

COPYRIGHT REGENERATED: HARNESSING GENAI TO MEASURE ORIGINALITY AND COPYRIGHT SCOPE

Uri Y. Hacoheh[†] & Niva Elkin-Koren[†]

ABSTRACT

The rise of Generative Artificial Intelligence (GenAI) models is revolutionizing the creative domain. By using models like Github Copilot, Open AI GPT, Stable Diffusion, Midjourney, or DeviantArt, 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 these 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

[†] Assistant Professor, Tel-Aviv University Faculty of Law; Faculty Associate at the Chief Justice Meir Shamgar Center for Digital Law and Innovation.

[†] Professor of Law, Tel-Aviv University Faculty of Law; Faculty Associate, Berkman Klein Center at Harvard University; Director of the Chief Justice Meir Shamgar Center for Digital Law and Innovation, and a co-director of the Algorithmic Governance Lab at TAU Innovation Lab. This paper is based on insights coming from interdisciplinary work at the Algorithmic Governance Lab at TAU Innovation Lab (TIL) in collaboration with Dr. Amit Bermano (the Blavatnik School of Computer Science in Tel-Aviv University), Dr. Roi Livni (The Iby and Aladar Fleischman Faculty of Engineering, Tel Aviv University), and Adi Haviv (the Blavatnik School of Computer Science in Tel-Aviv University). See Uri Hacoheh, Niva Elkin-Koren, Amit Beremano, Roi Livni, & Adi Haviv, *Measurable Copyright Similarity for Generative Models* (working paper, 2023); Uri Hacoheh, Niva Elkin-Koren, Roi Livni & Shai Moran, *Can Copyright be Reduced to Privacy?*, ARXIV (2023) https://www.researchgate.net/publication/371009258_Can_Copyright_be_Reduced_to_Privacy. For the thoughtful comments we would like to thank Prof. Assaf Hamdani, Prof. Oren Bracha, Prof. Dan Burk, Prof. Matthew Sag, Prof. Jane Ginsburg, Prof. Deven Desai and Rimon Weiss.

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 over-protection 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 to policy-makers as they grapple with adapting copyright law to meet the new challenges of an era of "cheap creativity" enabled by GenAI.

I.	INTRODUCTION.....	4
II.	DIVIDING ENTITLEMENTS AMONG AUTHORS	11
	A. Originality & Copyright Goals	12
	B. Genericity in Copyright Law	17
	1. Eligibility Doctrines	22
	2. Infringement Doctrines.....	29
III.	LEVERAGING GENAI TO QUANTIFY COPYRIGHT ORIGINALITY.....	33
	A. The Rise of GenAI	34
	B. GenAI and Copyright Disruption	37
	C. Measuring Originality by Generative Models	43
IV.	POLICY IMPLICATIONS	51
	A. Infringement	51
	B. Registration.....	57
	C. Signaling and Licensing.....	60
V.	CONCLUSION	62

I. INTRODUCTION

“I want to be a machine.”

Andy Warhol, 1963¹

Recently, in *Warhol v. Goldsmith*,² the Supreme Court faced a difficult dilemma. It had to decide whether Andy Warhol’s depiction of 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

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>. In reality, Warhol’s famous quote was probably a bit different. According to Gene Swenson who was responsible for the interview that produced Warhol’s quote, Warhol actually said “everybody should be a machine,” in the contexts of being nonjudgmental about gender. See *Did Andy Warhol want to be a machine?*, WARHOLSTARS NEWS (Dec. 2018), <https://warholstars.org/Did-Andy-Warhol-want-to-be-a-machine.html>.

² *Andy Warhol Foundation for the Visual Arts Inc. v. Lynn Goldsmith*, 598 U.S. 1 (2023) [Hereinafter *Warhol*].

to Goldsmith's exclusivity.³ The Court ended up siding with Goldsmith but not before revealing strong disagreement among the justices,⁴ the lower courts,⁵ and the numerous advising amici.⁶

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/opinion/andy-warhol-prince-series-op-ed-1962050> (“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 under the court's precedent 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) [Hereinafter *Campbell*].

⁴ The dissent accused the majority of “hamper[ing] creative progress and undermin[ing] creative freedom.” whereas the majority accused the dissent of reciting “theme[s]. . . familiar to any student of copyright law,” and “offers[ing] no theory of the relationship between transformative uses of original works and derivative works that transform originals.”, *Warhol*, *supra* note 2 dissenting opinion of Kagan, J., at 3-4, opinion of the court, at 36.

⁵ The district court's opinion was that “...it is plain that the Prince Series works are protected by fair use.” *Andy Warhol Foundation for the Visual Arts, Inc. v. Lynn Goldsmith*, 382 F.Supp.3d 312, 324 (S.D.N.Y. 2019). Conversely, in the 2nd Circuit disagreed, calling the transformativeness standard adopted by the District Court “overly liberal... [which] risks crowding out statutory protections for derivative works.” And noting that the District Court view would “...stretched the [pervious 2nd Circuit decisions] too far.” *The Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, No. 19-2420, at 18-20, 26 (2d Cir. 2021).

⁶ A “brief” update on the Warhol case – amicus briefs and the solicitor general's view, AUTHORS ALLIANCE (Aug. 25, 2022), <https://www.authorsalliance.org/2022/08/25/a-brief-update-on-the-warhol-case-amicus-briefs-and-the-solicitor-generals-view/#:~:text=In%20total%2C%20there%20were%2038,support%20of%20AWF%2C%20is%20here> (“Last week, the final set of amicus curiae (friend of the court) 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.”).



Figure 1: Andy Warhol’s depiction of Prince (left); Lynn Goldsmith’s photograph of Prince (right)

Recently, several pending class action lawsuits asked courts to make similar decisions. Courts had to determine whether works created with generative AI models (GenAI) such as *Stable Diffusion* (e.g., Figure 2, left side) constitute lawful independent creations or whether they are infringing derivatives of the copyrighted works which were used to train the models (e.g., Figure 2, right side).⁷ Such cases put the courts in an even more challenging position. Plaintiffs argue that GenAI models are “21st-century collage tools” that violate the rights of millions of authors.⁸ Defendants argue that GenAI models are “expanding the boundaries of human creativity.”⁹ Both sides have persuasive arguments.

⁷ J. DOE 1 v. GitHub, Inc., No. 3:22-cv-06823 (N.D. Cal. filed Nov. 03, 2022), Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023).

⁸ Complaint. Class Action & Demand for Jury Trial, at 3, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023).

⁹ Amended Notice of Motion, Motion to Dismiss & Memorandum of Points and Authorities in Support of Motion to Dismiss, at 10, Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. filed Jan. 13, 2023). See also Paul DelSignore, *AI Will Enhance Human Creativity*, MEDIUM (Feb. 1, 2023), <https://medium.com/generative-ai/ai-will-enhance-human-creativity-d9df54439171>.

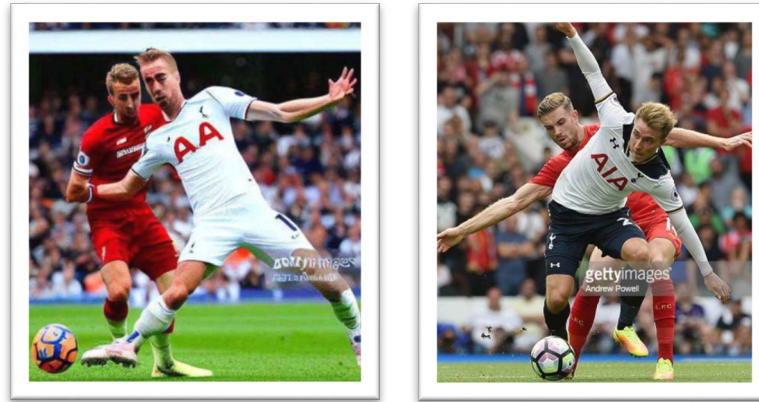


Figure 2: a photograph generated by *Stable Diffusion* (left), a picture owned by *Getty Images* (right)

Copyright law's ultimate goal is to foster the creation and dissemination of expressive works by granting authors limited rights to their respective expressions.¹⁰ Thus, copyright law's greatest challenge is allocating legal entitlements to expressive works by consecutive authors. This task has always been a challenge because authorship derives from cultural contexts. Authors routinely engage with preexisting materials to convey a meaningful message. This task grows more complicated when GenAI augments human creativity. Generative models draw upon an infinite number of preexisting works created by an often unidentifiable multitude of authors. Consequently, allocating legal entitlements among all possible claimants becomes an insurmountable objective.

Copyright scholars are only beginning to grapple with the ways in which GenAI exacerbates the challenges faced by copyright law.¹¹ In this paper, we argue that GenAI may also be able to address the same challenges that it creates. Copyright law applies to original works of authorship. Therefore, courts are often asked to assess the originality of given works and delineate their legal protections when

¹⁰ Pierre N. Leval, *Toward a Fair Use Standard*, 103 HAR. L. REV. 1105, 1107-9 (1990).

¹¹ See e.g., Pamela Samuelson, *Generative AI Meets Copyright*, 381 SCIENCE 158 (2023); Peter Henderson et al., *Foundation Models and Fair Use* (Mar. 29, 2023) (unpublished manuscript) (on file at arXiv:2303.15715v1); Matthew Sag, *Copyright Safety for Generative AI*, (forthcoming in the HOUSTON L. REV., 2023); See Oren Bracha, *The Work of Copyright in the Age of Machine Production* (work in progress) (on file with the authors).

deciding copyright disputes.¹² So far, courts have performed this task on an ad-hoc basis by applying numerous legal doctrines such as the idea/expression dichotomy, merger, Scènes à faire, substantial similarity, and fair use.¹³ These doctrines are notoriously vague and unpredictable¹⁴ and, in practice, lead to the overprotection of preexisting works.¹⁵

¹² See e.g., Gideon Parchomovsky & Alex Stein, *Originality*, 95 VIR. L. REV. 1497, 1517 (2009). See generally Robert D. Cooter & Uri Y. Hacoheh, *Progress in the Useful Arts: Foundations of Patent Law in Growth Economics*, 22 YALE J. L. & TECH 191 (2020).

¹³ SAS Institute Inc. v. World Programming Ltd., 64 F. Supp. 3d 755, 761 (E.D.N.C. 2014) (“...[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.”) [hereinafter *SAS v. WPL*]. See *infra* Section B.

¹⁴ See e.g., *SAS v. WPL*, *id.*, at 762 (“The scope of protection is ‘not constant’ across all literary works... Nor is it necessarily constant across all elements in a single work.”); *Walker v. Time Life Films*, 784 F.2d 44, 48 (2d Cir.), *cert. denied*, 476 U.S. 1159 (1986) (noting that the idea/expression dichotomy is “...[A] distinction easier to state than to apply.”); Jessica D. 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 719, 748 (2010); Richard H. Jones, *The Myth of the Idea/Expression Dichotomy in Copyright Law*, 10 PACE L. REV. 551, 553 (1990). For commentators that criticized the idea/expression dichotomy see: BENJAMIN KAPLAN, AN UNHURRIED VIEW OF COPYRIGHT (1967); HUGH LADDIE, PETER PRESCOTT & MARY VITORIA, THE MODERN LAW OF COPYRIGHT § 2.55 (1980); Charles B. Collins, *Some Obsolescent Doctrines of the Law of Copyright*, 1 S. CAL. L. REV. 127, 139 (1928); Steven Knowles & Ronald J. Palmieri, *Dissecting Krofft: An Expression of New Ideas in Copyright?*, 8 SAN FERN. VALLEY L. REV. 109, 124–9 (1980); Robert Y. Libott, *Round the Prickly Pear: The Idea-Expression Fallacy in a Mass Communications World*, 16 COPYRIGHT L. SYMP. 30, 32–5 (1968), revised and reprinted in 14 UCLA L. REV. 735, 737–9 (1967); Edward Samuels, *The Idea-Expression Dichotomy in Copyright Law*, 56 TENN. L. REV. 321 (1989); Kenneth B. Umbreit, *A Consideration of Copyright*, 87 U. PA. L. REV. 932, 950 (1939). *Sid Marty Krofft Tele. v. McDonald’s Corp.*, 562 F.2d 1157, 1163 n.6 (9th Cir. 1977); *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2nd Cir. 1930) (discussing levels of abstraction at which copyrightability may be tested); *Brandir Int’l, Inc. v. Cascade Pac. Lumber Co.*, 834 F.2d 1142 (2nd Cir. 1987); *Carol Barnhart Inc. v. Economy Cover Corp.*, 773 F.2d 411, 415–18, 420 (2d Cir. 1985). *Peter Pan Fabrics, Inc. v. Martin Weiner Corp.*, 274 F.2d 487, 489 (2nd Cir. 1960).

¹⁵ See e.g., James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 Yale L. J. 882, 884 (2007); Neil W. Netanel, *Introduction, Copyright’s Paradox* 54–80 (2008) 7 (UCLA School of Law, Research Paper

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. Consequently, GenAI may facilitate 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 are, therefore, not protected by copyright law.¹⁶ 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.¹⁷ 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. This article refers to this implicit dynamic in copyright law as the *genericity principle*.

