supra* note 23, at 152–53. However, making copyright owners automatically accountable for all forms of genericity might induce them to become overprotective of their rights and reduce public access to their copyrighted works, thwarting the objectives of copyright law. *Cf.* Desai & Rierson, *supra*, at 1839–40 (exploring how genericism doctrine incentivizes trademark holders to aggressively "police" use which negatively affects consumers).

70. *See* sources cited *supra* note 58 and accompanying text.

71. *See, e.g.*, SAS Inst., Inc. v. World Programming, Ltd., 64 F.4th 1319, 1325–26 (Fed. Cir. 2023) ("[C]opyright protection extends only to the expression of an idea, not to the underlying idea itself . . . . Additionally, other doctrines of copyright law detail what elements are not protectable, including scènes à faire elements, material in the public domain, factual material, and elements under the merger doctrine.").

1. Eligibility Doctrines

Expressive compositions are copyright eligible only if they are original and fixed in a tangible medium.[72] Courts interpret originality to require two conditions: independent authorship and a modicum of creativity.[73] Accordingly, facts and discoveries are never eligible for copyright protection because they are not independently authored.[74]

On the other hand, authored original expressions are eligible for copyright protection only if they satisfy a minimal threshold of creativity.[75] This requirement ensures that the expressive compositions originated from the author. A work must be original in that it does not simply reflect widely recognized compositions (thus preventing reasonable attribution to a singular creator) in order to justify restricted use by nonowners.

In the seminal case *Feist v. Rural*, the Supreme Court provided a negative definition of creativity.[76] This definition effectively equates creativity with nongenericity.[77] The Court found that creative expression is *not* an "age-old practice, firmly rooted in tradition and so commonplace that it has come to be expected as a matter of course."[78] The

---

72. 17 U.S.C. § 102(a); *see* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991) (noting that originality is "[t]he *sine qua non* of copyright"); L. Batlin & Son, Inc. v. Snyder, 536 F.2d 486, 489–90 (2d Cir. 1976) 51("'[T]he one pervading . . . prerequisite to copyright protection . . . ' is the requirement of originality — that the work be the original product of the claimant." (quoting 1 M. NIMMER, THE LAW ON COPYRIGHT § 10 (1975))); Litman, *supra* note 1819, at 974 (noting that "originality is a keystone of copyright law").

73. *See Feist*, 499 U.S. at 345.

74. *See id.* at 346. Justice O'Connor stated, for a unanimous court, that facts are categorically excluded from copyright protection because they never originate with the author but are "discovered" rather than "author[ed]." *Id.* at 347. However, this statement is not true with respect to "constructed" or "created" facts, which can be unique and thus original. *See, e.g.*, Litman, *supra* note 1819, at 996 (arguing that facts "do not exist independently of the lenses through which they are viewed"). These "facts" should be treated as regular expressive compositions and adhere to the genericity principle. *See* sources cited *supra* note 52 and accompanying text.

75. *See Feist*, 499 U.S. at 363. It is settled that *Feist* provides a constitutional, not statutory, creativity minimum. *See, e.g.*, Pamela Samuelson, *The Originality Standard for Literary Works Under U.S. Copyright Law*, 42 AM. J. COMPAR. L. 393, 394–95 (1994).

76. *See Feist*, 499 U.S. at 363 ("[T]here is nothing remotely creative about arranging names alphabetically in a white pages directory."); Howard B. Abrams, *Originality and Creativity in Copyright Law*, 55 L. & CONTEMP. PROBS. 3, 15 (1992) (noting that "*Feist* itself does not promulgate a definition or a test for determining creativity"); *see also* Barton Beebe, *Bleistein, the Problem of Aesthetic Progress, and the Making of American Copyright Law*, 117 COLUM. L. REV. 319, 366 (2017).

77. *Cf.* Miller, *supra* note 56, at 481–82 (arguing that *Feist*'s articulation of creativity resonates with patent laws' nonobviousness requirement); Russ VerSteeg & Paul K. Harrington, *Nonobviousness as an Element of Copyrightability? (or, Is the Jewel in the Lotus a Cubic Zirconia?)*, 25 U.C. DAVIS L. REV. 331, 379–81 (1992) (explaining Justice O'Connor's obviousness analysis in *Feist*). Other commenters would have taken this trend further. *See, e.g.*, Parchomovsky & Stein, *supra* note 37, at 1505 (offering to tailor protection to originality more zealously).

78. *Feist*, 499 U.S. at 362–63.

Court held that an alphabetical arrangement of telephone subscribers in a white-page directory is insufficiently creative or "too generic" to merit copyright protection. Indeed, "[t]his time-honored tradition does not possess the minimal creative spark required by the Copyright Act and the Constitution."[79]

Several scholars have argued that *Feist*'s definition of originality is implicitly concerned with the novelty of expressive compositions, not their genericity.[80] Applying a novelty standard in practice would mean that copyright law — like patent law — would evaluate the originality of expressive compositions of copyrighted works at the time of their creation.[81] For example, Joseph Miller suggested measuring the creativity prong of the originality requirement as "the degree to which [a] work moves away from conventional expression for [its] genre at the time the author authors it."[82] Similarly, William Fisher offered to measure whether "the overall impression that the work at issue produces on a reasonable person differs from the overall impression produced on such a person by any other single work that had been publicly available prior to the creation of the work at issue."[83]

However, genericity does not always settle by the time copyrighted works are created.[84] Sometimes, genericity creeps in gradually during the extremely long lifespan of copyrighted works.[85] As Jamie Lund explains, expressive compositions may become generic when overused to

---

79. *Id.*

80. *See, e.g.*, Oren Bracha & Talha Syed, *Beyond the Incentive–Access Paradigm? Product Differentiation & Copyright Revisited*, 92 TEX. L. REV. 1841, 1911 (2014) ("Product differentiation theory, however, turns the spotlight exactly on a substantial novelty requirement as a means for reducing rent dissipation."); Fisher, *Recalibrating Originality*, *supra* note 45, at 439 (noting that *Feist* repudiated the "sweat of the brow" doctrine and required at least a small degree of creativity for a work to be protectable); Miller, *supra* note 56, at 480 (same). The difference between genericity and novelty is that the former is dynamic and reevaluated throughout the life of copyrighted work, whereas the latter is investigated only once — at the time the work is created. Patent law focuses on novelty by measuring departure from the prior art at the time the work was created (this investigation may be done at the patent office or later during infringement litigation). Copyright law is different — it investigates the level of originality of the work at the time that the infringement took place.

81. As emphasized by Lord Jacob in the English patent case *Actavis UK v. Merck & Co.*, "obviousness must be determined as of a particular date . . . . Time can indeed change one's perspective. The perspective the court must bring to bear is that of the skilled man at the priority date and not any earlier time." [2008] RPC 26, ¶ 119 (UK). Indeed, patent law prescribes great importance to the priority date when evaluating novelty and non-obviousness because the consideration of more recent information may lead to hindsight bias and dilute the decision-maker's (i.e., the courts' or the Patent Office's) appreciation of the true value of the invention. *See, e.g.*, Baruch Fischhoff, *For Those Condemned to Study the Past: Heuristics and Biases in Hindsight*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 335, 341 (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982).

82. Miller, *supra* note 56, at 462.

83. Fisher, *Recalibrating Originality*, *supra* note 45, at 464.

84. Lund, *supra* note 23, at 139.

85. *See* sources cited *supra* note 60 and accompanying text.

become stock, standardized, or essential.[86] When this occurs, a host of scope-delineating doctrines such as useful article,[87] merger,[88] and scènes à faire[89] dissipate the legal protection afforded to expressive compositions that turn generic.[90] In other words, works (or portions of works) that were original when created may become generic over time.

As Justice Boudin of the United States Court of Appeals for the First Circuit explained in his concurrence in *Lotus v. Borland*, while software interfaces can be copyrighted ab initio,[91] they might lose their legal protection over time if they become popular enough to be considered standard.[92] Accordingly, the First Circuit held that the command hierarchy of the menu that the plaintiff, a software designer, used for its spreadsheet program, Lotus 1-2-3, was not copyrightable because it

---

86. Lund, *supra* note 23, at 131–32.

87. Copyright protections do not extend to the utilitarian aspects of a work. *See* 17 U.S.C. § 101; *see also* Baker v. Selden, 101 U.S. 99, 104–05 (1879) (refusing to afford copyright protection to a book-keeping system). *See generally* Lloyd L. Weinreb, *Copyright for Functional Expression,* 111 HARV. L. REV. 1149 (1998) (critically exploring the rule that excludes copyright protection from functional elements).

88. If there is essentially only one way to express an idea, the idea and its expression are considered to be merged and are one and the same. *See* Kregos v. Associated Press, 937 F.2d 700, 705 (2d Cir. 1991) (applying merger in the context of athletes' "stat sheets"); N.Y. Mercantile Exch., Inc. v. Intercontinental Exch., Inc., 497 F.3d 109, 116–18 (2d Cir. 2007) (settlement prices); Allen v. Acad. Games League of Am., Inc., 89 F.3d 614, 617–18 (9th Cir. 1996) (rules for academic games).

89. Under the scènes à faire doctrine, expressions indispensable and naturally associated with the treatment of a given idea "are treated like ideas and are therefore not protected by copyright." Apple Comput., Inc. v. Microsoft Corp., 35 F.3d 1435, 1444 (9th Cir. 1994); *see, e.g.*, Schwarz v. Universal Pictures Co., 85 F. Supp. 270, 275 (S.D. Cal. 1945) ("So in all dramatic works we find that situations which are identical call for scenes which are similar."); Paul J. Heald, *Reviving the Rhetoric of the Public Interest: Choir Directors, Copy Machines, and New Arrangements of Public Domain Music*, 46 DUKE L.J. 241, 260 (1996) ("Both the copyright law originality requirement and the patent law nonobviousness requirement focus on whether the derivative work is the result of conventions familiar to creators working in the relevant culture."); Litman, *supra* note 18, at 999 ("The lack of protection given to scènes à faire seems to lie more in their triteness than their necessity.").

90. Lund, *supra* note 23, at 132.

91. Lotus Dev. Corp. v. Borland Int'l, Inc., 49 F.3d 807, 819 (1st Cir. 1995) (Boudin, J., concurring). Indeed, several courts have found that computer programs are copyrightable, as long as they meet the other requirements of the Copyright Act. *See* Apple Comput., Inc. v. Franklin Comput. Corp., 714 F.2d 1240, 1247–49 (3d Cir. 1983) (stating that source and object code are copyrightable); Stern Elecs., Inc. v. Kaufman, 669 F.2d 852, 855 n.3 (2d Cir. 1982); Digit. Commc'ns Assocs., Inc. v. Softklone Distrib. Corp., 659 F. Supp. 449, 454 (N.D. Ga. 1987).

92. *Lotus*, 49 F.3d at 819–20 (Boudin, J., concurring). Judge Boudin nicely illustrated the point by using the QWERTY keyboard as an example. In his words: "Better typewriter keyboard layouts may exist, but the familiar QWERTY keyboard dominates the market because that is what everyone has learned to use." *Id.*; *see also* Lund, *supra* note 23, at 143; Menell, *supra* note 58, at 1066–67, 1101. *See generally* Paul A. David, *Clio and the Economics of QWERTY*, 75 AM. ECON. REV. 332 (1985) (explaining the network effects in the case of the QWERTY system).

had *become* a method of operation, even though it may not be considered as such ab initio.[93]

Similar dynamics apply to mundane statements,[94] ingrained cultural themes,[95] stock characters,[96] acknowledged artistic styles,[97] and common harmonic progressions[98] or chord arrangements.[99] For example, when it first appeared in the 1868 play *Under the Gaslight*, Augustin Daly's damsel in distress tied to a train track scene was the epitome of artistic creativity.[100] When the British playwright Dion Boucicault featured a similar scene a year later in his play *After Dark*, Daly sued for copyright infringement and won with a bang.[101]

Under today's standards, however, courts would consider popular scenes like Daly's railroad scene as unprotected scènes à faire, "incidents, characters or settings which are as a practical matter indispensable, or at least standard, in the treatment of a given topic."[102]

Similarly, in *Acuff-Rose v. Jostens*,[103] the United States District Court for the Southern District of New York employed a genericity logic in denying copyright protection for the phrase "You've Got to Stand for Something." The copyright owner of Aaron Tippin's famous cowboy song, which includes the lyric "[y]ou've got to stand for something, or you'll fall for anything," sued the defendant for using a similar

---

93. The *Lotus* majority opinion did not subscribe to the concurrence and believed the interfaces were unprotected ab initio. 49 F.3d at 815.

