

ARTICLE

OUTPUT TRANSPARENCY: DIFFERENTIATING HUMAN AND AI CREATIVITY THROUGH LEGAL REFORMS

*Haochen Sun**

The advent of artificial intelligence (AI) has presented our legal system with a profound new dilemma. Public confidence in the capacity of AI to generate content is growing quickly, nearly apace with advancements in the technology itself. Yet, as AI-generated content masquerading as human-authored increasingly floods online platforms, our ability to tell the difference is diminishing. This widespread concealment of AI's role in content creation has unleashed a cascade of harms for individuals as well as society at large.

This Article exposes the law's failure to regulate AI concealment, a legal vacuum that permits the proliferation of such concealment. To address this challenge, we must first theoretically interrogate why it is important to distinguish AI-generated content from human-authored works. Building on this theoretical foundation, we can then consider the feasibility of adopting a robust legal standard to regulate AI concealment.

To this end, the Article advances a theory of "differentiated creativity." This theory clarifies the fundamental differences between human creative processes—rooted in human intentionality, emotional expression, and control—and those of AI. Based on this theory, the Article then proposes "output transparency" as a new legal standard enabling clear differentiation between human and AI contributions.

* Professor of Law & Director, Program on Artificial Intelligence and the Law, University of Hong Kong Faculty of Law; Affiliated Fellow, Information Society Project, Yale Law School. I presented an earlier draft of this Article at Lunch Talk Series at the Yale Information Society Project, Boston University's Technology Policy Research Initiative Workshop, the "Reframing Intellectual Property Law in the Age of Artificial Intelligence" conference, the Second Annual US-Asia Comparative Copyright Law Roundtable, "Regulating AI in the Public Interest" conference, ATRIP Congress 2025, and Philosophy of AI Summer School (Paris 2025). I am grateful to the participants at these events for their feedback and also to Shyam Balganes, Jack Balkin, Herman Cappelen, Anupam Chander, Stacey Dogan, William Fisher, Ruth Okediji, Frank Pasquale, Martin Senftleben, Jessica Silbey, Madhavi Sunder, Scott Veitch, Christopher Yoo, and Peter Yu for their helpful comments and suggestions. This Article is an output of a research project funded by the Hong Kong General Research Fund (Project Number: 17613223).

To operationalize output transparency, the Article advocates two key legal reforms: requiring technology companies to watermark AI-generated content and obligating authors to provide proper explanations of the modifications they have made to such content. These proposed reforms would protect human adaptations of AI outputs, ensure a clear distinction between human and machine contributions, and enhance transparency across creative sectors. By implementing these reforms, the legal system can more effectively safeguard and foster the creative potential of both humans and AI, promoting balanced and ethical integration of technology in society.

INTRODUCTION	139
I. THE RISE OF AI OUTPUT OPACITY	148
A. <i>Three Forms of AI Output Opacity</i>	148
1. Output Opacity at the Technological Level.	148
2. Output Opacity at the Ethical Level	151
3. Output Opacity at the Legal Level	153
B. <i>Negative Consequences of AI Concealment</i>	156
C. <i>Algorithmic and Data Transparency’s Regulatory Vacuum</i>	158
D. <i>U.S. Copyright Office’s Failed Attempts to Regulate Output Opacity</i>	162
1. Human Authorship	162
2. Unsolved Problems	164
II. PROMOTING OUTPUT TRANSPARENCY: A THEORY OF DIFFERENTIATED CREATIVITY.	166
A. <i>Intent to Create</i>	167
1. AI’s Lack of Intent	168
B. <i>Expression of Emotions</i>	169
1. Human Emotions.	169
a. Aesthetic Emotions.	170
b. Intellectual Emotions	171
c. Moral Emotions	171
2. AI’s Lack of Emotions	172
C. <i>Control Over the Creative Process</i>	173
1. Control as Part of Creative Abilities	173
2. AI’s Lack of Control	175
D. <i>Benefits of Output Transparency</i>	177
1. Promoting Human Creativity	178
2. Promoting AI Creativity	180
III. ACHIEVING OUTPUT TRANSPARENCY: TWO PROPOSALS	182
A. <i>Watermarking AI-Generated Content</i>	182
1. Technological Feasibility	183
2. Legal Requirements.	188

B. Explanations about Modification of AI-Generated Content	192
1. Problems with Compilations	192
2. Explanations about Human Creativity	197
C. Addressing Problems	204
1. Potential Problems with AI Watermarking	204
2. Potential Problems with AI Explanations	206
CONCLUSION	209

INTRODUCTION

Debuting online in June 2025, a band named The Velvet Sundown took Spotify by storm, amassing over a million stream subscribers within weeks.¹ However, the band’s meteoric rise attracted close media scrutiny of its origins, conjecturing whether the music, promotional images, and even the band’s shadowy backstory were conjured by AI.² The “band” vehemently denied the allegations at first,³ only to later confess that every single aspect of its existence was, in fact, AI-generated.⁴ For many, the calculated concealment of the band’s AI origins amounted to a new form of consumer fraud,⁵ exploited the creative work of human musicians,⁶ and even raised broader ethical concerns about humanity’s “losing control of AI.”⁷

How does the law regulate concealment of AI use? In reality, it does not. There are no legal rules in place to address AI hoaxes like The Velvet

¹ Caitlin McCormack, *Rock Band with More Than 1 Million Spotify Listeners Reveals It’s Entirely AI-Generated—Down to the Musicians Themselves*, N.Y. POST (July 10, 2025), <https://nypost.com/2025/07/10/entertainment/rock-band-with-more-than-1-million-spotify-listeners-reveals-its-entirely-ai-generated-down-to-the-musicians-themselves/>.

² Rhys Blakely, *Meet Velvet Sundown, Spotify’s Hottest New Band. But Are They Real?*, THE TIMES (July 1, 2025), <https://www.thetimes.com/uk/technology-uk/article/velvet-sundown-not-real-ai-generated-spotify-m9z0sv0xs>; Erin McLaughlin, *An Indie Band Is Blowing Up on Spotify, but People Think It’s AI*, NBC NEWS (July 3, 2025), <https://www.nbcnews.com/tech/tech-news/the-velvet-sundown-indie-band-spotify-ai-accusations-speculation-rcna216728>.

³ *Id.* (“[T]he band pushed back against accusations that it isn’t ‘real,’ stating in one video that ‘you believed the lie, and danced to it anyway.’”).

⁴ Ethan Beck, *A ‘60s-Flavored Band Blew Up on Spotify. They’re AI.*, WASH. POST (July 8, 2025), <https://www.washingtonpost.com/entertainment/music/2025/07/08/velvet-sundown-artificial-intelligence-spotify/>.

⁵ Graham Barlow, *Apple and Spotify Are Sleepwalking into an AI Music Crisis – and The Velvet Sundown Mess Shows They Need to Act Fast*, TECHRADAR (July 3, 2025), <https://www.techradar.com/computing/artificial-intelligence/apple-and-spotify-are-sleepwalking-into-an-ai-music-crisis-and-the-velvet-sundown-mess-shows-they-need-to-act-fast>.

⁶ Tara Bellucci, *The Velvet Sundown: The AI Band Controversy Explained*, BERKLEE NOW (July 22, 2025), <https://www.berklee.edu/berklee-now/news/velvet-sundown-ai-band-controversy>.

⁷ Zoe Kleinman, *AI Claims and a Hoax Spokesman: Viral Band Confuses the World of Music*, BBC NEWS (July 3, 2025), <https://www.bbc.com/news/articles/cp8mjnn7eqno>.

Sundown.⁸ Our world has become increasingly saturated with AI-generated content passing itself off as human-made, with experts estimating that “as much as ninety percent of online content may be synthetically generated by 2026.”⁹ Circulating widely are news articles,¹⁰ books,¹¹ social media posts,¹² and videos¹³ produced by AI systems yet presented as the genuine work of human creators. Advertisers and influencers have taken this deception even further, deploying synthetic personas to deliver “real” testimonials that subtly mislead viewers into believing in their authenticity and foster trust in promoted products.¹⁴ Similarly, academia has been inundated with low-quality AI-generated papers masquerading as human-authored research, often evading plagiarism detection and peer review scrutiny.¹⁵

To ensure transparency across creative and social media sectors amid rampant AI concealment, legal reforms are urgently needed. As AI-generated content floods the internet¹⁶ the intentional obscurement of AI’s role poses grave dangers to both individuals and society at large. Among those sounding the alarm was the late Pope Francis, who called on lawmakers to act swiftly, warning that AI concealment undermines core human values including morality.¹⁷ Commentators have echoed this

⁸ One should note that the watermarking requirement under the laws such as the EU AI Act only apply to technology companies. They do not target AI users like The Velvet Sundown band.

⁹ Daniel Pereira, *By 2026, Online Content Generated by Non-humans Will Vastly Outnumber Human Generated Content*, OODA Loop (Mar. 6, 2024), <https://oodaloop.com/analysis/archive/if-90-of-online-content-will-be-ai-generated-by-2026-we-forecast-a-deeply-human-anti-content-movement-in-response/>.

¹⁰ Caitlin Harrington, *CNET Published AI-Generated Stories. Then Its Staff Pushed Back*, THE WIRED (May 16, 2023), <https://www.wired.com/story/cnet-published-ai-generated-stories-then-its-staff-pushed-back/>.

¹¹ Sanchari Sen, *Amazon’s AI Book Problem*, MEDIUM (Feb. 15, 2024), <https://medium.com/@sensanchari2018/amazons-ai-book-problem-31ce6ee59837>

¹² BBC News, *How to Spot AI Images on Soc. Media*, (Dec. 2024), <https://www.bbc.co.uk/bitesize/articles/z6s4239>.

¹³ Dee Naidoo, *AI Generated Videos are Destroying YouTube*, MEDIUM (July 13, 2023), <https://databudd.medium.com/ai-generated-videos-are-destroying-youtube-d6771badcce>.

¹⁴ Colin Campbell, et al., *How Deepfakes and Artificial Intelligence Could Reshape the Advertising Industry: The Coming Reality of AI Fakes and Their Potential Impact on Consumer Behavior*, 62 J. OF ADVERTISING RSCH. (2022).

¹⁵ Miryam Naddaf, *Low-Quality Papers Based on Public Health Data Are Flooding the Scientific Literature*, NATURE (July 15, 2025), <https://www.nature.com/articles/d41586-025-02241-2>; Cathleen O’Grady, *Low-Quality Papers are Surging by Exploiting Public Data Sets and AI*, SCIENCE (May 14, 2025) <https://www.science.org/content/article/low-quality-papers-are-surging-exploiting-public-data-sets-and-ai>.

¹⁶ Shelly Fan, *This Is What Could Happen if AI Content Is Allowed to Take Over the Internet*, SINGULARITYHUB (July 25, 2024), <https://singularityhub.com/2024/07/25/this-is-what-could-happen-if-ai-content-is-allowed-to-take-over-the-internet/>.

¹⁷ Nicole Winfield, *Pope, Once A Victim Of AI-Generated Imagery, Calls for Treaty to Regulate Artificial Intelligence*, AP NEWS (Dec. 14, 2023), <https://apnews.com/article/vatican-pope-ai-artificial-intelligence-9805fec11681adbf88d3a7c73bdf47de>; Joshua McElwee, *Vatican Says A.I. Has ‘Shadow*

urgency, stressing the critical need for legal regulation to address threats to national security and broader societal safety.¹⁸ They warn that if this AI deception continues unchecked, trust and ethical principles could be undermined on a global scale.¹⁹

This Article addresses the prevailing legal vacuum by introducing “output transparency” as a new legal standard for regulating AI concealment. I begin by analyzing the nature of practices that conceal AI involvement and the multifaceted problems these practices engender at individual, economic, and societal levels.²⁰ To tackle these challenges systematically, I propose a theory of “differentiated creativity”, which elucidates the fundamental distinctions between human and AI creative processes.²¹ This theory provides a robust foundation for promoting output transparency as an essential legal benchmark, thereby strengthening AI governance through mandatory disclosures that make AI contributions clearly distinguishable from human creative outputs.²²

Building on this theoretical framework, I suggest two legal mechanisms to operationalize the output transparency standard: mandatory AI watermarking and disclosure of human modifications to AI-generated works. Technology companies should be legally required to watermark. Alongside this, authors who adapt AI-generated content into new works should be legally required to explain their modifications. These mechanisms would maintain distinct boundaries between human and machine contributions, thereby encouraging the ethical use of AI to augment human creativity.²³

This Article seeks to make three original contributions to the scholarly discourse on law and AI regulation. First, the Article offers the first comprehensive theoretical analysis of the imperative to distinguish between AI-generated and human-created content. This theoretical exploration is necessary to justify the adoption of new legal standards that regulate the concealment of AI’s role in content creation.²⁴ Without a robust

of Evil,’ *Calls for Close Oversight*, REUTERS (Jan. 28, 2025), <https://www.reuters.com/technology/artificial-intelligence/vatican-says-ai-has-shadow-evil-calls-close-oversight-2025-01-28/>.

¹⁸ Wyatt Grantham-Philips, *Katy Perry and Rihanna Didn’t Attend the Met Gala. But AI-Generated Images Still Fooled Fans*, AP NEWS (May 7, 2024), <https://apnews.com/article/katy-perry-rihanna-met-gala-ai-3b3d2da6c57c0b0b563149c32851297f>.

¹⁹ Joe Coscarelli, *An A.I. Hit of Fake ‘Drake’ and ‘The Weeknd’ Rattles the Music World*, N.Y. TIMES (Apr. 24, 2023), www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html (“Aside from raising questions of legality, [AI concealment] can introduce knotty ethical concerns regarding race and identity.”).

²⁰ See *infra* Section I.B.

²¹ See *infra* Sections II.A-C.

²² See *infra* Section II.D.

²³ See *infra* Section III.B.

²⁴ Kevin Vallier, *Public Justification*, STAN. ENCYC. PHIL. (Dec. 2, 2022), <https://plato.stanford.edu/entries/justification-public/>.

theoretical foundation, such regulations risk being dismissed as arbitrary or unsophisticated, thereby eroding their legitimacy and effectiveness.²⁵

The European Union's Artificial Intelligence Act (EU AI Act)²⁶ exemplifies the pitfalls of implementing regulatory standards without a solid theoretical basis. The Act's mandate for technology companies to watermark AI-generated content was adopted hastily, without a thorough examination of why differentiating AI and human outputs is necessary.²⁷ This oversight has left fundamental questions unresolved, weakening the Act's conceptual foundation and legal efficacy.²⁸ As a result, critics have highlighted the Act's lack of a systematic rationale,²⁹ sparking resistance from technology companies³⁰ and further diminishing the Act's effectiveness.

Similarly, the absence of a robust theoretical framework justifying the regulation of AI concealment has disrupted legal policymaking in the U.S. In 2023, the Biden administration incorporated an AI watermarking requirement into its Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, aiming to promote transparency in AI-generated content.³¹ Upon resuming office in January 2025, President Trump rescinded this order, halting the implementation of the watermarking mandate and leaving the U.S. without a federal policy on this critical issue.³² This abrupt reversal highlights the urgent need to delve into the theoretical underpinnings of AI concealment regulation, which would clarify the policy imperatives for establishing new legal standards

²⁵ Tom R. Tyler & Jonathan Jackson, *Popular Legitimacy and the Exercise of Legal Auth.: Motivating Compliance, Cooperation and Engagement*, 20 PSYCH., PUB. POL'Y, & L. 78, 79 (2013) <https://law.yale.edu/sites/default/files/area/center/justice/document/ssrnpopularlegitimacy.pdf>.

²⁶ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689) 1.

²⁷ *Id.* art. 50.2 at 82 ("Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video, or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated").

²⁸ Gopal Ratnam, *Europe's Approach to Regulate AI Seen as Too Hasty for US*, HEINRICH BÖLL FOUNDATION (Sep. 11, 2023), <https://us.boell.org/en/2023/09/11/europes-approach-regulate-ai-seen-too-hasty-us>.

²⁹ *Id.* ("The EU's draft law on AI regulation leaves key questions unanswered, US policy-makers say.").

³⁰ Ram Iyer, *Meta Refuses to Sign EU's AI Code of Practice*, TECHCRUNCH (Jul. 18, 2025), <https://techcrunch.com/2025/07/18/meta-refuses-to-sign-eus-ai-code-of-practice/>.

³¹ "Exec. Order No. 14110, 88 Fed. Reg. 75191 (Nov. 1, 2023) ("[M]y Administration will help develop effective labeling and content provenance mechanisms, so that Americans are able to determine when content is generated using AI and when it is not."); *Voluntary Commitments from Leading A.I. Co.'s on July 21, 2023*, 137 HARV. L. REV. 1282, 1284 (2024) ("[The Executive Order] specifies watermarking, in particular, as a potential identification mechanism.").

³² Exec. Order No. 14179, 90 Fed. Reg. 8741 (Jan. 23, 2025).

and provide a more compelling rationale to inform future legislative efforts.³³

Despite the far-reaching consequences of AI concealment, few legal scholars have thoroughly examined the theoretical foundations for its regulation. Professor Jacob Noti-Victor initiated this conversation by justifying regulation primarily based on harms caused by hidden AI authorship. When AI authorship is concealed, it deceives consumers and violates their expectations and trust.³⁴ This concealment also disrupts the human-to-human dialogue that gives art its power to foster self-definition, empathy, and political engagement, ultimately reducing art to mere entertainment.³⁵

While this harm-based approach is both important and timely, a more theoretically robust analysis could better illuminate the fundamental problems posed by AI concealment and provide a stronger justification for regulation. By its very nature, AI concealment blurs the distinction between human and AI creativity, erasing the boundaries between two fundamentally different forms of content creation. It is therefore crucial to first examine the intrinsic qualities and values that make human creativity distinctive and meaningful.³⁶ From there, we can better articulate why safeguarding these elements remains imperative, not merely to prevent their dilution, but to explore how they can be enriched using AI tools. By addressing these foundational issues directly, such a theoretical analysis would not only provide deeper insight into the root causes of the harms Professor Noti-Victor identifies but also engage with broader questions about how legal regulation can better enhance both human and AI creativity.

Through proposing the theory of differentiated creativity, I aim to delve into these fundamental issues surrounding human and AI creativity. The theory, as I will show in the Article, provides a robust justification for legal mechanisms that mandate distinguishing between AI-generated and human-created content. Drawing from philosophical perspectives on creativity, the theory delineates three core conditions essential for human

³³ See Douglas A. Kysar, *Preferences for Processes: The Process/Prod. Distinction and the Reg. of Consumer Choice*, 118 HARV. L. REV. 525, 535 (2004) (“[P]olicymakers and commentators should discard the view that consumer preferences for processes are somehow less appropriate or worthy of governmental support than preferences for products.”).

³⁴ See Jacob Noti-Victor, *Regulating Hidden AI Authorship*, 111 VA. L. REV. 139, 172–173 (2025) (arguing that consumers prefer human-authored works due to ethical concerns—such as the displacement of artists by AI—as well as the values of aesthetic authenticity and emotional connection with creators).

³⁵ *Id.* at 181.