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.¹⁸ Indeed, genericity is a fundamental principle that GenAI operates on. During training, models learn recurrent patterns among basic elements in

No. 08-06, 2008) (discussing “copyright’s ungainly expansion”); Jessica D. Litman, *Billowing White Goo*, 31 COLUM. J. L. & ARTS 587, 587 (2008) (same); John Tehranian, *Infringement Nation: Copyright Reform and the Law/Norm Gap*, 2007 UTAH L. REV. 537, 543–8 (2007) (same). 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, 354–60 (1999). Pamela Samuelson, *The Copyright Grab*, WIRED (Jan. 1, 1996) <https://www.wired.com/1996/01/white-paper/>; SIVA VAIDHYANATHAN, COPYRIGHTS AND COPYWRONGS: THE RISE OF INTELLECTUAL PROPERTY (2003). See generally Michael J. Meurer & Peter Menell, *Notice Failure and Notice Externalities*, 5 J. LEGAL ANALYSIS 1 (2013) (explains the inherent vagueness in intellectual property scope).

¹⁶ Cf. *Boisson v. American County Quilts and Linens*, 273 F.3d 262, 272 (2001) (“original works broken down into their composite parts would usually be little more than basic unprotectable elements like letters, colors, and symbols. “). See generally Justin Hughes, *Size Matters (Or Should) in Copyright Law*, 74 FORDHAM L. REV. 575 (2005). ¹⁷ See *infra* note 47 and accompanying text.

¹⁷ See *infra* note 47 and accompanying text.

¹⁸ See *infra* Section II.B.

works included in the models' training set.¹⁹ During deployment, the models apply these patterns to generate new content in response to users' prompts.

Building on these insights, we demonstrate how new computational procedures can utilize GenAI models to produce originality scores for copyrighted works.²⁰ Specifically, we propose to rank the originality of works based on the genericity of their expressive compositions. When GenAI models treat such patterns as generic, their originality score will be low.²¹ 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 model could enable copyright law to more accurately distinguish between original works and those in the public domain. It may then insert more nuance into copyright analysis in deciding copyright scope.²²

These abilities are particularly crucial nowadays 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. To the extent such similarity arises from genericity, it may not result in copyright liability.²³

Our approach also opens up new opportunities for informing copyright registration and licensing practices more fairly and efficiently. For example, quantified originality measures could assist the Copyright Office in distinguishing "cheap creativity" from sufficiently original GenAI-augmented works worthy of legal protection.²⁴ Our approach could also support objective measures for

¹⁹ See *infra* Section III.C.

²⁰ See *infra* Section III.C.; Niva Elkin-Koren, Uri Y. Hacohen, Adi Haviv, Amit Beremano & Roi Livni, *Measurable Copyright Similarity for Generative Models Machines* (working paper 2023) (on file with authors).

²¹ See *infra* Part III. Our procedure essentially distills the works' protected "expressions" from its unprotected "ideas", or, more accurately, it ascertain how original, and, therefore protected, the expressive elements of each copyrighted works are. Cf. Jones, *supra* note 14 at 598 (Arguing that the dichotomy between ideas and expression is actually a spectrum with "[t]he only relevant criteria for distinguishing unprotectable from protectable expressions are originality and creativity of the expressions in a work.")

²² See *infra* Sections II.B., and IV.

²³ See *infra* Sections III.B., and IV.A.

²⁴ See *infra* Section IV.B.

remunerating authors of original works generated through interaction with GenAI systems.²⁵

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. It first explores the rise of GenAI technology, its disruptive effects on copyright law, and the shortcomings of contemporary approaches to remedy these effects. Then, building on these shortcomings, the last section in Part III proposes to use GenAI models to produce originality scores for expressive works of authorship.

Lastly, Part IV explores the policy implications of our quantification approach. By delineating copyright scope more precisely and predictably, our proposed methodology may benefit the copyright system throughout its lifecycle, from registration, licensing practices, and infringement litigation procedures.²⁶

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, 1845²⁷

Copyright law protects original works created by authors, such as original novels, articles, musical compositions, images, and software.²⁸ 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. In a creative ecosystem of interactive exchange

²⁵ See *infra* Sections IV.B., and IV.C.

²⁶ Cf. Meurer & Menell, *supra* note 15 (explains that vague copyright scope impact policy throughout the lifecycle of protected works).

²⁷ *Emerson v. Davies*, 8 F. Cas. 615 (C.C.D. Mass. 1845) (No. 4,436).

²⁸ 17 U.S. Code § 102(a) (1990).

where works and authors intermingle, these boundaries are often dif-fused. The organizing principle of copyright law, which protects au-thors' claims to entitlements to their expressions, is originality. To be copyrighted, a work must have originated with the person claim-ing the entitlement.

Yet, to be meaningful, work must also share some common ground with other expressions common in culture. This part exam-ines the tension between originality and the genericity of common expression in copyright law. Section A discusses copyright law's objectives and originality's role in serving copyright goals. Section B unfolds the inherent tension between originality and ubiquitous expression, and underscores the role of genericity in limiting the scope of copyright protections.

A. *Originality & Copyright Goals*

As the Constitution mandates, copyright law seeks “To pro-mote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respec-tive Writings and Discoveries.”²⁹ Accordingly, copyright law in-centivizes creation and dissemination of original works by granting authors exclusive rights to their respective works.³⁰ These rights ensure that authors can commercially exploit their works and sustain incentives to invest in creating future works. However, *promoting progress* is inconsistent with granting unlimited rights to control

²⁹ U.S. CONST. art. I, § 8, cl. 8.; *Mazer v. Stein*, 347 U.S. 201, 219 (1954); *Cooter & Hacothen*, *supra* note 12, at 197 (exploring how patent and copyright law tailor legal protections to encourage cumulative innovation and creativity); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 998 (1997) (same); Leval, *supra* note 10, at 1111 (same); PAUL A. SAMUELSON & WILLIAM D. NORDHAUS, *ECONOMICS*, at 48–9, 713–5 (12th ed. 1985); RICHARD W. TRESCH, *PUBLIC FINANCE: A NORMATIVE THEORY* 107–29 (1981); Meurer & Menell, *supra* note 15, at 3–4. *See generally* Paul A. Samuel-son, *The Pure Theory of Public Expenditure*, 36 REV. ECON. & STAT. 387, 389 (1954).

³⁰ Leval, *supra* note 10, at 1107. The Supreme Court has affirmed this goal in many occasions. *See e.g.*, *Harper & Row v. Nation Enters*, 471 U.S. 539, 545–6 (citation omitted) (quoting *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429, 477 (Blackmun, J., dissenting)); *Twentieth Century Music Corp. v. Aiken*, 422 U.S. II, 156 (1975); *Mazer v. Stein*, 347 U.S. 201, 219 (1954); *Fox Film Corp. v. Doyal*, 286 U.S. 123, 127 (1932).

copyrighted materials.³¹ Instead, it often requires setting limits on the rights granted to authors.³² 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, contexts, connections, and links. We often do this subconsciously. Once we have processed such new intakes, we make them “our own.” We often feel that our 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. However, when we are the copiers, we often describe our work with different terms: “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 interaction with multiple stakeholders rather than authorship in silos. Creating new works involves human capital trained on existing knowledge, engagement with preexisting materials, and sharing a cultural ‘language.’ A cultural language consists of common words, phrases, slogans, symbols, melodies, stanzas, and images, enabling effective communication within a culture.³³

³¹ The notion of absolute property rights is conceptually incoherent in general. 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* 229 (Henry Smith *et al.* eds., 2022). Full internalization is not a conceptually incoherent notion, but an extremely unattractive goal especially with respect to information goods. See Julie E. Cohen, *Lochner in Cyberspace: The New Economic Orthodoxy of “Rights Management”*, 97 MICH. L. REV. 462, 502 (1998); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1037-8 (2005); Oren Bracha, *Give us back our Tragedy: Nonrivalry in Intellectual Property Law and Policy*, 19 THEORETICAL INQUIRIES L. 633 (2018).

³² *Id.* See also, Peter S. Menell, *Intellectual Property and the Property Rights Movement*, RESEARCH & DEVELOPMENT (2007).

³³ William Fisher grounds this intuition in what he calls “the cultural theory of copyright law.” William W. Fisher, *Recalibrating Originality*, 54 HOUS. L. REV. 437, 452. also William Fisher, *Theories of Intellectual Property*, in *NEW ESSAYS*

Consequently, using copyrighted materials is an essential input in any creative process. This creative process involves learning from preexisting materials, applying existing styles, and referencing shared symbols and pervasive themes to communicate fresh meanings.³⁴ 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 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.³⁵

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.³⁶

IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 168 (Stephen R. Munzer ed., 2001); William W. Fisher, *When Should We Permit Differential Pricing of Information?*, 55 UCLA L. REV. 1, 33–4 (2007); William W. Fisher, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1746, 1752 (1988); William Fisher, *Cultural Theory: Premises*, COPYRIGHTX (last modified Jan. 24, 2016), <http://copyx.org/lectures/>.

³⁴ See generally JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE (2012); ROSEMARY J. COOMBE, THE CULTURAL LIFE OF INTELLECTUAL PROPERTY: AUTHORSHIP, APPROPRIATION AND THE LAW (1998); Niva Elkin-Koren, *Cyberlaw and Social Change: A Democratic Approach to Copyright Law in Cyberspace*, 14 CARDOZO ARTS & ENT. L. J. 215 (1996).

³⁵ Litman, *The Public Domain*, *supra* note 15, at 966–7.

³⁶ See MARTIN SENFTLEBEN, THE COPYRIGHT/TRADemark INTERFACE: HOW THE EXPANSION OF TRADEMARK PROTECTION IS STIFLING CULTURAL

From an economic perspective, copyright law achieves this balance by weighing the social benefits of restricting access to copyrighted expression (incentivizing initial creation) against the social costs incurred by such restriction (frustrating subsequent creation).³⁷ 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.³⁸

The social costs associated with protecting some expressive compositions are always prohibitive.³⁹ For this reason, copyright law never protects “idea[s], procedure[s], process[es], system[s], method[s] of operation, concept[s], principle[s], or discover[ies].”⁴⁰

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, focusing on the fundamental rights of the individual author. See ABRAHAM DRASSINOWER, *WHAT’S WORNG WITH COPYING?* (2015).

³⁷ William M. Landes, *Copyright*, in *A HANDBOOK OF CULTURAL ECONOMICS* 132, 132–4 (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 14, 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.”); U.S. CONST. ART. I, § 8, cl. 8. See 471 U.S. 539, at 580; 464 U.S. 417 at 429; *Zacchini v. Scripps-Howard Broadcasting Co.*, 433 U.S. 562, 577 n.13 (1977) (dictum); *New Era Publications Int’l v. Henry Holt Co.*, 873 F.2d 576, 584 (2nd Cir.), *reh’g denied*, 884 F.2d 659 (2nd Cir. 1989); *United Video, Inc. v. F.C.C.*, 890 F.2d 1173, 1191 (D.C. Cir. 1989); *Roy Export Co. Establishment of Vaduz v. Columbia Broadcasting Sys.*, 672 F.2d 1095, 1099–1100 (2d Cir.), *cert. denied*, 459 U.S. 826 (1982).

³⁸ Landes, *id.*; Jamie Lund, *Copyright Genericide*, 42 CREIGHTON L. REV. 132, 139 (2009) (“...[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...”).

³⁹ Copyright law usually does not refer to expressions that are never protected as “expressions.” Instead, they are called “ideas”. However, this binary distinction is illusory. Jones, *supra* note 14, at 567–8 (“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 (2016) (same); Lund, *supra* note 38 at 137 (noting that “courts interpreting Feist’s application of the merger doctrine have imputed a spectrum of copyright protection.”); Landes, *supra* note 37 at 350–51 (Ruth Towse ed., 2003); *Warren Publ’g, Inc. v. Microdos Data Corp.*, 115 F.3d 1509, 1515 n.16 (11th Cir. 1997) (en banc) (same).

⁴⁰ 17 U.S.C. § 102(b) (2006). This principle, known as the idea/expression dichotomy, traces back to the seminal Supreme Court case of *Baker v. Selden*, 101 U.S.

The same is generally true of facts.⁴¹ 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: it changes over time.⁴² Thus, expressive compositions worthy of strong legal pro-

99 (1879). On the importance of this principle see Samuels, *supra* note 15, at 322; Ashton Tate Corp. v. Ross, 728 F. Supp. 597, 601 (N.D. Cal. 1989); Narell v. Freeman, 872 F.2d 907, 910 (9th Cir. 1989); Warner Bros. v. Am. Broad. Cos., 654 F.2d 204, 208 (2nd Cir. 1981); 471 U.S. 539, at 589; 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] (1989); David E. Shipley, *Conflicts Between Copyright and the First Amendment After Harper & Row Publishers, Inc. v. Nation Enterprises*, 1986 B.Y.U. L. REV. 983, 987. The dichotomy of idea and expression is also accepted in English law. See WALTER A. COPINGER, F. E. SKONE JAMES ET AL., *COPINGER AND SKONE JAMES NN COPYRIGHT* 175-6 (12th ed. 1980).

⁴¹ 17 U.S.C. § 102(b) (2006), 471 U.S. 539, at 547 (“[N]o author may copyright facts or ideas.”); *Fin. Info., Inc. v. Moody’s Investors Serv., Inc.*, 751 F.2d 501, 504 (2d Cir. 1984); *Miller v. Universal City Studios, Inc.*, 650 F.2d 1365, 1368 (5th Cir. 1981); *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 974 (2nd Cir. 1980); NIMMER & NIMMER, *id.*, at § 2.11[A] (2013). 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 type 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 (2007) (advocating for an “essential facilities” type doctrine which would render created facts unprotectable when they become widely used as thus essential for follow-on authors); Wendy J. Gordon, *Reality as Artifact: From Feist to Fair Use*, 55 LAW & CONTEMP. PROBS. 93, 97 (1992) (similar). Genericity also explains why small linguistic features (such as slogans and neologisms) are unprotected ab initio: they immediately become essential for communication even if they are originated with the author. Cf. 37 C.F.R. § 202.1(a) (1989); Litman, *The Public Domain*, *supra* note 15, at 1013. For examples of unprotected neologisms see GYLES BRANDRETH, *THE JOY OF LEX: HOW TO HAVE FUN WITH 860,341,500 WORDS* 9 (1980). (aerosol; automation); WILLIAM SAFIRE, *ON LANGUAGE* 9 (2d ed. 1981) (deplane); *id.* at 282 (uptight); *id.* at 289 (pseudoevent). Consider also “palimony,” “yuppie,” “significant other,” and “Ms.”DD’).