94. *See* Acuff-Rose Music, Inc. v. Jostens Inc., 988 F. Supp. 289, 296 (S.D.N.Y. 1997) (finding that the statement "[y]ou've got to stand for something" is an unprotected cliché that belongs to the public domain); Matthews v. Freedman, 157 F.3d 25 (1st Cir. 1998) (finding that protection granted for trivial compilation of ideas, expressions, and images is thin and available only against complete literal copying); Litman, *supra* note 18, at 996–97.

95. *See, e.g.*, Nichols v. Universal Pictures Corp., 45 F.2d 119, 122 (2d Cir. 1930) (refusing to grant protection for general cultural narratives such as a quarrel between interracial married families).

96. *See* Walker v. Time Life Films, Inc., 784 F.2d 44, 50 (2d Cir. 1986) ("Elements such as drunks, prostitutes, vermin and derelict cars would appear in any realistic work about the work of policemen in the South Bronx. These similarities therefore are unprotectible as 'scènes à faire,' that is, scenes that necessarily result from the choice of a setting or situation.").

97. *See* Franklin Mint Corp. v. Nat'l Wildlife Art Exch., 575 F.2d 62, 67 (3d Cir. 1978) ("[W]e conclude that while the ideas are similar, the expressions are not. A pattern of differences is sufficient to establish a diversity of expression rather than only an echo.").

98. *See* Johnson v. Gordon, 409 F.3d 12, 23 (1st Cir. 2005) ("[T]his harmonic progression, which is a stereotypical building block of musical composition, lacks originality. Accordingly, it is unprotectable." (citations omitted)).

99. *See* Woods v. Bourne Co., 841 F. Supp. 118, 121 (S.D.N.Y. 1994), *aff'd in relevant part*, 60 F.3d 978, 991–93 (2d Cir. 1995) (noting that copyright protection requires that "there must be present more than mere cocktail pianist variations of the piece that are standard fare in the music trade by any competent musician").

100. Daly v. Palmer, 6 F. Cas. 1132, 1132 (C.C.S.D.N.Y. 1868) (No. 3,552).

101. The Circuit Court of the Southern District of New York reasoned that Boucicault's railroad scene "contains everything which makes the 'railroad scene' in the plaintiff's play attractive, as a representation on the stage." *Id.* at 1138.

102. Alexander v. Haley, 460 F. Supp. 40, 45 (S.D.N.Y. 1978).

103. Acuff-Rose Music, Inc. v. Jostens, Inc., 988 F. Supp. 289 (S.D.N.Y. 1997).

slogan, "U Got 2 Stand 4 Something," in a promotional campaign for selling class rings.[104]

While the district court acknowledged that the defendant indeed copied the expression from the plaintiff's song, it denied that copyright infringement occurred.[105] This is because the contested phrase had become a "cliché" — it was commonly used in sermons, political speeches, and newspaper articles that bore no connection to Tippin's song.[106]

Lastly, consider generic expressive compositions in music. Although there are infinite pitches, traditional Western music generally uses no more than twelve and often only seven pitches at a time.[107] As wittily captured in The Axis of Awesome's "Four Chords" medley, most commercially viable songs rely on standardized chord relationships and progressions.[108] For example, the popular I-V-vi-IV chord progression has been used, by some estimates, in thousands of songs, including Bob Marley's "No Woman, No Cry" and Green Day's "When I Come Around."[109] Like Daly's railroad scene, most courts consider these chord arrangements generic and thus copyright ineligible.[110]

In foreign jurisdictions, copyright law's originality standard and related eligibility doctrines are even more explicit in their adherence to the genericity principle.[111] Austrian copyright law, for example, affords photographs legal protection only if they differ significantly from preexisting photographs.[112] Similarly, until recently, copyright law in Germany granted legal protection to computer software only to the ex-

---

104. *Id.* at 290–92.

105. *Id.* at 293–94.

106. *Id.* at 294.

107. Although there are twelve distinct pitches (all the black and white keys on the piano in between two of the same notes), the major and minor scales, which are by far the most popular musical structures, only use seven of those notes at a time. *Cf.* Ronald P. Smith, *Arrangements and Editions of Public Domain Music: Originality in a Finite System*, 34 CASE W. RES. L. REV. 104, 104 (1983) (assuming that there are only twelve pitches involved in music protected by copyright).

108. THE AXIS OF AWESOME, *Four Chords*, *on* ANIMAL VEHICLE (2011).

109. Andrew Pouska, *The I-V-vi-IV Chord Progression*, STUDYBASS, https://www.study-bass.com/lessons/harmony/the-I-V-vi-IV-chord-progression/ [https://perma.cc/ZW79-GDHB].

110. *See supra* notes 99–101 and accompanying text.

111. Fisher, *Recalibrating Originality*, *supra* note 45, at 439.

112. *Id.*; Roman Heidinger, The Threshold of Originality Under EU Copyright Law, Presentation at the Hong Kong Workshop, Co-Reach Intellectual Property Rights in the New Media (Oct. 20, 2011), https://www.law.cuhk.edu.hk/en/research/cfred/download/CFRED_COREACH_IP_Workshop_Dr_Roman_Heidinger_20Oct2011.pdf [https://perma.cc/3BE8-LMFV].

tent that it embodied different expression from that of average computer programs.[113] Most notably, Swiss copyright law prescribes legal protection only to works that show "statistical uniqueness" compared to preexisting works.[114]

Last year, a Singaporean court attracted significant public interest after zealously applying the genericity approach. In that case, the court denied copyright protection for a photograph taken by the professional photographer Jingna Zhang for a 2017 cover shoot of *Harper's Bazaar Vietnam* magazine.[115] Although the photograph was detailed and highly expressive, the court denied legal protection because it considered the photographed model's pose and appearance widely common and, thereby, unoriginal.

2. Infringement Doctrines

Because the level of originality required for copyright eligibility is low, most works will pass this threshold even if they incorporate many generic expressive compositions.[116] For the reasons discussed above, works of partial expressive genericity merit only partial legal protection. Two copyright doctrines achieve this goal by limiting the scope of legal protection for partially generic works during infringement litigation. These doctrines are substantial similarity and fair use.

To succeed in a copyright infringement claim under substantial similarity, a plaintiff must establish ownership of a valid copyright and

---

113. Fisher, *Recalibrating Originality*, *supra* note 45, at 439; Bundesgerichtshof [BGH] [Federal Court of Justice] Sept. 5, 1985, 94 ENTSCHEIDUNGEN DES BUNDESGERICHTSHOFES IN ZIVILSACHEN [BGHZ] 276. For comments on *Inkasso* and comparison of the German and French positions with respect to originality in software, see *Commission Green Paper on Copyright and the Challenge of Technology*, at 187–88, COM (1988) 172 final (June 7, 1988). Today, a larger set of software programs are now eligible for copyright protection; the only things now excluded are simple, routine programs that ordinary programmers would write the same way. Bundesgerichtshof [BGH] [Federal Court of Justice] Oct. 6, 2005, 9 ENTSCHEIDUNGEN DES BUNDESGERICHTSHOFES IN ZIVILSACHEN [BGHZ] 111/02.

114. Fisher, *Recalibrating Originality*, *supra* note 45, at 448 (translated from "statistische Einmaligkeit"); Bundesgericht [BGer] [Federal Supreme Court] Apr. 1, 2010, 136 ENTSCHEIDUNGEN DES SCHWEIZERISCHEN BUNDESGERICHTS [BGE] III 225. The courts' stance on this issue derives at least in part from the unusual way in which the Swiss Federal Copyright Act of 1992 defined the set of works subject to copyright protection as "intellectual creations with an individual character." RECUEIL SYSTÉMATIQUE DU DROIT FÉDÉRAL [RS] [Systematic Compilation of Federal Law] Oct. 9, 1992, RS 231.1, art. 2(1).

115. Jacqueline Tobin, *Photographer Jingna Zhang Loses Plagiarism Case Against Artist*, RF + WPPI (Dec. 9, 2022), https://www.rangefinderonline.com/news-features/industry-news/photographer-jingna-zhang-loses-plagiarism-case-against-artist/ [https://perma.cc/J9NE-8X5M]; Espie Angelica de Leon, *Singaporean Photographer Loses in Luxembourg Plagiarism Trial Against Local Painter*, ASIA IP (Dec. 23, 2022), https://asiaiplaw.com/section/news-analysis/singaporean-photographer-loses-in-luxembourg-plagiarism-trial-against-local-painter [https://perma.cc/7NHJ-XELY].

116. Or other unprotected elements such as facts. *See* sources cited *supra* note 52 and accompanying text.

unauthorized copying of protected expression.[117] Because direct evidence of copying is rare, copying can be proven circumstantially by demonstrating access to the allegedly infringed copyrighted work plus substantial similarity of protected expression.[118]

Courts distill protected expression from the unprotected chaff by engaging in "analytic dissection," breaking the copyrighted work into its constituent parts and then evaluating the copyrightability of each of those parts.[119] The most sophisticated framework for executing this task was established in *Computer Associations International v. Altai* for computer programs.[120] According to the *Altai* framework, courts separate the copyrighted work into different levels of abstraction. Then, they filter out the non-original features and compare the remaining "golden nuggets" of protected expression — some more original than others — to the allegedly infringing work. A plaintiff can successfully prove copyright infringement only if she demonstrates substantial copying of protectable expression.[121]

When evaluating how much similarity is considered "substantial" in the final step of this analysis, courts sometimes consider the copyrighted work's originality (nongenericity) level.[122] The less original the

---

117. *See, e.g.*, Sid & Marty Krofft Television Prod., Inc. v. McDonald's Corp., 562 F.2d 1157, 1164 (9th Cir. 1977) ("There must be ownership of the copyright and access to the copyrighted work. But there also must be substantial similarity not only of the general ideas but of the expressions of those ideas as well.").

118. *See id.* at 1164; *see also* Brown Bag Software v. Symantec Corp., 960 F.2d 1465, 1476 (9th Cir. 1992) ("[T]he source of the similarity must be identified and a determination made as to whether this source is covered by plaintiff's copyright.").

119. Different circuits appraise substantial similarity differently. A complete consideration of the different approaches exceeds the scope of this Article. For an overview, see Lemley, *supra* note 18.

120. *See* Comput. Ass'n Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 706–10 (2d Cir. 1992). This framework was initially created for copyrighted computer programs. Most recently it was applied in *SAS Institute v. World Programming, Ltd.*, 64 F.4th 1319 (Fed. Cir. 2023), by the Federal Circuit in 2023. *Id.* at 1337.

121. Genericity plays a part in this filtration stage because some expressive elements that were nongeneric (i.e., original) at the time the copyrighted work was first created may become generic (i.e., nonoriginal) by the time of the infringement analysis. However, the current filtration practice is still binary and, therefore, limited. *See* sources cited *supra* note 60 and accompanying text.

122. *See* Mark A. Lemley, *Convergence in the Law of Software Copyright*, 10 HIGH TECH. L.J. 1, 28 (1995) (acknowledging the sliding scale between copyright scope and similarity but also noting that "[n]one of these gradations are captured by the *Altai* filtration test. This is not because the filtration test is flawed, but because the test is largely directed at identifying protectable expression (the copyrightability portion of the infringement analysis), not at comparing a copyrighted work to an accused work."); Apple Comput., Inc. v. Microsoft Corp., 35 F.3d 1435, 1443 (9th Cir. 1994) ("[T]he court must define the scope of the plaintiff's copyright — that is, decide whether the work is entitled to 'broad' or 'thin' protection."); *see also* Lund, *supra* note 23, at 147–48. *But see* 3 WILLIAM F. PATRY, PATRY ON COPYRIGHT § 9:166 n.9 (2007) (noting that "regardless of the relative creativity of the work, the test for all works is substantial similarity"; but, "[w]hile works having a 'thin' copyright due to a minimal amount of creative material may indeed only be infringed by close copying, this is because the majority of the work is unprotectible").

work — measured by the original vs. unoriginal (filtered) compositions ratio or by the originality level of the remaining compositions — the higher the degree of similarity that courts require for establishing infringement.[123] This similarity gradient effectively narrows the scope of protection for partially copyrighted generic works.[124] Works heavy with generic material will be infringed only if the allegedly infringing work is "virtually identical" to them.[125] Courts have indeed applied this more stringent similarity standard in cases involving computer operation systems with typical graphic user interfaces,[126] conventional photographs of commercial products,[127] and standardly structured stand-up jokes.[128]

The fair use doctrine also tailors copyright scope for partially generic works.[129] Even if a plaintiff successfully proves substantial copying of protectable expressive compositions, a defendant may still escape infringement if her copying constitutes fair use.[130] Courts may find fair use if the expressive compositions a defendant copied from the

---

123. *See* Lemley, *supra* note 122, at 31 (discussing the interaction of fair use with substantial similarity); Idema v. Dreamworks, Inc., 162 F. Supp. 2d 1129, 1178 (C.D. Cal. 2001) (explaining that stingier protection applies for copyrighted works of "largely 'unprotectable' elements").