³⁶ See Christopher S. Yoo, *Rethinking Copyright and Personhood*, 2019 U. ILL. L. REV. 1039, 1062 (2019) (“Promoting intrinsic motivation in turn encourages creativity, whereas introduction of external motivations reduces both the sense of self-determination and creativity.”); Mark Lemley, *How Generative AI Turns Copyright Law Upside Down*, 25 COLUM. SCI. & TECH. L. REV. 190, 212 (2024) (“AI makes us rethink what we value in creative work and whether and why we want it to be owned”).

creativity to flourish: the intent to create, which initiates the process with purposeful agency; the expression of emotions, rooted in human experiences and subjective understanding; and control over the creative process, enabling individuals to shape their work through deliberate decisions. In contrast, AI creativity functions mechanistically, governed by pre-programmed models and data inputs, lacking intent, emotional depth, and autonomous control.³⁷ These profound differences between human and AI creative processes underscore the critical need to differentiate their outputs,³⁸ justifying legal frameworks—such as watermarking—that ensure transparency and preserve the integrity of human authorship.³⁹

My second contribution is to put forward output transparency as a new legal standard for regulating AI concealment, grounded in the theory of differentiated creativity. While the expanding body of literature on AI transparency has made significant strides, it is yet to fully grapple with the challenges posed by deliberate obscuring of AI's role in content creation. Instead, scholarly focus has centered on two primary “black box” issues concerning the inner workings of AI technology: first, the complexity of AI models and their operations, which often render them incomprehensible to users,⁴⁰ and second, the opaque methods employed by AI companies to collect and leverage model training data.⁴¹ In response, scholars have advocated for the legal standards of “algorithmic

³⁷ See *infra* Section II.A-C.

³⁸ Micaela Mantegna, *ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI*, YALE L.J. FORUM (Apr. 22, 2024), at 1141 (“As a consequence, understanding what creativity is, namely what are creativity’s constitutive elements and possible definitions, becomes a key consideration in debates about GAI—particularly when trying to determine if creativity can be attributed to machines or if it is inherently an exclusively human feature.”); see also Eric J. Larson, *THE MYTH OF A.I.: WHY COMPUTERS CAN’T THINK THE WAY WE DO* 18 (2021) (“[T]he differences between minds and machines are more subtle and complicated than Turing imaged”).

³⁹ See *infra* Section II; see also Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI’s Copyright Crisis*, 110 VA. L. REV. ONLINE 207, 221 (2024) (“If uncompensated and uncontrolled expropriation of copyrighted works continues, many creatives are likely to be further demoralized and eventually defunded as AI unfairly outcompetes them or effectively drowns them out.”); David De Cremer et al., *How Generative AI Could Disrupt Creative Work*, HARV. BUS. REV. (Apr. 13, 2023), <https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work> (“If cheaply made generative AI undercuts authentic human content, there’s a real risk that innovation will slow down over time as humans make less and less new art and content.”).

⁴⁰ See Frank Pasquale, *THE BLACK BOX SOC’Y: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFO.* (Harvard Univ. Press 2015); Sonia K. Katyal, *Private Accountability in the Age of A.I.*, 66 UCLA L. REV. 54, 138 (2019).

⁴¹ See Meihe Xu et al., *A Cross-Cultural Analysis of Transparency: The Interplay of Law, Privacy Policies, And User Perceptions*, 14 INT’L DATA PRIV. L. 197, 204 (2024).

transparency”⁴² and “data transparency”⁴³ to regulate this input opacity of AI systems.

While algorithmic and data transparency standards effectively target the input opacity of AI systems, addressing the complexity of models and the obscurity of data practices,⁴⁴ they fall short in confronting output opacity and particularly the concealment of AI’s role in content production.⁴⁵ The sheer amount of AI-generated content, coupled with the growing trend of masking its artificial origins, is causing a cascade of unintended consequences for individuals as well as our economic, cultural, and political spheres, eroding trust, authenticity, and accountability.⁴⁶

In response, I demonstrate in this Article that implementation of output transparency would mitigate the growing challenge of distinguishing AI-generated content from human-made works—a difficulty that persists despite the use of AI content detection tools.⁴⁷ Furthermore, I argue that effective adoption of this legal standard will bolster human creativity by preserving incentives for authors to produce original content worthy of copyright protection.⁴⁸ Beyond this, output transparency could tackle threats to AI innovation and development,⁴⁹ such as model collapse⁵⁰ and

⁴² See Margot E. Kaminski, *Understanding Transparency in Algorithmic Accountability*, CAMBRIDGE HANDBOOK OF THE L. OF ALGORITHMS UNIV. PRESS (2020); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 33 (2019); Elizabeth A. Rowe & Nyja Prior, *Procuring Algorithmic Transparency*, 74 ALA. L. REV. 303, 307 (2022); Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J. & TECH. 103, 132 (2018); Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1269–70 (2019).

⁴³ See Christopher S. Yoo, *Beyond Algorithmic Disclosure For AI*, 25 COLUM. SCI. & TECH. L. REV. 314 (2024) (“Any transparency regime must also include key information about the data on which the algorithm was trainedFalse”); Shlomit Yanisky-Ravid & Sean K. Hallisey, “Equality and Privacy by Design:” A New Model of A.I. Data Transparency Via Auditing, Certification, and Safe Harbor Regimes, 46 FORDHAM URB. L.J. 428, 473 (2019).

⁴⁴ See Yoo, *supra* note 36 (“[A]lgorithms may be too complex for observers to understand fully”); Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1862 (2019) (“Machine learning’s inscrutability stems from the fact that the computational mapping from feature inputs to outcome prediction is often hard to explain in terms that are intuitively comprehensible to humans.”); Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 693 (2016) (concluding that large datasets complicating transparency and auditability likely emerges in the analysis of data mining and opacity).

⁴⁵ See *infra* Section I.A.

⁴⁶ See *infra* Section I.B.

⁴⁷ See *infra* Section I.A.

⁴⁸ See *infra* Section II.D.

⁴⁹ See *infra* Section II.D.

⁵⁰ Frank Pasquale & Haochen Sun, *Consent and Compensation: Resolving Generative AI’s Copyright Crisis*, 110 VA. L. REV. ONLINE 207, 220 (2024) (“Ironically, a policy of free appropriation of copyrighted work may even menace AI development itself [through causing model collapse].”); Michael Peel, *The Problem of ‘Model Collapse’: How A Lack of Human Data Limits AI Progress*, FINANCIAL TIMES (July 25, 2024), <https://www.ft.com/content/ae507468-7f5b-440b-8512-aea81c6bf4a5>.

hallucinations.⁵¹ By addressing these pitfalls, the legal standard of output transparency would lay the groundwork for sustained and enhanced AI creativity, fostering a balanced ecosystem for both human and artificial creative endeavors.

For my third contribution, I propose two specific legal reform to implement output transparency, offering practical solutions to the challenges of AI concealment. As I will show, these proposals can overcome the difficulties that other proposed approaches present. The U.S. Copyright Office⁵² and courts,⁵³ with the support of scholars,⁵⁴ have consistently denied efforts to register AI-generated content as copyrightable works, reaffirming that copyright law protects only works authored by humans.⁵⁵ However, reliance on this human authorship requirement alone, while doctrinally sound, proves inadequate amid the rising tide of AI concealment in content creation. As demonstrated in this Article, individuals and companies can readily circumvent this requirement—both technically and legally—due to the absence of mandatory technical identifiers marking content as AI-generated.⁵⁶ Without such labels, individuals and companies can falsely present themselves as the human creators of works produced by AI systems, facing minimal repercussions beyond the denial of copyright registration. This gap underscores the urgent need for more robust legal rules to enforce transparency and accountability.⁵⁷

By sidestepping those limitations, my proposals to implement output transparency would offer a more effective means of regulating AI concealment through law reform. They urge Congress to enact legislation mandating that technology companies embed watermarks in content produced by their AI systems. This watermarking mandate would enable the public

⁵¹ Dylan Brown, *Hallucinations Still the Top Concern for Law's Using AI*, LexisNexis Blogs (Oct. 9, 2024), <https://www.lexisnexis.co.uk/blog/future-of-law/mitigating-hallucinations-strategies-for-lawyers-to-ensure-accurate-ai-generated-information>

⁵² See U.S. Copyright Office, *Zarya of the Dawn* (Registration # VAu001480196) (Feb. 21, 2023); U.S. Copyright Office Review Board, *Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial* (SR # 1-11743923581; Correspondence ID: 1-5T5320R) (2023).

⁵³ See *Thaler v. Perlmutter*, 687 F. Supp. 3d 140, 147 (D.D.C. 2023) (ruling that the authorship requirement as “presumptively being human rests on centuries of settled understanding”).

⁵⁴ See Haochen Sun, *Redesigning Copyright Protection in the Era of A.I.*, 107 IOWA L. REV. 1213, 1250 (2022) (“[A]utonomous AI systems should not be recognized as owners of IP rights because they lack human agency.”); *A.I. and the Creative Double Bind*, 138 HARV. L. REV. 1585, 1585 (2025) (“It has long been taken for granted that creativity is a uniquely human ability.”); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343, 434 (2019); Daniel J. Gervais, *The Machine as Author*, 105 IOWA L. REV. 2053, 2079 (2020).

⁵⁵ See U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 88 Fed. Reg. 16190 (Mar. 16, 2023), https://www.copyright.gov/ai/ai_policy_guidance.pdf.

⁵⁶ See *infra* Section I.A & Part II.B.

⁵⁷ See *infra* Section II.B.

to distinguish between human-authored works and AI-generated ones, thereby promoting transparency and maintaining trust in creative outputs. To support this policy, I further propose specific legislative rules, including which types of AI-generated content should be watermarked first, the obligations of online platforms to support the watermarking regime, legal penalties for tampering with watermarks, and exceptions to the watermarking requirements for private use and communication.⁵⁸

My second proposal recommends that regulators, such as the U.S. Copyright Office, require authors seeking to register new works to provide detailed explanations of how they modified AI-generated content to create those works. This mechanism would effectively distinguish human authorship from AI contributions. The proposed watermarking mandates would facilitate this differentiation, as AI-generated works would be watermarked prior to human modifications. Through differentiating the roles of human and AI creators, this explanation mechanism would better legally recognize and protect the original contributions of human authors who edit, refine, adapt, or transform AI-generated material.⁵⁹ Over time, the mechanism would enhance the public documentation of creativity with more information about trends, influences, and the interplay between human and AI creativity, thereby enriching our collective understanding of artistic and intellectual development in the age of AI.⁶⁰

The remainder of this Article is structured as follows. In Part I, I explore the nature and extent of AI output opacity, a new phenomenon that conceals AI's role in content creation and thereby misrepresents AI-generated content as human-made. I then analyze the ethical and legal challenges stemming from the concealment of AI's role in content generation, examining its broader economic, social, and political ramifications.

In Part II, I argue that existing laws have failed to regulate AI concealment. Existing standards of algorithmic and data transparency, which target the opacity of AI systems' internal processes, are ill-equipped to tackle output-related issues. Despite its assertion of the human authorship requirement for copyright registration, the U.S. Copyright Office has adopted rules that fail to curb AI concealment actions.⁶¹ Although efforts have been made to address these challenges, I demonstrate that a reliance on existing frameworks has proven inadequate, highlighting the need to usher in a more effective regulatory framework.

⁵⁸ See *infra* Section III.A.2.

⁵⁹ See *infra* Section III.B.2.

⁶⁰ See Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1177 (2007) (“[C]reativity [exists] as an emergent property of social and cultural systems, continually shaped by and shaping other social changes”); Rebecca Crootof et al., *Humans in the Loop*, 76 VAND. L. REV. 429, 435-37 (2023).

⁶¹ See *infra* Section I.A.3

In Part III, I present the theory of differentiated creativity, examining how and why it is necessary to distinguish between human and AI creative processes and thereby pave the way for adoption of output transparency as a new legal standard for regulating AI concealment. Building on this theoretical foundation and its policy implications, I offer two legal reform proposals in Part III to promote output transparency. First, I recommend that Congress enact legislation requiring technology companies to implement watermarking for AI-generated content, ensuring its origins are properly documented. Second, I propose that the U.S. Copyright Office should require authors to provide proper explanations about their modifications of AI-generated works during the copyright registration process.

I. THE RISE OF AI OUTPUT OPACITY

A. *Three Forms of AI Output Opacity*

1. Output Opacity at the Technological Level

When generative AI emerged in the late 2010s, technology developers were eager to publicize and celebrate the AI-generated status of works. A notable example is The Next Rembrandt project, initiated in 2016 with the aim of creating a new piece of art that could pass as an undiscovered work by the Dutch master.⁶² To accomplish this, developers and data scientists employed machine learning-based AI to analyze and recreate the artistic elements found in Rembrandt's paintings.⁶³ From the outset, this use of AI technologies was announced to the public.⁶⁴ Similar spontaneous disclosures of AI-generated creation have occurred across various creative sectors, including literature⁶⁵ and music.⁶⁶

The spontaneous disclosure of AI creativity has evolved, with efforts now being made to have AI systems recognized by copyright protection authorities as authors of the works they generate.⁶⁷ These efforts aim to not only publicly disclose the AI-generated status of works but also to protect the legal status of both AI systems and their creations under copyright

⁶² Christina Campbell, *The Next Rembrandt: Originality and Authorship of AI Generated Works* (Oct. 31, 2017), <https://journals.library.columbia.edu/index.php/stlr/blog/view/176>.

⁶³ *Id.*

⁶⁴ Microsoft, *The Next Rembrandt* (Apr. 13, 2016), <https://news.microsoft.com/europe/features/next-rembrandt/> (“Blurring the boundaries between art and technology, The Next Rembrandt project is intended to fuel the conversation about the relationship between art and algorithms, between data and human design and between technology and emotion.”).

⁶⁵ Tristan Greene, *Microsoft's AI Can Convert Images into Chinese Poetry*, TNW (Aug. 10, 2018), <https://thenextweb.com/news/microsofts-ai-can-convert-images-into-chinese-poetry>.

⁶⁶ Rich Haridy, *AI-Generated Pop Song Puts Human Composers on Notice*, NEW ATLAS (Aug. 23, 2017), <https://newatlas.com/ai-pop-music-amper/51018/>.

⁶⁷ Haochen Sun, *Redesigning Copyright Protection in the Era of Artificial Intelligence*, 107 IOWA L. REV. 1213, 1213 (2021).

law.⁶⁸ In November 2018, Steven Thaler applied to register the artwork *A Recent Entrance to Paradise* with the U.S. Copyright Office.⁶⁹ Thaler argued that his Creativity Machine, an AI system using neural networks to generate creative works, should be deemed the author of the artwork since it was its sole creator, while he had only facilitated its creation.⁷⁰

Concealment of AI involvement in creation has begun to occur, however, owing to the rapid rise of much more powerful generative AI systems, such as Dall-E, ChatGPT, and Midjourney. The application of foundation models has revolutionized the capabilities of AI systems in content generation, with such systems able to employ advanced neural networks like generative adversarial networks, transformers, and variational encoders, and then train them with vast amounts of data.⁷¹ These two key functions enable AI systems to generate a wide array of outputs in response to human language instructions.⁷² For instance, in November 2022, OpenAI unveiled ChatGPT, an AI system built on large language models (LLMs). Ushering in a new era of AI, ChatGPT and the subsequent generative AI systems are capable of conducting research, providing writing assistance, and generating simulated videos, audio, and images.⁷³ Considering the high quality of the content produced by these new AI systems, users are increasingly inclined to present such content as human-created, claiming authorship for themselves and obscuring the involvement of AI.⁷⁴

More broadly, the unprecedented capacity of AI systems to generate content, coupled with social media's power to amplify its dissemination, highlights the urgent need to confront issues stemming from undisclosed AI involvement. One report indicates that mainstream image-generation AI models have produced over 15 billion images within a year of their

⁶⁸ *Id.* at 1215.

⁶⁹ U.S. COPY OFFICE REVIEW BD., DECISION ON SECOND REQUEST FOR RECONSIDERATION FOR REFUSAL TO REGISTER A RECENT ENTRANCE TO PARADISE 1 (Feb. 14, 2022), <https://copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>.

⁷⁰ *Id.* at 2 (“In his application, Thaler left a note for the Office stating that the Work ‘was autonomously created by a computer algorithm running on a machine’ and he was ‘seeking to register this computer-generated work as a work-for-hire to the owner of the Creativity Machine.’”).

⁷¹ *What is a Foundation Model?*, AMAZON WEB SERVICES, <https://aws.amazon.com/what-is/foundation-models/>.

⁷² *What is artificial intelligence (AI)?*, INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, <https://www.iso.org/artificial-intelligence/what-is-ai> (“AI uses natural language processing . . . the ability of computers to understand, interpret and generate human language, using text analysis, sentiment analysis and machine translation.”).

⁷³ *What is Generative AI?*, MCKINSEY & COMPANY (Apr. 2, 2024), <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>.

⁷⁴ Noti-Victor, *supra* note 34, at 153 (“[P]roducers are likely to choose to conceal the role of generative AI in a work’s creation, rather than disclose it.”).

launch—a milestone that took human photographers 150 years to achieve.⁷⁵ Given social media’s central role in modern life, it’s unsurprising that AI-generated works are flooding these platforms. For instance, a study examining over 100 Facebook pages found that many were posting dozens of AI-generated images daily, while algorithms recommended even more such content to the researchers involved.⁷⁶

Meanwhile, new technologies have been developed enabling people to conceal the role of AI in creation and to make its outputs even more difficult to differentiate from human-created works. These technologies remove or alter watermarks associated with AI-generated content, edit away blemishes in AI-generated images and videos, and “humanize” AI-generated text.⁷⁷ Digital services aimed at the latter, for instance, have mushroomed online, with programs such as Humanize AI, Quillbot, and StealthGPT⁷⁸ using the LLM natural language processing technique to rewrite and restructure AI-generated text multiple times until it passes an internal AI detection system.⁷⁹ Free AI content humanizer HideMyAI, for example, states that it can “[b]ypass AI detectors, and transform AI generated content to humanlike copy[.]”⁸⁰ WriteHuman, meanwhile, allegedly the world’s most powerful AI content humanizer, promises even more, utilizing “sophisticated algorithms to produce humanized text that simulates human writers”, allowing users to “turn AI into natural writing with the world’s most powerful AI Humanizer.”⁸¹

Nevertheless, methods for identifying and disclosing AI-generated content have been unable to keep up with the rapidly growing capacities of AI systems and the sheer quantity of such content being produced. At its core, concealment of AI involvement exploits a challenge society already faces in distinguishing between human and AI creativity based solely on output.⁸² For example, recent studies discovered that participants struggled

⁷⁵ *AI Has Generated 150 Years Worth of Photographs in Less Than 12 Months, Study Shows*, DESIGNBOOM (Aug. 21, 2023), <https://www.designboom.com/technology/ai-has-generated-150-years-worth-of-photographs-in-less-than-12-months-study-shows-08-21-2023>.

⁷⁶ Renee DiResta, Abhiram Reddy & Josh A. Goldstein, *From Shrimp Jesus to Fake Self-Portraits, AI-Generated Images Have Become the Latest Form of Soc. Media Spam*, THE CONVERSATION (Apr. 24, 2024), <https://theconversation.com/from-shrimp-jesus-to-fake-self-portraits-ai-generated-images-have-become-the-latest-form-of-social-media-spam-226903>.

⁷⁷ Lauren Leffer, *Tech Companies’ New Favorite Solution for the AI Content Crisis Isn’t Enough*, SCI. AM. (Aug. 8, 2023), <https://www.scientificamerican.com/article/tech-companies-new-favorite-solution-for-the-ai-content-crisis-isnt-enough/>;

⁷⁸ Kripesh Adwani, *Eleven Best AI Humanizer Tools – Bypass AI Detection With These Tools*, KRIPESH ADWANI (May 16, 2025), <https://kripeshadwani.com/best-ai-humanizers/>.