⁴² But there are also other views. See e.g., Brief for Arthur R. Miller as Amicus curiae supporting respondent. *Google LLC v. Oracle Am. Inc.* 593 U.S. 1 (2021), at 4 (“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.”).

tection at one point in time may deserve weaker or no legal protection at a later point, and vice versa.⁴³ This notion is reflected in the limited duration of copyright, which eventually expires, unlike other propriety rights.⁴⁴ However, given the extraordinary length of copyrights, these changes may even occur during the lifespan of individual copyrighted works.⁴⁵

In practice, copyright law tailors the scope of legal protection it affords expressive compositions by their originality, applying numerous 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 prevalent. 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 associate genericity with trademark law. Trademark law protects the distinctive power of a symbol to identify the signified product or service. Therefore, in trademark law, the genericity principle dismisses legal protection from marks commonly used to become “generic,” namely synonymous with a general class of products or services.⁴⁶

⁴³ Lund, *supra* note 38, at 139; William M. Landes & Richard A. Posner *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 326, 347 (1989).

⁴⁴ Copyrights protection extends for a period of the life of the author plus seventy years. 17 U.S.C. § 302(a) (1994 & Supp. V 1999). The term for anonymous works and works made for hire is 120 years from the date of creation, or 95 years from the date of first publication, whichever expires first. *Id.* § 302(c). Prior to the passage of the Copyright Term Extension Act in 1999, the term of copyright was the life of the author plus 50 years, or, for anonymous works and works made for hire, 100 years from the date of creation or 75 years from the date of first publication. *Id.* §§ 302(a), (c) (amended 1998).

⁴⁵ *Cf.* Joseph S. Miller, *Hoisting Originality*, 31 CARDOZO L. REV. 452, 467 (2009) (suggesting that extensions in copyright length over the years justifies hoisting the threshold of originality). *See also* Cooter & Hacoen, *supra* note 12, at 214.

⁴⁶ *See, e.g.,* Miller Brewing Co. v. Falstaff Brewing Corp., 655 F.2d 5 (1st Cir. 1981) (“Lite” beer); King-Seely Thermos Co. v. Aladdin Indus., Inc. 321 F.2d 577 (2nd Cir. 1963) (“Thermos” bottle). For analysis *see* William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J. L. & ECON.

In this section, we explain that the notion of originality in copyright law, and other scope delineating copyright doctrines, also adheres to similar genericity dynamics.⁴⁷ The more widely expressive compositions of copyrighted works are embedded in a culture and incorporated into existing expressions, the less vigorously copyright law will likely protect them.⁴⁸ Thus, copyright law will protect works that incorporate many “generic” expressive compositions less zealously than works that contain few or no such compositions, and vice versa.⁴⁹

265, 291-6 (1987). *See generally* J. THOMAS MCCARTHY, TRADEMARKS AND UNFAIR COMPETITION § 12 (2nd ed. 1984).

⁴⁷ Others have observed that there is a common thread underling seemingly disparate copyright doctrines. *Cf.* Samson Vermont, *The Sine Qua Non of Copyright is Uniqueness, not Originality*, 20 TEXAS INTLL. PRO. L. J. 327 (2012) (asserting that uniqueness is copyright’s main theory); Robert A. Gorman, *Fact or Fancy? The Implications for Copyright*, 29 J. COPYRIGHT SOC’Y 560, 560–61 (1982) (asserting there is a common ground to copyrightability, infringement, and defenses of privilege in copyright law); John Shepard Wiley, Jr., *Copyright at the School of Patent*, 58 U. CHI. L. REV. 119–20 (1991) (same). 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, 1066-7, 1101 (1989). Other have since applied this dynamic more broadly. *See e.g.*, Stephen Preonas, *Mergicide, 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) (calling it “mergicide”); Timothy S. Teter, *Merger and the Machine: An Analysis of the Pro-Compatibility Trend in Computer Software Copyright Cases*, 45 STAN. L. REV. 1061, 1088-97 (1993) (calling it “dynamic merger”). *See also* 1 PAUL GOLDSTEIN, GOLDSTEIN ON COPYRIGHT § §2.3.2.1 at 2:41 (3rd ed. 2005) (discussing application of merger to de facto standards); Lund, *supra* note 38, at 153; Vermont, *id.*, at 357 (calling it the dominance principle). 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 more detailed overview *see*, Hacoen et. al., *supra* signature note.

⁴⁸ Lund, *supra* note 38, at 153 (“Copyright genericide occurs when diminishing means of expression cause a thin copyright to get thinner, possibly to the point of nonexistence.”).

⁴⁹ 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 *The Wizard of Oz* such as “*Toto, I’ve a feeling we’re not in Kansas anymore,*” will grow generic faster the Fleming’s entire masterpiece. Nevertheless, over time the scope of protection for Fleming’s entire work would also be

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.⁵⁰ 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-being. Society has no interest in protecting generic expressions that affect no advancements in art, literature, education, or other creative endeavors.⁵¹ Thus, at least when expressions become generic before copyright protection is sought, the social benefits associated with protecting them are nil.⁵²

Legal protection for non-generic or *unique* expressive compositions is justified but only if and to the extent that their non-genericity is preserved. The more widely expressive compositions are used

reduced. 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. *Cf.* Hughes, *supra* note 16 at 605; GOLDSTEIN, *supra* note 47, § 2.7.3 at 2:96 (noting that “the shorter a phrase is. . . the more likely it is to constitute an idea rather than an expression”); at 2:96 (“The shorter a phrase is, the less likely it is to be original . . .”). *See* 37 C.F.R. § 202.1(a) (2004) (simple expressive compositions such as names, titles, and other pure designators are categorically excluded from copyright protection). *Cf.* Justin Hughes, *Fair Use Across Time*, 50 UCLA L. REV. 775, 775–8 (2003) (suggest narrowing copyright scope for complex works over time); *Warner Bros. v. Am. Broad. Co.*, 720 F.2d 231, 236, 242 (2nd Cir. 1983) (citing *D.C. Comics, Inc.*, 205 U.S.P.Q. (BNA) 1177) (refusing to find a “generic” term unprotected, but nevertheless finding that the use of the term was lawful as it constituted fair use); *Alberto-Culver Co. v. Andrea Dumon, Inc.*, 466 F.2d 705 (7th Cir. 1972) (examining copyright on short phrases as well as copyright on different labels and aspects of labels); *Perma Greetings, Inc. v. Russ Berrie & Co.*, 598 F. Supp 445, 448–49 (E.D. Mo. 1984) (finding that phrases such as “hang in there,” “message,” and “along the way take time to smell the flowers” are unprotected); *Alexa der v. Haley*, 460 F. Supp. 40, 45–46 (S.D.N.Y. 1978).

⁵⁰ A cost benefit analysis is implicit in the antitrust flavor of some of the conceptual configuration of the genericide doctrine. *See e.g.*, Fisher, *Recalibrating Originality*, *supra* note 33 at 449 (conceptualize originality as “the degree to which the work reflects the exercise of choice by the author.”); Vermont, *supra* note 47, at 357 (conceptualizing originality as limited by market dominance). *See generally* Lund, *supra* note 38.

⁵¹ *See e.g.*, Miller, *supra* note 45, at 464; Parchomovsky & Stein, *supra* note 12, at 1517; Fisher, *Recalibrating Originality*, *supra* note 33 at 454.

⁵² *See infra* notes 68–71, and accompanying text.

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 interoperability⁵³ or participate in established expressive discourses.⁵⁴ In other words, copyrighted expressions generate network externalities; they become more valuable to users the more they are used.⁵⁵

⁵³ In the seminal case of *Baker v. Selden*, the Supreme court recognized that once Selden designed his successful bookkeeping system, Baker's design choices for arranging columns and headings to implement that same system were constrained by the choices that Selden had made. *Baker*, at 101. See also Brief for Pamela Samuelson as Amicus curiae supporting Petitioner. *Google LLC v. Oracle Am. Inc.* 593 U.S. 1 (2021), at 26; FINAL REPORT OF THE NATIONAL COMMISSION ON TECHNOLOGICAL USES OF COPYRIGHTED WORKS, CONTU 20 (1978) Report at 20; *Apple Computer, 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 (1st Cir. 1995); GOLDSTEIN, *supra* note 47, § 2.3.2.1 (analogizing the merger to trademark genericide); See also Menell, *supra* note 47, at 1101; Teter, *supra* note 47, at 1088. But see NIMMER & NIMMER, *supra* note 40, at § 13.03[F][2][e] (rejecting de facto standards as limit on copyright).

⁵⁴ Litman, *The Public Domain*, *supra* note 15, at 996–97 (noting that “Some authors have suggested that ideas are not protected because of the strictures imposed on copyright by the first amendment.”); 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 40, at 987–91; Rochelle C. Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation* 65 NOTRE DAME L. REV. 397, 413–14 (1990). See generally Robert Denicola, *Copyright and Free Speech: Constitutional Limitations on the Protection of Expression*, 67 CAL. L. REV. 283 (1979); Celia Goldwag, *Copyright Infringement and the First Amendment*, 79 COL. 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. (ASCAP) 43 (1969); James L. Swanson, *Copyright Versus the First Amendment: Forecasting an End to the Storm*, 7 LOY. L.A. ENT. L. REV. 263 (1987); Comment, *Copyright and the First Amendment: Where Lies the Public Interest?*, 59 TUL. L. REV. 135 (1984).

⁵⁵ Glynn S. Lunney Jr, *Fair Use and Market Failure: Sony Revisited*, 82 B.U. L. REV. 975, 1024 (2002); Vermont, *supra* note 47, at 357.

Generic expressions are also costlier to administer compared to non-generic expressions.⁵⁶ 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.⁵⁷

While courts rarely invoke the genericity principle explicitly, this principle is nevertheless baked into copyright law's originality standard and related scope-delineating doctrines.⁵⁸

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

⁵⁶ See e.g., Lund, *supra* note 38, at 133 n.32; Parchomovsky & Stein, *supra* note 12, at 1520 (discussing negotiation costs).

⁵⁷ Litman, *The Public Domain*, *supra* note 15, at 996–97. 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. Cf. Brief for Arthur R. Miller, *supra* note 45, 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. Rierison, *Confronting the Genericism Conundrum*, 28 *CARDOZO L. REV.* 1789, 1826 (2007) (invoking such criticism in the trademark context). See generally Lund, *supra* note 38, For example, copyrighted expressions may become generic as a result of permissible licensing schemes which benefits the copyright owners, as well as by a widespread infringing activity which damages them. Cf. Rebeca Tushnet, *Copy This Essay; How Fair Use Doctrine Harms Free Speech and How Copying Serves it*, 114 *YALE L. J.* 535 (2004). It is unclear whether or not this 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. Lund, *supra* note 38, at 151. On the other hand, making copyright owners automatically accountable for all forms of genericity might induce them to become overprotective of their rights. Accordingly, copyright owners may reduce public access to their copyrighted works, thwarting the objectives of copyright law. Cf. Desai & Rierison, *id.* at 1826.

⁵⁸ *Supra* note 47.

to copyrighted works incorporating generic or partially generic expressive compositions.⁵⁹

1. Eligibility Doctrines

Expressive compositions are copyright eligible only if they are original and fixed in a tangible medium.⁶⁰ Courts interpreted originality to require two conditions: independent authorship and a modicum of creativity.⁶¹ Accordingly, facts and discoveries are never eligible for copyright protection because they are not independently authored.⁶²

⁵⁹ See e.g., *SAS v. WPL* (“copyright 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.”).

⁶⁰ 17 U.S.C. § 102(a) (1990) (“original work[] of authorship fixed in any tangible medium of expression.”); *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.”) [hereinafter *Feist*]. See also Litman, *The Public Domain*, *supra* note 15, at 996-97 (noting that “originality is a keystone of copyright law.”); GOLDSTEIN, *supra* note 47, § 1.2.2.3; Leon R. Yankwich, *Originality in the Law of Intellectual Property (Its Meaning from a Legal and Literary Standpoint)*, 11 F.R.D. 457, 457 (1951) (“[O]riginality is at the basis of the recognition of the rights of the author. It is the measure and boundary of protection.”). See also *L. Batlin & Son, Inc. v. Snyder*, 536 F.2d 486, 489–90 (2nd Cir. 1976) (quoting *NIMMER & NIMMER*, *supra* note 40, §10, at 32) (“[T]he one pervading . . . prerequisite to copyright protection. . . is the requirement of originality—that the work be the original product of the claimant.”)

⁶¹ *Feist*, *supra* note 60 at 345. The originality standard, guarantees that the work is attributed to an author who contributed something of his own. GOLDSTEIN, *supra* note 47, § 2.2.1.4.

⁶² *Feist*, *supra* note 60 at 1288–89 (O’Connor for a unanimous court (stating that facts are categorically excluded from copyright protection because they are never originated with the author. They are “discovered” rather than “authored.”). This statement is not true with respect to “constructed” or “created” facts which can be unique and thus original. See e.g., Litman, *The Public Domain*, *supra* note 15, 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 *supra* note 41.