124. Not all courts prescribe to this gradient-of-protection analysis. But even adherents of the more conservative approach to tailoring legal protection during the infringement analysis may implicitly consider genericity during the filtration stage. *See* PATRY, *supra* note 122, § 9:166 n.9. Original compositions of generic elements are protectable, but they too undergo a genericity process. In a sense, any work can be thought of as a compilation of smaller parts. The more complex the work is, the slower the genericity process. *See* sources cited *supra* note 60 and accompanying text.

125. *Apple Computer*, 35 F.3d at 1442. *See generally* Lemley, *supra* note 18 (arguing for a substantial similarity test based on filtration and dissection of unprotectable elements).

126. *Apple Computer*, 35 F.3d at 1443, 1445.

127. Ets-Hokin v. Skyy Spirits, Inc., 323 F.3d 763, 766 (9th Cir. 2003) (noting the limited "conventions of [a] commercial product shot").

128. Kaseberg v. Conaco, LLC, 260 F. Supp. 3d 1229, 1245 (S.D. Cal. 2017) (acknowledging that one of the limiting features of the jokes was the necessity to be structured in a way that would provide "mass appeal"); *see also* Matthew L. Pangle, *The Last Laugh: A Case Study in Copyright of Comedy and the Virtual Identity Standard*, 28 TEX. INTELL. PROP. L.J. 183, 201 (2020) ("The joke structure at issue relied on social commentary of factual material, which necessarily limited the number of variations possible to achieve humor.").

129. Copyright law allows copying of protected expression if it qualifies as fair use. The U.S. fair use doctrine, as codified in § 107 of the U.S. Copyright Act of 1976, is yet another legal standard to carve out an exception for an otherwise infringing use after weighing a set of four statutory factors. The four statutory factors are: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work. 17 U.S.C. § 107(1)–(4). The primary factor that considers the level of the underlying work's originality (or nongenericity) is the second factor (the "nature of the copyrighted work"). *Id.* § 107(2).

130. Bracha & Syed, *supra* note 80, at 1894–95.

plaintiff are considered at the turning point of genericity.[131] This was the case in *Google v. Oracle*.[132]

There, Oracle claimed that Google infringed upon its copyrights in the Java program by copying thirty-seven "declaring code" packages, which was part of the Java application program interfaces ("APIs"), without authorization.[133] Several amici urged the Court to recognize that the Java declaring code was generic by the time of the litigation and that, accordingly, it should not be copyright protected.[134] This argument was similar to Judge Boudin's concurrence in Lotus v. Borland.[135]

The Supreme Court disagreed. Instead, the Court assumed, arguendo, that the Java declaring code was original enough to merit copyright protection.[136] Then, the Court assessed whether the legal protection afforded to the Java declaring code was broad enough to enable the exclusion of Google's unauthorized copying.[137] The Court held that it was not; therefore, Google's copying was fair.

In reaching that conclusion, the Supreme Court relied on the fact that Java's declaring code was becoming generic at the time that the action was filed due to its established popularity among software programmers.[138] Here, genericity emphasized the functional importance of the declaring code. According to the Court, this popularity elevated the declaring code to the level of a functional interface because programmers had no real choice but to use the same code if they wished to participate in the established programming ecosystem.[139] For this reason, the Court decided that weakening Oracle's exclusivity rights and allowing Google to use the Java declaring code was socially beneficial

---

131. Lemley, *supra* note 122, at 30.

132. Google, LLC, v. Oracle Am., Inc., 141 S. Ct. 1183 (2021).

133. *Id*. at 1194.

134. *See, e.g.*, Brief for Electronic Frontier Foundation as Amicus Curiae Supporting Petitioner at 24, Google, LLC, v. Oracle Am., Inc., 141 S. Ct. 1183 (2021) (No. 18-956) (arguing that affording copyright protection to the standardized Java API declarations would effectively require millions of software developers "to learn a whole new language to be able to use" the API labels (quoting Google's expert Owen Astrachan)); Brief for Peter S. Menell & David Nimmer as Amici Curiae Supporting Petitioner at 16, *Google*, 141 S. Ct. 1183 (No. 18-956) (similar).

135. However, Judge Boudin's concurrence endorsed a fair use approach similar to the one the Court took here. *See* Lotus Dev. Corp. v. Borland Int'l, Inc., 49 F.3d 807, 821 (9th Cir. 1994) (Boudin, J., concurring).

136. *Google*, 141 S. Ct. at 1190 ("[W]e assume, for argument's sake, that the material was copyrightable.").

137. *Id.* at 1197–99.

138. *Id.* at 1202 ("Unlike many other programs, its value in significant part derives from the value that those who do not hold copyrights, namely, computer programmers, invest of their own time and effort to learn the API's system.").

139. *Id.* at 1205 (noting that Google "copied [the Java declaring code] because programmers had already learned to work with the Sun Java API's system, and it would have been difficult, perhaps prohibitively so, to attract programmers to build its Android smartphone system without them.").

enough to outweigh the benefits of broad copyright exclusion in this context.[140]

III. LEVERAGING GENAI TO QUANTIFY COPYRIGHT ORIGINALITY

> THE COPYRIGHT IS LIMITED TO THOSE ASPECTS OF THE WORK — TERMED 'EXPRESSION' — THAT DISPLAY THE STAMP OF THE AUTHOR'S ORIGINALITY.
>
> — SANDRA DAY O'CONNOR[141]

As demonstrated in Part II, genericity is an implicit principle of copyright law. Accordingly, pervasive expressive compositions are less likely to be vigorously protected by copyright law. Until recently, measuring the genericity of expressive compositions was not feasible. Although courts consider the prevalence of expressive compositions in copyright disputes, this assessment was done, thus far, in a rather loose, intuitive, and imprecise manner.[142]

With the rise of GenAI models, the prevalence of expressive compositions in a given cultural domain could be measured far more rigorously. GenAI models can extract complex probability distributions by extrapolating reoccurring relationships between input and output values in training samples.[143] GenAI models seek to generalize expressive pat-

---

140. *Id.* at 1202 (reasoning that Oracle's "declaring code is, if copyrightable at all, further than are most computer programs (such as the implementing code) from the core of copyright"). Note that the level of originality of the allegedly infringing work is also considered in a fair use analysis. 17 U.S.C. § 107(2); *see also* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578–79 (1994) (asking whether a use was transformative as part of determining fair use); Bill Graham Archives v. Dorling Kindersley, Ltd., 448 F.3d 605, 608 (2d Cir. 2006) (analyzing whether or not a use was transformative); Kelly v. Arriba Soft Corp., 336 F.3d 811, 818 (9th Cir. 2003) (noting that "the more transformative a use is, the less important other factors are in determining fair use"); Suntrust Bank v. Houghton Mifflin Co., 268 F.3d 1257, 1269 (11th Cir. 2001) (noting the importance of transformative value in determining fair use); Leval, *supra* note 12, at 1111 (asserting that justification of a fair use turns largely on whether the use is transformative). Indeed, if Google was copying Oracle's code verbatim to create a mere imitation rather than a transformative new ecosystem, the Court would likely have found that Oracle's appropriation rights were strong enough to exclude Google's use. *Cf.* Cooter & Hacohen, *supra* note 39, at 202 (explaining how patent policy weakens patent-holders' rights against innovative uses but strengthens these rights against consumption and production uses).

141. Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 547 (1985).

142. For example, when the court in *Acuff-Rose Music v. Jostens*, 988 F. Supp. 289 (S.D.N.Y. 1997), found that a statement had become cliché, they used anecdotal evidence but did not engage in a comprehensive empiric inquiry into the history and evolution of the phrase. *See id.* at 291; *see also* sources cited *supra* note 18.

143. For a deeper exploration of the concepts of generalization and memorization, see Dan Ventura, *Mere Generation: Essential Barometer or Dated Concept?*, 7TH PROC. INT'L CONF. ON COMPUTATIONAL CREATIVITY, June 2016, at 17, 19–20.

terns from examples and apply these insights to tasks such as autocompleting sentences or visual images and generating visual outputs in response to a textual prompt.[144]

Generalization by GenAI models, by indicating the expressive compositions that are prevalent in preexisting works, could be helpful in measuring these compositions' genericity. Such measurements could inform legal analysis when prescribing rights to copyrighted works and when enforcing those rights against allegedly infringing uses.

Section III.A explores the rise of GenAI technology. Section III.B provides a brief overview of the disruptive effect of GenAI on copyright law and the shortcomings of current proposals to address this disruption. Lastly, Section III.C explores how GenAI models can be harnessed to measure the genericity of expressive compositions and generate originality scores to copyrighted works.

A. The Rise of GenAI

GenAI models are rapidly expanding in popularity and reach. They can generate copyrightable materials such as text, image, music, and code based on expressive input.[145] Some images and text generated with the aid of GenAI models even won prestigious awards.[146] These systems piggyback on the impressive capability of foundation models, such as OpenAI's Generative Pre-trained Transformer ("GPT") or Google's Bidirectional Encoder Representations from Transformers ("BERT"), to extrapolate patterns and structures from granular data.[147] Foundation models are large-scale neural network models pre-trained

---

144. Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee et al., *Extracting Training Data from Large Language Models*, 30 USENIX Ass'n 2633–50 (2021) ("The appeal of generative diffusion models is rooted in their ability to synthesize novel images that are ostensibly unlike anything in the training set.").

145. *See* Bracha, *supra* note 11, at 10.

146. Boris Eldagsen, *I Used AI to Beat Real Photos in a Major Competition*, Newsweek (Apr. 26, 2023, 8:30 AM), https://www.newsweek.com/ai-photography-contest-sony-art-1796455 [https://perma.cc/3WX7-CTA6].

147. *See generally* Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora & Sydney von Arx, On the Opportunities and Risks of Foundation Models 5 (Aug. 16, 2021) (unpublished manuscript), https://arxiv.org/abs/2108.07258 [https://perma.cc/B3WT-3B2K] ("Foundation models have led to an unprecedented level of homogenization: Almost all state-of-the-art NLP [natural language processing] models are now adapted from one of a few foundation models, such as BERT, RoBERTa, BART, T5, etc."); *see also* Uri Hacohen, Adi Haviv, Shahar Sarfaty, Bruria Friedman, Niva Elkin-Koren, Roi Livni et al., Not All Similarities Are Created Equal: Leveraging Data-Driven Biases to Inform GenAI Copyright Disputes 1–2 (Mar. 26, 2024) (unpublished manuscript), https://arxiv.org/pdf/2403.17691.pdf [https://perma.cc/XD7H-MPWS].

on colossal amounts of unlabeled data by self-supervised learning, often on surrogate tasks.[148] These models learn generalizable and adaptable data representations that can be used for multiple downstream tasks.

In addition, the models' ability to transfer learning, or take the "knowledge" learned from one task and apply it to another, cuts the high cost of training a model from scratch.[149] Adapting to new tasks sometimes involves techniques like fine-tuning (training a foundation model on a much smaller and specific new dataset related to the task) or prompt engineering.[150] ChatGPT, for instance, is built on OpenAI's foundation models GPT3.5 and GPT4 to enable bot-human interaction. ChatGPT could be fine-tuned further for more nuanced natural language processing tasks, such as language translation, classification, and text summarization.

While much of the hype associated with generative models is currently focused on Large Language Models ("LLMs"), computer vision has been at the forefront of deep learning research in AI for many years.[151] As with text, foundation models in computer vision are pretrained on raw data of different types (e.g., text, audio, colors) from diverse sources (e.g., uploaded from sensors, cameras, or scraped from the Internet).[152] By extracting visual knowledge from such data, image models have demonstrated the ability to transfer learning to new tasks such as image classification and object detection.[153]

Generative models vary in their technical approach, including Generative Adversarial Networks ("GANs"), Variational Auto Encoders ("VAEs"), and diffusion models (e.g., DALL-E 2 and Stable Diffusion)

---

148. The practice that allows models to learn pretraining tasks automatically from their training data is called self-supervised learning. "Interestingly, self-supervised learning was dominant in the early days of deep learning, but was for a decade largely overtaken by pure supervised learning as labeled datasets became larger." *Id.* at 4 n.3. "Self-supervised tasks are not only more scalable, only depending on unlabeled data, but they are designed to force the model to predict parts of the inputs, making them richer and potentially more useful than models trained on a more limited label space." *Id.* at 4.

149. *Id.*

150. Bommasani et al., *supra* note 147, at 3 ("A foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks . . . .").

151. *Id.* at 14 ("Computer vision led the adoption of deep learning in AI . . . ."); *see also* Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma et al., ImageNet Large Scale Visual Recognition Challenge (Jan. 30, 2015) (unpublished manuscript), https://arxiv.org/pdf/1409.0575.pdf [https://perma.cc/N85T-E4ZK].