⁷⁹ Thierry Lavergne, *Do AI Text Humanizers Work? Can They Bypass AI Detectors?*, WINSTON AI (Sep. 1, 2023), <https://gowinston.ai/do-ai-text-humanizers-work-can-they-bypass-ai-detectors/>.

⁸⁰ HideMyAI, *HideMyAI makes AI Content Undetectable and Humanlike*, <https://hidemy.ai/>

⁸¹ WriteHuman, *Elevate AI to Human Perfection*, <https://writehuman.ai/>.

⁸² See Hazel Anne E Elumba, et al., *Impact Analysis of AI-Created Content on the Decline of Public Trust and the Breakdown of Information Integrity in the Digital Information*

to differentiate haiku written by human poets from those generated by a specialized haiku-creating model.⁸³ Even experienced users of AI systems struggle to detect AI-generated images. For instance, only 1.9% of 31,118 participants in Humanize AI's *Human or AI?* game were able to correctly distinguish between AI-generated and real images or human faces.⁸⁴

2. Output Opacity at the Ethical Level

The concealment of AI use has given rise to unprecedented ethical concerns regarding content authenticity and academic integrity. These concerns underscore the need for an ongoing discourse on the ethical implications and authorship challenges that AI-generated content present within the academic community. The first challenge is that such content might be considered plagiarized, as the work is not entirely the product of the named authors.⁸⁵ Second, it could mislead readers and other researchers who assume the work is the result of human intellectual labor.⁸⁶

There have been numerous reports of scholars using AI applications to write academic papers without disclosing this fact. For example, in March 2023, three British scholars published in a prestigious academic journal a research paper exploring the potential risks and rewards of ChatGPT's growing influence on scholarly pursuits. Remarkably, the entire paper was written by ChatGPT, a fact known only to the journal's editors at the time of publication.⁸⁷ Readers and journalists were initially deceived by ChatGPT's ghostwriting capabilities.⁸⁸ Moreover, a recent investigation has found that nearly 200 research papers published in various academic

Environment, 5 COGNIZANCE JOURNAL OF MULTIDISCIPLINARY STUDIES 72, 73 (2025) (“[P]articipants only achieved an accuracy of 61 to 63 percent in identifying authentic and AI-created images. This would mean that approximately 40 percent of all AI-generated visual images are perceived as genuine by familiar viewers.”); Thomas Roca et al., *How Good Are Humans at Detecting AI-Generated Images? Learnings from an Experiment*, arXiv:2507.18640, at 3 (2025) (“When looking specifically at AI-generated images, 193,779 were seen, of which 121,735 were accurately identified, resulting in a success rate of 63%. These results suggest that, in this setting, our users’ odds at detecting AI images are only slightly higher than flipping a coin.”).

⁸³ Jimpei Hitsuware et al., *Does AI-Human Collaboration Lead to More Creative Art? Aesthetic Evaluation of Human-Made and AI-Generated Haiku Poetry*, 139 COMPUT.’S IN HUM. BEHAVIOR 6 (2023).

⁸⁴ EIN Presswire, *Only 1.9% of People are Able to Differentiate between AI Generated and Human Faces — Deepfake Study by HumanOrAI.io*, EIN PRESSWIRE (Feb. 8, 2024), <https://www.einpresswire.com/article/686492596/only-1-9-of-people-are-able-to-differentiate-between-ai-generated-and-human-faces-deepfake-study-by-humanorai-io>.

⁸⁵ Carrie Spector, *What Do AI Chatbots Really Mean for Students and Cheating?*, STANFORD (Oct. 31, 2023), <https://ed.stanford.edu/news/what-do-ai-chatbots-really-mean-students-and-cheating>.

⁸⁶ Paulina Okunytė, *Google Search Exposes Acad.’s Using ChatGPT in Rsch. Papers*, CYBER-NEWS (last updated Nov. 15, 2023), <https://cybernews.com/news/academic-cheating-chatgpt-openai/>.

⁸⁷ *Id.*

⁸⁸ *Id.*

journals were entirely or partially generated by ChatGPT.⁸⁹ Given AI algorithms' propensity to hallucinate information, disclosure of when and how AI has been used in academic research is highly pertinent to reliability of a paper's findings.⁹⁰ The current system of peer-reviewing papers is not sufficiently rigorous,⁹¹ and given the difficulties in differentiating AI works from human works, academic institutions may need to rely more on AI-detecting technologies to ensure academic integrity.

Universities are increasingly grappling with the challenge of students using AI tools to write papers without disclosure.⁹² This practice raises concerns about plagiarism and academic dishonesty, and it also makes it difficult for instructors to accurately assess students' abilities. Data from Turnitin's AI detection tool reveals that AI-generated content appears in roughly 1 in 10 assignments, with about 3 in 100 being predominantly AI-produced.⁹³ However, another study indicates that self-reported cheating rates in anonymous surveys remained largely unchanged before and after ChatGPT's emergence.⁹⁴ While data has not substantiated fears of rampant AI-enabled cheating, the technology may have heightened the ease of misconduct even if the overall number of cheaters holds steady—a view supported by a survey of over 2,000 students, where 47% agreed that AI simplifies cheating.⁹⁵ Instructors have turned to AI detection tools to combat this, but their reliability is dubious, essentially pitting AI against AI in a counterproductive cycle that may not be the optimal safeguard for upholding students' academic integrity.⁹⁶

In the art world, AI-generated pieces have infiltrated competitions, securing prizes before their synthetic origins come to light. In 2023, Boris Eldagsen submitted an AI-created image titled *Pseudomnesia: The Electrician* to the Sony World Photography Awards, clinching the creative open category, only to decline the honor upon revealing its AI genesis.⁹⁷ Despite his post-submission disclosure, the World Photography

⁸⁹ Peter Hess, *AI Scandal Rocks Academia as Nearly 200 Studies Are Found to Have Been Partly Generated by ChatGPT*, DAILY MAIL (Mar. 19, 2024), <https://www.dailymail.co.uk/sciencetech/article-13211523/ChatGPT-scandal-AI-generated-scientific-papers.html>.

⁹⁰ Okunytė, *supra* note 86.

⁹¹ *Id.*

⁹² Arianna Prothero, *New Data Reveal How Many Students Are Using AI to Cheat*, EDUC. WEEK (Apr. 25, 2024), <https://www.edweek.org/technology/new-data-reveal-how-many-students-are-using-ai-to-cheat/2024/04>.

⁹³ *Id.*

⁹⁴ Spector, *supra* note 85.

⁹⁵ Lauren Coffey, *Students and Professors Believe AI Will Aid Cheating*, INSIDE HIGHER ED (2024), <https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2024/07/29/students-and-professors-expect-more>.

⁹⁶ Tara García Mathewson, *Cal. Coll.'s Spend Millions to Catch Plagiarism and AI. Is the Faulty Tech Worth It?*, THE MARKUP (June 26, 2025), <https://themarkup.org/artificial-intelligence/2025/06/26/ai-detector-california>.

⁹⁷ Alex Greenberger, *Artist Wins Photography Contest After Submitting AI-Generated Image, Then Forfeits Prize*, ARTNEWS (Apr. 17, 2023), <https://www.artnews.com/art-news/>

Organization severed ties with him, declaring, “We no longer feel we are able to engage in a meaningful and constructive dialogue with him.”⁹⁸ This incident underscores the contentious status of AI in artistic realms, highlighting stark shifts in valuation and reception once a work’s AI involvement is exposed, which in turn incentivizes creators to conceal such usage to shield their reputations and evade backlash from the art community.

3. Output Opacity at the Legal Level

Beyond ethical concerns, the concealment of AI has already resulted in legal controversies and penalties. First to loom large were copyright law problems. Some artists have intentionally hidden the fact that their works were produced by AI systems in order to meet the human authorship requirement under copyright law when applying for copyright registrations with the U.S. Copyright Office.⁹⁹

In September 2022, Kris Kashtanova attempted to register her graphic novel *Zarya of the Dawn* with the Copyright Office.¹⁰⁰ During the book’s creation, she input prompts into the generative AI program Midjourney, which produced a series of images she then incorporated into the novel.¹⁰¹ Kashtanova did not disclose her use of Midjourney to the Copyright Office, instead identifying herself as the author of the book. Her application made no mention of Midjourney being used in any part of the work, and nor did it disclaim any portion of it, explicitly stating that “the text of the Work was written entirely by Kashtanova without the help of any other source or tool, including any generative AI program.”¹⁰²

Unaware of Midjourney’s involvement in the creation process, the Copyright Office registered the entirety of *Zarya of the Dawn* in September 2022.¹⁰³ However, when Kashtanova revealed on social media that she had used Midjourney to create images for the book, the Copyright Office deemed the application inaccurate, or “substantively incomplete” at the very least.¹⁰⁴

news/ai-generated-image-world-photography-organization-contest-artist-declines-award-1234664549/.

⁹⁸ *Id.*

⁹⁹ Letter from Robert J. Kasunic, Assoc. Reg. of Copyrights & Dir. of the Off. of Registration Pol’y & Prac., U.S. Copy. Off., to Van Lindberg, Taylor English Duma LLP (Feb. 21, 2023) [hereinafter Lindberg Letter] <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>.

¹⁰⁰ *Id.* at 1-2.

¹⁰¹ Tony Analla, *Zarya of the Dawn: How AI is Changing the Landscape of Copyright Protection*, HARV. J. OF LAW & TECH. (Mar. 6, 2023), <https://jolt.law.harvard.edu/digest/zarya-of-the-dawn-how-ai-is-changing-the-landscape-of-copyright-protection>.

¹⁰² U.S. Copyright Office, *Letter re: Zarya of the Dawn* (Reg. No. VAu001480196) (Feb. 21, 2023), <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>.

¹⁰³ *Id.* at 2.

¹⁰⁴ *Id.* at 2-3.

While the *Zarya* case involved one attempt at concealment, Jason Allen committed two: first, to enter a fine arts competition, and second in an application to the Copyright Office. Allen used Midjourney to generate the artwork *Théâtre D'opéra Spatial* and submitted it the Colorado State Fair fine art competition.¹⁰⁵ Competing against 20 other artists in the digitally manipulated photography category, Allen's artwork was met with acclaim and won first prize in September 2022.¹⁰⁶

Allen subtly hinted at, but did not openly disclose, Midjourney's significant role in creating his artwork.¹⁰⁷ The work, featuring three figures in flowing robes gazing towards a radiant horizon, was rendered with intricate detail.¹⁰⁸ The competition judges claimed they were unaware of Allen's AI use until after awarding him the prize for his category.¹⁰⁹ By concealing his AI use in the creative process, Allen ignited a debate about the implications of AI-generated art in competitions, copyright protection, and the broader sphere of the art world.¹¹⁰

Shortly after winning, Allen submitted an application to the U.S. Copyright Office to register the artwork.¹¹¹ However, his application did not disclose that Midjourney generated the artwork to a significant extent.¹¹² Allen withheld this information to ensure the Copyright Office would consider his artwork for copyright protection, as generally human authorship is a prerequisite for granting copyrights—as in the cases of *A Recent Entrance to Paradise* and *Zarya of the Dawn*.¹¹³ However, due to *Théâtre D'opéra Spatial*'s national recognition as the first AI-generated image to win an art competition, the Copyright Office became aware of Midjourney's role in its creation and requested Allen submit further details about the work.¹¹⁴ Allen then revealed Midjourney's involvement,

¹⁰⁵ John Wenzel, *Global Controversy After AI Art-Win Prompts Rule Changes at Colo. State Fair*, THE DENV. POST (Sep. 1, 2023), <https://www.denverpost.com/2023/09/01/ai-art-colorado-state-fair-2023-rule-changes-global-controversy/>.

¹⁰⁶ Drew Harwell, *He Used AI to Win a Fine-Arts Competition. Was it Cheating?*, WASH. POST (Sep. 2, 2022), <https://www.washingtonpost.com/technology/2022/09/02/midjourney-artificial-intelligence-state-fair-colorado/>.

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*; Wenzel, *supra* note 105.

¹⁰⁹ Wenzel, *supra* note 105.

¹¹⁰ Kevin Roose, *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.*, N.Y. TIMES (Sep. 2, 2022), <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>; Sarah Kuta, *Art Made With A.I. Won a State Fair Last Year. Now, the Rules Are Changing*, SMITHSONIAN MAG. (Sep. 8, 2023), <https://www.smithsonianmag.com/smart-news/the-state-fair-changed-its-rules-after-a-piece-made-with-ai-won-last-year-180982867/>.

¹¹¹ Letter from Copyright Rev. Bd. to Tamara Pester, 1 (Sep. 5, 2023), <https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf>.

¹¹² *Id.*

¹¹³ *Id.* at 2, 3; Letter from Copyright Rev. Bd. to Ryan Abbott, 1, 2, 3 (Feb. 14, 2022), <https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>; Lindberg Letter, *supra* note 99.

¹¹⁴ Letter from Copyright Rev. Bd. to Tamara Pester, 2 (Sep. 5, 2023), <https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf>.

disclosing he “input numerous revisions and text prompts at least 624 times to arrive at the initial version of the image.”¹¹⁵

AI concealment has also led to instances of criminal fraud. In 2024, a North Carolina musician was charged with orchestrating a scheme to inflate his music streaming numbers on platforms such as Spotify and Apple Music.¹¹⁶ The musician artificially boosted his streaming figures by employing AI-driven bots.¹¹⁷ Mimicking genuine listener behavior, fraudulent activity was initially difficult to detect.¹¹⁸ By hiding AI’s involvement, the musician aimed to deceive streaming platforms, manipulate their algorithms, and unjustly benefit from increased visibility and higher royalty payments.¹¹⁹

Moreover, AI concealment has facilitated the creation and spread of disinformation.¹²⁰ For instance, AI imitations of people’s voices and likenesses have been widely used to facilitate scams. In 2024, a finance worker at a multinational firm paid out \$25 million to fraudsters after participating in a video conference call.¹²¹ Deepfake technology was also used to create several other employees who appeared to be present at the meeting as well. Recently, the voice of Italy’s defense minister Guido Crosetto was cloned using AI to facilitate phone scams of multiple Italian business leaders, asking them for money to free kidnapped Italian journalists in the Middle East.¹²² Romance scammers have also utilized AI to build closer connections with victims using deepfake videos and photos.¹²³ Additionally, AI has been used for false advertising in order to generate reviews and endorsements. The U.S. Federal Trade Commission has already announced rules banning the use of consumer reviews and testimonials fabricated by AI.¹²⁴

¹¹⁵ *Id.*

¹¹⁶ Maia Coleman, *The Bands and the Fans Were Fake. The \$10 Million Was Real*, N.Y. TIMES (Sep. 5, 2024), <https://www.nytimes.com/2024/09/05/nyregion/nc-man-charged-ai-fake-music.html>.

¹¹⁷ *Id.*

¹¹⁸ *Id.*

¹¹⁹ *Id.*

¹²⁰ John Villaseñor, *A.I. Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth*, BROOKINGS (Feb. 14, 2019), <https://www.brookings.edu/articles/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/>; Buffett Institute for Global Affairs, *The Rise of Artificial Intelligence and Deepfakes*, BUFFETT INST. FOR GLOB AFFS. 1 BRIEF (Sep. 2025), https://buffett.northwestern.edu/documents/buffett-brief_the-rise-of-ai-and-deepfake-technology.pdf.

¹²¹ Heather Chen & Kathleen Magramo, *Finance Worker Pays Out \$25 Million After Video Call With Deepfake “Chief Financial Officer,”* CNN: WORLD (Feb. 4, 2024), <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>.

¹²² Angela Giuffrida, *AI Phone Scam Targets Italian Business Leaders Including Giorgio Armani*, THE GUARDIAN (Feb. 10, 2025), <https://www.theguardian.com/world/2025/feb/10/ai-phone-scam-targets-italian-business-leaders-including-giorgio-armani>.

¹²³ Laura Goodwin, *Romance Scammer Duped £17k From Me With Deepfakes*, BBC (Dec. 19, 2024), <https://www.bbc.com/news/articles/cdr0g1em52go>.

¹²⁴ Sidley Austin, *U.S. FTC’s New Rule on Fake and AI-Generated Reviews and Social Media Bots*, SIDLEY (Aug. 28, 2024) <https://www.sidley.com/en/insights/newsupdates/2024/08/us-ftcs-new-rule-on-fake-and-ai-generated-reviews-and-social-media-bots>.

B. *Negative Consequences of AI Concealment*

The unprecedented volume of AI-generated content, projected to constitute up to 90% of online material by 2026,¹²⁵ amplifies the perils of concealing AI's role in its production, unleashing a host of unintended consequences that harm individuals while infiltrating economic, cultural, and political spheres. Fundamentally, such concealment capitalizes on the existing challenge of distinguishing human from AI creativity based solely on the output, eroding trust, spreading misinformation, and undermining societal authenticity.

Concealing the role of AI in the creation of works can undermine human creativity by causing a host of copyright-related problems for copyright owners, such as authorship ambiguity and moral rights violations. Traditionally, copyright law extends protection to works created by human authors.¹²⁶ When the contribution of AI is hidden, clarity of authorship is lost, making it challenging to ascertain if the work qualifies for copyright protection. This lack of clarity can result in disputes or challenges to the work's copyright status. Further, moral rights, which encompass the right of attribution and integrity, are pivotal components of copyright law, serving to protect the intimate bond between an author and their creation.¹²⁷ When the contribution of AI is obscured, moral rights can be compromised through false attribution of the work entirely to AI systems, without acknowledgement of the contributions of human authors.

If left unchecked, concealment of AI use could jeopardize the interests of artists, deterring them from producing and sharing their work, and ultimately leading to the collapse of segments within the creative industry.¹²⁸ It is technically straightforward to present AI-generated works as human-created. Consequently, driven by a desire for profit, enhanced reputation, and even attempts to disguise plagiarism, the incidence of false attribution by both individuals and corporations has the potential to increase rapidly in the near future.

Second, concealing the contribution of AI to the creation of works could hamper AI creativity and development, contributing to degradation in training data quality and model collapse.¹²⁹ AI model effectiveness is directly proportional to training data quality.¹³⁰ Concealment of

¹²⁵ Pereira, *supra* note 9.

¹²⁶ Sun, *supra* note 67.

¹²⁷ Rebecca Tushnet, *Naming Rights: Attribution and Law*, 2007 UTAH L. REV. 789, 791 (2007) ("According to authorial high-protectionists, there exists "the dignity and integrity of the unique relationship between the author and her creations").

¹²⁸ Pasquale & Sun, *supra* note 50, at 209.

¹²⁹ Ilia Shumailov et al., *AI Models Collapse When Trained on Recursively Generated Data*, 631 NATURE 755 (2024).

¹³⁰ Cem Dilmegani, *Data Quality in AI: Challenges, Importance & Best Practices*, AIMULTIPLE (Jan. 3, 2024), <https://research.aimultiple.com/data-quality-ai/>.

AI involvement in content creation could lead to AI models inadvertently being trained with AI-generated content, which may lack the nuance and creativity that is inherent in human-generated content. Over time, this can degrade the quality of the AI model and the creativity of its outputs.