On the other hand, authored original expressions are eligible for copyright protection only if they satisfy a minimal threshold of creativity.⁶³ This requirement ensures that the expressive compositions originated from the author. A work must be original in that it does not simply reflect such widely recognized compositions that would prevent it from being reasonably attributed to a singular creator, justifying restricted use by non-owners.

In the seminal case *Feist v. Rural*, the Supreme Court provided a negative definition of creativity.⁶⁴ This definition effectively equates creativity with non-genericity.⁶⁵ 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.”⁶⁶ The 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.”⁶⁷

Several scholars have argued that *Feist*’s definition of originality is implicitly concerned with the novelty of expressive compositions, not their genericity.⁶⁸ Applying a novelty standard in practice

⁶³ *Feist*, *supra* note 60, at 363. It is settled that *Feist* provides a constitutional, not a statutory, creativity minimum. *See, e.g.*, Pamela Samuelson, *The Originality Standard for Literary Works Under U.S. Copyright Law*, 42 AM. J. COMP. L. 393, 395 (1994).

⁶⁴ Howard B. Abrams, *Originality and Creativity in Copyright Law*, 55 LAW & 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, 334 (2017).

⁶⁵ *Cf.* Miller, *supra* note 45, at 481–2 (arguing that *Feist* articulation of creativity resonates with patent-laws’ non-obviousness requirement); EDWARD C. WALTERSCHEID, *THE NATURE OF THE INTELLECTUAL PROPERTY CLAUSE: A STUDY IN HISTORICAL PERSPECTIVE* 398–400 (2002) (similar); 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) (similar). Other commenters would have taken this trend further. *See e.g.*, Parchomovsky & Stein, *supra* note 12 (offering to tailor protection to originality more zealously).

⁶⁶ *Feist*, *supra* note 60, at 362–3.

⁶⁷ *Id.*

⁶⁸ *See e.g.*, Oren Bracha & Talha Syed, *Beyond the Incentive–Access Paradigm? Product Differentiation & Copyright Revisited* 92 Tex. L. Rev. 1841, 1910 (2014); Fisher, *Recalibrating Originality*, *supra* note 33 at 460; Miller, *supra* note

would mean that copyright law—similarly to patent law—would evaluate the originality of expressive compositions of copyrighted works at the time of their creation.⁶⁹ 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.”⁷⁰ 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.”⁷¹

However, genericity does not always settle by the time copyrighted works are created.⁷² Sometimes, genericity creeps in gradually during the extremely long lifespan of copyrighted works.⁷³ As Jamie Lund explains, expressive compositions may become generic when overused to become stock, standardized, or essential.⁷⁴ When

45, at 477. 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 (whether if this investigation is done at the patent office or later during infringement litigation). Copyright law is different, it investigated originality during infringement at the time of the infringement.

⁶⁹ As emphasized by Lord Jacob in the English patent case *Actavis v. Merck* [2008] R.P.C. 26 at para 119, “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. . .” 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 (courts or the Patent Office) appreciation of the true value of inventions. *See e.g.*, Baruch Fischhoff, *For Those Condemned to Study the Past: Heuristics and Biases in Hindsight*, in DANIEL KAHNEMAN, PAUL SLOVIC, & AMOS TVERSKY, EDS, *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 335, 341 (Cambridge 1982).

⁷⁰ Miller, *supra* note 45, at 462.

⁷¹ Fisher, *Recalibrating Originality*, *supra* note 33 at 464.

⁷² Lund, *supra* note 38, at 139.

⁷³ *See supra* notes 44–45, and accompanying text.

⁷⁴ Lund, *supra* note 38, at 131–32.

this occurs, a host of scope-delineating doctrines such as useful article,⁷⁵ merger,⁷⁶ and *scènes à faire*⁷⁷ dissipate the legal protection afforded to expressive compositions that turn generic.⁷⁸

As Justice Boudin of the United States Court of Appeals for the First Circuit explained in *Lotus v. Borland*,⁷⁹ while software interfaces can be copyrighted *ab initio*,⁸⁰ they might lose their legal protection over time if they become popular enough to be considered standard.⁸¹ Accordingly, the Court held that the plaintiff software

⁷⁵ 17 U.S.C. §101 (2010) (defining “useful article”). Copyright protection do not extend to the utilitarian aspects of the work. *Id. Baker*, at 104–05, *superseded by statute*, 17 U.S.C. § 102 (2006), *as recognized in* *Richmond Homes Mgmt. v. Raintree, Inc.*, 862 F. Supp. 1517, 1524 (W.D. Va. 1994). *See generally* Lloyd L. Weinreb, *Copyright for Functional Expression*, 111 HARV. L. REV. 1149 (1998).

⁷⁶ *Kregos v. Assoc. Press*, 937 F.2d 700, 705 (2nd Cir. 1991); *New York Mercantile Exch. Inc. v. Intercontinental Exchange, Inc.*, 497 F.3d 109, 116–8 (2nd Cir. 2007); *Allen v. Academic Games League of America, Inc.*, 89 F.3d 614, 617-18 (9th Cir. 1996).

⁷⁷ 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.” *See e.g.*, *Schwarz v. Universal Pictures Co.*, 85 F.Supp. 270, 278 (S.D.Cal.1945); *Rice v. Fox Broad. Co.*, 330 F.3d 1170, 1175 (9th Cir. 2003); *Gaiman v. McFarlane*, 360 F.3d 644, 659 (7th Cir. 2004); 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) (“[b]oth the copyright law originality requirement and the patent law non-obviousness requirement focus on whether the derivative work is the result of conventions familiar to creators working in the relevant culture.”); Litman, *The Public Domain*, *supra* note 15, at 996–7 (noting that “scenes a faire seems to lie more in their triteness than their necessity.”).

⁷⁸ Lund, *supra* note 38, at 132 (2009) (noting that courts, implicitly “apply the principles of genericide primarily through the dynamic application of limiting doctrines such as the idea/expression dichotomy, scenes a faire, and the merger doctrine.”).

⁷⁹ 49 F.3d 807 (1st Cir. 1995).

⁸⁰ Indeed, Computer programs are copyrightable as long as they meet the other requirements of the Copyright Act. *See Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1247–49 (3rd Cir. 1983) (stating source and object code are copyrightable). *See also* *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852, 855 n.3 (2nd Cir. 1982); *Williams Elecs. Inc. v. Artic Int'l, Inc.*, 685 F.2d 870, 877 (3rd Cir. 1982); *Digital Commc'ns Assocs., Inc. v. Softklone Distrib. Corp.*, 659 F. Supp. 449 (N.D. Ga. 1987); *Hubco Data Prods., Corp. v. Mgmt. Assistance, Inc.*, 219 U.S.P.Q. (BNA) 450, 454 (D. Id. 1983); *Midway Mfg. Co. v. Strohon*, 564 F. Supp. 741, 750 (N.D. Ill. 1983); *GCA Corp. v. Chance*, 217 U.S.P.Q. (BNA) 718, 720 (N.D. Cal. 1982).

⁸¹ 49 F.3d 807, 815–18 (1st Cir. 1995), Lund, *supra* note 38, at 143; Menell, *supra* note 47, at 1066–67, 1101.

designer Lotus's menu command hierarchy for its spreadsheet program, Lotus 1-2-3, was not copyrightable because it had become a *method of operation*.⁸²

Similar dynamics apply to mundane statements,⁸³ ingrained cultural themes,⁸⁴ stock characters,⁸⁵ acknowledged artistic styles,⁸⁶ and common harmonic progressions⁸⁷ or chord arrangements.⁸⁸ 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.⁸⁹ 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.⁹⁰

⁸² The majority opinion did not prescribe to the concurrence and believed the interfaces were unprotected *ab initio*. 49 F.3d 807, 815 (1st Cir. 1995), *Lotus Dev. Corp. v. Borland Int'l (Lotus III)*, 516 U.S. 233 (1996).

⁸³ *Acuff-Rose Music v. Jostens Inc.*, 988 F. Supp. 289, 291 (S.D.N.Y. 1997) (finding that the statement "You'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 the protection granted for a trivial compilation of ideas, expression and images is thin, available only against complete literal copying); *Litman, The Public Domain*, *supra* note 15, at 996–97.

⁸⁴ *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 122 (2nd Cir. 1930), *cert. denied*, 282 U.S. 902 (1930); *Shipman v. R.K.O. Radio Pictures, Inc.*, 100 F.2d 533 (2nd Cir. 1938); *Warshawsky v. Carter*, 132 F.Supp. 758 (D.D.C. 1955); *Roe-Lawton v. Hal E. Roach Studios*, 18 F.2d 126 (S.D.Cal. 1927); *Ware v. Columbia Broad. Sys.*, 155 U.S.P.Q. (BNA) 413 (Cal. App. 1967).

⁸⁵ *Walker v. Time Life Films, Inc.*, 784 F.2d 44, 50 (2nd 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 'scenes a faire,' that is, scenes that necessarily result from the choice of a setting or situation.").

⁸⁶ *See, e.g., Franklin Mint Corp. v. National Wildlife Art Exch.*, 575 F.2d 62 (3d Cir.), *cert. denied*, 439 U.S. 880 (1978).

⁸⁷ *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.") (internal citations omitted).

⁸⁸ *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 (2nd Cir. 1995); *Smith v. George E. Muehlenbach Brewing Co.*, 140 F. Supp. 729, 731 (W.D. Mo. 1956).

⁸⁹ *Daly v. Palmer*, 6 F. Cas. 1132, 1132 (C.C.S.D.N.Y. 1868) (No. 3552).

⁹⁰ 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 plaintiffs play attractive, as a representation on the stage." *Daly*, 6 F. Cas. at 1138.

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."⁹¹

Similarly, in *Acuff-Rose v. Jostens*,⁹² 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 "You've got to stand for something, or you'll fall for anything," sued the defendant for using a similar slogan, "U Got 2 Stand 4 Something," in a promotional campaign for selling class rings.⁹³

While the Court acknowledged that the defendant indeed copied the expression from the plaintiff's song, the Court denied that copyright infringement occurred.⁹⁴ 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.⁹⁵

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 at a time.⁹⁶ As wittily captured in The Axis of Awesome's *Four Chords* medley, most commercially viable songs rely on standardized chord relationships and progression combinations.⁹⁷ The *Wikipedia* page for the popular I–V–vi–IV progression, for example, lists more than one hundred and fifty songs containing multiple, repeated uses of this one string of popular chords, including Bob Marley's *No Woman, No Cry*, and

⁹¹ *Alexander v. Haley*, 460 F. Supp. 40, 45 (S.D.N.Y. 1978).

⁹² 155 F.3d 140 (2d Cir. 1998).

⁹³ *Acuff-Rose Music, Inc. v. Jostens, Inc.*, 155 F.3d 140 (2d Cir. 1998).

⁹⁴ *Id.* at 293–95.

⁹⁵ *Id.*

⁹⁶ Although there are 12 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 7 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 12 pitches involved in music protected by copyright).

⁹⁷ THE AXIS OF AWESOME, *FOUR CHORDS*, (Not on Label 2011).

Justin Bieber’s *Ghost*.⁹⁸ Like Daly’s railroad scene, most courts consider these chord arrangements generic and thus copyright ineligible.

Copyright law’s originality standard and related eligibility doctrines in foreign jurisdictions are even more explicit in their adherence to the genericity principle.⁹⁹ Austrian copyright law, for example, affords photographs legal protection only if they differ significantly from preexisting photographs.¹⁰⁰ Similarly, until recently, copyright law in Germany granted legal protection to computer software only to the extent that it embodied different expression from that of average computer programs.¹⁰¹ Most notably, Swiss copyright law prescribes legal protection only to works that show “statistical uniqueness” (“statistische Einmaligkeit”) compared to preexisting works.¹⁰²

⁹⁸ *I-V-vi-IV progression*, WIKIPEDIA, https://en.wikipedia.org/wiki/I%E2%80%93V%E2%80%93vi%E2%80%93IV_progression (last visited Jul. 31, 2023).

⁹⁹ Fisher, *Recalibrating Originality*, *supra* note 33 at 439.

¹⁰⁰ Fisher, *id.*; Roman Heidinger, *The Threshold of Originality Under EU Copyright Law*, HONG KONG WORKSHOP 1, 3 (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>].

¹⁰¹ Fisher, *Recalibrating Originality*, *supra* note 33 at 439; Bundesgerichtshof [BGH] [Federal Court of Justice] Sep. 5, 1985, 94 Entscheidungen des Bundesgerichtshofes in Zivilsachen [BGHZ] 276, 1985 (Ger.). 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—Copyright Issues Requiring Immediate Action*, at 187–88, COM (1988) 172 final (Jun. 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] Mar. 3, 2005, IX ZR 111/02, (Ger.), <http://www.rechtsberaterhaftung.de/PDF/442.pdf> [<https://perma.cc/EEM8-5TUE>].

¹⁰² Fisher, *Recalibrating Originality*, *supra* note 33 at 448. See Bundesgericht [BGer] [Federal Supreme Court] Apr. 1, 2010, 136 ENTSCHEIDUNGEN DES SCHWEIZERISCHEN BUNDESGERICHTS [BGE] III 225 (Switz.); BGer Feb. 13, 2008, 134 BGE III 166; BGer Sept. 5, 2003, 130 BGE III 168. The courts’ stance on this issue derives at least in part from the unusual way in which the Swiss copyright statute of 1992 defined the set of works subject to copyright protection: “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).