152. *See, e.g.*, Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator*, WAXY (Aug. 30, 2022), https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/ [https://perma.cc/9G9M-DHB6] (explaining how Stable Diffusion collected and trained upon images scraped from webpages across the Internet).

153. Bommasani et al., *supra* note 147, at 14.

which raised recent controversies in copyright law.[154] Diffusion models, which underlie some of the most popular foundation models, aim to learn the latent structure of a dataset by modeling how data points diffuse through a latent space.[155] These models have both a forward process and a reverse process. The forward diffusion process gradually adds noise of different magnitudes (e.g., random pixels or distortion changes that affect the original image) to a clean data point.[156] In the reverse process, the model learns to remove the added noise to generate a clear image.[157] Figure 5 in Section III.C below visualizes the latter process.

Foundation models were initially aimed at learning about data without supervision[158] but were increasingly deployed for generative applications.[159] Downstream diffusion applications, such as Midjourney and Stable Diffusion, which have been the focus of recent copyright disputes, involve text-to-image tools for creating and editing visual works. These applications enable users to generate original expressive outputs.[160] Like the foundation models on which they are based, diffusion applications do not rely on formal instruction to generate original outputs. Instead, these models' learning is extrapolated from their preexisting training examples.

The value of GenAI models is in their ability to simplify how humans create and interact with preexisting expressive materials. Humans have always engaged in creative processes within a cultural context.[161] We derive inspiration from preexisting genres; we implicitly reference

---

154. *See, e.g.*, Complaint & Demand for Jury Trial, *supra* note 9. *See generally* Viet Anh, *An Overview of Generative AI in 2023*, AI CURIOUS (May 1, 2023), https://aicurious.io/notes/2023-05-02-overview-of-generative-ai [https://perma.cc/NP69-KG8C] (detailing the development of different types of GenAI models).

155. Hacohen et al., *supra* note 147, at 5–6 (explaining the diffusion process); *see also* Kailash Ahirwar, *A Very Short Introduction to Diffusion Models*, MEDIUM (Sept. 26, 2023), https://kailashahirwar.medium.com/a-very-short-introduction-to-diffusion-models-a84235e4e9ae [https://perma.cc/96BT-JYE4].

156. Hacohen et al*., supra* note 147, at 5–6.

157. *See* Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum & Jonas Geiping, Universal Guidance for Diffusion Models 2 (Feb. 14, 2023) (unpublished manuscript), https://arxiv.org/abs/2302.07121 [https://perma.cc/AG5N-JQJL].

158. Sam Bond-Taylor, Adam Leach, Yang Long & Chris G. Willcocks, *Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models*, 44 IEEE TRANSCON. PATTERN ANALYSIS & MACH. INTEL. 7327, 7327 (2022).

159. *Id.* ("The central idea of generative modelling stems around training a generative model whose samples $\tilde{x} \sim p\theta(\tilde{x})$ come from the same distribution as the training data distribution, $x \sim pd(x)$.").

160. *See* Bracha, *supra* note 11, at 10 ("The main purpose of GenAI, by contrast, is to generate new informational goods . . . ."). For a technical definition, *see* Bond-Taylor et al., *supra* note 158.

161. *See* JULIE E. COHEN, CONFIGURING THE NETWORKED SELF 81–83 (2012); DRASSINOWER, *supra* note 47, at 76 (arguing that authorship does not occur in isolation, but rather through a discursive creative process, where authors constantly interact and exchange ideas).

prior works and use shared cultural components. Like other modes of communication, human creations are intended to communicate discursive meanings to prospective audiences. Therefore, learning from previous works has been not only necessary for acquiring expressive skills, but also essential to the creation process itself.[162]

Before GenAI technology, human creators interacted with preexisting expressive content primarily through memory, skill, and instruction. For example, when an artist experimented with an impressionist style, she had to observe impressionist paintings. She had to extract rules regarding the painting technique, the configuration of figures, the depiction of light, or the composition. Then she could apply these rules to her own painting. Today, these processes are increasingly mediated by GenAI models trained on massive corpora of preexisting expressive content from various domains.[163] For example, Copilot and Midjourney are trained on giant corpora of prewritten code and images from the GitHub open source code repository and the LAION 5B database, respectively.[164] These models assist users in generating code and images in response to "prompts," namely textual inputs. These tools augment artists' creativity and democratize the creative profession.[165]

B. GenAI and Copyright Disruption

Since the 1990s, the Internet and digital distribution have radically decentralized the power to reproduce and disseminate copyrighted materials. Once exclusive to publishers, the ability to copy and distribute works on a large scale now rests in the hands of any individual connected to the Internet.[166] Today, GenAI systems empower users to generate high-quality text, image designs, music, and code, challenging the dominance of professional creatives.

---

162. *See* Jessica Litman, *Readers' Copyright*, 58 J. COPYRIGHT SOC'Y U.S.A. 325, 341 (2010); Tushnet, *supra* note 69, at 571–72; Julie E. Cohen, *The Place of the User in Copyright Law*, 74 FORDHAM L. REV. 347, 362 (2005).

163. *See* Bommasani et al., *supra* note 147, at 3–4; *see also* Bracha, *supra* note 11, at 2.

164. *See* 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/Z6M7-2D2Z] (explaining the Laion-5B project); *see also* Doe 1 v. GitHub, Inc., No. 22-CV-06823, 2023 WL 3449131, at *2–3 (N.D. Cal. May 11, 2023); Complaint & Demand for Jury Trial, *supra* note 9, at 16.

165. *Cf.* Rachel Metz, *AI Won an Art Contest, and Artists Are Furious*, CNN BUS. (Sept. 3, 2022), https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html [https://perma.cc/ELC3-V9W4] (reporting that a game designer with no artistic training won a fine art competition using GenAI).

166. The issue of authorship in GenAI is highly contested. Some argue that GenAI are assisting tools that enhance the creative capacity of human users through sophisticated prompt engineering strategies. Others believe that the creativity is entirely encompassed by the models themselves, and users simply benefit from autonomously generated AI works. *See* Edward Lee, *Prompting Progress: Authorship in the Age of AI*, 76 FLA. L. REV. (forthcoming 2024) (manuscript at 4–5, 20, 23).

GenAI technology disrupts copyright law by forcing it to make three fundamental legal decisions.[167] First, the law must decide whether to afford authorship and ownership interests to expressive works created with the aid of GenAI models.[168] Second, the law must decide whether GenAI models can be trained on copyrighted works without authorization from their owners.[169] Third, the law must decide whether (and to what extent) works created with the aid of GenAI models constitute copyright violations if the models used copyrighted works at the training stage.[170]

All three questions are debated in academic circles and among copyright practitioners. From a practical perspective, the first question was preliminarily resolved by recent formal guidance issued by the U.S. Copyright Office.[171] The Office considers most GenAI outputs unprotected by copyright law for lack of human authorship.[172] It also solicited public comments and scheduled multiple "public listening sessions" to further debate this topic and related issues.[173]

The other two questions were recently invoked in several pending class action lawsuits. In *Doe 1 v. GitHub, Inc.*, plaintiffs claimed that Copilot, GitHub's model for generating computer code, infringes copyrights in the licensed code on which the model was trained.[174] Specifically, they claimed that the model infringes copyrights because it was (1) trained on copyrighted code without authorization, and (2) generated snippets of that same copyrighted code.[175]

In another class action lawsuit, the plaintiffs argued that Stable Diffusion, Midjourney, and DeviantArt infringed copyrights in the images

---

167. 11*See generally* Samuelson, *supra* note 11 (discussing these issues).

168. *See generally* Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1200, 1224–25 (1986) (exploring copyright challenges introduced by GenAI); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 445 (2019).

169. 11*Cf.* Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 767 (2021) (suggesting that GenAI training is usually fair use); Bracha, *supra* note 11, at 1 (suggesting that GenAI training should not even be considered an infringing act).

170. 11Samuelson, *supra* note 11. S11*ee generally* Bracha, *supra* note 11, at 38 (considering the question of GenAI copyrightability).

171. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

172. *Id.* at 16191. While the Copyright Office's guidelines do not get deference from the courts, they at least provide a detailed rationale for withholding protection to GenAI outputs. Courts in other jurisdictions provide additional (sometimes contradictory) guidance. *See, e.g.*, Andres Guadamuz, *Chinese Court Declares that AI-Generated Image Has Copyright*, TECHNOLLAMA (Dec. 9, 2023), https://www.technollama.co.uk/chinese-court-declares-that-ai-generated-image-has-copyright [https://perma.cc/6R85-XR5D].

173. *See Copyright Office Launches New Artificial Intelligence Initiative*, U.S. COPYRIGHT OFF. (Mar. 16, 2023), https://www.copyright.gov/newsnet/2023/1004.html [https://perma.cc/AUR3-YM5L].

174. *See* Complaint Class Action & Demand for Jury Trial at 5–6, 12, Doe 1 v. GitHub, Inc., No. 22-CV-06823, 2023 WL 3449131 (N.D. Cal. Nov. 3, 2022).

175. *Id.* at 21–24.

used for their training. Similar to *GitHub, Inc.*, the plaintiffs claimed that the models infringed copyrights both by training on copyrighted images in violation of licensing terms, and by generating allegedly unauthorized derivative images.[176] A third class action invoked similar claims against OpenAI's flagship application ChatGPT.[177]

On the academic front, most of the legal scholarship has focused on the first two questions. Discussions have largely considered whether works created with the aid of GenAI models can be regarded as proprietary and whether training GenAI models with copyrighted material without authorization by the rights holders should be permitted. Some scholars believe that training with copyrighted material does not constitute infringement,[178] while others are more skeptical.[179] Similarly, some scholars are sympathetic to the idea that certain GenAI models' outputs should remain unowned,[180] while others hold a more assertive stance.[181] The approach introduced in this Article has important policy implications for both these questions, which we explore in Part IV.

The third legal question — whether and to what extent GenAI models can produce infringing outputs — has not been seriously debated in legal scholarship. Scholars addressing this question have often

---

176. *See* Complaint & Demand for Jury Trial, *supra* note 9, at 1, 31.

177. *See* Complaint Class Action & Demand for Jury Trial at 1, 11–12, Tremblay v. OpenAI, Inc., No. 23-CV-03223 (N.D. Cal. June 28, 2023).

178. 11 *See* Bracha, *supra* note 11, at 26–27, 30 (arguing that such GenAI model training is non-infringement because it is considered non-expressive use).

179. *See* Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 1, 48 (2018) (arguing that machine learning is different than what courts have traditionally labeled non-infringing non-expressive use). *See generally* Martin Senftleben, A Tax on Machines for the Purpose of Giving a Bounty to the Dethroned Human Author: Towards an AI Levy for the Substitution of Human Literary and Artistic Works 2 (Jan. 28, 2022) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4123309 [https://perma.cc/4H28-G2UH] (proposing imposing levy — "a new legal obligation to pay a single equitable remuneration" — on the users of GenAI models to compensate human authors).

180. *See generally* Burk, *supra* note 31 (posing that if GenAI reduce the costs of creativity, the justification for intellectual property protection is reduced).

181. *See, e.g.*, Jani McCutcheon, *The Vanishing Author in Computer-Generated Works: A Critical Analysis of Recent Australian Case Law*, 36 MELB. U. L. REV. 915, 958, 965 (2013) (suggesting a sui generis regime for protection of "authorless" computer-generated works); Ana Ramalho, *Will Robots Rule the (Artistic) World?: A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems*, 21 J. INTERNET L. 1, 21–22 (2017) (arguing that GenAI outputs which lack a human author should fall into the public domain, but advocating for the establishment of a "disseminator's right"); Ryan Abbott & Elizabeth Rothman, *Disrupting Creativity: Copyright Law in the Age of Generative Artificial Intelligence*, 75 FLA. L. REV. 1141, 1195–99 (2023) (recommending attribution of authorship to AI owners); Shlomit Yanisky-Ravid & Luis A. Velez-Hernandez, *Copyrightability of Artworks Produced by Creative Robots and Originality: The Formality-Objective Model*, 19 MINN. J.L. SCI. & TECH. 1, 49–50 (2018) (same); Shlomit Yanisky-Ravid, *Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era — The Human-Like Authors Are Already Here — A New Model*, 2017 MICH. ST. L. REV. 659 (recommending a statutory amendment to create AI works made for hire).

dismissed it by resorting to the standard copyright infringement tests.[182] This gap in scholarship was filled rapidly by interdisciplinary studies in computer science and law.[183] These studies employed computational procedures to establish whether GenAI models' outputs infringe copyrights in the works used in these models' training.