Model collapse is a situation in AI training where the model, despite diverse inputs, has a tendency to produce similar or identical outputs. An article published in *Nature* in July 2024 highlighted that an influx of AI-generated content used for training these systems could result in model collapse due to the compounding and magnification of errors inherent in synthetic data.¹³¹ The Article warned that model collapse could “halt the improvement of large language models as they run out of human-derived training data and as increasing amounts of AI-generated text pervade the Internet.”¹³² The inability to distinguish between human-created and AI-generated content due to concealment can make it difficult to identify and rectify this problem.¹³³

Third, AI output opacity can adversely affect the interests of consumers engaging with such content. Specifically, it obstructs consumers’ ability to provide informed consent when interacting with or consuming material. They have the right to know the true nature of the works they engage with, whether for entertainment, educational, or professional purposes.¹³⁴ Presenting AI-generated content as human-created prevents users from considering the potential limitations and errors that may arise from AI systems. As a result, users might place unwarranted trust in the content’s quality and reliability, leading to misunderstandings or ill-informed decision-making.¹³⁵ Such concealment can also impact consumers’ economic interests, as they may be willing to pay a premium for human-created content over AI-generated content.¹³⁶ If the work’s nature is misrepresented, consumers might unwittingly pay more for content they would value less if they were aware of the involvement of AI.

Last but not least, AI output opacity has given rise to larger social and political problems. AI-generated content is susceptible to inherent biases

¹³¹ Shumailov et al., *supra* note 129.

¹³² Elizabeth Gibney, *AI Models Fed AI-Generated Data Quickly Spew Nonsense*, *NATURE* (July 24, 2024), <https://www.nature.com/articles/d41586-024-02420-7>.

¹³³ Alistair Knott et al., *Generative AI Models Should Include Detection Mechanisms as a Condition for Public Release*, 25 *ETHICS INFO. TECH.* 55, 3 (2023).

¹³⁴ Dan L. Burk, *Cheap Creativity and What It Will Do*, 57 *GA. L. REV.* 1669, 1682 (2023) (“[A]s the supply of such [AI-generated] artifacts becomes routine and settled, society begins to place an increasing value on the fruits of human production, imperfect and inefficient as they may be.”).

¹³⁵ Kysar, *supra* note 33, at 531 (“Consumer process preferences instead offer an important vehicle through which individuals influence the world, express their views on public issues, and fashion their moral identity in an era of extraordinary interconnectedness, complexity, and dynamism in the market.”).

¹³⁶ Noti-Victor, *supra* note 34, at 154, 170 (“[Consumers may have] a preference for works by a specific author born out of a sense of fan loyalty.”).

derived from the data used for training.¹³⁷ Concealment of AI involvement could inadvertently magnify and propagate these biases, thereby misleading users into assuming the work is a balanced and impartial representation of human creativity and thought. Further, the ability of AI to generate deepfakes could have devastating consequences for politics by disrupting political dialogue, planting disinformation in voters' minds, and ultimately influencing election results.¹³⁸

C. Algorithmic and Data Transparency's Regulatory Vacuum

Despite the growing severity of issues stemming from AI concealment, regulators have yet to establish comprehensive legal standards to address these challenges. Instead, they have prioritized addressing the inner opaque nature of AI systems, which they identify as a primary source of risks and harms.¹³⁹ Advanced AI systems, such as deep neural networks, consist of multiple layers with countless parameters, interacting in highly complex and nonlinear ways. This complexity also makes it challenging to trace how input data transforms into outputs.¹⁴⁰ Moreover, these systems learn from vast datasets containing subtle, intricate patterns that influence their decisions—patterns often imperceptible or difficult for humans to comprehend.¹⁴¹

To counter the opacity of AI systems, scholars and policymakers have proposed “algorithmic transparency” and “data transparency”¹⁴² as primary legal standards for AI regulation. Algorithmic transparency would require AI companies to disclose details about their systems' inner workings, including model weights, neuron activation functions, and training processes (such as training duration, hardware specifications, and

¹³⁷ See, e.g., Jorge Luis Morton, *On Inscription And Bias: Data, Actor Network Theory, and the Social Problems of Text-to-Image AI Models*, AI ETHICS (2024), <https://doi.org/10.1007/s43681-024-00431-8> (“[G]enerative AI platforms are trained on massive amounts of text data from the Internet . . . [T]hey may generate outputs that are biased . . . Data bias . . . inevitably exists in the text-image pair data collected for training these large models.”); MIT Sloan Teaching & Learning Technologies, *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, <https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/> (“[Generative AI models] can reproduce any falsehoods or biases present in [their training data]”); Betsy Morris, *AI from AI: a Future of Generic and Biased Online Content?*, UCLA ANDERSON REV. (Nov. 8, 2023), <https://anderson-review.ucla.edu/ai-from-ai-a-future-of-generic-and-biased-online-content/> (“AI systems are inherently biased — intentionally or not — by the people who build them and the choice of databases they’re trained on.”).

¹³⁸ Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and Nat’l Sec.*, 107 CALIF. L. REV. 1753, 1771-1784 (2019).

¹³⁹ See Warren J. Von Eschenbach, *Transparency and the Black Box Problem: Why We Do Not Trust AI*, 34 PHILOSOPHY & TECHNOLOGY 1607, 1613 (2021).

¹⁴⁰ *Id.*

¹⁴¹ Andreas Holzinger, et al., *Toward Human-Level Concept Learning: Pattern Benchmarking for AI Algorithms*, 4 PATTERNS 1, 2 (2023).

¹⁴² See *supra* notes 47-48 and accompanying text.

model hyperparameters).¹⁴³ Data transparency, on the other hand, focuses on revealing information about the training data, including its sources, labelling methods, dataset size, and representative samples.¹⁴⁴

Algorithmic and data transparency standards promote clarity in understanding AI systems’ inner workings. They are designed to create openness about an AI’s model structure, data, and training details, thereby laying the groundwork for articulation, in terms that humans can understand, of why an AI produced a specific result.¹⁴⁵ The information disclosed as required by those legal standards illuminates AI systems’ inner workings, such as their processes for producing outputs, making decisions, and handling inputs, in terms that users, developers, and regulators can better comprehend.¹⁴⁶ Without transparency, we are in the dark; with it, the “why” of AI actions comes to light.

Recognizing the need to address the opaque nature of AI system inputs, lawmakers across various jurisdictions have worked to establish legal frameworks that enhance algorithmic and data transparency. For instance, the EU AI Act mandates that high-risk AI systems, deployed in fields such as employment, healthcare, and law enforcement, document their decision-making processes and incorporate human oversight.¹⁴⁷ A hiring AI, for example, must record how it evaluates candidates, revealing to regulators critical factors like the weighting of experience.¹⁴⁸ The Colorado Artificial Intelligence Act,¹⁴⁹ enacted in 2024 as the first comprehensive state-level AI law in the U.S., also focuses on high-risk systems like recruitment tools, requiring annual audits to check for algorithmic bias and clear explanations for significant decisions, such as “Your application was denied due to a specific score.”¹⁵⁰

On the data transparency front, the EU AI Act mandates that providers of high-risk AI systems detail their training datasets, covering sources and potential biases, and perform risk assessments. General-purpose AI, such as ChatGPT, face less stringent requirements but must still provide data summaries and adhere to EU copyright laws. In California, the Artificial

¹⁴³ Yoo, *supra* note 36, at 314 (“Understanding AI’s impact . . . also requires disclosure of many aspects of the data on which the algorithm was trained, including its source, scope, quality, and inner correlations.”); Gyandeep Chaudhary, *Unveiling the Black Box: Bringing Algorithmic Transparency to AI*, 18 MASARYK UNIV. J. L. & TECH. 93, 95 (2024).

¹⁴⁴ Yoo, *supra* note 36, at 330.

¹⁴⁵ *Id.*

¹⁴⁶ See Aniek F. Markus, et al., *The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of The Terminology, Design Choices, And Evaluation Strategies*, 113 J. OF BIOMEDICAL INFORMATICS 1, 2 (2021).

¹⁴⁷ See 2024 O.J. (L 1689) 7.

¹⁴⁸ See *id.* at 66, 71.

¹⁴⁹ Consumer Protections for Artificial Intelligence, COLO. REV. STAT. §§ 6-1-1701—1707 (2024).

¹⁵⁰ See *id.* § 6-1-1701(1), (2), (3), and (9); § 6-1-1703(1), (3), (4), and (7).

Intelligence Training Data Transparency Act,¹⁵¹ signed into law in 2024, requires developers of generative AI available in the state to disclose training data origins, types, and relevance for models trained since 2022.¹⁵² A publicly accessible “high-level summary” is also mandated, fostering greater openness about the data driving these systems.¹⁵³

While algorithmic and data transparency standards effectively clarify the inputs and operational frameworks of AI systems, they do not resolve the output opacity caused by AI concealment. In a regulatory vacuum, such concealment may undermine ethics and accountability in the age of AI.

Recognizing these limitations, regulators have begun addressing the regulatory gap surrounding AI concealment. For example, the EU AI Act introduces a preliminary requirement for technology companies to watermark AI-generated content. Article 50.2 of the Act mandates that “providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video, or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.”¹⁵⁴ Similarly, the California AI Training Data Transparency Act of 2024 imposes comparable watermarking obligations to enhance transparency in AI outputs.¹⁵⁵

However, the watermarking requirement introduced by the EU AI Act and the California AI Transparency Act has raised significant challenges. Critics argue that, in their haste to implement these regulations, EU and California regulators overlooked the technical complexities and limitations of AI watermarking.¹⁵⁶ The lack of a clear conceptual framework leaves uncertainty about whether watermarking technology can advance or will reach a dead end.¹⁵⁷ Critics highlight the vague language in the legal documents regarding watermarking capabilities, emphasizing the need for regulations that align with current technological possibilities and

¹⁵¹ Artificial Intelligence Training Data Transparency, CAL. CIV. CODE §§ 3110–3110.2 (West 2024).

¹⁵² *Id.* §3111.

¹⁵³ *Id.* § 3111(a).

¹⁵⁴ EU AI Act, *supra* note 26, art. 50(2).

¹⁵⁵ CAL. CIV. CODE §§ 3110–3111 (West 2024).

¹⁵⁶ Justyna Lisinska & Daniel Castro, *The AI Act’s AI Watermarking Requirement Is a Misstep in the Quest for Transparency*, CTR. FOR DATA INNOVATION (Jul. 9, 2024), <https://datainnovation.org/2024/07/the-ai-acts-ai-watermarking-requirement-is-a-misstep-in-the-quest-for-transparency/>.

¹⁵⁷ Hodan Omaar, *California’s AI Transparency Law Is a Misstep Other States Should Avoid*, CENTER FOR DATA INNOVATION (Nov. 4 2024), <https://datainnovation.org/2024/11/californias-ai-transparency-law-is-a-misstep-other-states-should-avoid/>, (“The law’s biggest flaw is its reliance on watermarks—a unique signal embedded in AI-generated content. However, watermarks are vulnerable to manipulation and circumvention. For instance, cropping an AI-generated image can remove a visible watermark, while more sophisticated editing can erase even the most robust invisible watermarks. Unless watermarking technology advances significantly, mandating their use risks creating a gap between regulatory intentions and technical feasibility.”).

constraints.¹⁵⁸ They also contend that the requirement was adopted without a thorough theoretical exploration of its significance, leaving fundamental questions unresolved and undermining the Acts’ conceptual foundation.¹⁵⁹ In response to these issues, companies like Meta have publicly stated they will not adopt the EU’s General-Purpose AI Code of Practice, which includes the watermarking requirement, as part of its regulatory framework for implementing the AI Act.¹⁶⁰

Focusing solely on technology companies, the watermarking requirements under both the EU AI Act and the California AI Transparency Act fail to address how to prevent or stop AI users from concealing the role of AI in content creation. For instance, the rise of AI-generated academic papers or examples like The Velvet Sundown’s music demonstrate that concealment extends beyond visual and audio media into text and creative industries. Relying on watermarking alone does not capture the full spectrum of AI concealment, such as fraudulent authorship claims in academia or misleading advertisements, which demand broader transparency measures. Mandatory disclosure of human modifications to AI-generated content—an approach I will propose in this Article—can more effectively hold AI users accountable and promote genuine transparency.

Furthermore, a significant setback to the AI watermarking requirement occurred in the U.S. In 2023, the Biden Administration issued an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which included a mandate for watermarking AI-generated content to enhance transparency.¹⁶¹ However, in January 2025, President Trump rescinded this order upon taking office, halting the watermarking mandate and leaving the U.S. without a federal policy on this matter.¹⁶² This abrupt reversal not only stalled the implementation of AI watermarking in the U.S. but also likely hindered its global adoption, given the U.S.’s leadership in the AI development market.

¹⁵⁸ Xuandong Zhao, et al., SoK: *Watermarking for AI-Generated Content*, ARXIV PRE-PRINT ARXIV:2411.18479 (2024); Mihai Christodorescu, et al., *Securing the future of GenAI: Policy and technology*, ARXIV PREPRINT ARXIV:2407.12999 (2024).

¹⁵⁹ Mihai Christodorescu et. al., *Securing the Future of GenAI: Policy and Technology*, DEF. ADVANCED RSCH. PROJECTS AGENCY (2024) (“The field lacks a clear metric for alignment [beyond the results of loss minimization when preference tuning and preference optimization], and thus at this point it is not feasible to rank models in terms of alignment Overly broad regulation risks stifling the very innovation policymakers seek to govern. Current regulation aims for highly desirable guarantees from GenAI, including privacy, fairness, interpretability, but only defines them with broad strokes.”).

¹⁶⁰ Ram Iyer, *supra* note 30.

¹⁶¹ Exec. Order No. 14110, 88 Fed. Reg. 75191 (Oct. 30, 2023).

¹⁶² Exec. Order No. 14179, 90 Fed. Reg. 8741 (Jan. 23, 2025).

D. U.S. Copyright Office's Failed Attempts to Regulate Output Opacity

To address challenges posed by generative AI, the U.S. Copyright Office has introduced three requirements, detailed below, stemming from legal disputes where parties obscured AI's role in content creation or blurred the lines between human and AI authorship. Despite these efforts, the Copyright Office has struggled to effectively counter the output opacity resulting from the growing concealment of AI involvement in the creative process.

1. Human Authorship

The U.S. Copyright Office has mainly utilized the notion of human authorship to defy efforts to register AI-generated content as copyrightable works. Pursuant to this human-centric notion, copyright law only protects works created by humans.¹⁶³ Hence, the door to copyright protection should never be opened to AI systems, no matter how creative their outputs are.

The Copyright Office first applied the human authorship requirement to Steven Thaler's abovementioned registration application for *A Recent Entrance to Paradise*.¹⁶⁴ In August 2019, the application was rejected on the grounds that the artwork lacked human authorship.¹⁶⁵ Thaler demonstrated his AI system's involvement in creating the artwork, but failed to provide evidence of any human author contributing to its creative aspects.¹⁶⁶ The Copyright Office's Review Board upheld the initial ruling, citing convincing evidence from statutory text, judicial precedents, and administrative practices.¹⁶⁷

In dealing with Thaler's lawsuit against the Copyright Office, the District Court for the District of Columbia backed the administrative decision to deny the copyright registration of the AI-generated image due to the absence of human authorship. The court employed the human-centric approach to support its determination of the legal concept that "[h]uman authorship is a bedrock requirement of copyright."¹⁶⁸ According to the court, this concept is supported by the plain text and legislative intent

¹⁶³ See *Thaler v. Perlmutter*, 130 F.4th 1039, 1045 (D.C. Cir. 2025).

¹⁶⁴ Zachary Small, *As Fight Over A.I. Artwork Unfolds, Judge Rejects Copyright Claim*, THE NEW YORK TIMES (Aug. 22, 2023), <https://www.nytimes.com/2023/08/21/arts/design/copyright-ai-artwork.html>.

¹⁶⁵ U.S. Copyright Office, *Initial Letter Refusing Registration from U.S. Copyright Office to Ryan Abbott* (Aug. 12, 2019) (deciding that the work in question "lacked the required human authorship necessary to sustain a claim in copyright").

¹⁶⁶ Copyright Review Board, *Second Request for Reconsideration for Refusal to Register a Recent Entrance to Paradise* (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071), at 2 (Feb. 14, 2022) (concluding that Thaler "provided no evidence on sufficient creative input or intervention by a human author in the Work").

¹⁶⁷ *Id.* at 4–5.

¹⁶⁸ *Thaler v. Perlmutter*, 687 F. Supp. 3d 140, 146 (D.D.C. 2023).

of the Copyright Act,¹⁶⁹ as well as previous judicial rulings,¹⁷⁰ all of which regard human authorship as a prerequisite for copyright protection.¹⁷¹

As discussed above, after approving the registration of the book *Zarya of the Dawn* in September 2022, the Copyright Office learned from social media that its author had publicly acknowledged that Midjourney generated the book's comic images.¹⁷² Given this lack of disclosure and realizing the role of AI in the creative process, the Office reevaluated the case and ultimately reversed its initial approval of registration, adhering to its stance on the necessity of human authorship for copyright protection.¹⁷³

In a statement of policy released in March 2023, the Office thereafter required applicants to disclose sufficient information with respect to AI-generated works, specifically, the presence of AI-generated content and a brief description of the human author's contributions.¹⁷⁴ As stated in the Copyright Act, such disclosures are considered relevant information regarding the preparation, identification, existence, ownership, or duration of the copyright.¹⁷⁵

Another critical issue addressed by the Copyright Office is whether an AI system, or its user providing input prompts, controls the generation of a work. In the February 2023 statement of policy, the Office concluded that the AI system in question primarily exercises control over the work's creation. The reasons for this determination were explained as follows:

[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

¹⁶⁹ *Id.* at 147–48 (“The 1976 Act’s ‘authorship’ requirement as presumptively being *human* rests on centuries of settled understanding.”).

¹⁷⁰ *Id.* at 148 (“The human authorship requirement has also been consistently recognized by the Supreme Court when called upon to interpret the copyright law. . . . Accordingly, courts have uniformly declined to recognize copyright in works created absent any human involvement False”).

¹⁷¹ In the earlier, related case of *Naruto v. Slater*, the court ruled that animals do not have legal standing to sue. *Naruto v. Slater*, 888 F.3d 418, 426 (9th Cir. 2018) (ruling that “if an Act of Congress plainly states that animals have statutory standing, then animals have statutory standing[, but i]f the statute does not so plainly state, then animals do not have statutory standing.”). This case has been widely interpreted as reaffirming human agency as the foundational criterion for claiming authorship and copyright protection. *See, e.g.*, Shyamkrishna Balganesh, *Authoring the Law*, 68 J. COPYRIGHT SOC’Y U.S.A. 353, 357–58 (2021) (“Copyright has long embodied a fairly distinctive conception of authorship, one that seeks to identify the source of a work and its creative elements as a precondition to protection . . . Personalization — through identification — is therefore central to copyright . . .”).

¹⁷² U.S. Copyright Office, *Zarya of the Dawn* (Registration # VAu001480196) at 2 (Feb. 21, 2023).

¹⁷³ *Id.* at 2–3.

¹⁷⁴ U.S. Copyright Office, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190, 37 C.F.R. 202 (Mar. 15, 2023).