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 the cover shoot of *Harper's Bazaar Vietnam* magazine in 2017.¹⁰³ 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.¹⁰⁴ For the reasons discussed above, works of partial expressive genericity merit only partial legal protection.¹⁰⁵ Two copyright doctrines achieve this goal by limiting the scope of legal protection for partially generic works during infringement litigation.¹⁰⁶ 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 unauthorized copying of protected expression.¹⁰⁷ 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.¹⁰⁸

¹⁰³ Jacqueline Tobin, *Photographer Jingna Zhang Loses Plagiarism Case Against Artist*, RF + WPPI (Dec. 9, 2022) <https://www.range-finderonline.com/news-features/industry-news/photographer-jingna-zhang-loses-plagiarism-case-against-artist/>; Espie Angelica de Leon, *Singaporean Photographer Loses in Luxembourg Plagiarism Trial Against Local Painter*, ASIAIP (23 Dec. 2022) <https://asiaiplaw.com/section/news-analysis/singaporean-photographer-loses-in-luxembourg-plagiarism-trial-against-local-painter>.

¹⁰⁴ Or other unprotected elements such as facts. *See supra* note 41.

¹⁰⁵ *See supra* notes 48–55, and accompanying text.

¹⁰⁶ *See e.g.*, *Brown Bag Software v. Symantec Corp.*, 960 F.2d 1465, 1475 (9th Cir.) (noting that “analytic dissection” during the infringement analysis serves “the purpose of defining the scope of plaintiff's copyright.”).

¹⁰⁷ *See generally* Lemley, *Our Bizarre System for Proving Copyright Infringement*, *supra* note 14.

¹⁰⁸ *See e.g.*, *Sid & Marty Krofft TV Prod., Inc. v. McDonald's Corp.*, 562 F.2d 1157, 1164 (9th Cir. 1977); *Brown Bag Software v. Symantec Corp.*, 960 F.2d 1465, 1472 (9th Cir. 1992).

Courts sometimes 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.¹⁰⁹ The most sophisticated framework for executing this task was established in *Computer Associations Int’l, Inc. v. Altai, Inc.*¹¹⁰ 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.¹¹¹

When evaluating how much similarity is considered “substantial” in the final step of this analysis, courts sometimes consider the copyrighted work’s originality (non-genericity) level.¹¹² The less

¹⁰⁹ Different circuits apprise substantial similarity differently. A complete consideration of the different approaches exceeds the scope of this article. For an overview see Lemley, *Our Bizarre System for Proving Copyright Infringement*, *supra* note 14.

¹¹⁰ 982 F.2d 693 (2d Cir. 1992). This framework was initially created for copyrighted computer programs. Most recently it was applied in *SAS v. WPL*. See also *Eng’g Dynamics*, 26 F.3d at 1335 (5th Cir. 1994); *Gates Rubber*, 9 F.3d at 823, 834 (10th Cir. 1993); *Computer Mgmt. Assistance Co. v. Robert F. DeCastro, Inc.*, 220 F.3d 396, 399–400 (5th Cir. 2000).

¹¹¹ Genericity plays a part in this filtration stage because some expressive elements that were non-generic (original) at the time the copyrighted work was first created, and may become generic (non-original) by the time of the infringement analysis. However, the current filtration practice is still binary and therefore limited. See *infra* notes 39, 204 and accompanying text.

¹¹² Mark A. Lemley, *Convergence in the Law of Software Copyright*, 10 HIGH TECH. L. J. 1, 28 (1995) (Acknowledging the sliding scale between 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 Computer v. Microsoft Corp.*, 35 F.3d 1435, 1443 (9th Cir. 1994) (“the 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 38, at 147; *Satava v. Lowry*, 323 F.3d 805, 812 (9th Cir. 2003); *Beaudin v. Ben & Jerry’s Homemade, Inc.*, 95 F.3d 1, 2 (2nd Cir. 1996). *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.”) He further explains, however, that “[w]hile works having a ‘thin’ copyright due

original the 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.¹¹³ This similarity gradient effectively narrows the scope of protection for partially copyrighted generic works.¹¹⁴ Works heavy with generic material will be infringed only if the allegedly infringing work is “virtually identical” to them.¹¹⁵ Courts have indeed applied this more stringent similarity standard in cases involving computer operation systems with typical graphic user interfaces,¹¹⁶ conventional photographs of commercial products,¹¹⁷ and standardly structured stand-up jokes.¹¹⁸

The fair use doctrine also tailors copyright scope for partially generic works.¹¹⁹ Even if a plaintiff successfully proves substantial copying of protectable expressive compositions, a defendant may still escape infringement if her coping constitutes fair use.¹²⁰ Courts may find fair use if the expressive compositions a defendant copied

to a minimal amount of creative material may indeed only be infringed by close copying, this is because the majority of the work is unprotectable.”).

¹¹³ Lemley, *Convergence in the Law of Software Copyright*, *supra* note 112, at 31 (1995) (discussing the interaction of fair use with substantial similarity); *Idema v. Dreamwork, Inc.*, 162 F. Supp. 2d 1129, 1178 (C.D. Cal. 2001) (explaining that a more stingiest similarity applies for copyrighted works of “largely ‘unprotectable’ elements.”).

¹¹⁴ 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 112, at § 9:166 n.9.

¹¹⁵ *See generally* Lemley, *Our Bizarre System for Proving Copyright Infringement*, *supra* note 14.

¹¹⁶ *Apple Computer v. Microsoft Corp.*, 35 F.3d 1435, 1443 (9th Cir. 1994); *Lund*, *supra* note 38, at 153–55.

¹¹⁷ *Ets-Hokin v. Skyy Spirits, Inc.*, 323 F.3d 763, 765–66 (9th Cir. 2003) (noting the limited “conventions of [a] commercial product shot.”).

¹¹⁸ *Kaseberg v. Conaco, LLC*, 260 F. Supp. 3d 1229, 1245 (S.D. Cal. 2017) (noting that some 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.”).

¹¹⁹ The primary lever that consider the level of the underlining work originality (or non-genericity) is the third factor (the “nature of the copyrighted work”). 17 U.S.C. § 107(2) (2012).

¹²⁰ *Bracha & Syed*, *supra* note 68, at 1894–95.

from the plaintiff are considered at the turning point of genericity.¹²¹ This was the case in *Oracle v. Google*.¹²²

In that case, Oracle claimed that Google infringed its copyrights in the Java program by copying 37 “declaring code” packages, which was part of the Java application program interfaces (APIs), without authorization.¹²³ 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.¹²⁴ This argument was similar to Judge Boudin’s concurrence in *Lotus*.¹²⁵

The Supreme Court disagreed. Instead, the Court assumed, *arguendo*, that the Java declaring code was original enough to merit copyright protection.¹²⁶ 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. 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 action was filed due to its established popularity among software programmers.¹²⁷ 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

¹²¹ Lemley, *Convergence in the Law of Software Copyright*, *supra* note 112, at 30.

¹²² 593 U.S. 1 (2021).

¹²³ *Id.* at 8.

¹²⁴ *See e.g.*, Brief for EFF (Electronic Frontier Foundation) as Amicus curiae supporting respondent, *Google LLC v. Oracle Am. Inc.* 593 U.S. 1, 23 (2021) [hereinafter *EFF brief*]; Brief for Peter S. Menell & David Nimmer as Amici curiae supporting petitioner. *Google LLC v. Oracle Am. Inc.* 593 U.S. 1 (2021), at 16. Affording copyright protection to the standardized Java API declarations, would effectively require millions of software developers “to learn a while new language to be able to use” the API labels. *EFF brief*, at 24 (citing Google’s expert Astrachan).

¹²⁵ Although Judge Boudin’s concurrence endorsed a fair use approach similar to the one taken in *Oracle v. Google*. *see Lotus Dev. Corp. v. Borland Inf.*, 49 F.3d 807, 821 (9th Cir. 1994) (Boudin, J., concurring); Lemley, *Convergence in the Law of Software Copyright*, *supra* note 112, at 31.

¹²⁶ *Oracle v. Google* at 1 (“we assume, for argument’s sake, that the material was copyrightable.”).

¹²⁷ *Id.* at 24 (“[u]nlike 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.”).

to participate in the established programming ecosystem.¹²⁸ For this reason, the Court decided that weakening Oracle’s exclusivity rights and allowing Google to use the Java declaring code was socially beneficial enough to outweigh the benefits of broad copyright exclusion in this context.¹²⁹

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.”

Justice O’Connor, 1985¹³⁰

As demonstrated in Part II, genericity is an implicit principle of copyright law. Accordingly, pervasive expressive compositions are

¹²⁸ *Id.* at 29 (noting the 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.”).

¹²⁹ *Id.* at 24 (In the court’s words, 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 the analysis. *See* 17 U.S.C. § 107 (1) (2012); *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 (2nd 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 10, at 1111 (asserting that justification of a fair use turns largely on whether the use is transformative). Indeed, if Google’s was copying Oracle’s code verbatim to create a mere imitation rather than a transformative new ecosystem, the court would likely find that Oracle’s appropriation rights are strong enough to exclude Google’s use. *Cf.* *Cooter & Hacoen*, *supra* note 12 (explaining that patent policy weakens patent rights against innovative uses, but strengthens these rights against consumption and production uses).

¹³⁰ *Harper & Row, Publishers, Inc. v. Nation Enter.*, 471 U.S. 539, 547 (1985).

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.

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.¹³¹ GenAI models seek to generalize expressive patterns 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.¹³²

Generalization by GenAI models indicates the expressive compositions that are prevalent in preexisting works and, therefore 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.

Subsection A explores the rise of GenAI technology. Subsection 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, subsection 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 (e.g., text, image, music, code) based on expressive input. Some images and text generated

¹³¹ For a deeper exploration of the concepts of generalization and memorization see e.g., Dan Ventura, *Mere Generation: Essential Barometer or Dated Concept?*, in 7 PROC. INT'L CONF. ON COMPUT CREATIVITY, 17–24 (2016).

¹³² Nicholas Carlini, et al., *Extracting Training Data from Large Language Models*, in 30 USENIX SEC. SYMP., 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.”).

with the aid of GenAI models even won prestigious awards.¹³³ These systems piggyback on the impressive capability of foundational models, such as OpenAI's GPT (*Generative Pre-trained Transformer*) or Google's BERT, to extrapolate patterns and structures from granular data.¹³⁴ Foundational models are large-scale neural network models pre-trained on colossal amounts of unlabeled data by self-supervised learning, often on a surrogate task.¹³⁵ These models learn generalizable and adaptable data representations that can be used for multiple downstream tasks.

In addition, the models' ability to *transfer learning*, namely, taking the "knowledge" learned from one task and applying it to another, cuts the high cost of training a model from scratch.¹³⁶ 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. ChatGPT, for instance, is built on OpenAI's foundational models GPT3.5 and GPT4 to enable bot-human interaction. ChatGPT could be fine-tuned further for a variety of more nuanced natural language processing tasks including language translation, classification, and text summarization.

While much of the hype associated with generative models is currently focused on Large Language Models (LLM), computer vision has been at the forefront of deep learning research in AI for many years.¹³⁷ As with text, foundational models in computer vision are pre-trained on raw data of different types (e.g., text, audio, colors) from diverse sources (e.g., uploaded from sensors, cameras, or scraped from the web). 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.¹³⁸

¹³³ Newsweek (April 2023) Boris Eldagsen a Germany Artist who used Artificial Intelligence won the SONY world photography photo award . [Alt. Newsweek citation: Boris Eldagsen, 'I Used AI to Beat Real Photos in a Major Competition', NEWSWEEK (Apr. 26, 2023, 08:30 AM).

¹³⁴ See generally Rishi Bommasani *et al.*, *On the Opportunities and Risks of Foundation Models*, CENTER RSCH. FOUND. MODELS (CRFM) (2021).

¹³⁵ *Id.* at 4.

¹³⁶ *Id.*

¹³⁷ Richmond Alake, *What AlexNet Brought To The World Of Deep Learning*, MEDIUM (Jul. 10, 2020), <https://towardsdatascience.com/what-alexnet-brought-to-the-world-of-deep-learning-46c7974b46fc>.

¹³⁸ Bommasani, *supra* note 134, at 4.

Generative models vary in their technical approach, including Generative Adversarial Networks (GANs), Variational Auto Encoders (VAEs), and diffusion models, like *DALL-E 2* and *Stable Diffusion*, which raise recent controversies in copyright law.¹³⁹ 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. These models have both a forward process and a reverse process. The forward diffusion process gradually adds Gaussian noise of different magnitudes (random pixels or distortion changes that affect the original image) to a clean data point. In the reverse process, the model learns to remove the added noise to generate a clear image.¹⁴⁰ Figure 5 below visualizes the latter process.

Foundational models were initially aimed at learning about data without supervision¹⁴¹ but were increasingly deployed for generative applications.¹⁴² Downstream 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.¹⁴³ 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.

¹³⁹ Viet Anh, *An overview of Generative AI in 2023*, AI CURIOUS (May 1, 2023), <https://aicurious.io/blog/2023-05-02-overview-of-generative-ai>.

¹⁴⁰ See generally Arpit Bansal et al., *Universal Guidance for Diffusion Models*, *ArXiv* abs/2302.07121 (2023).

¹⁴¹ Sam Bond-Taylor et al., *Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models*, *IEEE TRANSCON. PATTERN ANALYSIS & MACH. INTELL.* 1, 1 (2021).

¹⁴² *Id.* at 1 (“The central idea of generative modelling stems around training a generative model whose samples $x \sim p_{\theta}(x)$ come from the same distribution as the training data distribution, $x \sim p_d(x)$.”).