However, a major limitation of these contemporary attempts to measure GenAI copyright infringement is that they often presume that an accused work used the copyrighted (rather than uncopyrighted) materials in the training set. Indeed, several studies presume copyright infringement when GenAI produces outputs that significantly resemble copyrighted training data.[184] Accordingly, these studies employ various data-extraction attacks to show that GenAI models can reconstruct specific copyrighted works that appear in the GenAI models' datasets.[185]

While intuitively appealing, this approach is legally flawed. From the perspective of copyright law, the mere fact that a certain output of a GenAI system is similar to a copyrighted work in the model's training sets does not necessarily imply that the former was copied from the latter. For example, an allegedly infringing output can derive similar expressive compositions that arise in multiple works in the training set, not just the copyrighted one. Indeed, the models may have copied generic compositions which are bound to appear in multiple works and thus may be insufficient to establish copyright infringement.[186]

Furthermore, the similarity between the outputs of GenAI models and copyrighted works in the models' datasets might be due to the use of similar facts or discoveries rather than protected expression. Copyright law permits copying of unprotected compositions without authorization. Consequently, an approach that seeks to establish copyright infringement based exclusively on access to unlicensed copyrighted materials during training fails to tackle an essential issue for establishing copyright infringement, namely, whether the unauthorized reproduction involves protected expression.

---

182. 11*See, e.g.*, Bracha, *supra* note 11, at 33 ("[C]opying of a single work's style should be analyzed under the standard infringement test."); Samuelson, *supra* note 11 ("Infringement can be found only if there is a close resemblance between expressive elements of a stylistically similar work and original expression in particular works by that artist.").

183. 144 *See, e.g.*, Carlini et al., *supra* note 144, at 2633; Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping & Tom Goldstein, *Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models*, 2023 IEEE/CVF CONF. ON COMPUT. VISION & PATTERN RECOGNITION 6048, 6048.

184. *See, e.g.*, Nicholas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr et al., *Extracting Training Data from Diffusion Models*, 32 USENIX ASS'N 5253, 5258 (2023) (noting that "Stable Diffusion thus memorizes numerous copyrighted and non-permissive-licensed images, which the model may reproduce without the accompanying license"). *See generally* Somepalli et al., *supra* note 183 (accusing the GenAI models of "digital forgery").

185. *See* Carlini et al., *supra* note 184; 183Somepalli et al., *supra* note 183, at 6049.

186. As explained in Section II.B, the legal protection of such expressive compositions is thin.

A reverse critique can be leveled against studies that use computational procedures to prove the opposite assentation, namely that GenAI outputs do not infringe the copyrights of works in their models' datasets. According to this approach, there is no infringement provided that the models can also generate the same outputs without "seeing" the allegedly infringed work.

This assumption takes a step closer in the direction of the proposal suggested in this Article by drawing its inspiration from the genericity principle.[187] Expressive compositions that GenAI models can extrapolate from multiple works, even without accessing the allegedly infringed work, are likely generic. As explained in Section II.B, copyright law's protection of such generic compositions is thin, so GenAI models could probably copy parts of these patterns without risking copyright infringement.

However, the approach taken in these studies is also incomplete.[188] While genericity narrows the scope of legal protection that copyright law affords widespread compositions, it cannot serve as a definitive binary test for copyright infringement.[189] To see why, consider a hypothetical extension of *Warhol v. Goldsmith*.[190] Assume, for example, that after training a GenAI model on Andy Warhol's "Prince Series,"[191] the model generated an output highly similar to Lynn Goldsmith's photograph of Prince. Assume further that this model did not "see" Goldsmith's original photograph in its dataset. Can we positively say that the model's output does not "access" and then infringe on Goldsmith's copyright?[192] The answer is not straightforward.

---

187. From a computer science perspective, these approaches piggyback on the field of differential privacy. *See* Hacohen et al., *supra* note 25.

188. *Id.*

189. The most similar copyright application to the differential privacy scenario is the clean room procedure used in *Google v. Oracle*. There, to prove that Google did not copy the implementation code from Oracle's Java programing language, Google made sure that programmers did not "see" these elements during their construction of the Android code. *See* Google, LLC, v. Oracle Am., Inc., 141 S. Ct. 1183, 1203–04 (2021). If the Google team would have seen the entire Java code during the construction of the Android code, the court would have been less likely to rule that Google did not engage in unlawful copying, even if Google could prove that their team *could have* constructed the Android code without seeing the Java code in its entirety.

190. Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258598 U.S. 508 (2023).

191. The Prince Series was a collection of sixteen derivative works that Andy Warhol created based on Lynn Goldsmith's photograph of Prince. *Id.* at 1261.

192. Situations in which plaintiffs successfully accuse infringers of copying from derivative works based on their original works are common in music copyright disputes. In these cases, plaintiffs usually establish access to their copyrighted works (i.e., musical composition) by showing access to a derivative work which is based on their work (e.g., sound recording). Plaintiffs are not required to demonstrate that the defendant also had access to the original sheet music nor that they could actually read musical notes. We thank Jane Ginsburg for this example.

Indeed, the plurality of the Prince derivatives may indicate that Goldsmith deserves narrower legal protection for her work.[193] Nevertheless, this fact alone does not necessarily absolve the GenAI model's output of infringing Goldsmith's copyright.[194] Narrower legal protection is not the absence of legal protection, and courts must consider the specific circumstances of each case. For example, thin copyright protection would still likely cover commercial uses of derivative outputs, especially if these outputs were themselves generic (incorporate no additional originality).[195]

In other words, attempts to harness computational approaches to establish binary tests for copyright infringement overlook the need for more nuanced analysis. As discussed, resemblance to copyrighted materials alone is insufficient to establish copyright infringement. At the same time, resemblance to generic expressive compositions alone is inadequate to negate copyright infringement.

For this reason, in lieu of binary standards, this Article proposes to harness GenAI models to measure the originality of copyrighted works. These measurements will not provide a definitive answer to whether GenAI models' outputs infringe the copyrights of the works used for these models' training. Nor will such measurements answer the other two pending legal questions mentioned above. Nevertheless, originality measurements would help provide valuable input for courts and policymakers to inform their analysis when grappling with these issues. Part IV explores how this could be done.

C. Measuring Originality by Generative Models

GenAI models offer a unique opportunity to assess the originality of creative works. These models extrapolate knowledge through self-supervised learning on a massive scale and with a level of nuance that was previously unimaginable. Thus, GenAI models provide invaluable lenses into the latent dimensions of interconnected expressive compositions. This capacity offers new opportunities for measuring originality and delineating the legal scope of copyright works more objectively and predictably.

As humans, we routinely engage with the corpus of preexisting materials, learning from images, styles, themes, colors, compositions, and the like. Humans memorize impressions, extract principles, and generalize from new materials they observe, deconstruct, and reconstruct. All

---

193. *See supra* Section II.B.
194. *See infra* Section IV.A.
195. *See supra* note 140 and accompanying text140.

these processes take place in the silo of the human mind.[196] GenAI also learns from engagement with preexisting materials, but with greater nuance and on a much grander scale. This capacity to learn from data at different levels of granularity reveals some underlying shared patterns in preexisting works, which have been difficult to measure accurately thus far.

During learning, GenAI models distill and rank expressive compositions based on their prevalence in the models' datasets. The more often certain expressive compositions appear in the GenAI models' datasets (i.e., the more "generic" they are), the more likely GenAI models are to apply those expressive compositions when generating new works. Conversely, GenAI models are less likely to apply more rarely expressed (i.e., "original") compositions in their datasets.[197]
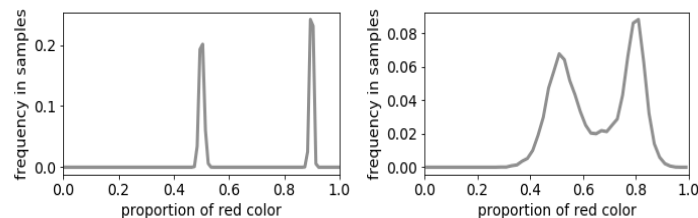
Figure 3: Input-Output Relationship of a Single Feature (Number of Occurrences

This data-driven "bias" is a fundamental feature of inductive machine learning and affects the ways GenAI models generalize.[198] For example, GenAI models trained only on images featuring red dots and other shapes with other colors would probably be able to generalize and generate dots of different colors but would be biased toward generating images of red dots. Figure 3 from Zhao et al. visualizes this dynamic.[199]

The same dynamics may be demonstrated using an inpainting technique, which requires GenAI models to reconstruct missing parts from

---

196. These processes of learning are often termed "generalizations." *See* MARK A. GUCK, EDUARDO MERCADO & CATHERINE E. MYERS, LEARNING AND MEMORY: FROM BRAIN TO BEHAVIOR 209 (2d ed. 2013).

197. Hacohen et al., *supra* note 147, at 2.

198. Yunzhe Liu, Rinon Gal, Amit H. Bermano, Baoquan Chen & Daniel Cohen-Or, Self-Conditioned Generative Adversarial Networks for Image Editing 2 (Feb. 8, 2022) (unpublished manuscript), https://arxiv.org/pdf/2202.04040.pdf [https://perma.cc/B5CB-C9SH].

199. Shengjia Zhao, Hongyu Ren, Jiaming Song & Stefano Ermon, *Bias and Generalization in Deep Generative Models*, GITHUB: ERMON GRP. BLOG, https://ermongroup.github.io/blog/bias-and-generalization-dgm/ [https://perma.cc/N6JJ-M9BT].

images.[200] As shown in Figure 4, when we tasked Stable Diffusion with completing the apple from René Magritte's famous painting *The Son of Man*,[201] the model reconstructed the image with human male faces rather than with apples.[202] This result indicates that the model was trained (unsurprisingly) on many more images of men's faces rather than men with apples in front of their faces.[203]
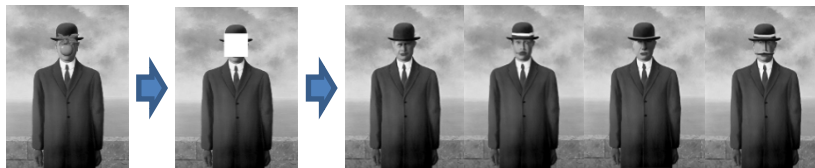

Figure 4: An Inpainting Exercise to Reconstruct René Magritte's Painting *The Son of Man*

In a rough analogy to Nobel laureate Daniel Kahneman's famous cognitive thesis, one can envision GenAI models as "thinking fast" when generating generic expressive compositions.[204] GenAI models apply these compositions "quickly" and "instinctively" when generating new works.[205] Conversely, one can envision GenAI models as "thinking slow" when asked to generate original expressive compositions. [206] GenAI models apply these compositions only after considerable "processing" and "deliberation."[207] Thus, much like humans, GenAI models tend to be more familiar with the things they see often compared to the things they see rarely or do not see at all. As James Bridle of *The Guardian* put it:

---

200. *See* Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser & Bjorn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models 8 (Apr. 13, 2021) (unpublished manuscript), https://arxiv.org/abs/2112.10752 [https://perma.cc/N2NM-J3K9] ("Inpainting is the task of filling masked regions of an image with new content either because parts of the image are . . . corrupted or to replace existing but undesired content within the image.").

201. René Magritte, *The Son of Man*, 1964, oil on canvas, 116 x 89 cm, private collection, https://www.riseart.com/article/2699/masterpiece-in-the-spotlight-the-son-of-man-magritte [https://perma.cc/A7ZN-DTES].

202. We also tried this exercise with adding to the model the textual prompt "Magritte The Son of Man" but the outputs were still dominated by men's faces.

203. This outcome may also demonstrate that the model's training data might be biased in favor of images of men wearing a suit and hat, rather than images of women. But it is difficult to tell. Given that the body shape is clearly masculine, and so are other cues in the outfit, the model may properly recognize the appropriate gender.

204. *See generally* DANIEL KAHNEMAN, THINKING, FAST AND SLOW 19 (2013) (contrasting two primary ways of human thinking and reasoning).

205. *Id.*

206. *Id.*

207. We are currently undertaking a series of experiments to validate this proposition. *See* Hacohen et al., *supra* note 25.

[A]lthough it's very, very hard to imagine the way the machine's imagination works, it is possible to imagine it having a shape. This shape is never going to be smooth or neatly rounded: rather, it is going to have troughs and peaks, mountains and valleys, areas full of information and areas lacking many features at all. Those areas of high information correspond to networks of associations that the system 'knows' a lot about. One can imagine the regions related to human faces, cars and cats, for example, being pretty dense, given the distribution of images one finds on a survey of the whole Internet.