¹⁷⁵ *Id.* at 4–5.

the human user. Based on the Office’s understanding of the generative AI technologies currently available, users do not exercise ultimate creative control over how such systems interpret prompts and generate material. Instead, these prompts function more like instructions to a commissioned artist—they identify what the prompter wishes to have depicted, but the machine determines how those instructions are implemented in its output.¹⁷⁶

The Office further clarified its stance with a hypothetical scenario: Suppose a user instructs a text-generating AI to “write a poem about copyright law in the style of William Shakespeare.”¹⁷⁷ While the user could expect the AI to generate text that resembles poetry, incorporates copyright law concepts, and mirrors Shakespeare’s unique style, it is the AI that determines the poem’s rhyme scheme, the vocabulary used in each line, and the overall structure of the text.¹⁷⁸ Hence, when an AI system independently makes these creative decisions, the resulting work is not a product of human creation. Therefore, such work does not qualify for copyright protection, and any copyright registration application should explicitly state this.¹⁷⁹ The Copyright Office decided that the images in *Zarya of the Dawn*¹⁸⁰ and the other two images¹⁸¹ were not copyrightable works, mainly on the basis of the applicant’s lack of control over the creative process.

2. Unsolved Problems

The requirements of human authorship, disclosure, and creative control proposed by the Copyright Office and courts may not be sufficient to mitigate factors contributing to AI output opacity and thereby promote transparency in the case of AI-generated content. The requirements for human authorship and disclosure are ineffectual in dissuading or inhibiting individuals from obscuring AI involvement in creative works. The efficacy of the human authorship requirement is largely dependent on how accurately and honestly applicants adhere to the disclosure requirement. If individuals choose to withhold or misrepresent information about the role of AI in creating works, this effectively undermines the human authorship requirement.

¹⁷⁶ *Id.* at 4.

¹⁷⁷ *Id.*

¹⁷⁸ *Id.*

¹⁷⁹ *Id.* (“When an AI technology determines the expressive elements of its output, the generated material is not the product of human authorship. As a result, that material is not protected by copyright and must be disclaimed in a registration application.”).

¹⁸⁰ U.S. Copyright Office, *supra* note 174, at 8.

¹⁸¹ U.S. Copyright Office, *Second Request for Reconsideration for Refusal to Register SURYAST* (SR # 1-11016599571; Correspondence ID: 1-5PR2XKJ), at 1, 3 (Dec. 11, 2023).

Despite being designed to ensure transparency, the authorship disclosure requirement lacks the necessary enforcement mechanisms to effectively prevent individuals from concealing the role of AI in creative works. As discussed earlier, various self-serving motives may lead people to obscure the creative contributions of AI systems.¹⁸² Rather than acting as a deterrent to the registration of AI-generated works, the human authorship requirement could inadvertently incentivize individuals to obscure AI involvement in order to secure copyright protection for their AI-produced works.

Prominent cases of AI concealment, including Kashtanova's graphic novel *Zarya of the Dawn* and Allen's artwork *Théâtre D'opéra Spatial*, highlight the ease with which people can bypass the human authorship requirement.¹⁸³ Given the extensive media coverage of the *Thaler* case and both artists' engagement with AI technology, they should have been aware of the human authorship requirement and its purpose of denying copyright protection for AI-generated works. Thus, they may have deliberately chosen not to disclose sufficient information about their use of AI systems when submitting registration applications to the Copyright Office.

Moreover, the Copyright Office faces significant challenges in identifying instances where applicants fail to comply with the disclosure requirement. Before the artist publicly revealed her use of AI, the Office approved the *Zarya of the Dawn* application.¹⁸⁴ Following its inability to detect this concealment of information, the Office implemented the disclosure requirement.¹⁸⁵ However, it did not establish specific penalties for applicants who fail to comply with it beyond the existing \$2,500 fine for making "a false representation of material fact," diminishing its effectiveness as a deterrent against concealment.¹⁸⁶ The absence of such penalties likely emboldened Allen to withhold information about his use of AI in his initial *Théâtre D'opéra Spatial* application to the Office, even though the disclosure requirement was already in place.

The control requirement also poses significant challenges, much like the human authorship requirement, due to its reliance on compliance with the disclosure requirement. If an applicant chooses to evade the disclosure requirement, the control requirement would not be activated, rendering it ineffective.

It is worth highlighting that the concept of "control" is a subject of considerable debate in copyright law. The Copyright Office has posited

¹⁸² See *supra* Part I. B.

¹⁸³ Noti-Victor, *supra* note 34, at 163 ("Indeed, the ease with which registrants can hide information regarding a work's AI origins can be seen in both the *Théâtre D'opéra Spatial* and *Zarya of the Dawn* disputes.").

¹⁸⁴ See U.S. Copyright Office, *supra* note 52, at 1.

¹⁸⁵ See U.S. Copyright Office, *supra* note 55, at 4–5.

¹⁸⁶ See Noti-Victor, *supra* note 34, at 163.

that an AI system's control over a work it generates is paramount. However, there is a compelling argument that the user of an AI system also exerts control over the creation process, as they can input prompts to guide the system. The recent globally noted case *Li v. Liu*¹⁸⁷ saw the Beijing Internet Court ruling in favor of a user of Stable Diffusion, recognizing them as the copyright owner of images generated by this AI system. This ruling was based on the premise that the user's numerous prompts effectively controlled the image generation process.¹⁸⁸ This has sparked vigorous debate among commentators over whether it is the AI systems or their users who truly exercise control over content generation.¹⁸⁹

Finally, it is important to note that the three requirements might not apply beyond the scope of the Copyright Office's registration mechanism, as a significant portion of AI-generated content may not be registered with the Copyright Office. In the United States, authors primarily register their works with the Copyright Office in case legal disputes arise concerning their creations. Registering a work establishes a public record of the copyright claim and provides legal evidence of the author's ownership, which is particularly valuable in cases of copyright disputes or infringement allegations.¹⁹⁰ This is why copyright registration is a prerequisite for filing a copyright infringement lawsuit: to enforce their rights and pursue legal action for copyright infringement, authors must register their works with the Copyright Office beforehand.¹⁹¹ However, the majority of works do not face copyright infringement lawsuits, so their authors may not find it necessary to register them with the Copyright Office.

II. PROMOTING OUTPUT TRANSPARENCY: A THEORY OF DIFFERENTIATED CREATIVITY

Given the challenges posed by AI output opacity and the limitations of existing efforts to address it, I propose a new theoretical framework that establishes "output transparency" as a new legal standard for regulating AI-generated content. Grounded in the philosophical underpinnings of

¹⁸⁷ *Li v. Liu*, Jing 0491 Min Chu No. 11279 (Beijing Internet Court Nov. 27, 2023) (China).

¹⁸⁸ *Id.* at 12.

¹⁸⁹ See, e.g., Edward Lee, *Prompting Progress: Authorship in the Age of AI*, 76 FLA. L. REV. 1445, 1581 (2024) (arguing that "creators should receive copyrights for their AI-prompted works that embody a minimally creative selection or arrangement of elements."); Johannes Fritz, *The Notion of 'Authorship' Under EU Law—Who Can be an Author and What Makes One an Author? An Analysis of the Legislative Framework and Case Law*, 19 J. OF INTEL. PROP. L. & PRAC. 552 (2024), <https://academic.oup.com/jiplp/article/19/7/552/7614897>.

¹⁹⁰ See U.S. COPYRIGHT OFFICE, *CIRCULAR 1: COPYRIGHT BASICS* 5 (2021), <https://www.copyright.gov/circs/circ01.pdf> ("Registration establishes prima facie evidence of the validity of the copyright and facts stated in the certificate when registration is made before or within five years of publication.").

¹⁹¹ *Id.* ("Before an infringement suit may be filed in court, registration (or refusal) is necessary for U.S. works.").

creativity, the theory of differentiated creativity explains the necessity of distinguishing between human-authored and AI-generated works and identifies three core conditions of human creative activity: intentionality in creation, emotional expression, and control over the creative process. While human creativity flourishes under these conditions, AI-generated outputs fall short. This fundamental distinction necessitates clear differentiation of their respective outputs to ensure transparency.

A. *Intent to Create*

The intent to create constitutes a triggering condition for human creativity. It encompasses the purpose, aspiration, or objective that sparks and fuels the human creative process. It also acts as a compass for the creative journey, with plans molding the resulting work.¹⁹²

Intent can take myriad forms. It may be a profound message, stirring emotion or unique ideas which the creator seeks to articulate or convey through their work. This intention, formulated within the creative's mind, is the foundation of their creative endeavor.¹⁹³ For example, artistic creation requires an intention to communicate or express an emotion, concept or experience, exemplifying the vital role of human agency. Without a conscious decision to create, there would be no artistic product.¹⁹⁴ While intentionality in creativity does not necessarily require conscious deliberation or meticulous planning, creatives usually begin their work with a proximal intention to produce some new ideas or works.¹⁹⁵

In contrast, aesthetic value can exist independently of human creative intent. Natural phenomena, such as a uniquely shaped snowflake, a vibrant sunset with cascading red-orange hues, or intricate dune patterns sculpted by desert winds, are often admired for their beauty. However, these are not products of human creativity. Their aesthetic appeal arises naturally, without any deliberate intent to create.¹⁹⁶

The intent to create triggers not only the creative process but also the formation of plans to carry it forward.¹⁹⁷ Rather than adhering to a

¹⁹² Intention can be broadly defined as “an executive attitude towards plans, where this attitude is further characterized in terms of the various functions it performs in our lives.” PAISLEY LIVINGSTON, *ART AND INTENTION: A PHILOSOPHICAL STUDY* 14 (2005).

¹⁹³ See, e.g., Dan L. Burk, *Thirty-Six Views of Copyright Authorship*, by Jackson Pollock, 58 Hous. L. Rev. 263, 268–69 (2020) (contending that “concepts of actual and proximate cause, intent, and volition” have “have received surprisingly little attention in the literature on copyright”); Shyamkrishna Balganesh, *Causing Copyright*, 117 COLUM. L. Rev. 1 (2017); Mark Bartholomew, *Copyright and the Creative Process*, 97 NOTRE DAME L. Rev. 357 (2021).

¹⁹⁴ See LIVINGSTON, *supra* note 192, at 12–13.

¹⁹⁵ See *id.*, at 42–43.

¹⁹⁶ See Kendall L. Walton, *Categories of Art*, 79 PHILOS. Rev. 334, 355 (1970).

¹⁹⁷ See Mark Bartholomew, *Copyright and the Creative Process*, 97 NOTRE DAME L. Rev. 357, 360 (2021) (“Neuroscience confirms that creative works are the product of a particular process that involves lengthy planning, deliberation, and focus.”).

single specific plan from inception to completion, the creative process often entails developing a series of plans, which are continuously refined as the work progresses under the guidance of a more general plan.¹⁹⁸ Thus, the process encompasses the formulation and execution of both macro and micro plans.¹⁹⁹ While it is true that the creative process frequently involves spontaneously occurring and unforeseeable moments, these are significantly influenced by earlier and subsequent episodes of planning and reasoning.²⁰⁰ Therefore, if an individual intends to write a business letter, they are unlikely to inadvertently create a beautiful sonnet instead.²⁰¹

Imagine a snowboarder carving through fresh powder on a pristine slope, unknowingly leaving behind a striking pattern in the snow that is visible only from above. Although the snowboarder later notices the pattern, he leaves the slope without claiming credit for its creation.²⁰² Despite the pattern's beauty, it is not the result of the snowboarder's creative process as he neither intended to create the pattern nor planned its formation. This visually captivating pattern, however, with its unintended aesthetic appeal, catches the eye of a photographer, who captures it in a photograph. In contrast to the snowboarder, the photographer actively engages in the creative process, deliberately intending and planning to photograph the pattern.

1. AI's Lack of Intent

By their nature, machines seem incapable of possessing intent. As demonstrated above, intent is commonly seen as a mental determination to achieve a specific action or result, a concept that poses immediate challenges when applied to machines like AI systems.²⁰³ Some scholars argue that machines cannot form intentions due to their lack of a "first-person perspective necessary for agency" and the self-consciousness required to recognize themselves or their thoughts as their own.²⁰⁴ Consequently, AI-generated works often lack the symbolism or personal narratives that characterize human art. Unlike human artists, who infuse their creations

¹⁹⁸ See LIVINGSTON, *supra* note 192, at 14–15 (defining intent as “an executive attitude towards plans, where this attitude is further characterized in terms of the various functions it performs in our lives”).

¹⁹⁹ *Id.* at 44.

²⁰⁰ *Id.* at 46.

²⁰¹ See *id.* at 38.

²⁰² See Elliot Samuel Paul & Dustin Stokes, *Creativity*, STAN. ENCYCLOPEDIA OF PHIL. 7 (Feb. 16, 2023), <https://plato.stanford.edu/entries/creativity/>.

²⁰³ Lance Eliot, *On the Beguiling Question Of Whether AI Can Form Intent, Including the Case of Self-Driving Cars*, FORBES (Jun. 6, 2020), <https://www.forbes.com/sites/lanceeliot/2020/06/06/on-the-beguiling-question-of-whether-ai-can-form-intent-including-the-case-of-self-driving-cars/>.

²⁰⁴ Lynne Rudder Baker, *Why Computers Can't Act*, 18 AM. PHIL. Q. 157, 157 (1981).

with intention, intuition, and purpose, AI does not truly comprehend the meaning behind the works it produces.²⁰⁵

Some scholars challenge the notion that AI lacks intent, advocating for a broader interpretation that reconciles the apparent contradiction of programmed machine intention.²⁰⁶ This expanded view might consider intent present when reasons trigger an action, provided those reasons cause the action appropriately.²⁰⁷ Such a definition could potentially apply to AI if a system's action stems from a reason within its opaque processes, even if that reason remains inaccessible to us.²⁰⁸ Alternatively, if an AI model could articulate the rationale behind its output, it might be seen as demonstrating intent.²⁰⁹ However, even under this more inclusive definition, the intent behind an AI-generated output is arguably not the AI's own but rather a product of its programming.

AI systems lack intent due to their fundamental design as tools to enhance human productivity and execute tasks. They rely on human input to generate outputs and do not operate autonomously. An AI with its own intentions would be less effective as a tool. In AI-generated art, human intervention drives the creative process by defining the macro plan through prompts that specify the core elements and characteristics of the work, ensuring the output aligns with a particular category. While users may refine outputs through additional prompts, this iterative tweaking reflects micro-level planning, with the overarching vision originating in the user's mind, independent of the AI. Although AI contributes to micro plans by determining how to execute the user's specifications, its role remains mechanical, lacking creative intent.

B. Expression of Emotions

1. Human Emotions

While the intent to create serves as the triggering condition for human creativity, the expression of emotions through creative activities constitutes its shaping force. Humans, endowed with sentient biological bodies, derive emotions from experiences, such as tasting an apple, seeing

²⁰⁵ *AI-Generated Art*, INTERACTION DESIGN FOUND. (Jul. 23, 2024), <https://www.interaction-design.org/literature/topics/ai-generated-art>.

²⁰⁶ Don Berkich, *Machine Intentions*, 18 THE AM. PHIL. ASS'N NEWSLETTER ON PHILO. AND COMPUTERS 3, 3, 5 (2018).

²⁰⁷ See *id.* at 6 (“Suppose . . . Smith gets on the plane marked ‘London’ with the intention of flying to London, England. Without alarm and without Smith’s knowledge, a shy hijacker diverts the plane from its London, Ontario destination to London, England. Smith’s beliefs and pro-attitudes caused him to get on the plane marked ‘London’ so as to fly to London, England. Smith’s intention is satisfied, but only by accident, as it were. So it must be that Smith’s reasons cause his action in the right way, thereby avoiding so called wayward causal chains.”).

²⁰⁸ See Eliot, *supra* note 203.

²⁰⁹ *Id.*

the color red, or interacting with others, through perceptions, impressions and associations.²¹⁰ These emotions, in turn, provide the motivational force that drives human actions.²¹¹ Human creativity, whether lowbrow or highbrow, is fundamentally an expressive process that channels aesthetic, intellectual, or moral emotions into the resulting work.

a. Aesthetic Emotions

Aesthetic emotions are a vital aspect of the creative process. They serve as a source of inspiration for artists and contribute to the emotional depth and resonance of their works, enriching the experience of the audience.²¹² As reactions to works of art, aesthetic emotions may be triggered by the appreciation or assessment of beauty and artistic expression in creative forms such as visual arts, music, literature, and dance.²¹³ Spanning an emotional spectrum from joy, awe, and wonder, to fascination, and even melancholy and sadness,²¹⁴ aesthetic emotions stem from recognition and admiration of the sensory, symbolic, and expressive aspects of an artistic piece.²¹⁵

Vincent Van Gogh's masterpiece *Starry Night* epitomizes the expression of aesthetic emotions through creative work. The swirling patterns of the stars and the vibrant colors used in the painting evoke a sense of wonder and fascination, reflecting Van Gogh's own aesthetic emotions upon observing the night sky.²¹⁶ In literature, authors can evoke aesthetic emotions through their choice of words, imagery, and narrative structure. A poignant example is the classic novel *To Kill a Mockingbird* by Harper Lee, which elicits a wide range of aesthetic emotions through its exploration of themes such as innocence, morality, and social justice.²¹⁷

²¹⁰ See DAVID HUME, A TREATISE OF HUMAN NATURE bk. I, pt. I, § I, at 1 (L.A. Selby-Bigge ed., 2d ed. 1978) (1739) ("All the perceptions of the human mind resolve themselves into two distinct kinds, which I shall call IMPRESSIONS and IDEAS. The difference betwixt these consists in the degrees of force and liveliness, with which they strike upon the mind, and make their way into our thought or consciousness.").

²¹¹ See *id.* bk. II, pt. III, § III, at 415 ("Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.").

²¹² Rui Xu, *Discussing the Aesthetic Emotion of Artworks by AI and Human Artists with the Mediating Variable of Aesthetic Fluency*, in CROSS-CULTURAL DESIGN. APPLICATIONS IN ARTS, LEARNING, WELL-BEING, AND SOCIAL DEVELOPMENT – HCII 2021 LECTURE NOTES IN COMPUTER SCIENCE 86 (Rau, PL.P. eds., 2021).

²¹³ *Id.*

²¹⁴ Ursula Beermann et al., *Dimensions and Clusters of Aesthetic Emotions: A Semantic Profile Analysis*, 12 FRONTIERS IN PSYCH. 1 (2021), <https://doi.org/10.3389/fpsyg.2021.667173>.

²¹⁵ John Dilworth, *Artistic Expression as Interpretation*, 44 BRITISH J. AESTHETICS 1, 24 (2004).

²¹⁶ RICHARD THOMSON, VINCENT VAN GOGH THE STARRY NIGHT 3, 12 (The Museum of Modern Art, New York 2008).

²¹⁷ Jeffrey Somers, *'To Kill a Mockingbird' Themes, Symbols, and Literary Devices*, THOUGHTCO. (Jan. 28, 2020), <https://www.thoughtco.com/to-kill-a-mockingbird-themes-4693699>.

b. Intellectual Emotions

Intellectual emotions arise from cognitive and emotional responses during activities such as learning, problem-solving, reasoning, and evaluating ideas. These emotions are crucial for nurturing creativity across various fields, including scientific research and innovation. Specifically, epistemic emotions, such as curiosity, interest, confusion, surprise, and awe, are triggered by cognitive tasks like tackling complex problems or generating novel ideas.²¹⁸ For instance, curiosity and interest can drive scientists to explore new research questions, while confusion and surprise may inspire them to question established assumptions and develop alternative hypotheses.