¹⁴³ See Bracha, *supra* note 11, at 10 (“The main purpose of [GenAI], by contrast, is to generate new informational goods”); MOHAK AGARWAL, *GENERATIVE AI FOR ENTREPRENEURS IN A HURRY* 5 (2023) (“While traditional AI is designed to recognize or classify existing data, generative AI is able to generate novel and diverse outputs based on a given set of input parameters or conditions”). For a technical definition see Bond-Taylor et al., *supra* note 141, at (“The central idea of generative modelling stems around training a generative model whose samples $x \sim p_{\theta}(x)$ come from the same distribution as the training data distribution, $x \sim p_d(x)$.”).

The value of GenAI models manifests 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. We derive inspiration from preexisting genres; we implicitly reference 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 not only been necessary for acquiring expressive skills but has also been essential to the creation process itself.

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.¹⁴⁴ For example, *Copilot* and *Midjourney* are trained on giant corpora of pre-written code and images from the *GitHub* open source code repository and the LAION 5B database, respectively.¹⁴⁵ These models assist users in generating code and images in response to “prompts,” namely textual inputs.

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. Fast forward to the 2020s, GenAI systems empower users to generate high-quality text, image designs, music, and code, challenging the dominance of professional creatives.

¹⁴⁴ Bommasani, *supra* note 134, at 4.

¹⁴⁵ Romain Beaumont, *Laion-5B: A New Era of Open Large-Scale Multi-Modal Datasets*, LAION (Mar. 31, 2022), <https://laion.ai/blog/laion-5b/>.

GenAI technology disrupts copyright law by forcing it to make three fundamental policy decisions.¹⁴⁶ First, the law must decide whether to afford authorship and ownership interests to expressive works created with the aid of GenAI models.¹⁴⁷ Second, the law must decide whether GenAI models can be trained on copyrighted works without authorization from their owners.¹⁴⁸ 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.¹⁴⁹

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 US Copyright Office.¹⁵⁰ The Office considers most GenAI outputs unprotected by copyright law for lack of human authorship.¹⁵¹ It also solicited public comments and scheduled multiple “public listening sessions” to further debate this topic and related issues.¹⁵²

The other two questions were recently invoked in several pending class action lawsuits. In one such lawsuit, plaintiffs claimed that the code-generating model *Copilot* infringes copyrights in the licensed code that the model was trained on.¹⁵³ Specifically, they claimed that the model infringes copyrights both because it was trained on copyrighted code without authorization and because it generated snippets of that same copyrighted code.

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

¹⁴⁶ See generally, Samuelson, *supra* note 11.

¹⁴⁷ See e.g. Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1200, 1224–5 (1986); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L. J. 343, 445 (2019).

¹⁴⁸ See e.g., Mark A. Lemley & Bryan Casey, *Fair Learning*, 99(4) TEX. L. REV. 743 (2021); Bracha, *supra* note 11.

¹⁴⁹ Samuelson, *supra* note 11; See Bracha, *supra* note 11, at 38.

¹⁵⁰ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. ??, 16190 (Mar. 16, 2023) (to be codified at 37 C.F.R pt. 202).

¹⁵¹ *Id.* at 16191.

¹⁵² *Copyright Office Launches New Artificial Intelligence Initiative*, COPYRIGHT.GOV (Mar. 16, 2023), <https://www.copyright.gov/newsnet/2023/1004.html>.

¹⁵³ Complaint Class Action & Demand for Jury Trial, Doe 1 v. GitHub, Inc., No. 06823 (N.D. Cal. Nov. 3, 2022).

images used for their training. Similar to the *Copilot* suit, the plaintiffs claim is that the models infringed copyrights both by training on copyrighted images in violation of licensing terms, and by generating allegedly unauthorized derivative images.¹⁵⁴ A third class action invoked similar claims against OpenAI’s flagship application *ChatGPT*.¹⁵⁵

On the academic front, most of the legal scholarship has focused on the first two questions mentioned above. 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 should be authorized. Some scholars believe that training with copyrighted material does not constitute infringement,¹⁵⁶ while others are more skeptical.¹⁵⁷ Similarly, some scholars are sympathetic to the idea that certain GenAI models’ outputs should remain unowned,¹⁵⁸ while others hold a more assertive stance.¹⁵⁹ The approach introduced in this article has

¹⁵⁴ Complaint Class Action & Demand for Jury Trial, *Andersen v. Stability AI Ltd.*, No. 00201 (N.D. Cal. Jan. 13, 2023).

¹⁵⁵ *Tremblay vs. OpenAI, Inc.*, No. 03223 (N.D. Cal. Jun. 28, 2023).

¹⁵⁶ See e.g., Lemley & Casey, *supra* note 148 (arguing that machine learning is fair use); See Bracha, *supra* note 11 (arguing that it is non-infringement).

¹⁵⁷ See generally, Benjamin L. W. Sobel, *Artificial Intelligence’s Fair Use Crisis*, 41 COLUM. J.L. & ARTS 1 (2017–2018) (arguing that machine learning is different than what courts have traditionally labeled non-infringing non-expressive use); 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*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4123309 (proposing imposing levy on the users of GenAI models to compensate human authors).

¹⁵⁸ See generally Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 GA. L. REV. 1669, (2023) (posing that if GenAI reduce the costs of creativity the justification for intellectual property protection is reduced).

¹⁵⁹ 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, Part VIII (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* (Jan. 23, 2023) (recommending attribution of authorship to AI software) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4185327; Shlomit Yanisky-Ravid &

important policy implications for both these questions, which we explore in Part IV.

The third policy 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 dismissed it by resorting to the standard copyright infringement tests.¹⁶⁰ This gap in scholarship was filled rapidly by interdisciplinary studies in computer science and law.¹⁶¹ These studies employed computational procedures to establish whether GenAI models' outputs either infringe or do not 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 overlook the non-infringing aspects of copyrighted materials. Indeed, several studies presume copyright infringement when GenAI produces outputs that significantly resemble copyrighted training data.¹⁶² 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.¹⁶³

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

Luis A. Velez- Hernandez, *Copyrightability of Artworks Produced by Creative Robots and Originality: The Formality-Objective Model*, 19 MINN. J.L. SCI. & TECH. 1 (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 (2017) (recommending a statutory amendment to create AI works made for hire).

¹⁶⁰ See e.g., Bracha, *supra* note 11, at 34 (“copying of a single work’s style should be analyzed under the 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 “)

¹⁶¹ See e.g., Carlini *et al.*, *supra* note 132; Gowthami Somepalli, *et al.*, *Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models*, ARXIV (2022).

¹⁶² Carlini *et al.*, *supra* note 132, at 6–7 (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.*, *id.* (accusing the GenAI models of “digital forgery”).

¹⁶³ Carlini *et al.*, *supra* note 132; see also Somepalli *et al.*, *supra* note 161.

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.¹⁶⁴

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 expressions. Copyright law permits copying of such unprotected compositions and no authorization of right holders is necessary for their use. 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.

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.¹⁶⁵ 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 these patterns without risking copyright infringement.

However, the approach taken in these studies is also incomplete.¹⁶⁶ While genericity narrows the scope of legal protection that copyright law affords widespread compositions, it cannot serve as a

¹⁶⁴ As explained in Part II.B., the legal protection of such expressive compositions is thin.

¹⁶⁵ From a computer science perspective these approaches piggyback on the field of differential privacy. See Hacoen et al., *supra* signature note.

¹⁶⁶ See Hacoen et al., *supra* signature note.

definitive binary test for copyright infringement.¹⁶⁷ To see why consider a hypothetical extension of *Goldsmith vs. Prince*, the case that opened this article.¹⁶⁸ Assume, for example, that after training a GenAI model on Andy Warhol’s “Prince Series,”¹⁶⁹ 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 infringe on Goldsmith’s copyright?¹⁷⁰ The answer is not straightforward.

Indeed, the plurality of the Prince derivatives may indicate that Goldsmith deserves narrower legal protection for her work.¹⁷¹ Nevertheless, this fact alone does not necessarily avail the GenAI model’s output of infringing Goldsmith copyright.¹⁷² 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 to cover commercial uses of derivative outputs, especially if these outputs were themselves generic (incorporate no additional originality).¹⁷³

In other words, attempts to harness computational approaches to establish binary tests for copyright infringement overlook the need

¹⁶⁷ The most similar copyright application to the differential privacy scenario was the clean room procedure used in *Oracle v. Google*. There, to prove that Google did not copy the implementation code from Oracle’s JAVA programming language, Google made sure that programmers did not “see” these elements during their construction of the Android code. If the Google team would have seen the entire JAVA code during the construction of the Android code, the court was less likely to rule that Google did not engage in unlawful copying even if Google could prove that their team *could have* construct the Android code even without seeing the JAVA code in its entirety.

¹⁶⁸ Andy Warhol Foundation for the Visual Arts, Inc. v. Lynn Goldsmith, 382 F.Supp.3d 312, 324 (S.D.N.Y. 2019).

¹⁶⁹ The Prince Series was a collection of 16 derivative works that Andy Warhol created based on Lynn Goldsmith photograph of Prince. *Id.*

¹⁷⁰ 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 (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.

¹⁷¹ *Supra* Part II.B.

¹⁷² *Infra* Part IV. A.

¹⁷³ *See supra* note 129.

for more nuanced analysis. As stated, resemblance to copyrighted materials alone is insufficient to establish copyright infringement. Similarly, resemblance to only 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 policy questions mentioned above. Nevertheless, originality measurements would be a valuable input for courts and policymakers, and would inform their legal analysis when grappling with these issues.¹⁷⁴

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 these processes take place exclusively in the silo of the human mind. 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 commonly expressive compositions appear in the GenAI models' datasets (the more "generic" they are), the more likely GenAI

¹⁷⁴ *Infra* Part IV.

models are to apply them when generating new works. Conversely, GenAI models are less likely to apply more rarely expressed (“original”) compositions in their datasets.

This data-driven “bias” is a fundamental feature of inductive machine learning and affects the way in which GenAI models generalize.¹⁷⁵ 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.¹⁷⁶

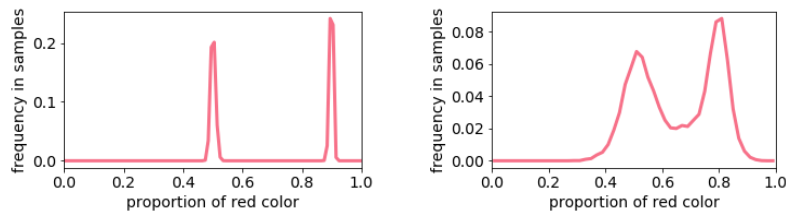


Figure 3: Input-output relationship in a single feature variance (number of occurrences). *Zhao et. al.*

The same dynamics may be demonstrated using an inpainting technique, which requires GenAI models to reconstruct missing parts from images.¹⁷⁷ As shown in Figure 4, when we tasked *Stable Diffusion* with completing the apple from René Magritte’s famous painting *The Son of Man*, the model reconstructed the image with human male faces rather than with apples.¹⁷⁸ This result indicates

¹⁷⁵ Yunzhe Liu (or: *et al.*), Rinon Gal, Amit H. Bermano, Baoquan Chen & Daniel Cohen-Or, *Self-Conditioned Generative Adversarial Networks for Image Editing*, ARXIV (2022).

¹⁷⁶ Shengjia Zhao, *et al.*, *Bias and Generalization in Deep Generative Models*, GITHUB: ERMONGROUP BLOG, (2019), <https://ermongroup.github.io/blog/bias-and-generalization-dgm/>.

¹⁷⁷ Robin Rombach et al., *High-Resolution Image Synthesis with Latent Diffusion Models*, ARXIV 8 (2021) (“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.”)

¹⁷⁸ We tried this exercise also with adding to the model the textual prompt “Magritte The Son of Man” but the outputs were still dominated by men’s faces.

that the model was trained (unsurprisingly) on many more images of men’s faces rather than men with apples in front of their faces.¹⁷⁹



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.¹⁸⁰ GenAI models apply these compositions “quickly” and “instinctively” when generating new works. Conversely, one can envision GenAI models as “thinking slow” when asked to generate original expressive compositions. GenAI models apply these compositions only after considerable “processing” and “deliberation.” 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:¹⁸¹

“[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

¹⁷⁹ This outcome may also demonstrates that the model’s training data might be biased in favor of men images wearing a suit and a 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.

¹⁸⁰ See generally, DANIEL KAHNEMAN, THINKING, FAST AND SLOW, (Farrar, Straus and Giroux, 2013).

¹⁸¹ Shengjia Zhao *et al.*, *Bias and Generalization in Deep Generative Models: An Empirical Study.* *Neural Information Processing Systems*, ARXIV (2018); James Bridle, *The Stupidity of AI*, THE GUARDIAN (Mar. 16, 2023), <https://www.theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt>

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. . . 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.”

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 overlooked.¹⁸² Because genericity confines the legal protection that copyright law affords expressive works, courts and policymakers could quantify originality and delineate the scope of copyright protections by measuring genericity.