It is these regions that an AI image generator will draw on most heavily when creating its pictures. But there are other places, less visited, that come into play when negative prompting — or indeed, nonsense phrases — are deployed. In order to satisfy such queries, the machine must draw on more esoteric, less certain connections, and perhaps even infer from the totality of what it does know what its opposite may be.[208]

The positive correlation between an expressive composition's genericity and a GenAI model's likelihood of recreating that composition is crucial to copyright law. Unfortunately, this fundamental insight is grossly misunderstood.[209] Because genericity confines the legal protec-

---

208. James Bridle, *The Stupidity of AI*, GUARDIAN (Mar. 16, 2023), https://www. theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt [https://perma.cc/74SJ-BWK3]; *see also* Shengjia Zhao, Hongyu Ren, Jiaming Song & Stefano Ermon, Bias and Generalization in Deep Generative Models: An Empirical Study 1 (2018) (unpublished manuscript), https://arxiv.org/abs/1811.03259 [https://perma.cc/7U3D-5JAE].

209. For example, some legal scholars have used the technical term "memorization" in association with copyright infringement. *See* Sag, *supra* note 11, at 295 ("[M]odels suggest that they are capable of memorizing and reconstituting works in the training data, something that is incompatible with non-expressive use."); *see also* Jannat Un Nisa, *ChatGPT Is About To Face Some Copyright Issues After 'Memorizing' These Books*, WONDERFUL ENG'G (May 4, 2023), https://wonderfulengineering.com/chatgpt-is-about-to-face-some-copyright-issues-after-memorizing-these-books/ [https://perma.cc/XK4H-6YKP] (noting that "AI-generated output is too similar to the input data, which [is] refer[red] to as 'memorization'" and citing Professor Tyler Ochoa saying that "when that occurs, there will be lawsuits, and it will almost certainly constitute copyright infringement"); Ken K. Chang, Machenzie Cramer, Sandeep Soni & David Bamman, Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4 (Oct. 20, 2023) (unpublished manuscript), https://arxiv.org/abs/2305.00118

tion that copyright law affords expressive works, courts and policymakers could quantify originality and delineate the scope of copyright protections by measuring genericity.

Computer scientists have long tinkered with computational measurements for assessing creativity.[210] For instance, researchers Giorgio Franceschelli and Mirco Musolesi proposed using generative learning techniques to assess creativity based on Margaret Boden's definition of value, novelty, and surprise.[211] They also created a tool, called DeepCreativity, that executed this assessment.[212] We propose a different computational approach to measure originality that is more consistent with copyright principles. Building on the work of Rinon Gal et al.,[213] we propose using a computational procedure called textual inversion to measure the originality of specific works ("targeted images") compared to the cumulative learning of a particular GenAI model.[214]

Text-to-image GenAI models, such as Stable Diffusion, generate images from textual prompts.[215] Textual inversion performs the same task but in reverse: it inverts images into textual prompts.[216] These textual prompts are not readable to humans. They are newly invented "pseudo-words" that most accurately capture the expressive configurations of the inverted image as they are represented in the machine's embedding space — meaning they are vectors of a few dozens of numbers.[217] These prompts can then be entered into GenAI models, which

---

[https://perma.cc/2APB-MHAP] (exploring "[t]he ability of these models to memorize an unknown set of books"). However, this characterization is misleading. Assuming that the data on which GenAI models are trained is unbiased and "natural" (i.e., reflects organic real-world usage), the tendency of GenAI models to "memorize" expressive compositions correlates with the compositions' genericity. In other words, the more generic the expression, the more likely a model will "memorize" it. Thus, such memorizations by GenAI models often do not constitute infringement, as generic expressions receive thin to no copyright protection. *See supra* Section II.B.

210. *See* Giorgio Franceschelli & Mirco Musolesi, Creativity and Machine Learning: A Survey 1 (July 5, 2022) (unpublished manuscript), https://arxiv.org/abs/2104.02726 [https://perma.cc/P8PG-FQYQ].

211. *See* Giorgio Franceschelli & Mirco Musolesi, DeepCreativity: Measuring Creativity with Deep Learning Techniques (Jan. 16, 2022) (unpublished manuscript), https://arxiv.org/abs/2201.06118 [https://perma.cc/5FZN-8FKU]; *see also* Margaret A. Boden, *Creativity in a Nutshell*, 5 THINK 83, 83 (2007) ("Creativity is the ability to come up with ideas or artefacts that are *new, surprising, and valuable*." (emphasis added)).

212. *See* Franceschelli & Musolesi, *supra* note 210.

213. Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechick et al., An Image Is Worth One Word: Personalizing Text-to-Image Generation Using Textual Inversion (Aug. 2, 2022) (unpublished manuscript), https://arxiv.org/abs/2208.01618 [https://perma.cc/L9VW-UN77].

214. *Cf.* Weihao Xia, Yulun Zhang, Yujiu Yang, Jin-Hao Xue, Bolei Zhou & Ming-Hsuan Yang, GAN Inversion: A Survey (Mar. 22, 2022) (unpublished manuscript), https://arxiv.org/abs/2101.05278 [https://perma.cc/YR7C-CWHK] (exploring the inversion methodology).

215. Hacohen et al., *supra* note 147, at 5.

216. Gal et al., *supra* note 213, at 3.

217. *Id.* at 2–3.

use cumulative learning to reconstruct the inverted images. Textual inversion was initially used to personalize text-to-image GenAI models.[218] This procedure empowers GenAI models to generate variations of newly introduced images or transfer a new image's style to other images.[219]

Here we argue that textual inversion can also be used to score the originality of the reconstructed (and, by extension, the targeted) images.[220] There are different ways to achieve this goal. The most intuitive one is learning the originality of reconstructed/targeted images from the length of the textual inversion prompts that formulate these images. In other words, this approach equates prompts' length with images' originality: the longer the prompts associated with reconstructed images, the more original these images are, and vice versa.[221]

The logic that guides that relationship should, by now, be intuitive. The more familiar GenAI models are with particular images, the less guidance they will need to generate them. It is similar to the ease with which we envision the appearance of a cat once we hear the word "cat" compared to the difficulty of imagining the appearance of an armadillo once we hear the word "armadillo." More mental labor is needed before we can fully grasp concepts with which we are less familiar.[222] Figure 5 below visualizes the main steps of our proposed procedure.
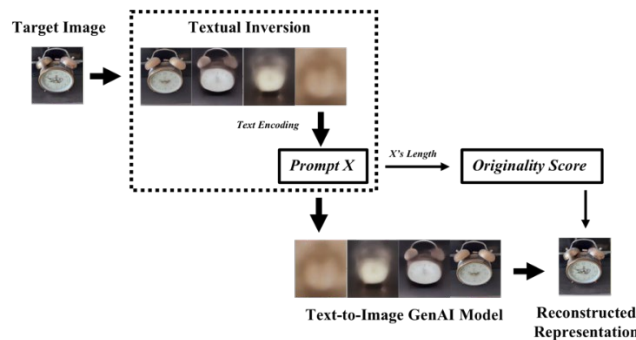
Figure 5: Scoring Image Originality Using Textual Inversion

---

218. *Id.* at 2.

219. *Id.* at 1–2, 9.

220. For images that are not perfectly reconstructed by the model, we also consider an alternative procedure consisting of fine tuning the model from the reconstruction to the targeted image and measuring the amount of training needed to achieve this goal as approximating the targeted image originality. *See* Hacohen et al., *supra* note 25, at 3–4.

221. In this approach, another benchmark that must be determined is what constitutes the golden standard for (a lossless) "reconstruction." The reconstructed image must be identical to the targeted image so that we could learn from the former's originality to the latter.

222. Less mental labor is needed to grasp concepts that are familiar or in machine learning lexicon "generalizable." For an overview of generalization, see generally Liu et al., *supra* note 198.

Table 1 reveals preliminary results from our textual inversion studies with the visual GenAI model Stable Diffusion.[223] As it indicates, a single-word textual prompt embedding "<S>" was adequate to capture the expressive compositions of highly prevalent images such as Barack Obama's portrait (row 1, column A) and Vincent Van Gogh's famous *Starry Night* (row 2, column A). In other words, when using "<S>" as a textual prompt in Stable Diffusion, the model was able to accurately reconstruct and edit these targeted images as depicted in columns B and C, respectively.

Conversely, a single-word textual prompt embedding "<S>"" was unable to accurately capture the expressive compositions of less common images, such as the portrait in row 3, column A or the floating female robot in row 4, column A. In other words, when using "<S>" as a textual prompt in these cases, Stable Diffusion was able to create and edit images that have some expressive compositions similar to the targeted images (rows 3–4, columns B and C, respectively), but none that can properly be labeled as accurate reconstructions.[224] Results were even worse for images that Stable Diffusion did not "see" at all in its dataset, as depicted in rows 5–6. These initial findings support our hypothesis that image originality (or nongenericity) is positively correlated with the length of textual inversion prompts.[225]

Table 1: Textual Inversion Prompts' Length and Inverted Images' Originality

	Targeted Image	One-Word Embedding Textual Prompt ("<S>")	
		Reconstruction	Edit
1.	Prevalent	Prompt: "<S>"	Prompt: "<S> with sunglasses"

223. For further elaboration, see Hacohen et al., *supra* note 25.

	A.	B.	C.
2.	Prevalent	Prompt: "<S>"	Prompt: "cat in <S>"
3.	Rare	Prompt: "<S>"	Prompt: "<S> with sunglasses"
4.	Rare	Prompt: "<S>"	Prompt: "<S> in apocalypse"
5.	Unseen	Prompt: "<S>"	Prompt: "<S> in the morning"
6.	Unseen	Prompt: "<S>"	Prompt: "<S> in the park"

Our research in this area is still preliminary. More work is needed before GenAI models can practically be used to quantify copyright originality. Nevertheless, the theoretical contribution of our suggested approach is to make the generic dimension of copyright law more explicit and calculable. Once refined, the ability to quantitatively measure generics would allow copyright law to move away from binary dichotomies (i.e., idea/expression, copy/non-copy), to facilitate a more nuanced analysis of copyright scope, and to apply microdistributive measures as we discuss next.

IV. POLICY IMPLICATIONS

CREATIVITY IS JUST CONNECTING THINGS.

— STEVE JOBS[226]

The ability to harness GenAI to measure copyright originality has groundbreaking implications for copyright law. So far, jurists have lacked a rigorous means of distinguishing between expressions and ideas. Courts have not been able to differentiate between expressions strictly originated from the author (and therefore considered original) and those which have already become generic, making their use in creative content no longer sufficiently original.[227] This inherent vagueness in copyright doctrines has led to the systematic over-protection of copyrighted works.[228]

The originality scores proposed in this Article could empower courts to delineate copyright scope more efficiently and fairly. Originality scores may also inform the U.S. Copyright Office's registration practices and facilitate market licensing transactions. All these benefits would serve to realign copyright law with its constitutional foundation.[229] This Section explores some potential implications that the originality scores may have on current copyright doctrines and practices.

---

226. Gary Wolf, *Steve Jobs: The Next Insanely Great Thing*, WIRED (Feb. 1, 1996), https://www.wired.com/1996/02/jobs-2/ [https://perma.cc/ZS5C-K9AT].

227. *See* sources cited *supra* note 18 and accompanying text.

228. *See* Parchomovsky & Stein, *supra* note 37, at 1513; *see also* Benkler, *supra* note 19, at 358 (discussing how "copyright and similar laws tend to concentrate information production" and thus raise First Amendment concerns); Gibson, *supra* note 19, at 895; Pamela Samuelson, *The Copyright Grab*, WIRED (Jan. 1, 1996), https://www.wired.com/1996/01/white-paper/ [https://perma.cc/J3JQ-RHQA] (discussing how proposed pre-DMCA laws were so vague as to include benign activity).

229. *Cf.* Cooter & Hacohen, *supra* note 39, at 195 n.13 ("[T]he constitutional purpose requires focusing on innovation, not other values.").

A. Infringement

Copyright scores could have the most dramatic impact on copyright infringement litigation. As explained in Part II, courts consider the originality of copyrighted works when evaluating whether allegedly infringing works use original subject matter without authorization and thus infringe copyrights. This consideration is baked into the substantial similarity inquiry under the *Altai* framework and the fair use analysis under the second factor of § 107 of the Copyright Act.[230] However, courts' ability to appraise copyright originality under these doctrines remains quite limited in practice.[231]

When evaluating substantial similarity, courts often resort to a binary standard that merely approximates the optimal gradient standard articulated above.[232] Courts begin by filtering out the expressive elements from the allegedly infringed work that they consider "fully nonoriginal" under various copyright doctrines, such as functionality, merger, or scènes à faire.[233] Courts then evaluate the similarity of these remaining expressions to the allegedly infringing works without considering how original these remaining patterns actually are.[234] By ignoring the originality ranking of the expressive compositions that survived filtration, courts insufficiently tailor the similarity standard to the originality level of allegedly infringed works.[235]

Courts also apply a binary approximation standard when evaluating originality under the second fair use factor.[236] Courts distinguish "expressive or creative" works from those that are "factual or informational," but they hardly assess the precise level of these works' creativity or functionality.[237] For example, while the Supreme Court in *Google*

---

230. *See* Comput. Ass'n Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 706–10 (2d Cir. 1992); 17 U.S.C. § 107(2); *see also supra* Section II.B.2.