Building on this, achievement emotions relate to the outcomes of intellectual pursuits, encompassing feelings of success or failure.²¹⁹ These emotions significantly influence researchers' and inventors' motivation and perseverance. For example, pride and satisfaction from successfully completing a research project or creating a groundbreaking invention can strengthen their commitment to their work. Conversely, frustration or disappointment from setbacks may either discourage further efforts or motivate individuals to overcome challenges, refine their hypotheses, and enhance their innovations.²²⁰

c. Moral Emotions

Moral emotions, particularly self-conscious emotions and empathy, play a significant role in human creativity. Self-conscious emotions—such as shame, guilt, embarrassment, and pride—arise from self-reflection and self-evaluation, whether consciously experienced or subtly influencing behavior.²²¹ These emotions evolved to help individuals navigate group dynamics,²²² fostering conformity to social norms and preventing disapproval from others.²²³ By promoting adherence to rules, self-conscious emotions help maintain social order. In contrast, empathy focuses on others, serving as an emotional process that drives moral

²¹⁸ See Dilworth, *supra* note 215, at 16–17.

²¹⁹ Reinhard Pekrun, *The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice*, 18 EDUC. PSYCH. REV. 315, 317 (2006).

²²⁰ See *id.* (discussing the role of achievement emotions for students in the educational context).

²²¹ June Price Tangney et al., *Moral Emotions and Moral Behavior*, 58 ANNU. REV. PSYCH. 345, 347 (2007).

²²² JONATHAN HAIDT, THE MORAL EMOTIONS, in HANDBOOK OF AFFECTIVE SCIENCES 852, 853, 855 (R. J. Davidson, K. R. Scherer, & H. H. Goldsmith eds., 2003).

²²³ *Id.* at 861.

behavior. An empathetic response to another's distress often motivates actions to alleviate suffering or refrain from causing harm.²²⁴

These moral emotions directly influence creative processes, as seen in responses to societal events. For example, Flood and Frank note that after the September 11 attacks on the World Trade Center, filmmakers took empathetic and socially conscious steps: the teaser trailer for *Spider-Man* (2001) featuring the Twin Towers was withdrawn, the climactic scene of *Men in Black II* (2002) was reshot, and several other productions removed the Twin Towers in post-production.²²⁵ These decisions reflect both empathy for those affected and a desire to align with social expectations, demonstrating how moral emotions shape creative outputs.

2. AI's Lack of Emotions

Do AI systems express emotions through their engagement in creative activities such as generating works? The preceding section shows that human creativity is first a physical process triggered by our biological bodies; it is a profound communicative medium of human experience that gives rise to three types of emotions. Given that AI systems lack such biological bodies and real-life experience, they are unable to express these emotions.

While the rise of ChatGPT and other generative models has prompted some debate around whether these algorithms could be sentient, it is crucial to remember that machines lack the experience necessary for communication.²²⁶ Although a large language model (LLM) like ChatGPT can generate statements such as "I'm hungry," it does not have physiological needs.²²⁷ It is simply not capable of experiencing hunger, pain, love, or any sensation derived from a physiological state.²²⁸ In essence, while AI systems like ChatGPT can mimic human language and expression, they remain fundamentally different from humans in their inability to experience genuine emotions and sensations.²²⁹

Ultimately, AI is not engaging in the expression of emotion, as a generative AI system's process of "creation" is one of imitation. LLMs are trained using vast datasets of real human expression to be able to "generate highly realistic and coherent text."²³⁰ When an LLM generates the statement

²²⁴ *Id.* at 862.

²²⁵ Maria Flood & Michael C. Frank, *How 9/11 Changed Cinema*, THE CONVERSATION (Sept. 10, 2021), <https://theconversation.com/how-9-11-changed-cinema-167323>

²²⁶ Fei-Fei Li & John Etchemendy, *No, Today's AI Isn't Sentient. Here's How We Know*, TIME (May 22, 2024), <https://time.com/6980134/ai-llm-not-sentient/>.

²²⁷ *Id.*

²²⁸ *Id.*

²²⁹ *Id.*

²³⁰ Henrik Skaug Sætra, *Generative AI: Here to Stay, but for Good?*, 75 TECH. IN SOC'Y 1, 1 (2023).

“I’m hungry,” “it is simply generating the most probable completion of the sequence of words in its current prompt” and when it responds to another prompt with “I am not hungry,” it is engaging in exactly the same process.²³¹ The same is true of other forms of generative AI. For instance, training diffusion models to generate images requires a large dataset which the system will learn to emulate by adding and removing “noise” from the sample images.²³² No matter what prompt a diffusion model is responding to, when providing an output, it is simply generating noise so that what has been asked of it resembles what it has learned.²³³

While human creativity is an expressive endeavor, the imitative nature and purpose of generative AI means that artificial creativity is only ever evident in outputs. Artificial creativity is successful when an AI-generated work convincingly emulates human expression, and companies commonly market their generative models on the convincingness of this imitation. For instance, when OpenAI launched GPT-4, the company stressed its ability to exhibit “human-level performance on various professional and academic benchmarks.”²³⁴ News stories on technological advancements in the AI sector are also often premised on new capacities for emulation.²³⁵ Highlighting the success of these efforts to imitate, a recent study on evaluation of AI art found that respondents had “clear preferences for artwork labelled as human-created as compared to AI-created, despite all art being AI-created.”²³⁶ Although the products of human and artificial creativity are now often indistinguishable, the expressive process which shapes the former remains diametrically opposed to the mechanical process of imitation behind the latter.

C. Control Over the Creative Process

1. Control as Part of Creative Abilities

To complete a creative process, an individual must have the ability to control it. Control over a creation process is the finalizing condition of

²³¹ Li & Etchemendy, *supra* note 226.

²³² Ziyi Chang et al., *On the Design Fundamentals of Diffusion Models: A Survey*, 179 PATTERN RECOGNITION 1, 2 (2025), <https://arxiv.org/pdf/2306.04542>.

²³³ *Id.*

²³⁴ GPT-4, OPENAI (Mar. 14, 2023), <https://openai.com/research/gpt-4>.

²³⁵ See e.g. Alyson Klein, *AI Can Mimic Students’ Writing Styles. How Are Teachers Supposed to Catch Cheaters Now?*, EDUC. WEEK (Dec. 13, 2023), <https://www.edweek.org/technology/ai-can-mimic-students-writing-styles-how-are-teachers-supposed-to-catch-cheaters-now/2023/12>; Abdul Samad, *Review – The Language of Deception*, E-INT’L RELS. (Mar 5, 2024), <https://www.e-ir.info/2024/03/05/review-the-language-of-deception/>.

²³⁶ See Lucas Bellaiche et al., *Humans Versus AI: Whether and Why We Prefer Human-Created Compared to AI-Created Artworks*, COGNITIVE RSCH. PRINCIPLES AND IMPLICATIONS (July 4, 2023), <https://doi.org/10.1186/s41235-023-00499-6>.

human creativity, embodying human authority to exercise final decision-making over the planning and execution of an creative output, such as a film, song, performance, or other artistic creation.²³⁷ In the creative process, the mind plays an indispensable originating and causative role, making it inseparable from the overall creative endeavor.²³⁸

Human creativity involves “cognitive and neural processes that support the generation of novel and contextually appropriate ideas.”²³⁹ From a cognitive science viewpoint, “[‘control’] refers to a set of neurocognitive processes, typically associated with frontal lobe activity, that maximize the adaptability of the human cognitive system to contextual demands.”²⁴⁰ When an individual engages in creative thinking, these processes do more than just generate ideas. They also play a vital role in discerning which ideas are original, adequate, or novel.²⁴¹ This is crucial as the creation of valuable ideas requires the suppression of tendencies to produce commonplace ones.²⁴² Thus, this cognitive control not only facilitates the generation of ideas but also ensures their quality and uniqueness, reinforcing its central role in the creative process.²⁴³

The capacity for creative thinking in humans is demonstrated through their cognitive control at various stages. For example, at the idea generation stage, the efficiency of cognitive control can dictate accessibility to limited cognitive resources when encountering novel tasks. Simultaneously, effective control governs the “suppression of an automatized response to produce a novel, original, unexpected solution” when dealing with well-established tasks.²⁴⁴ At the evaluation stage, cognitive control can determine the efficiency of suppressing irrelevant response tendencies. This process may partially depend on information stored in memory. Thus, cognitive control plays a pivotal role in both the generation and evaluation stages of creativity, influencing the efficiency and originality of the creative process.²⁴⁵

Individuals can maintain control over the creative process even when using tools or machinery, as long as they remain responsible for generating and evaluating ideas.²⁴⁶ There is a distinction between creation

²³⁷ See MICHAEL L. SICILIANO, *CREATIVE CONTROL: THE AMBIVALENCE OF WORK IN THE CULTURE INDUSTRIES*, 12 (2021).

²³⁸ Howard W. Hintz, *Causation, Will, and Creativity*, 55 J. PHIL. 514, 519 (1958).

²³⁹ Evangelia G. Chrysikou, *Creativity in and out of (Cognitive) Control*, 27 CURRENT OP. IN BEHAV. SCIS. 94, 94 (2019).

²⁴⁰ *Id.*

²⁴¹ Magdalena Groborz & Edward Necka, *Creativity and Cognitive Control: Explorations of Generation and Evaluation Skills*, 15 CREATIVITY RSCH. J. 183, 194–95 (2003).

²⁴² *Id.*

²⁴³ *Id.*

²⁴⁴ *Id.* at 195.

²⁴⁵ *Id.*

²⁴⁶ United States Copyright Office, *supra* note 55.

employing tools and creation deferring to tools.²⁴⁷ In the former case, individual creators use tools “to recast, transform, or adapt their expressive authorship.”²⁴⁸ Instances include a photographer using a camera to take photos, “a visual artist [using] Adobe Photoshop to edit an image. . . , and a [musician using] effects such as guitar pedals when creating a sound recording.”²⁴⁹ What truly matters is the extent to which the individual maintains “creative control over the work’s expression.”²⁵⁰ In the cases above, the person utilizing the camera, editing software or guitar pedal still ultimately generates, evaluates, and controls the creative components of their work, ensuring that their creative control is preserved throughout the process.²⁵¹

In the absence of significant intervention in a human’s creative process, such as a second party performing the ultimate act of creation, the individual is generally seen as the originator of the creative work they have set out to create.²⁵² Intervention sufficient for creative control to be seized can be less direct, such as when coercion leads an individual to make creative choices they would not otherwise have made.²⁵³ In the production of the 1974 documentary, *General Idi Amin Dada*, for example, the film’s subject threatened to execute 150 French citizens being held hostage if certain scenes were not removed from the film, leading the director to comply but reject creative ownership of the work.²⁵⁴

2. AI’s Lack of Control

Based on a cognitive account of creative thinking alone, the distinction between human and artificial creative processes is not immediately apparent. Generative AI systems, in the process of generation through their neural networks and other machine learning models, learn to evaluate potential outputs in order to serve their purpose effectively.²⁵⁵ However, the distinction becomes more evident when attempting to establish these processes as causally responsible for the creation of a work. Despite outwardly comparable “neural” processes controlling the creation of human

²⁴⁷ See *id.*

²⁴⁸ *Id.*

²⁴⁹ *Id.*

²⁵⁰ *Id.*

²⁵¹ See *id.*

²⁵² Balganes, *supra* note 193, at 5.

²⁵³ See Livingston, *supra* note 192, at 81–82.

²⁵⁴ *Id.* at 82 (The intended director Barbet Schroeder “complied, but distributed leaflets saying that the work had become a film by Idi Amin Dada, with Barbet Schroeder as an assistant.”).

²⁵⁵ See, e.g., Alberto Romero, *GPT-3 – A Complete Overview*, TOWARDS DATA SCI. (May 24, 2021), <https://towardsdatascience.com/gpt-3-a-complete-overview-190232eb25fd>; Molly Ruby, *How ChatGPT Works: The Model Behind the Bot*, MEDIUM: TOWARDS DATA SCI. ARCHIVE (Jan. 30, 2023), <https://medium.com/data-science/how-chatgpt-works-the-models-behind-the-bot-1ce5fca96286>.

and artificial outputs, can it be said that both outputs are “the achievement of a sovereign . . . author or artist”?²⁵⁶

A generated work cannot be considered the sovereign creative achievement of an AI model, as the creative process is not an exercise of control or a decision-making authority. While AI lacks a mind, more important in this context is the extreme difficulty in establishing a generative model as the ultimate cause for a creative output. Human intervention, by both programmers and users, is apparent throughout the creative processes of today’s generative AI models because AI is ultimately bound by the rules established by human programmers. This is most apparent in the case of “narrow” or “weak” AI models which “are skilled in a very specific task, where they can often perform at superhuman levels.”²⁵⁷

A model trained solely for the purpose of playing chess is a classic example of narrow AI,²⁵⁸ whereas examples of generative AI include systems trained to write code or draft emails.²⁵⁹ Systems trained for such singular purposes are unsuitable for “tasks or task environments that offer little structure, consistency, rules or guidance”, as “[k]nowing and following fixed procedures usually does not lead to proper solutions in these varying circumstances.”²⁶⁰ A model programmed only to write code for a specific purpose could not, for example, independently identify a new purpose and generate code that would fulfill it “because it cannot reason from a general perspective and adapt accordingly.”²⁶¹ Therefore, at least for narrow AI systems, the process of artificial creativity itself is arguably an insufficient intervention to break the causal connection between human programmers and generated outputs.

More advanced generative AI models pose a significant challenge to the notion of human-led creativity. Systems like ChatGPT have a wide array of potential applications, as they can learn to generate any data distribution and create novel outputs, akin to real-world examples across various modalities including images, music, speech, and prose.²⁶² However, despite the seemingly independent creativity resulting from the learning process, generative AI is still bound by rules like any other algorithmic

²⁵⁶ See Livingston, *supra* note 192, at 62.

²⁵⁷ J. E. (Hans). Korteling *et al.*, *Human- Versus Artificial Intelligence*, 4 FRONTIERS A.I. 1, 8 (2021).

²⁵⁸ Bernard Marr, *The Difference Between Generative AI And Traditional AI: An Easy Explanation for Anyone*, FORBES (Aug. 23, 2023), <https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/>.

²⁵⁹ *Id.*

²⁶⁰ See Korteling *et al.*, *supra* note 257, at 8.

²⁶¹ *Id.*

²⁶² J. Mark Bishop, *Artificial Intelligence is Stupid and Causal Reasoning Will Not Fix It*, 11 FRONTIERS PSYCH. 1, 4 (2021).

process.²⁶³ These rules are derived from the training data after the fact, meaning that all products of artificial creativity are already implicit within the combination of the algorithmic process and the data on which it has been trained.²⁶⁴

The concept of artificial general intelligence (AGI) envisions systems with increased autonomy and human-level intelligence,²⁶⁵ which could potentially possess the creative control necessary to sever the causal connection between human programmers, human authors of training data, and generated works. However, the feasibility of AGI's existence is highly debated,²⁶⁶ and regardless of how intelligent and autonomous AI agents become in some respects, they will likely remain unconscious machines lacking genuine sovereignty.²⁶⁷

Assuming artificial sentience is possible, AI agents might be likened to commissioned artists. Despite the constraints placed on them by the commissioning party, these artists are still recognized as engaging in the creative process and maintaining a degree of control over the work.²⁶⁸ Therefore, the fact that an AI adheres to the directives of another does not necessarily deprive it of creative control. The crux of the issue is not the AI's compliance with the user's prompts, but rather its inability to exercise discretion or judgment due to its need to adhere to programmed rules.

D. *Benefits of Output Transparency*

The theory of differentiated creativity, as I put forward in the preceding sections, posits that human creativity is contingent upon three aspects: intent to create, expression of human emotions, and control over the creative process. In contrast, AI creativity can flourish without fulfilling these conditions, relying instead on the automatic operation of AI models and human data. While it is essential to investigate whether AI has reached a level of creativity comparable to that of humans, which could be measured by the famous Turing test,²⁶⁹ it is equally vital to scrutinize how the creative process of AI differs from that of humans.

²⁶³ Luke Tredinnick & Claire Laybats, *Black-Box Creativity and Generative Artificial Intelligence*, 40 BUS. INFO. REV. 98, 99 (2023).

²⁶⁴ *Id.*

²⁶⁵ See Korteling et al., *supra* note 257, at 2; see, e.g., Lauren Leffer, *In the Race to Artificial General Intelligence, Where's the Finish Line?*, SCI. AM. (Jun. 25, 2024), <https://www.scientificamerican.com/article/what-does-artificial-general-intelligence-actually-mean/>.

²⁶⁶ See Leffer, *supra* note 265.

²⁶⁷ See Korteling et al., *supra* note 257, at 2.

²⁶⁸ See Katerina Bantinaki, *Commissioning the (Art)Work: From Singular Authorship to Collective Creatorship*, 50 J. AESTHETIC EDUC. 16, 16 (2016).

²⁶⁹ See generally A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433 (1950); see also Melanie Mitchell, *The Turing Test and Our Shifting Conceptions of Intelligence*, SCIENCE (Aug. 15, 2024), DOI:10.1126/science.adq9356 ("Turing's point was that if a

This theory does not focus on the output of AI creativity itself. Rather, it closely examines the processes through which humans and AI systems engage in creative activities and how these diverge.²⁷⁰ From this process-based perspective, the theory demonstrates the need to differentiate between AI-generated and human-created content, which subsequently gives rise to a need for “output transparency,” requiring disclosure of AI involvement in generating content.²⁷¹

Building on the theory of differentiated creativity, I advocate for output transparency as a new legal standard to regulate AI technology, alongside existing algorithmic and data transparency frameworks. While models and data are critical to AI content generation, the outputs of these systems deserve equal scrutiny. Concealing AI’s role in content creation often leads to parties claiming proprietary ownership over such works. To address this, technology companies and users should be required to disclose sufficient information about the AI-generated nature of their works, thereby promoting output transparency. This new legal standard for AI regulation could effectively address the major problems with AI concealment outlined in Part I, providing better avenues for promoting the creativity of both human activities and AI systems.²⁷²

1. Promoting Human Creativity

Disclosing the involvement of AI in content generation could enhance human creativity by providing robust protection for authors’ rights, particularly their right to consent and claim compensation. The supply of new human data may also decline as artists become unmotivated due to the widespread use of AI works in place of human works. Drove of copyright infringement cases have arisen as a result of the widespread “scraping” of copyright images by tech companies, with over a dozen suits pending in the United States alone.²⁷³ These suits allege copyright infringement in the

computer seems indistinguishable from a human (aside from its appearance and other physical characteristics), why shouldn’t we consider it to be a thinking entity?”).

²⁷⁰ For arguments supporting a closer look into the creative process, see Balganesch, *supra* note 193 (“Originality, however, remains a rather poor fit for the causal question, given that it focuses almost entirely on the work itself rather than on the process of creation.”); Bartholomew, *supra* note 197 (“Yet copyright law blinds itself to information on the creative *process*, judging creativity by exclusive reference to the final *product*—the allegedly creative work itself—and repeatedly insisting that even accidental and unconscious conduct can be creative.”).

²⁷¹ For more about the importance of transparency, see generally Oliver Schilke & Martin Reimann, *The Transparency Dilemma: How AI Disclosure Erodes Trust*, 188 ORG. BEHAV. & HUM. DECISION PROCESSES 104405 (2025).

²⁷² See *supra* Part I. B.