¹⁸² For example, some legal scholars have used the technical term “memorization” in association with copyright infringement. *See e.g.*, Sag, *supra* note 11, at 1 (“models suggest that they are capable of memorizing and reconstituting works in the training data, something that is incompatible with non-expressive use”); Bracha, *supra* note 11, at 39 (“Professionals sometimes talk about the system ‘memorizing’ a specific work”); Jannat Un Nisa, *ChatGPT Is About To Face Some Copyright Issues After ‘Memorizing’ These Books*, WONDERFUL ENGINEERING (May 4, 2023) <https://wonderfulengineering.com/chatgpt-is-about-to-face-some-copyright-issues-after-memorizing-these-books/> (noting that “AI-generated output is too similar to the input data, which [is] refer[red] to as ‘memorization’” And citing Law professor Tyler Ochoa saying that “when that occurs, there will be lawsuits, and it will almost certainly constitute copyright infringement.”). *See also* Ken K. Chang, et. al., *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4*, ARXIV 1 (2023) (noting “[t]he ability of these models to memorize an unknown set of books.”). However, this reference is misleading. Because GenAI models’ tendency to memorize expressive composition is correlated with these compositions’ genericity, assuming that the data on which GenAI models are trained is unbiased and “natural” (namely reflects organic real world usage), memorization would often be lawful and will not infringe copyrights.

Computer scientists have long tinkered with computational measurements for assessing creativity.¹⁸³ For instance, Franceschelli et al.,¹⁸⁴ proposed using generative learning techniques to assess creativity based on Margaret Boden’s definition of value, novelty, and surprise.¹⁸⁵ They also created a tool that executed this assessment called *DeepCreativity*.¹⁸⁶ We propose a different computational approach to measure originality that is more consistent with copyright principles. Building on the work of Gal et al. (2022),¹⁸⁷ 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.¹⁸⁸

Text-to-image GenAI models, such as *Dall-E-2* and *Stable Diffusion*, generate images from textual prompts.¹⁸⁹ Textual inversion performs the same task but in reverse: it inverts images into textual prompts.¹⁹⁰ These prompts can then be entered into GenAI models, which use cumulative learning to reconstruct the inverted images. Textual inversion was initially used to personalize text-to-image GenAI models.¹⁹¹ This procedure empowers GenAI models to generate variations of newly introduced images or transfer a new image’s style to other images.¹⁹²

¹⁸³ See generally Giorgio Franceschelli, et. al., *Creativity and Machine Learning: A Survey*, ARXIV 1 (2022).

¹⁸⁴ See Giorgio Franceschelli & Mirco Musolesi, *DeepCreativity: Measuring Creativity with Deep Learning Techniques*, ARXIV (2022).

¹⁸⁵ Margaret A. Boden, *Creativity in a Nutshell*, THINK (2009) (“Creativity is the ability to come up with ideas or artifacts that are *new, surprising, and valuable.*”)

¹⁸⁶ *Id.*

¹⁸⁷ Rinon Gal et al., *An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion*, ARXIV (2022).

¹⁸⁸ Weihao Xia, et al.), *GAN Inversion: A Survey*, ARXIV (2021); Jun-Yan Zhu (et al.), Philipp Krähenbuhl, Eli Shechtman & Alexei A Efros, *Generative visual manipulation on the natural image manifold*, EUR. CONF. COMP. VISION 597 (2016).

¹⁸⁹ Alberto Romero, *DALL-E 2, Explained: The Promise and Limitations of a Revolutionary AI*, MEDIUM (Jun, 16, 2022).

¹⁹⁰ These textual prompts are not readable to humans. These new “pseudo-words” are in the machine’s embedding space – meaning they are vectors of a few dozens of numbers.

¹⁹¹ Gal, *supra* note 187, at 3.

¹⁹² *Id.*

Here we argue that textual inversion can also be used to score the originality of the reconstructed (and, by extension, the targeted) images.¹⁹³ 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.¹⁹⁴

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 in 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 we are less familiar with. Figure 5 visualizes the main steps of our proposed procedure.

Table 1 reveals preliminary results from our textual inversion studies with the visual GenAI model *Stable Diffusion*. As it indicates, a single-word textual prompt embedding “<S>” was adequate to capture the expressive compositions of highly prevalent images such as Barak Obama’s portrait (row 1, column A) and Van Gogh’s famous *Starry Night* (rows 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 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

¹⁹³ For images that are not perfectly reconstructed by the model, we also consider an alternative procedure consisted of fine tuning the model from the reconstruct to the targeted image, and measuring the amount of training needed to achieve this goal as approximating the targeted image originality. See Hachohen, et al, *supra* signature note.

¹⁹⁴ In this approach, another benchmark that must be determined is what constitute 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.

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.¹⁹⁵ 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 non-genericity) is positively correlated with the length of textual inversion prompts.

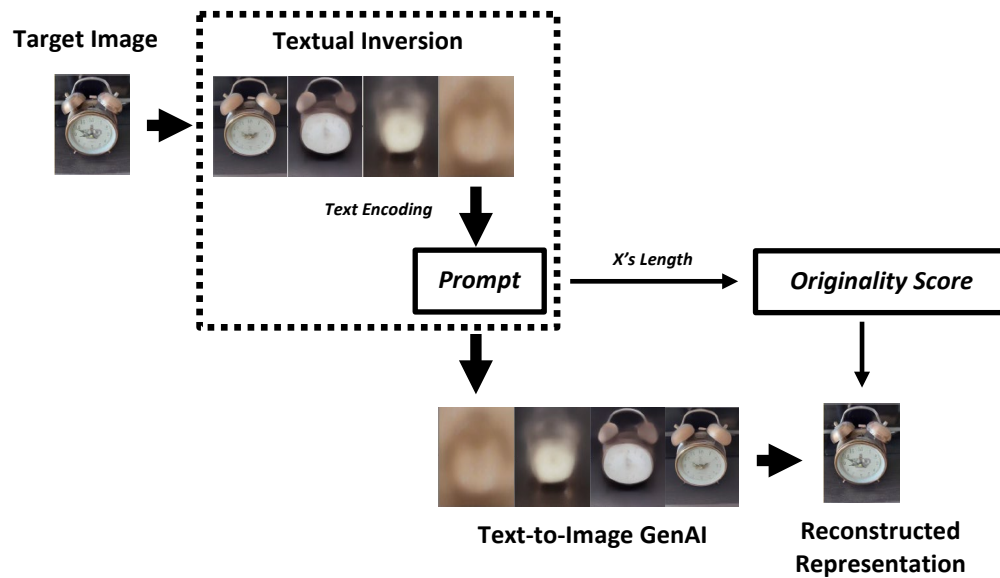


Figure 5: Scoring Image Originality Using Textual Inversion

¹⁹⁵ *Id.*










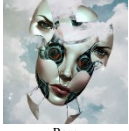


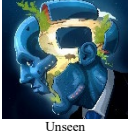





	Targeted Image	One-Word Embedding Textual Prompt (“<S>”)	
		Reconstruction	Edit
1.	 Prevalent	 Prompt: “<S>”	 Prompt: “<S> with sunglasses”
2.	 Prevalent	 Prompt: “<S>”	 Prompt: “cat in <S>”
3.	 Rare	 Prompt: “<S>”	 Prompt: “<S> with sunglasses”
4.	 Rare	 Prompt: “<S>”	 Prompt: “<S> in apocalypse”
5.	 Unseen	 Prompt: “<S>”	 Prompt: “<S> in the morning”
6.	 Unseen	 Prompt: “<S>”	 Prompt: “<S> in the park”
	A.	B.	C.

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

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 micro distributive measures as we discuss next.

IV. POLICY IMPLICATIONS

“Creativity is just connecting things . . . [to] synthesize new things.”

Steve Jobs¹⁹⁶

The ability to harness GenAI to measure copyright originality has groundbreaking implications for copyright law. So far, jurists have lacked rigor in 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. This inherent vagueness in copyright doctrines has led to the systematic over-protection of copyrighted works.¹⁹⁷

The originality scores proposed in this article could empower courts to delineate copyright scope more efficiently and fairly. Originality scores may also inform the Copyright Office registration practices and facilitate market licensing transactions. All these benefits would serve to realign copyright law with its constitutional foundation.¹⁹⁸ This section explores some potential implications that the originality scores may have on current copyright doctrines and practices.

A. Infringement

¹⁹⁶ Gary Wolf, *Steve Jobs: The Next Insanely Great Thing*, WIRED, (Feb 1, 1996), <https://www.wired.com/1996/02/jobs-2/>

¹⁹⁷ Gideon Parchomovsky & Alex Stein, *Originality* 95 VA. L. REV. 1505, 1509 (2009); 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, 354–60 (1999); Pamela Samuelson, *The Copyright Grab*, WIRED (Jan. 1, 1996); James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 887–906 (2007).

¹⁹⁸ Cf. Cooter & Hacoen, *supra* note 12.

Copyright scores could have the most dramatic impact on copyright infringement litigation. As explained in Section II, courts consider the originality of copyrighted works when evaluating whether allegedly infringing works use original subject matter without authorization and thus infringe copyrights.¹⁹⁹ This consideration is baked into the substantial similarity inquiry under the *Altai* framework and the fair use analysis under the second factor of Section 107 of the Copyright Act.²⁰⁰ However, courts' ability to appraise copyright originality under these doctrines is quite limited in practice.

When evaluating substantial similarity, courts often resort to a binary standard that merely approximates the optimal gradient standard articulated above.²⁰¹ Courts begin by filtering out the expressive elements from the allegedly infringed work that they consider "fully" generic under various copyright doctrines, such as functionality, merger, or Scènes à faire. Then courts evaluate the similarity of these remaining expressions to the allegedly infringing works without considering how original these remaining patterns actually are.²⁰² 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.²⁰³

Courts also apply a binary approximation standard when evaluating originality under the second fair use factor. 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.²⁰⁴ For example, while the Supreme Court in *Oracle v. Google* ruled that the Java declaring code

¹⁹⁹ See *supra* Part II.

²⁰⁰ See *supra* note 119, and accompanying text.

²⁰¹ See *supra* note 112, and accompanying text.

²⁰² See *supra* note 112, and accompanying text.

²⁰³ The notion that there is no true binary relationship between protected expression and unprotected ideas was emphasized by numerous courts and commentators. See *supra* note 39.

²⁰⁴ *Blanch v. Koons*, 467 F.3d 244, 256 (2006); 2 HOWARD B. ABRAMS, THE LAW OF COPYRIGHT, § 15:52 (2006) ("Two types of distinctions as to the nature of the copyrighted work have emerged that have figured in the decisions evaluating the second factor: (1) whether the work is expressive or creative, such as a

was functional under the second fair use factor, it did not explain at what point that functionality would render it utterly unprotected.²⁰⁵

Courts apply a binary standard even when considering the allegedly infringing uses’ “transformativeness” under the first fair use factor.²⁰⁶ Nearly all derivative works “transform” to the works from which they are derived to some extent, but this does not necessarily discharge them of copyright infringement.²⁰⁷ Nevertheless, courts

work of fiction, or more factual, with a greater leeway being allowed to a claim of fair use where the work is factual or informational”). See e.g., *Andy Warhol Foundation for the Visual Arts v. Goldsmith*, 11 F.4th 26, 117 (2nd Cir. 2021), (“Having recognized the Goldsmith Photograph as . . .creative. . .the district court should have found this factor to favor Goldsmith irrespective of whether it adjudged the Prince Series works transformative within the meaning of the first factor.”).

²⁰⁵ *Google LLC v. Oracle America, Inc.*, 593 U.S. ___ 1 (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” factor. The more functional the Java declaring code is, the more likely it is that Google use of it will be considered fair even if it’s level of transformativeness is low, and vice versa. See *supra* note 129.

²⁰⁶ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 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); *GOOGLE LLC V. ORACLE AMERICA, INC.*, BRIEF OF PROFESSORS PETER S. MENELL, SHYAMKRISHNA BALGANESH, & JANE C. GINSBURG AS AMICI CURIAE IN SUPPORT OF RESPONDENTS, 17, 25-26 (Aug. 11, 2022) (criticizing courts for not “understanding transformativeness as a matter of degree rather than a binary switch.”); Mark A. Lemley, *The Economics of Improvements in Intellectual Property Law*, 75 TEX. L. REV. 989, 1077 (1997) (noting that fair use analysis “requires [] a more careful balancing of the relative contributions made by the original copyright owner and the improver.”).

²⁰⁷ 17 U.S.C. § 106(2) (2002) (granting authors the exclusive right “to prepare derivative works based on 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).

often treat transformativeness as a binary switch rather than a matter of degree by equating transformative derivatives with fair use.²⁰⁸

Originality scores would avail courts of such pitfalls and inaccuracies.²⁰⁹ Courts could rely on originality scores to devise a more precise differential similarity standard for copyright infringement: the higher the score, the lower the similarity burden that plaintiffs must satisfy to prove infringement.²¹⁰ Similarly, courts could apprise of the originality of specific copyrighted works even without resorting to categorical proxies. For example, courts differentiate an original photograph from an unoriginal painting, even though, as a class, the latter is considered more “creative” and the former more

²⁰⁸ See Menell, Balganes, & Ginsburg, *supra* note 206, at 17 (“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.”). *Cf. Cariou v. Prince*, 714 F.3d 694, 708 (2nd 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 v. Koons*, 467 F.3d 244, 254 (2nd Cir. 2006) (summarily concluding that the defendant’s use was “substantially transformative” to discount its commercial nature).

²⁰⁹ In this vein, originality scores would challenge Justice Learned Hand 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) (a case involving an alleged film adaptation of a stage play, a derivative work). See also Warhol, *supra* note 2, at 36 (the majority 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. 251–52 (Justice Oliver Holms famously noting that, “[i]t 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.”).

²¹⁰ *Cf. Lemley, Convergence in the Law of Software Copyright*, *supra* note 112.