231. Indeed, the doctrines of fair use and substantial similarity are notoriously vague. *See* Meurer & Menell, *supra* note 33, at 12.

232. *See* sources cited *supra* notes 120–22 and accompanying text.

233. *See* Rentmeester v. Nike, Inc., 883 F.3d 1111, 1118 (9th Cir. 2018) ("Before that comparison can be made, the court must "filter out" the unprotectable elements of the plaintiff's work — primarily ideas and concepts, material in the public domain, and scènes à faire . . . The protectable elements that remain are then compared to corresponding elements of the defendant's work to assess similarities in the objective details of the works.").

234. *See* Funky Films, Inc. v. Time Warner Ent. Co., 462 F.3d 1072, 1077 (9th Cir. 2006) (describing how courts assess "whether articulable similarities" exist between remaining elements, like "plot, themes, dialogue, mood, setting, pace, characters, and sequence of events in the two works" (citation omitted)); *see also* cases cited *supra* note 121.

235. The notion that there is no true binary relationship between protected expression and unprotected ideas has been emphasized by numerous courts and commentators. *See* sources cited *supra* note 50 and accompanying text.

236. *See supra* note 129 and accompanying text.

237. *See, e.g.*, Blanch v. Koons, 467 F.3d 244, 256 (2d Cir. 2006); Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 11 F.4th 26, 117 (2d Cir. 2021) ("Having recognized the Goldsmith Photograph as . . . creative . . . the district court should have found this factor

v. Oracle ruled that the Java declaring code was functional under the second fair use factor, it did not explain at what point that functionality would render it utterly unprotected.[238]

Courts apply a binary standard even when considering the "transformativeness" of the allegedly infringing uses under the first fair use factor.[239] Nearly all derivative works "transform" the works from which they are derived to some extent, but this does not necessarily discharge them of copyright infringement.[240] Nevertheless, courts often treat transformativeness as a binary switch rather than a matter of degree by equating transformative derivatives with fair use.[241]

Originality scores would avail courts of such pitfalls and inaccuracies.[242] Courts could rely on originality scores to devise a more precise

to favor Goldsmith irrespective of whether it adjudged the Prince Series works transformative within the meaning of the first factor."), *aff'd*, 598 U.S. 508 (2023).

238. *See* 141 S. Ct. 1183, 1190 (2021) ("In reviewing that decision, we assume, for argument's sake, that the material was copyrightable."). In addition, the degree of functionality also impacts the fair use analysis of Google's conduct under the first "transformativeness" fair use factor. The more functional the Java declaring code is, the more likely it is that Google's use of it will be considered fair, even if its level of transformativeness is low, and vice versa. *See supra* note 140 and accompanying text.

239. *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 569 (1994) (implicitly referring to transformativeness as a scale by noting that "[t]he *more* transformative the new work, the less will be the significance of other factors" (emphasis added)); Brief of Professors Peter S. Menell et al. as Amici Curiae in Support of Respondents at 12, Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508 (2023) (No. 21-869), 2022 WL 3371308, at *26 (criticizing courts for not "understanding transformativeness as a matter of *degree* rather than as a binary switch" (emphasis in original)); Lemley, *supra* note 39, at 1077 (noting that fair use analysis "requires . . . a more careful balancing of the relative contributions made by the original copyright owner and the improver").

240. *See* 17 U.S.C. § 106(2) (granting authors the exclusive right "to prepare derivative works based upon the copyrighted work"). The Act defines a "derivative work" as "a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, *transformed*, or adapted." *Id.* § 101 (emphasis added).

241. *See* Brief of Professors Menell et al., *supra* note 239, at 18 ("The District Court below adopted this erroneous approach and concluded that since the defendant's works were 'transformative' — in a purely binary sense and without specifying the *degree* of their transformativeness — 'the import of their (limited) commercial nature [wa]s diluted.'" (alteration in original)); *cf.* Cariou v. Prince, 714 F.3d 694, 708 (2d Cir. 2013) ("Although there is no question that Prince's artworks are commercial, we do not place much significance on that fact due to the transformative nature of the work."); *Blanch*, 467 F.3d at 254 (2d Cir. 2006) (summarily concluding that the defendant's use was "substantially transformative" to discount its commercial nature).

242. In this vein, originality scores would challenge Justice Learned Hand's famous statement that "[n]obody has ever been able to fix that boundary [of legal protection against allegedly infringing uses], and nobody ever can." Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (making this proposition in a case involving an alleged film adaptation of a stage play, a derivative work); *see also Warhol*, 598 U.S. at 548 (2023) (complaining that the dissent "offers no theory of the relationship between transformative uses of original works and derivative works that transform originals"); *cf.* Bleistein v. Donaldson Lithographic Co., 188 U.S. 239, 251 (1903) ("It would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of pictorial illustrations, outside of the narrowest and most obvious limits.").

differential similarity standard for copyright infringement: the higher the score, the lower the similarity burden that plaintiffs must satisfy to prove infringement.[243] Similarly, courts could appraise the originality of specific copyrighted works even without resorting to categorical proxies. For example, courts could differentiate an original photograph from an unoriginal painting, even though, as a class, the latter is considered more "creative" and the former more "factual."[244] Courts could also distinguish famous works from original ones, although currently, they sometimes confuse the two concepts.[245]

To illustrate the value of originality scores, let's imagine they were available in *Warhol*. To evaluate whether the licensing of Warhol's depiction of Prince constituted fair use, the Court might consider the comparable originality of Warhol's and Goldsmith's works by allowing radical transformations to outweigh "even a showing of direct harm to the original copyright owner."[246]

While making such estimations in the abstract is difficult, there are reasons to believe that Goldsmith's photograph would likely have received a relatively low originality score at the time of Warhol's creation.[247] First, all photographs are works of limited originality as they are essentially compilations of unprotected facts.[248] Second, portraits are even less original than most photographs because many of their expressive choices are constrained by the rules of their given genre.[249] Third, Prince was a famous individual subject of numerous portraits and photographers. Prince's popularity and familiarity might further

---

243. *Cf.* Lemley, *supra* note 122, at 28–29.

244. *See* Justin Hughes, *The Photographer's Copyright — Photograph as Art, Photograph as Database*, 25 HARV. J.L. & TECH. 339, 342 (2012) ("[C]opyright protects far fewer photographs than is commonly understood and, as with the thin copyright of a database, offers less protection to those photographs that are copyrighted." (citation omitted)); *see also supra* Section IV.B. (arguing that originality scoring could also allow courts to appraise the originality of GenAI-augmented works which, as a class, are currently considered by the Copyright Office unprotected for the lack of human authorship).

245. *See Warhol*, 598 U.S. at 524 (criticizing the lower court for implicitly suggesting that fame can be used as an objective proxy for originality and transformativeness and noting that this approach would effectively "create a celebrity-plagiarist privilege"); Brief for the United States as Amici Curiae in Support of Respondents at 23, Google, LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021) (No. 18-956), 2020 WL 1028353, at *23–24 (same).

246. Lemley, *supra* note 39, at 1078.

247. This Article's proposal relies on the argument that originality (nongenericity) should be measured at the time of infringement, not the time of creation, because of the dynamic nature of expressive genericity. *See supra* Section II.B.

248. Hughes, *supra* note 244, at 342.

249. *Cf.* Tobin, *supra* note 115 (discussing Singaporean case about photographer Jingna Zhang's cover shoot for *Harper's Bazaar Vietnam* magazine). Although, unlike traditional portrait photography, modern forms of portrait photography permit photographers greater artistic leeway and, accordingly, make this expressive genre far more diverse and complicated. *See, e.g.*, Kyle DeGuzman, *What is Portrait Photography — Types, Styles, Concepts & More*, STUDIOBINDER (Nov. 20, 2022), https://www.studiobinder.com/blog/what-is-portrait-photography-definition/ [https://perma.cc/9X7Z-N7BH] (exploring different forms of portrait photography).

degrade the originality of Goldsmith's photograph by rendering some of Prince's facial expressions ubiquitous and thus generic.[250]

Unlike Goldsmith's photograph, the originality of Warhol's depiction at the time the Andy Warhol Foundation for Visual Arts licensed it is harder to estimate.[251] Today, Warhol's work would most certainly have received a relatively low originality score, partially due to Warhol's own influence and success.[252] Warhol had an enormous cultural impact, and his work motivated numerous artists to employ similar expressive patterns in their works.[253] The prevalence of such patterns in today's expressive environment would likely render Warhol's work as original as "an Instagram filter, and a simple one at that."[254]

Nevertheless, at the time his work was licensed, Warhol's expressive patterns had a more limited cultural impact, which might have awarded his work a higher originality score.[255] If Warhol's originality score had been significantly higher than Goldsmith's, the Court might have been more inclined to weigh the first and second fair use factors in Warhol's favor.

Originality scores would not be singularly dispositive to Warhol's fair use analysis.[256] The Court would still need to weigh other elements

250. *See, e.g.*, *supra* Section III.C. tbl.1 (showing the results of textual inversion for a portrait of Barack Obama). The fact that the likenesses of more famous individuals are more generic than that of lesser known individuals follows from the general logic of the genericity of expressive compositions. However, this logical intuition is difficult to measure without our computational approach. For example, users' free speech interests (which weigh in favor of prescribing weaker exclusivity rights) are clearly stronger with respect to famous public figures than with respect to private individuals. *See* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 567 (1985) (discussing the newsworthiness of public figures). In addition, the property interests of celebrities are also protected by other legal schemes such as publicity rights. *See* Robert C. Post & Jennifer E. Rothman, *The First Amendment and the Right(s) of Publicity*, 130 YALE L.J. 86, 107 (2020). Even in this domain, some scholars claim that the more famous a celebrity becomes, the narrower the legal protection afforded by their publicity rights should be. *See, e.g.*, Michael Madow, Private Ownership of Public Image: Popular Culture and Publicity Rights, 81 CALIF. L. REV. 125, 134 (1993) (arguing that the scope of publicity rights should reflect the fact that prominent celebrities shape social and cultural discourse, and that the public should be allowed to use celebrity personas for these purposes).

251. Because originality (nongenericity) is dynamic, a work might be entitled to a different score at the time it was created compared to when the allegedly infringing use was conducted, i.e., the time of licensing. *See supra* Section II.B.

252. *See supra* Section II.B.

253. *See, e.g.*, Andy Warhol Found. for the Visual Arts Inc. v. Goldsmith, 598 U.S. 508, 561 (2023) (recognizing Warhol's influence and noting how "he changed modern art"); Art Works Advisory Ed., *Why is Andy Warhol Significant in the Contemporary Art World?*, ART WORKS ADVISORY (May 13, 2020), https://www.artworks.com.sg/news/why-is-andy-warhol-significant-in-the-contemporary-art-world/ [https://perma.cc/9VMM-XAG7] ("Andy Warhol wasn't just influential; he created a whole new genre of contemporary art — pop art.").

254. *Warhol*, 598 U.S. at 574.

255. Note that this logic is the opposite of the view that fame is associated with high transformativeness. In reality, fame dilutes transformativeness by making expressive composition generic. *See supra* note 250 and accompanying text.

256. *See supra* note 129 and accompanying text.

(e.g., commercialism) and other factors (e.g., the amount of expression taken and the market impact of the allegedly infringing use). However, originality scores would empower the Court to conduct a clear and predictable originality assessment supporting and informing the legal analysis.[257]

B. Registration

Works eligible for copyright protection can also be registered in the U.S. Copyright Office. Registration is not required for copyright eligibility but is necessary for filing copyright infringement lawsuits.[258] Because the originality threshold for copyright eligibility is low, registration with the Copyright Office was traditionally assumed to be nearly automatic.[259] This assumption was robust, enduring many technological waves including those that birthed works of modest originality, such as digital photography.[260]

Today, GenAI technology poses a new threat to this long-accepted assumption. While GenAI-augmented works can be highly creative, the Copyright Office does not consider these works original because it does not attribute their creation to the human users who "authored" them.[261]

In this vein, the Copyright Office recently refused to register graphic images that a comic book writer Ms. Kris Kashtanova produced with the aid of the GenAI system Midjourney.[262] After determining that

---

257. *Cf. Warhol*, 598 U.S. at 548 (criticizing the dissent for not providing a "theory of the relationship between transformative uses of original works and derivative works that transform originals").

258. *See* Fourth Est. Pub. Benefit Corp. v. Wall-Street.com, 139 S. Ct. 881, 886 (2019); Zvi S. Rosen, *Examining Copyright*, 69 J. COPYRIGHT SOC'Y U.S.A. 481, 482 (2023), https://ssrn.com/abstract=4099976 [https://perma.cc/SFN4-YEX6].