²⁷³ David M. McIntosh et al., *AI and the Copyright Liability Overhang: A Brief Summary of the Current State of AI-Related Copyright Litigation*, ROPES & GRAY (Apr. 2, 2024), <https://www.ropesgray.com/en/insights/alerts/2024/04/ai-and-the-copyright-liability-overhang-a-brief-summary-of-the-current-state-of-ai-related>.

use of copyrighted images for training, as well as in the works produced by the AI.²⁷⁴ Because the outcomes of these cases have yet to be determined, artists have found other ways to protect their works against AI companies. Data poisoning tools such as Glaze and Nightshade have been created to implement invisible and unperceivable visual changes to works that, when fed into AI algorithms, confuse them and cause them to misinterpret the image.²⁷⁵ These rising tensions between artists and tech companies are unsustainable; without human creativity, AI development cannot progress. A compromise is necessary to ensure that an appropriate balance is struck between AI development and the protection of human artists and artistic works.

Implementing a disclosure requirement for AI works may tip the balance in artists' favor without completely stunting AI development. Researchers have observed that people generally prefer works labelled human-created over AI-created, even when works are in fact created by AI.²⁷⁶ As a result, the disclosure requirement will have a significant influence on consumer behavior, protect the market for human-made works and human creativity, and enable artists to leverage the authenticity of their works to increase their value.²⁷⁷ This would also open opportunities for artists to license their works to tech companies for training, allowing them to cooperate with tech companies rather than oppose them.

Acknowledging AI's role in content generation streamlines the process of determining appropriate compensation for the original author. When AI-generated works draw upon or are based on a human author's work, such disclosure facilitates the establishment of proper royalty payment and revenue-sharing channels. Consequently, authors could secure their rightful compensation for the use of their works in AI-generated content. Although some artists may continue to oppose the idea of AI art and refuse to contribute to its development, the advancement of AI is a reality that all artists need to accept, and cooperation with tech companies may be the best way forward for them. Tech companies should also ensure that they take steps to welcome this cooperation by financially compensating artists for their contributions. Otherwise, although AI companies can avoid using AI-generated works for training, the supply of human works may not be sufficient to sustain AI development.

²⁷⁴ *Id.*

²⁷⁵ Lauren Leffer, *Artists Are Slipping Anti-AI 'Poison' into Their Art. Here's How It Works*, SCI. AM. (Mar. 1, 2024), <https://www.scientificamerican.com/article/art-anti-ai-poison-heres-how-it-works/>.

²⁷⁶ Eric W. Dolan, *New Psychology Research Reveals Why People Prefer Human-Created Artwork to AI Created Artwork*, PSYPOST (Aug. 3, 2023), <https://www.psypost.org/new-psychology-research-reveals-why-people-prefer-human-created-artwork-to-ai-created-artwork/>.

²⁷⁷ See Bellaiche et al., *supra* note 236.

It is possible that a remuneration system for artists contributing to AI training may have undesirable side effects. Realistically, for this system to be financially viable for tech companies the remuneration provided will be modest.²⁷⁸ Artists may begin producing large quantities of low-quality works to maximize financial gain, which would fail to resolve the problems surrounding hallucinations and model collapse, and would also not contribute to enhancing or protecting creativity. Despite this, such instances are not likely to be common. It appears that income is not the main motivation for many artists.²⁷⁹ Nonetheless, some kind of remuneration or at least permission from artists is necessary to give them the acknowledgment they deserve, and tech companies must understand that their algorithms' future depends on artists' continuous contributions.²⁸⁰

2. Promoting AI Creativity

Adopting a legal standard of output transparency has great potential in promoting AI creativity. First, this legal standard would reduce the risk of model collapse, a phenomenon in which an AI system, such as a neural network, converges to a suboptimal solution or is unable to learn diverse representations from its training data.²⁸¹ The danger of model collapse has sent AI developers racing to secure high-quality data sources, particularly from social media sites which can ensure a steady flow of data.²⁸² For example, OpenAI has signed a deal with Reddit to train ChatGPT on its user content.²⁸³ The same deal was also made with Google to train Gemini.²⁸⁴ Nonetheless, these measures are unlikely to be sufficient to maintain AI models' current trajectory of growth. According to Epoch, it is predicted that such models will "utilize the full supply of public human text data at some point between 2026 and 2032."²⁸⁵ Furthermore, training

²⁷⁸ See, e.g., Martin Senftleben, *Generative AI and Author Remuneration*, 54 INT'L REV. INTELL. PROP. & COMPETITION L. 1535, 1548 (2023).

²⁷⁹ See, e.g., JESSICA SILBEY, THE EUREKA MYTH: CREATORS, INNOVATORS, AND EVERYDAY INTELLECTUAL PROPERTY 27–28 (2014); Daniel J. Gifford, *Innovation and Creativity in the Fine Arts: The Relevance and Irrelevance of Copyright*, 18 CARDOZO ARTS & ENT. L.J. 569, 602 (2000), https://scholarship.law.umn.edu/faculty_articles/347.

²⁸⁰ See Frank Pasquale & Haochen Sun, *supra* note 50, at 221.

²⁸¹ *Id.* ("LLMs are *language* models, not *knowledge* models, and have no ability to independently reason about what is in the human-generated articles or images they process.").

²⁸² Matt O'Brien, *AI 'Gold Rush' For Chatbot Training Data Could Run out of Human-Written Text*, AP NEWS (Jun. 6, 2024), <https://apnews.com/article/ai-artificial-intelligence-training-data-running-out-9676145bac0d30ecce1513c20561b87d>.

²⁸³ Michelle Chapman, *OpenAI, Reddit Teaming in Deal that Will Bring Reddit's Consent to ChatGPT*, AP NEWS (May 17, 2024), <https://apnews.com/article/reddit-openai-chatgpt-bd2291fcc226bc737a44dbef4a31563f>.

²⁸⁴ Adam Smith, *What are AI 'Hallucinations' and can they be Stopped?*, CONTEXT (July 8, 2024), <https://www.context.news/ai/what-are-ai-hallucinations-and-can-they-be-stopped>.

²⁸⁵ Pablo Villalobos et al., *Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data*, arXiv:2211.04325, at 12 (2024), <https://arxiv.org/abs/2211.04325>.

data supplied to AI models has been increasing by approximately two and a half times annually, while computing power has been increasing by about four times each year.²⁸⁶ This issue is exacerbated to a certain extent by the growing prevalence of AI content in the media, which makes it difficult to find reliable, high-quality data sources that will not lead to model collapse. Many human works are also now being created with assistance from AI, making it even harder to differentiate AI data from human data.²⁸⁷ OpenAI CEO Sam Altman stated that OpenAI has begun experimenting with generating synthetic data for training AI; Altman suggests that although there is low-quality synthetic data, there is also low-quality human data.²⁸⁸ Feeding AI with low-quality data such as those with errors, biases, or insufficient diversity can cause the AI to hallucinate more often.²⁸⁹ Data with errors may cause the AI to reproduce the errors, and biased data can cause the AI to develop a skewed or incomplete understanding.²⁹⁰ Some AI systems may even fabricate information in order to fill in gaps when asked about topics underrepresented in its training data.²⁹¹ Another risk is overfitting, where an AI model learns the details and noise in the training data too well, making it too tailored to the training data such that it becomes unable to make predictions on information outside of its training set.²⁹² All these factors demonstrate how much care and attention are required to train a competent AI.²⁹³

Under the output transparency standard, disclosure of AI involvement would provide much-needed information as to the source of hallucinations and help technology companies refine their algorithms to avoid making mistakes. Since AI's inability to truly understand the information, it is fed is a major factor leading to hallucinations, human oversight is required

²⁸⁶ Robi Rahman & David Owen, *The Training Compute of Notable AI Models Has Been Doubling Roughly Every Six Months*, EPOCH AI (Jun. 19, 2024), <https://epoch.ai/data-insights/compute-trend-post-2010>.

²⁸⁷ Pasquale & Sun, *supra* note 50, at 221–22.

²⁸⁸ Matt O'Brien, *AI 'Gold Rush' for Chatbot Training Data Could Run out of Human-Written Text*, PBS NEWS (Jun. 6, 2024), <https://www.pbs.org/newshour/economy/ai-gold-rush-for-chatbot-training-data-could-run-out-of-human-written-text-as-early-as-2026>.

²⁸⁹ Wan Zhang & Jing Zhang, *Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review*, 13 MATHEMATICS 856, 894 (2025); Yihan Li, et al., *Mitigating Hallucination in Large Language Models (LLMs): An Application-Oriented Survey on RAG, Reasoning, and Agentic Systems*, ARXIV PREPRINT ARXIV:2510.24476 1, 19 (2025).

²⁹⁰ See European Union Agency for Fundamental Rts., *Data Quality and Artificial Intelligence – Mitigating Biases and Error to Protect Fundamental Rights*, 2 (2019).

²⁹¹ See Noman Nayeem, *The Fabrication Problem: How AI Models Generate Fake Citations, URLs, and References*, MEDIUM (Jun. 12, 2025), <https://medium.com/@nomannayeem/the-fabrication-problem-how-ai-models-generate-fake-citations-urls-and-references-55c052299936>; Faiz Surani & Daniel Ho, *AI on Trial: Legal Models Hallucinate 1 out of 6 (or More) Benchmarking Queries*, STAN. INST. FOR HUM.-CENTERED ARTIF. INTEL (May. 23, 2024), <https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries>.

²⁹² IMANOL BILBAO & JAVIER BILBAO, *OVERFITTING PROBLEM AND THE OVER-TRAINING IN THE ERA OF DATA: PARTICULARLY FOR ARTIFICIAL NEURAL NETWORKS* (2017).

²⁹³ See *id.*

to ensure that AI algorithms make the right interpretations and do not reinforce incorrect ones. Only humans can instill a human understanding and perspective into AI algorithms.²⁹⁴ Such an understanding is not something AI can inherently possess and therefore teach itself.²⁹⁵

Hallucinations create a significant risk of misinformation for AI users, especially for those working in fields like public healthcare and legal services. A notorious example of AI hallucinating is the citation of non-existent legal cases.²⁹⁶ The reduction of hallucinations is important for the longevity of AI's use as a tool, as it will allow users to have more trust in AI models and improve AI technology adoption. In terms of AI art, ensuring that AI algorithms are trained with high-quality human data may also lead to a higher appreciation of AI works and AI creativity. Therefore, an effective way to reduce the risk of such hallucinations is to enhance the quality of training data as promoted by the output transparency requirement. With the disclosure of AI's involvement in content generation, human experts could verify and correct AI outputs and offer feedback on how models could function better.²⁹⁷

III. ACHIEVING OUTPUT TRANSPARENCY: TWO PROPOSALS

The theory of differentiated creativity advanced above underscores the necessity of distinguishing between human and AI creative processes. This framework supports adopting output transparency as a new legal standard to regulate actions that obscure AI's role in content creation, effectively addressing the resulting challenges.

To achieve output transparency, I propose two legal reforms to clearly delineate AI-generated and human-created works. First, a new federal law should mandate that technology companies watermark content produced by AI systems. Second, a legal requirement should compel authors to provide proper explanations of how their new works have modified AI-generated works. These reforms would not only enhance human creativity but also promote innovation in AI-driven content creation.

A. Watermarking AI-Generated Content

Watermarking AI-generated content enables tracking and identification, addressing AI concealment, content manipulation, or misuse. In this section, I will explore the nature of watermarking technology, evaluate its feasibility for promoting output transparency, and then propose strategies

²⁹⁴ Cornerstone OnDemand, *The Crucial Role of Humans in AI Oversight*, Cornerstone (Oct. 16, 2025).

²⁹⁵ *Id.*

²⁹⁶ See Rebecca Crootof et al., *Humans in the Loop*, 76 VAND. L. REV. 429, 435-37 (2023) (exploring the role of human intervention in AI decision-making processes).

²⁹⁷ *Id.* at 475.

for legally mandating technology companies to implement watermarking for AI-generated content.

1. Technological Feasibility

Watermarking AI-generated content has been adopted as a straightforward means of differentiating it from human content.²⁹⁸ There are many different approaches to watermarking AI content, but they can generally be categorized into two types of watermarks: visible and invisible.²⁹⁹

Visible watermarks are patterns or images superimposed on the AI-generated content, making it immediately identifiable as AI-created.³⁰⁰ Examples include the use of stock images of watermarks, which usually take the form of a translucent logo or pattern spread across the entire work.³⁰¹ A watermark's purpose tends to be copyright protection, as users may be unable to access unwatermarked images until they agree to pay a fee.³⁰² As of now³⁰³, many AI models have already implemented visible watermarks in their outputs, such as the colored squares on the bottom of Dall-E works.³⁰⁴

Invisible watermarks are not immediately perceptible to the naked eye and require a detector for identification.³⁰⁵ There are currently several kinds of invisible watermarking techniques, such as statistical watermarks.³⁰⁶

Statistical and cryptographic watermarks embed a statistically unusual pattern in AI-generated content that can only be decoded with a specific key.³⁰⁷ These patterns are embedded by manipulating the pixels of an image or adding specific syntax patterns (such as particular word choices

²⁹⁸ Madiega, *Generative AI and Watermarking*, at 2 (2023), [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI\(2023\)757583_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf).

²⁹⁹ Peter Henderson, *Understanding Watermarks for AI-Generated Content*, AI Policy Exchange (Mar. 3, 2023), <https://aipolicyexchange.org/watermarks-ai-content>.

³⁰⁰ See Gustaf Björkstén and Daniel Leufer, *Identifying Generative AI Content: When and How Watermarking Can Help Uphold Human Rights*, ACCESSNOW (Sep. 2023), <https://www.accessnow.org/wp-content/uploads/2023/09/Identifying-generative-AI-content-when-and-how-watermarking-can-help-uphold-human-rights.pdf>.

³⁰¹ See generally Rafal Pytel, *What is AI Watermarking?*, SOFTWAREMILL (Jul. 2, 2024) <https://softwaremill.com/what-is-ai-watermarking/> (archived Oct. 23, 2025) (using the lawsuit between Getty and Stability AI as an example where the AI generated watermark is visible).

³⁰² Photography-Ish, *The Legal and Ethical Implications of Removing a Photographers Watermark*, PICS OF COLO. (Apr. 27, 2023), <https://picsofcolorado.com/blogs/photography-101/the-legal-and-ethical-implications-of-removing-a-photographers-watermark>.

³⁰³ *Voluntary Commitments from Leading Artificial Intelligence Companies on July 21, 2023*, 137 HARV. L. REV. 1283, 1284 (2024).

³⁰⁴ Siddharth Srinivasan, *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*, BROOKINGS (Jan. 4 2024), <https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/>.

³⁰⁵ Takaaki Yamada, *Use of both invisible and emergable watermarks to deter illegal copying of images*, SCIENCE DIRECT (2013), <https://www.sciencedirect.com/science/article/pii/S1877050913009113>.

³⁰⁶ Siddharth Srinivasan, *supra* note 304.

³⁰⁷ *Id.*

or punctuation) to AI-generated text.³⁰⁸ A separate algorithm then analyzes the content to determine whether the observed patterns are randomly produced or the result of a watermark.³⁰⁹ For instance, DALL-E might adjust the low-order bits in an image's color channels in a way that is invisible to the human eye while carrying a signed marker indicating "AI-generated, January 2025."³¹⁰ These imperceptible changes also ensure that the visual quality of the image is preserved. Regarding AI text, watermarking can be achieved by categorizing tokens as "restricted" or "promoted," and increasing the probability of "promoted" tokens.³¹¹ For example, a paragraph generated by Grok 3 might incorporate synonym pairs that are statistically unlikely to appear in human writing, thereby encoding a signature that only xAI's verification tool can decode.³¹² This gives the AI a particular, identifiable style based on the algorithm's preferred use of words.³¹³

Metadata is another means of detecting AI generated content besides invisible watermarks.³¹⁴ It offers a secure way to store provenance data without directly affecting the content's visual or textual elements.³¹⁵ Attaching provenance information using metadata is an existing practice used by news companies like The BBC,³¹⁶ as well as large tech companies like Google, Amazon, Microsoft, and Adobe.³¹⁷ The Coalition for Content Provenance and Authenticity, or C2PA, is a widely known standard for

³⁰⁸ Lev Craig, *AI watermarking*, TechTarget (Oct. 31, 2023), <https://www.techtarget.com/searchenterpriseai/definition/AI-watermarking>.

³⁰⁹ Siddarth Srinivasan, *supra* note 304.

³¹⁰ Jacob Hoffman-Andres, *AI Watermarking Won't Curb Disinformation*, Elec. Frontier Found. (Jan. 5 2024), <https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation>.

³¹¹ Sasha Luccioni, Yacine Jernite, Derek Thomas, Emily Witko, Ezi Ozoani, Josef Fukano, Vaibhav Srivastav, Brigitte Tousignant, Margaret Mitchell, *AI Watermarking 101: Tools and Techniques*, (Feb. 26 2024), <https://huggingface.co/blog/watermarking>.

³¹² GPT Clean Up, *Grok Watermark Detector*, <https://www.gptcleanup.com/grok-watermark-detector>.

³¹³ Jacob Hoffman-Andres, *supra* note 310.

³¹⁴ Christine Deschaseaux, *Invisible Watermark Technology: Resilient Against Gemini 2.0's Watermark Removal Capabilities*, Imatag (Mar. 18 2025), <https://www.imatag.com/blog/invisible-watermark-technology-resilient-against-gemini-2-0s-watermark-removal-capabilities>; Angel Articles, *Metadata and Watermarking to Reduce Piracy*, (last updated July 2025), <https://angelarticles.gitbook.io/articles/metadata-and-watermarking-to-reduce-piracy>; Digimarc, *About Digital Watermarks*, (Oct. 30, 2023), <https://www.digimarc.com/blog/about-digital-watermarks#:~:text=Metadata%20and%20Manifests%20are%20Not,of%20being%20difficult%20to%20remove>.

³¹⁵ Annie Badman & Matthew Kosinski, *What is Metadata?*, IBM, <https://www.ibm.com/think/topics/metadata> [<https://perma.cc/T6D5-YDQA>] (last visited Nov. 6, 2025).

³¹⁶ Grant Hatchimonji, *What is the Coalition for Content Provenance and Authenticity (C2PA)?*, TechTarget (Jan. 14, 2025), <https://www.techtarget.com/whatis/definition/Coalition-for-Content-Provenance-and-Authenticity-C2PA> [<https://perma.cc/4UX2-RYBA>].

³¹⁷ Kyle Wiggers, *Google Will Begin Flagging AI-Generated Images in Search Later This Year*, TECHCRUNCH (Sep. 17, 2024), <https://techcrunch.com/2024/09/17/google-will-begin-flagging-ai-generated-images-in-search-later-this-year/> [<https://perma.cc/TF9L-LDD5>].

metadata of this kind.³¹⁸ Under C2PA, a manifest will be attached to a file containing information about its source of origin, all the changes made to it over time, as well as any changes made to the manifest.³¹⁹ The manifest is viewable by anyone, however, it is also removable by anyone.³²⁰ Simply taking a screenshot of the image will produce a copy without the metadata.³²¹ Overall, this method is quite easy to implement but not very robust, and thus it cannot guarantee extensive watermark coverage.