“factual.”²¹¹ Courts could also distinguish famous works from original ones, although currently, they sometimes confuse the two concepts.²¹²

To illustrate the value of originality scores, let’s imagine they were available when the Supreme Court heard the case that opened this article, *Warhol v. Goldsmith*.²¹³ To evaluate whether the licensing of Warhol’s depiction of Prince constituted fair use, the Court had to consider the comparable originality of Warhol’s and Goldsmith’s works under the first and second fair use factors, respectively.²¹⁴

While making such estimations in the abstract is difficult, there are reasons to believe that Goldsmith’s photograph would likely have gotten a relatively low originality score at the time of Warhol’s creation.²¹⁵ First, all photographs are works of limited originality as they are essentially compilations of unprotected facts.²¹⁶ Second, portraits are even less original than most photographs because many of their expressive choices are constrained by the rules of their given

²¹¹ Justin Hughes, *The Photographer’s Copyright—Photograph As Art, Photograph As Database*, 25 HARV. J. L. & TECH. 328, 330 (2012) (“copyright 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.”). See also, Section IV.B., *infra* (arguing that originality scoring could also allow courts to apprise the originality of GenAI-augmented works which as a class, are currently considered by the Copyright Office unprotected for the lack of human authorship).

²¹² The Majority in *Warhol*, *supra* note 2, at 10 (criticizing the lower court for implicitly suggesting that fame can be used as an objective proxy for originality and transformativeness. Noting that this approach would effectively “create a celebrity-plagiarist privilege.” *Id.*); GOOGLE LLC V. ORACLE AMERICA, INC., BRIEF FOR THE UNITED STATES AS AMICI CURIAE IN SUPPORT OF RESPONDENTS, 23 (Aug. 11, 2022) (same).

²¹³ *Warhol*, *supra* note 2.

²¹⁴ Lemley, *The Economics of Improvements in Intellectual Property Law*, *supra* note 206, at 1077.

²¹⁵ We propose that originality (non-genericity) should be measured at the time of infringement not the time of creation because of the dynamic nature of expressive genericity. See Section II.B. *supra*.

²¹⁶ Hughes, *supra* note 211, at 330.

genre.²¹⁷ Third, Prince was a famous individual subject of numerous portraits and photographers. Prince's familiarity might further degrade the originality of Goldsmith's photograph by rendering some of Prince's facial expressions ubiquitous and thus generic.²¹⁸

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.²¹⁹ Today, Warhol's work would most certainly have gotten a relatively low originality score, partially due to Warhol's own influence and success.²²⁰ Warhol had an enormous cultural impact, and his work motivated numerous artists to employ similar expressive patterns in their works.²²¹ The prevalence of such

²¹⁷ Cf. Tobin, *supra* note 103 (discussing the Singaporean case). Although, unlike traditional portrait photography, modern forms of portrait photography allows more liberty to photographers, and accordingly made 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/> (exploring different forms of portrait photography).

²¹⁸ Cf. Barak Obama portrait in Table 1. The fact the likeness of more famed individuals is more generic than that of lesser known individuals follows from the general logic of expressive compositions' genericity although it 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. Cf. *Harper & Row v. Nation Enterprises*, 471 U.S. 539, 557 (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 generally, Robert C. Post & Jennifer E. Rothman, *The First Amendment and the Right(s) of Publicity*, 130 YALE L.J. 1 (2020).

²¹⁹ Because works' originality is dynamic it might be entitled to a different score at the time the work was created compared to when the allegedly infringing use was conducted – the time of licensing. See *supra* note 42, and accompanying text.

²²⁰ See Section II.B. *supra*.

²²¹ See e.g., Art Works Advisory Editor, *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/> (“Andy Warhol wasn't just influential; he created a whole new genre of contemporary art – pop art.”); Warhol *supra* note 2, at 4 (Justice Kagan recognizing Warhol influence and noting how “he changed modern art.”).

patterns in today’s expressive environment will likely render Warhol’s work as original as “an Instagram filter, and a simple one at that.”²²²

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.²²³ 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.²²⁴ The Court would still need to weigh other elements (such as commercialism) and other factors (such as 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.²²⁵

B. Registration

Works eligible for copyright protection can also be registered in the Copyright Office. Registration is not required for copyright eligibility but is necessary for filing copyright infringement lawsuits.²²⁶ Because the originality threshold for copyright eligibility is low, Copyright Office registration was traditionally assumed to be nearly automatic.²²⁷ This assumption was robust. It endured many

²²² Cite the court. Cite the Obama case.

²²³ Note that this logic is the opposite of the view that fame is associated with high transformativeness. *Supra* note 218. In reality, fame dilutes transformativeness by making expressive composition generic.

²²⁴ See *supra* notes 170–173, and accompanying text.

²²⁵ Cf. Warhol *supra* note 2, at 36 (criticizing the minority for not providing a “theory of the relationship between transformative uses of original works and derivative works that transform originals.”).

²²⁶ See *Fourth Estate Public Benefit Corp. v. Wall-Street.com*, 586 U.S. ____ (2019); Zvi S. Rosen, *Examining Copyright* (Forthcoming, Journal of the Copyright Society of the USA) (May 4, 2022), <https://ssrn.com/abstract=4099976>

²²⁷ See *Cosmetic Ideas, Inc. v. IAC/InteractiveCorp*, 606 F.3d 612, 621 (9th Cir. 2010) (“... the Register’s decision of whether or not to grant a registration certificate is largely perfunctory”); Susan M. Richey, *The Troubling Role of Federal*

technological waves, including those that birthed works of modest originality, such as digital photography.²²⁸

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.²²⁹

In this vein, the Office recently refused to register graphic images that Ms. Kashtanova produced with the aid of the GenAI system Midjourney.²³⁰ After determining that “it was Midjourney—not Kashtanova—that originated the ‘traditional elements of authorship’ in the images,” the Copyright Office ruled that “the images generated by Midjourney. . . are not original works of authorship protected by copyright.”²³¹ 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 the aid of GenAI tools can be authorless.²³² This outcome might also be socially desirable to the extent that GenAI models lower the creation cost, allowing authors

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.”).

²²⁸ Rosen, *supra* note 226, at 69 (noting that “the 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, 56–8 (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.”).

²²⁹ 37 CFR Part 202, Federal Register, Vol. 88, No. 51, 16191–2 (Mar. 16, 2023) 16192 (“[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. . . .”) (internal citations omitted).

²³⁰ See U.S. Copyright Office, *Cancellation Decision re: Zarya of the Dawn* (VAu001480196) at 2 (Feb. 21, 2023), <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>.

²³¹ *Id.*

²³² See *supra* note 147.

to recoup their costs without legal intervention.²³³ 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.²³⁴ In such cases, society might lose from categorically excluding GenAI-augmented works from the realm of copyrightability.²³⁵ 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.²³⁶ 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.²³⁷ 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.

²³³ Burk, *supra* note 158, at 1676–79. See also Greg Bensinger, *Focus: ChatGPT Launches Boom in AI-written E-books on Amazon*, REUTERS (Feb. 21, 2023, 10:43 PM), <https://www.reuters.com/technology/chatgpt-launches-boom-ai-written-e-books-amazon-2023-02-21/>; 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>.

²³⁴ 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 formation,” and “prompt engineering,” respectively. See Oguz A. Acar, *AI Prompt Engineering Isn’t the Future*, HARV. BUS. REV. (Jun. 6., 2023), <https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future>. See also Kate Whiting, *3 new and emerging jobs you can get hired for this year*, WORLD ECONOMIC FORUM (Mar. 2, 2023) <https://www.weforum.org/agenda/2023/03/new-emerging-jobs-work-skills/> (calling prompt engineering the “job of the future”).

²³⁵ 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 formation of new laws that would support appropriations of such outputs. See *supra* note 159.

²³⁶ Cf. Abbott, Ryan Benjamin, *Everything is Obvious*. 66 UCLA L. REV. 2 (2018), Available at SSRN: <https://ssrn.com/abstract=3056915>.

²³⁷ Oppenlaender et al, *Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering*, ARXIV (May 13, 2023) <https://arxiv.org/pdf/2303.13534>.

The Copyright Office has recognized that human authors may sometimes contribute sufficient original expression to make GenAI-augmented works eligible for registration.²³⁸ 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.²³⁹

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.²⁴⁰ 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. The higher the score, the more likely the work reflects a modicum of creativity compared to the state-of-the-art learned by the model.²⁴¹

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 as GenAI calibrates Copyrighted works with higher originality scores would be valued more than copyrighted works with low originality scores.²⁴²

²³⁸ 37 CFR Part 202, *supra* note 229, at 16192.

²³⁹ *Id.*

²⁴⁰ In addition, our approach could also empower the copyright office to apprise 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.” 37 CFR Part 202, *supra* note 229, at 16192. Still, the degree to which the prompt is original may serve as a useful proxy to the user’s original contribution.

²⁴² *Cf.* Clarisa Long, *Patent Signals*, 69 *UNI. OF CHI. L. REV.* 625, 643 (2002) (explore the singling function of patent).

²⁴² *Cf.* Clarisa Long, *Patent Signals*, 69 *UNI. OF CHI. L. REV.* 625, 643 (2002) (explore the singling function of patent).

GenAI enables users to draw upon existing knowledge embedded within the model automatically, rather than learn and create everything from scratch. This reduces the barriers to entry to the production of creative works, allowing individuals with basic language, graphic, or technical skills to generate high-quality content.²⁴³ Such a flow of ‘cheap’ content may require those marketing the works to adequately signal to 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 singling value may thus facilitate market licensing transactions. For example, copyright owners could advertise their works’ originality scores to attract potential consumers.²⁴⁴ Similarly, GenAI companies could rely on originality scores to solicit original works to improve their models’ performance.²⁴⁵

²⁴³ David De Cremer, Nicola Morini Bianzino, & Ben Falk, *How Generative AI Could Disrupt Creative Work*, Harv. Bus. Rev. (Apr. 13, 2023).

²⁴⁴ In the patent context consumers sometimes use a “patent pending” or “patented” disclosures to single to 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/>. Originality scores could even more effectively communicate the degree to which copyrighted works are valuable.

²⁴⁵ Many scholars have argued, and we agree, that GenAI leaning does not constitute copyright infringement, and, as such, should not be bargained for. *See supra* note 148. Nevertheless, authors can still refuse to make their intellectual output available for learning. In addition, our proposed originality scoring system could also help policymakers devise various remuneration or taxation schemes. As mentioned, we do not necessarily endorse these approaches (at least to the extent that they embrace the notion that machine learning constitute copyright infringement), but some scholars and policymakers do consider them. For example, Martin Senftleben offers to impose an equitable tax on GenAI users based on esthetic theories which assign intrinsic value to human authorship. Senftleben, *supra* note 157. 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.

In addition, courts might use originality scores to assess lost licensing opportunities when considering actual damages in infringement litigation.²⁴⁶ Currently, courts consider the assessment of damages to be one of the most challenging tasks they handle.²⁴⁷ By simplifying this task, originality scores might even lead courts to be more liberal in prescribing monetary awards in copyright infringement lawsuits instead of injunctive relief, which some scholars have long considered advisable.²⁴⁸

V. CONCLUSION

This article has demonstrated how GenAI capacities could be levered to gain more nuanced insights into the genericity of specific expressions on a significantly large scale. Our novel approach to measuring originality is based on interdisciplinary computer science and law research. This approach employs data-driven bias to evaluate the genericity of expressive compositions in preexisting works. The more generic some compositions are, the more likely GenAI models will 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.

²⁴⁶ Courts are required to appraise loss licensing fees when they prescribe actual damages in copyright infringement claims. 17 U.S.C. § 504(b) (2010); *On Davis v. Gap, Inc.*, 246 F.3d 152, 166 (2nd Cir. 2001).

²⁴⁷ *Cf. Cincinnati Car Co. v. New York Rapid Transit Corp.* 66 F.2d 592, 593 (2d Cir. 1933) (Justice Learned Hand noting that, "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. Properly, the question is in its nature unanswerable.... It is generally impossible to allocate quantitatively the shares of the old and the new, and the party on whom that duty falls, will usually lose.").

²⁴⁸ *See e.g., Parchomovsky & Stein, supra* note 12, at 1533 (introducing "the added-value doctrine" which would allow copyright owners to secure injunctive relief against putative infringers only in cases the latter allegedly infringing work is less original than the former work. In all other cases, they argue, the court should allow the use and prescribe compensatory damages); Peter S. Menell & Ben Depoorter, *Using Fee Shifting to Promote Fair Use and Fair Licensing*, 102(1) CAL. L. REV. 53, 58 (2014) (proposing a delicate approach to incentivize parties to bargaining and to inform the courts of reasonable licensing fees in cases of infringement).

This scale for measuring originality could impact all the major phases in the lifecycle of copyrighted works, from registration and licensing to copyright infringement litigation. In the context of copyright infringement, for instance, scholars have argued that while generalization by the model is non-infringing 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.²⁴⁹ 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 new 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 failed to implement them mainly due to institutional incompetency.²⁵⁰ 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 policymakers as they grapple with adapting copyright law to meet the new challenges of an era characterized by “cheap creativity” enabled by GenAI.

²⁴⁹ EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE 5 (2013). As argued by Morozov, technological solutionism derives from a “never-ending quest to ameliorate”, while being oblivious to complex social situations and conditions.

²⁵⁰ Cf. Parchomovsky & Stein, *supra* note 12, at 1534 (accepting the criticism of the courts' institutional incapacities to appraise originality but suggesting that courts should rely more on expert testimony); Miller, *supra* note 45, at 477 (reciting the traditional scholarly skepticism about the ability of jurists to appraise copyright originality); Fisher, *Recalibrating Originality*, *supra* note 33 at 458 (2016) (same); Bracha & Syed, *supra* note 68, at 1912 (same).