259. *See* Cosmetic Ideas, Inc. v. IAC/InteractiveCorp, 606 F.3d 612, 621 (9th Cir. 2010) ("[T]he Register's decision of whether or not to grant a registration certificate is largely perfunctory . . . ."); Susan M. Richey, *The Troubling Role of Federal Registration in Proving Intellectual Property Crimes*, 50 AM. CRIM. L. REV. 455, 465 (2013) ("[T]he Copyright Office adopts a cursory examination process and generally confines its inquiry to a determination of whether filing and deposit requirements have been met.").

260. *See* Rosen, *supra* note 258, at 538 ("[T]he rejection rate for photos is substantially lower than the rate for visual arts generally."); *see also* Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 56–58 (1884) (holding that there was "no doubt" the Constitution's Copyright Clause permitted photographs to be subject to copyright, "so far as they are representatives of original intellectual conceptions of the author").

261. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202) ("[W]hen an AI technology receives solely a prompt from a human and produces complex written, visual, or musical works in response, the 'traditional elements of authorship' are determined and executed by the technology — not the human user . . . . As a result, that material is not protected by copyright . . . ." (footnotes omitted)).

262. *See* Letter from Robert J. Kasunic, Assoc. Reg. of Copyrights & Dir. of the Off. of Registration Pol'y & Prac., U.S. Copyright Off. to Van Lindberg, Taylor English Duma, LLP at 1 (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf [https://perma.cc/5V4K-859G].

"it was Midjourney — not Kashtanova — that originated the 'traditional elements of authorship' in the images," the Office ruled that "the images generated by Midjourney . . . are not original works of authorship protected by copyright."[263] Accordingly, the Office refused to register the Midjourney images, and instead afforded Ms. Kashtanova only a limited copyright interest in how the images were compiled.

This decision is not necessarily mistaken. Scholars have long recognized that works created with GenAI tools can be authorless.[264] This outcome might also be socially desirable to the extent that GenAI models lower the creation cost, allowing authors to recoup their costs without legal intervention.[265] However, the decision to exclude GenAI-augmented works from copyrightability should not be definite.

While some GenAI-augmented works may be created "cheaply," others may involve substantial user input.[266] In such cases, society might lose by categorically excluding GenAI-augmented works from the realm of copyrightability.[267] This fear will grow as GenAI technology democratizes and penetrates more creative fields.

In addition, as GenAI models become increasingly sophisticated and capable, original GenAI-augmented outputs will become increasingly harder to produce.[268] We are already seeing the rise of new creative skills, such as the art of prompt engineering, where users iteratively craft prompts to generate and improve their creative output.[269] By limiting copyright registration only to original GenAI-augmented outputs, society could better tailor copyright law's incentives to the production of these valuable but increasingly scarce information goods.

---

263. *Id.* at 8.

264. *Cf.* sources cited *supra* note 168.

265. *See* Burk, *supra* note 31, at 1679–80; *see also* Greg Bensinger, *Focus: ChatGPT Launches Boom in AI-Written E-books on Amazon*, REUTERS (Feb. 21, 2023), https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/ [https://perma.cc/5RV3-PGSD]; Travis Diehl, *Mimicking the 19th Century in the Age of A.I.*, N.Y. TIMES (May 3, 2023), https://www.nytimes.com/2023/05/03/arts/design/ai-makes-nostalgic-images.html [https://perma.cc/K8YF-XSS8].

266. User input is manifested both in the creative guidance of GenAI models and in the technical skills when executing this guidance. These two types of input are sometimes called "problem formulation" and "prompt engineering," respectively. *See* Oguz A. Acar, *AI Prompt Engineering Isn't the Future*, HARV. BUS. REV. (June 6, 2023), https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future [https://perma.cc/26TK-LLMR]; *see also* Kate Whiting, *3 New and Emerging Jobs You Can Get Hired for This Year*, WORLD ECON. F. (Mar. 2, 2023), https://www.weforum.org/agenda/2023/03/new-emerging-jobs-work-skills/ [https://perma.cc/BKS6-DLNN] (calling prompt engineering the "job[] of the future").

267. Indeed, fears that, in the absence of intellectual property protection to AI-generated works, society may lose creative outputs have urged scholars to propose amendments to existing laws or enactment of new laws that would support appropriations of such outputs. *See* sources cited *supra* note 181.

268. *Cf.* Ryan Benjamin Abbott, *Everything Is Obvious*, 66 UCLA L. REV. 2, 34 (2019) (suggesting that when machines aid in innovation, the standards of novelty will rise).

269. *See* Jonas Oppenlaender, Rhema Linder & Johanna Silvennoinen, Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering 1 (May 13, 2023) (unpublished manuscript), https://arxiv.org/pdf/2303.13534 [https://perma.cc/59LF-CH93].

The Copyright Office has recognized that human authors may sometimes contribute sufficient original expression to make GenAI-augmented works eligible for registration.[270] However, the Office also recognized that deciding whether to attribute the originality of GenAI-augmented works to the users who prompted them is a difficult task.[271]

Originality scores could aid the Copyright Office in reaching such decisions. Our approach may enable the Office to bypass the need to rely on assessing creativity in prompt engineering and instead compare the originality of the work to synthesized content captured by the model.[272] Specifically, the Copyright Office could consider the originality scores of works generated by GenAI vis-à-vis the preexisting materials from which the model has already learned.[273] The higher the score, the more likely the work reflects a modicum of creativity compared to the aggregated expressive knowledge learned by the model.

C. Signaling and Licensing

Originality scores may facilitate copyright licensing practices by providing objective indicia for the market value of expressive works. This signal might be particularly crucial for GenAI developers which would value copyrighted works with higher originality scores more than copyrighted works with low originality scores.[274]

GenAI enables users to draw upon existing knowledge embedded within the model automatically, rather than learn and create everything from scratch. This functionality may reduce the barriers to entry for the production of creative works, allowing individuals with basic language, graphic, or technical skills to generate high-quality content.[275] Such a

---

270. *See* Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 16192 (Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

271. *Id.*

272. In addition, our approach could also empower the Copyright Office to appraise the originality of the textual prompts themselves. The Copyright Office stated in its guideline that "prompts may be sufficiently creative to be protected by copyright, [but] that does not mean that material generated from a copyrightable prompt is itself copyrightable." *Id.* at 16192 n.27. Still, the degree to which the prompt is original may serve as a useful proxy to the user's original contribution.

273. Future research should clarify the relationship between the originality of textual prompts and the originality of the image these textual prompts produce. The Copyright Office already recognized that sufficiently original prompts can be protected as literary works but refuses to extend this protection to the images GenAI systems produce in reaction to such prompts. *See id.* If there is a strong link between the originality of textual prompts and the images they produce, perhaps applying our approach only to the textual domain — namely, measuring the originality of textual prompts – will suffice for determining works' eligibility.

274. *Cf.* Clarisa Long, *Patent Signals*, 69 U. CHI. L. REV. 625, 643–64 (2002) (exploring the signaling function of patents).

275. *See* David De Cremer, Nicola Morini Bianzino & Ben Falk, *How Generative AI Could Disrupt Creative Work*, HARV. BUS. REV. (Apr. 13, 2023), https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work [https://perma.cc/3V86-U88X].

flow of "cheap" content may require those marketing the works to adequately signal the added value generated by each interaction of a human creator with GenAI. If the generated output consists of generic clichés, the originality score is likely to be lower, indicating a lower economic value to potential licensees who can easily generate it themselves.

Such signaling value may thus facilitate market licensing transactions. For example, copyright owners could advertise their works' originality scores to attract potential consumers.[276] Similarly, GenAI companies could rely on originality scores to solicit original works to improve their models' performance.[277]

In addition, courts might use originality scores to assess lost licensing opportunities when considering actual damages in infringement litigation.[278] Currently, courts consider the assessment of damages to be one of the most challenging tasks they handle.[279] By simplifying this task, originality scores might even lead courts to be more liberal in prescribing monetary awards instead of injunctive relief for copyright infringement lawsuits, which some scholars have long considered advisable.[280]

---

276. In the patent context, consumers sometimes use a "patent pending" or "patented" disclosure to signal the potential value of their technology. *See, e.g.*, *Patent Pending: What It Means and How to Protect Your Invention*, THOUGHTS TO PAPER BLOG (Mar. 14, 2023), https://www.thoughtstopaper.com/blog/patent-pending-what-it-means-and-how-to-protect-your-invention/ [https://perma.cc/4M6Y-GBP9]. Originality scores could even more effectively communicate the degree to which copyrighted works are valuable.

277. Scholars have argued (and we agree) that GenAI learning does not constitute copyright infringement, and, as such, should not be bargained for. *See* sources cited *supra* note 169. Nevertheless, authors can still refuse to make their intellectual output available for learning. In addition, our proposed originality scoring system could help policymakers devise various remuneration or taxation schemes. While we do not necessarily endorse these approaches (at least to the extent that they embrace the notion that machine learning constitutes copyright infringement), other scholars and policymakers do consider them. For example, Martin Senftleben offers to impose an equitable tax on GenAI users based on aesthetic theories which assign intrinsic value to human authorship. *See* Senftleben, *supra* note 179, at 3. Originality scores could align Senftleben's and similar proposals with their policy objectives by enabling policymakers to adjust AI levies, taxes, and other remuneration schemes to the originality scores of GenAI-augmented works.

278. Courts are required to appraise loss licensing fees when they prescribe actual damages in copyright infringement claims. *See* On Davis v. Gap, Inc., 246 F.3d 152, 166 (2d Cir. 2001).

279. *Cf.* Cincinnati Car Co. v. N.Y. Rapid Transit Corp., 66 F.2d 592, 593 (2d Cir. 1933) ("The difficulty of allocating profits in such cases has plagued the court from the outset, and will continue to do so, unless some formal and conventional rule is laid down, which is not likely.").

280. *See, e.g.*, Parchomovsky & Stein, *supra* note 37, at 1533–34. (The authors introduce "the added-value doctrine," which would allow copyright owners to secure injunctive relief against putative infringers only in cases where the latter allegedly infringing work is less original than the former work. In all other cases, they argue, the court should allow the use of the work and prescribe compensatory damages.); Peter S. Menell & Ben Depoorter, *Using Fee Shifting to Promote Fair Use and Fair Licensing*, 102 CALIF. L. REV. 53, 58–59 (2014) (proposing a delicate approach to incentivize parties to bargain and to inform the courts of reasonable licensing fees in cases of infringement).

V. CONCLUSION

This Article has demonstrated how GenAI capacities could be leveraged to gain more nuanced insights into the genericity of specific expressions on a large scale. Our novel approach to measuring originality is based on interdisciplinary computer science and legal research. This approach employs data-driven bias to evaluate the genericity of expressive compositions in preexisting works. The more generic the compositions, the more likely GenAI models will utilize them when generating new works. Conversely, the rarer the expressive compositions in GenAI models' datasets (indicating their "original" nature), the less likely GenAI models are to utilize them.

This scale for measuring originality could impact all major phases in the lifecycle of copyrighted works, from registration and licensing to copyright infringement litigation. For instance, in the context of copyright infringement, scholars have argued that while generalization by the model is noninfringing since it only utilizes ideas, memorization by the model constitutes a copyright infringement. Our approach offers a more nuanced analysis. Memorization by the model will be considered infringing only to the extent that it is inconsistent with the model's data-driven bias.

Our approach does not subscribe to "technological solutionism" — the belief that all human problems can be solved by technological intervention alone — and should not be interpreted as such.[281] In fact, we assume that choices regarding originality reflect normative tradeoffs, which should be decided by social institutions (e.g., courts, regulators, standard-setting bodies) using acceptable procedures. Nevertheless, such choices could now be better informed by evidence.

Originality scores could empower policymakers to go beyond ensuring compliance. Policymakers could use originality scores to devise policies and doctrines that better calibrate copyright protections to the originality of expressive works. Scholars have long advocated for such proposals, but the judicial system has failed to implement them.[282] Originality scores could revive these old proposals and make implementation feasible.

By harnessing GenAI to measure originality at scale, our approach can offer valuable insights to policymakers. These insights could assist

---

281. *See* EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE 5 (2013) (arguing that technological solutionism derives from a "never-ending quest to ameliorate," while being oblivious to complex social situations and conditions).

282. *Cf.* Parchomovsky & Stein, *supra* note 37, at 1534 (accepting the criticism of the courts' institutional incapacities to assess originality and suggesting that courts should rely more on expert testimony); Miller, *supra* note 56, at 477 (reciting the traditional scholarly skepticism about the ability of jurists to appraise copyright originality); Fisher, *Recalibrating Originality*, *supra* note 45, at 458 (arguing that originality should better assess creative contribution); Bracha & Syed, *supra* note 80, at 1912 (same).

policymakers as they grapple with adapting copyright law to meet the new challenges of an era characterized by "cheap creativity" enabled by GenAI.