Watermarking technology has evolved significantly, driven by needs in copyright protection, media authentication, and now AI regulation. Some major AI companies have begun applying it to their generative systems, for example, Google, which has introduced a watermarking method called SynthID to identify images,³²² videos, and even text³²³ generated by its AI systems. TikTok has also begun disclosing AI effects in audio tracks used on their platform,³²⁴ and their community guidelines require users to label content that “uses AI or includes significant edits to show realistic-looking people or scenes” using a label provided by TikTok, or by providing their own clear caption, watermark, or sticker.³²⁵ Similarly, since early 2024, OpenAI has embedded C2PA metadata into images generated by ChatGPT and DALL·E 3.³²⁶ In 2024, to bolster trust and accountability, Meta also unveiled its plan to incorporate invisible watermarks into its text-to-image generation systems.³²⁷ These steps demonstrate coinciding efforts from technology companies to enhance AI transparency, suggesting that they are, to a certain extent, open to the idea of watermarking regulations.

³¹⁸ Kaushal Rathi et al., *Insights into Coalition for Content Provenance and Authenticity (C2PA)*, INFOSYS, <https://www.infosys.com/iki/techcompass/content-provenance-authenticity.html> [<https://perma.cc/Z426-8MFE>] (last visited Nov. 6, 2025).

³¹⁹ *Id.* C2PA Explainer, COALITION FOR CONTENT PROVENANCE AND AUTHENTICITY, <https://spec.c2pa.org/specifications/specifications/1.4/explainer/Explainer.html> [<https://perma.cc/HEZ8-3CXW>] (last visited Nov. 6, 2025).

³²⁰ Hatchimonji, *supra* note 316.

³²¹ *C2PA in ChatGPT Images*, OPENAI, <https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images> [<https://perma.cc/KC7V-FRZP>] (last visited Nov. 6, 2025).

³²² Google DeepMind, *Identifying AI-generated Images with SynthID* (Aug. 29, 2023), <https://deepmind.google/discover/blog/identifying-ai-generated-images-with-synthid/>.

³²³ *Watermarking AI-generated Text and Video with SynthID*, GOOGLE DEEPMIND (May 14, 2024), <https://deepmind.google/discover/blog/watermarking-ai-generated-text-and-video-with-synthid/>.

³²⁴ Nihal Krishan, *AI Watermarking Could Be Exploited by Bad Actors to Spread Misinformation. But Experts Say the Tech Still Must Be Adopted Quickly*, FEDSCOOP (Jan. 3, 2024), <https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/>.

³²⁵ TikTok, *Integrity and Authenticity*, TIKTOK COMMUNITY GUIDELINES (Sep. 13, 2025), <https://www.tiktok.com/community-guidelines/en/integrity-authenticity>.

³²⁶ *C2PA in ChatGPT Images*, OPENAI, <https://help.openai.com/en/articles/8912793-c2pa-in-dall-e-3>.

³²⁷ Nick Clegg, *Labeling AI-Generated Images on Facebook, Instagram and Threads*, META (Feb. 6, 2024), <https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/>.

With available technologies and those yet to be advanced, AI watermarking promotes output transparency by providing content source attribution. Whereas a lack of transparency strips people of control and autonomy, weakening their ability to determine reality, watermarking closes this gap by tagging AI origins, putting users back in charge of how they engage with what they see, hear, or read.

First, watermarks link AI content to the model by which it was created—giving people context to better evaluate the quality and purpose of the content. Due to the degree of realism that AI can achieve, even AI images published in good faith can be mistaken by viewers as depictions of real life.³²⁸ Watermarking provides greater transparency as to the intentions of the author and allows users to assess the reliability of the information presented to them based on the AI’s reputation or intended use.³²⁹ For example, watermarking can help them distinguish between an AI-powered research tool and an AI-powered meme generator, the latter of which may suggest that the content is intended as a joke. A watermarked AI article on a social media platform would indicate that it is not human journalism, prompting readers to evaluate its credibility accordingly.³³⁰ Likewise, we might think twice about using digital art from models that we know have been trained on copyrighted materials due to the risk of potential copyright liability.

Second, AI watermarking also provides a powerful tool to combat deepfakes and disinformation by embedding verifiable markers into AI-generated content. Although AI-generated content is not disinformation per se, generative AI has made the production of fake content significantly easier and more efficient.³³¹ Doctored images have been possible in the past using tools like Photoshop, but AI platforms have substantially reduced the need for the investment of skill, effort, and time. Watermarking can help us navigate the media landscape in an era where AI can convincingly mimic reality and AI content is ubiquitous.³³² By clearly flagging content as artificial, users, platforms, and regulators are provided a means to combat misinformation and mitigate its harms.³³³ Watermarking empowers users

³²⁸ See e.g., Amy Dawel et al., *Can you spot the AI impostors? We found AI faces can look ‘more real’ than actual humans*, THE CONVERSATION (Nov. 13, 2023), <https://theconversation.com/can-you-spot-the-ai-impostors-we-found-ai-faces-can-look-more-real-than-actual-humans-215160>.

³²⁹ Olivia Burrus et al., *Unmasking AI: Informing Authenticity Decisions by Labeling AI-Generated Content*, INTERACTIONS, at 38, 42 (2024), <https://doi.org/10.1145/3665321>.

³³⁰ Soham Nandi, *Enhancing AI Transparency: Researchers Develop Advanced Watermarking Techniques*, AZOAI (Dec. 5, 2024), <https://www.azoai.com/news/20241205/Enhancing-AI-Transparency-Researchers-Develop-Advanced-Watermarking-Techniques.aspx>.

³³¹ See Alexander Loth et al., *Blessing or curse? A survey on the Impact of Generative AI on Fake News*, arXiv:2404.03021 (Dec. 2024), <https://arxiv.org/pdf/2404.03021>.

³³² Xuandong Zhao et al., *SoK: Watermarking for AI-Generated Content*, arXiv:2411.18479, (June 12, 2025), at 1, <https://arxiv.org/abs/2411.18479>.

³³³ *Id.* at 4.

by equipping them with the tools and knowledge to identify, evaluate, and respond to AI-generated content.³³⁴

Without these markers, AI creations—whether a celebrity deepfake, a chatbot’s essay, or a synthetic artwork—blend effortlessly with human work, leaving users unable to distinguish reality from fabrication.³³⁵ Misinformation is particularly damaging in contexts where quick responses are necessary.³³⁶ For example, a deepfake video of a CEO can have significant and instant market repercussions, causing investors to sell their shares and causing economic instability. Likewise, fabricated footage of natural disasters or an impending catastrophe can cause panic and disorder. Although these watermarks cannot inherently verify the truthfulness of content, they at least encourage the viewer to critically think about whether what they are seeing is fabricated and to not act rashly.³³⁷ During elections, watermarked AI-generated speeches could also alert voters to the presence of synthetic content, reducing the impact of unmarked fakes and helping maintain the integrity of information and democracy.³³⁸

Third, the watermarking requirement would also incentivize technology companies to take greater responsibility in protecting copyrighted works and promoting human creativity. Watermarking facilitates the identification of AI-generated content, playing a pivotal role in detecting potential copyright infringements.³³⁹ Generative AI models are frequently trained on extensive datasets that may encompass copyrighted materials, leading to outputs that mimic or integrate protected works without authorization.³⁴⁰ For instance, an AI-produced image or song could unintentionally replicate elements from a copyrighted artwork or musical composition.³⁴¹ By serving as a distinct marker of AI origin, watermarks empower copyright holders to ascertain if their creations were utilized in training datasets or replicated in generated outputs. This enhanced traceability incentivizes companies to procure training data legally via licensed or public domain sources, thereby mitigating legal risks. As an example, should a watermarked AI-generated image bear resemblance to a copyrighted painting, rights owners can probe

³³⁴ *Id.* at 5.

³³⁵ Ashutosh P. Singh, *Safeguarding Authenticity in the Digital Realm: A Holistic Approach Integrating Content Provenance, Secure Watermarking, and Transparent Labeling to Combat Deepfakes*, 6 INT’L J. FOR MULTIDISCIPLINARY RSCH. 1, 2 (2024).

³³⁶ Daniel I. Weiner & Lawrence Norden, *Regulating AI Deepfakes and Synthetic Media in the Political Arena*, BRENNAN CTR. FOR JUSTICE (Dec. 5, 2023), <https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena>.

³³⁷ Singh, *supra* note 335, at 2.

³³⁸ Weiner & Norden, *supra* note 336, at 4.

³³⁹ Kateryna Militsyna, *Can Copyright Law Benefit from the Marking Requirement of the AI Act?*, 56 IIC (2025).

³⁴⁰ Pasquale & Sun, *supra* note 50, at 223; Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUSTON L. REV. 101, 295, 300 (2023).

³⁴¹ Sag, *supra* note 340.

the AI's training procedures for violations, prompting firms to adopt more rigorous data governance practices.

2. Legal Requirements

The advanced state of watermarking technology makes it practical to mandate its use for promoting AI output transparency. The EU AI Act and the California AI Transparency Act of 2024 have introduced preliminary requirements for technology companies to watermark AI-generated content.³⁴² Besides their limitations as discussed in Part II,³⁴³ these laws still lack specificity, providing only broad mandates without clear guidelines for compliance.

To remedy these deficiencies, Congress should enact comprehensive legislation delineating technology companies' obligations for watermarking AI-generated content. Informed by rigorous research and stakeholder input, this law should address key areas such as defining watermarked content, clarifying responsibilities of online platforms, establishing exceptions to watermarking mandates, and imposing penalties for unauthorized removal or tampering with watermarks. By instituting detailed and standardized requirements, such legislation would surmount the shortcomings of the EU AI Act and California's AI Transparency Act, thereby ensuring reliable and consistent implementation across sectors.

First, legislators should examine the types of AI-generated content that should be watermarked. The current technical challenges to watermarking largely revolve around robustness to removal and alteration. Watermarking for text and voice presents unique challenges due to their fluid nature and ease of manipulation.³⁴⁴ For the time being, it is technologically feasible and reliable to watermark both AI-generated images and videos as discussed above.³⁴⁵ Voice watermarking, though feasible with frequency-based signals, struggles with re-recording. Resemble AI, a generative audio company, introduced PerTh (Perceptual Threshold) in 2023. It embeds inaudible data packets into AI-generated speech using machine learning, placing tones in frequency bands masked by louder sounds.³⁴⁶ Text watermarking, on the other hand, remains difficult. Google DeepMind's

³⁴² See *supra* Part I.A.

³⁴³ *Id.*

³⁴⁴ As to text, watermarks should not alter the readability or meaning of the text in question. This can be difficult because text is highly sensitive to changes in word choice, syntax, and grammar. Also, they need to survive common text transformations such as paraphrasing, summarization, or translation. Watermarks for AI-generated voice should not be perceptible to human listeners. They must remain intact despite common audio processing techniques such as compression, filtering, or noise addition.

³⁴⁵ Huixin Luo et al., DIGITAL WATERMARKING TECHNOLOGY FOR AI-GENERATED IMAGES: A SURVEY, Preprints (2025). (Jan. 6, 2025) <https://www.preprints.org/manuscript/202501.0399>.

³⁴⁶ *Introducing Neural Speech AI Watermarker*, (Jan. 31, 2023), <https://www.resemble.ai/neural-speech-watermarker/>.

SynthID, which relies on token probability tweaks, remains vulnerable to paraphrasing and is unreliable for short passages.³⁴⁷

A more focused approach targeting visual content should be implemented first, but there must be a concerted effort to ensure that watermarking text and voice is implemented within a reasonable time. Excluding text and voice from the AI watermarking requirement leaves massive vulnerabilities. A 2023 Pew study found 58% of people worry about AI text misinformation, which is higher than the number of people worried about misinformation from images (52%).³⁴⁸ Voice deepfakes, though less common, hit harder emotionally as they are often employed to impersonate family members and friends in scams. Hence, laws mandating watermarking for AI-generated content should ideally cover all major forms—images, videos, text, and voice—to maximize transparency and combat deception comprehensively.

Another concern is about the compatibility of watermarking techniques with different AI models. In the case of statistical watermarks, the pattern embedded into the AI content must be kept confidential, which may conflict with some AI developers' open-source policies.³⁴⁹ Additionally, some techniques like statistical watermarks will directly influence the weights of the model, and thus will be unique to each AI model.³⁵⁰ If some content is not flagged by the watermark detector of one AI model, it may still be generated by another model. A user must cross-reference the image with every detection software to conclusively determine whether something is AI-generated or not, and even this is made even more difficult if the image was generated by some obscure and lesser-known model. Imposing invisible watermarks on all AI companies may not be particularly effective if it produces such a large number of detection software that it is impractical for users to use them all.

I propose that legislators adopt a tiered and phased legislative strategy for mandating AI watermarking, tailoring requirements to the readiness of technology and the urgency of each content type's risks. AI-generated images and videos should be categorized as high risk because they have a visceral and more immediate impact on viewers. In particular, the use of deepfakes in disrupting politics or falsifying evidence is a serious concern that needs to be addressed immediately.³⁵¹ According

³⁴⁷ *Watermarking AI-generated text and video with SynthID*, GOOGLE DEEPMIND (May 14, 2024), <https://deepmind.google/blog/watermarking-ai-generated-text-and-video-with-synthid/>.

³⁴⁸ See Christian Pheby, *Half of Americans think AI art will be used to spread fake news*, YouGov (May 30, 2023), <https://yougov.com/en-gb/articles/45765-half-americans-think-ai-art-will-be-used-spread-fa>.

³⁴⁹ Srinivasan, *supra* note 304, at 2.

³⁵⁰ *Id.*

³⁵¹ Chloe Wittenberg, et al., *Labeling AI-Generated Content: Promises, Perils, and Future Directions*, MASS. INST. OF TECH. (Mar. 27, 2024), <https://mit-genai.pubpub.org/pub/hu71se89/release/1>.

to the “Picture-Superiority Effect,” people remember pictures better than words.³⁵² Additionally, the majority of social media content is visual, and oftentimes, images and videos may not be accompanied by any caption at all. These factors make visual media a more effective vehicle for misinformation.

Initially, the new law would require both visible and invisible watermarks for AI-generated images and videos, capitalizing on the mature state of statistical watermarks using metadata, cryptographic hashes or pixel-based embeddings. The combined use of both types of watermarks ensures that legislators can leverage the instantaneous communicative effects of visible watermarks whilst bolstering its robustness with more sophisticated invisible techniques. Additionally, the issue of watermarking standardization can be solved by establishing a platform to centralize the AI models’ detection software so that users can analyze content in a more streamlined way.

For AI-generated voice and text, however, legislators may opt for a softer approach, encouraging voluntary watermarking without an immediate legal obligation. To bridge this gap, legislators could establish a 3–5-year phase-in period, providing a runway for technological maturation—say, refining cryptographic audio markers or robust text embedding—before enforcing a mandate. During this time, incentives like tax credits or research grants could spur innovation,³⁵³ ensuring that by the deadline, watermarking for all media types is both practical and effective, balancing innovation with accountability in a dynamic AI landscape.

Second, online platforms should be put under an obligation to verify whether content published on their platform contains any indicators that it is AI-generated and provide a corresponding watermark to inform users. There are three kinds of indicators that platforms must give to users: a watermark stating that the work is AI-generated if the content contains an AI watermark in its metadata, a watermark stating that the content is likely to be AI generated if there is no metadata but the publisher of the work themselves has declared it AI-generated, and a watermark stating that the content is suspected to be AI generated if the platform detects an explicit logo or traces of generative AI from other kinds of invisible watermarks. These three tiers of indicators address the different levels of reliability in each watermarking technique.

Legislators can identify metadata as the most conclusive watermark, while finding user declaration and physical watermarks to be less

³⁵² Sara Paul, *The Picture-Superiority Effect: Harness the Power of Visuals*, Nielsen Norman Group (Apr. 26, 2024), NIELSEN NORMAN GROUP, <https://www.nngroup.com/articles/picture-superiority-effect/>.

³⁵³ See Xuan-Thao Nguyen & Jeffrey A. Maine, *Incentivizing Innovation*, 75 TAX. LAW. 351, 362 (2022); Camilla A. Hrdy, Patent Nationally, Innovate Locally, 32 Berkeley Tech. L.J. 3273271301, 1371 (2017).

convincing, likely due to the increased risk of alteration and fabrication. Other kinds of invisible watermarks like statistical watermarks should be treated with a degree of caution as they are also vulnerable to attacks and provide users with the probability that the work is AI-generated rather than making conclusive determinations. Additionally, these platforms should be required to add to the metadata of the content when labelling it as AI-generated, which ensures that its transmission can be traced more effectively.³⁵⁴

Third, legislators should craft carefully defined exceptions to watermarking mandates in order to balance transparency with practical considerations. A one-size-fits-all watermarking requirement could impose undue burdens or impede legitimate AI applications in specific scenarios. For example, lawmakers might introduce exemptions for personal use of AI outputs. AI tools like chatbots are increasingly employed in private exchanges—such as drafting personal emails, text messages, or social media posts—or for individual creative endeavors, including generating artwork for personal enjoyment. Mandating watermarks in these settings could interrupt fluid interactions, introduce technical hurdles for users, or encroach on privacy by labelling content not meant for public dissemination.³⁵⁵ For instance, someone using an AI tool like Grok to compose a private letter should not be obligated to embed a watermark, given the absence of risks like public deception or copyright infringement.³⁵⁶ To mitigate abuse, technology firms should mandate user confirmations that exempted content remains private and non-commercial, reinforced by terms of service that bar public sharing without watermarks.

Another viable exemption could apply to real-time interactive systems in business operations. AI deployed in dynamic environments—such as virtual assistants, customer service chatbots, or live translation services—emphasizes smooth, efficient exchanges. Inserting watermarks into every output (e.g., a chatbot’s reply or a translated sentence) might degrade user experience, introduce delays, or perplex recipients unacquainted with watermark indicators. For example, compelling watermarks in all responses from a virtual assistant like Siri during informal dialogues would prove cumbersome and superfluous, especially when users generally recognize the AI’s involvement.³⁵⁷ Instead, companies should provide an initial disclosure of AI usage (e.g., “You are engaging with an AI assistant”) to inform users without relying on watermarks.

³⁵⁴ Nick Clegg et al., *supra* note 327.

³⁵⁵ Sylvia Lu, *Data Privacy, Human Rights, and Algorithmic Opacity*, 110 CAL. L. REV. 2087, 2099 (2022).

³⁵⁶ Adam Buick, *Copyright and AI Training Data—Transparency to the Rescue?*, 20 J. INTELL. PROP. L. & PRAC. 182 (2025).

³⁵⁷ See generally Miriam Vogel et al., *Is Your Use of Ai Violating The Law? An Overview of the Current Legal Landscape*, 26 N.Y.U. J. LEGIS. & PUB. POL’Y 1029, 1103–1104 (2023).

B. *Explanations about Modification of AI-Generated Content*

My second proposal would require authors who modify AI-generated works to provide detailed explanations of their contributions when applying for copyright registration. This requirement is both technologically and legally feasible, especially if AI outputs are watermarked from the outset. Enhancing transparency in this way is essential for intellectual property law: it clarifies the extent of human involvement, distinguishes human input from AI-generated content, and ensures accountability as AI tools become more deeply integrated into creative processes.

1. Problems with Compilations

The Copyright Office recognizes compilations incorporating AI-generated content as copyrightable works, as detailed in its 2023 policy statements.³⁵⁸ While purely AI-generated works are excluded from copyright protection due to the absence of human authorship, a collective work that compiles AI-generated works in a certain way may qualify for protection if it features human authorship.³⁵⁹ Such authorship can arise, for example, when a human creatively selects, arranges, or modifies works initially produced by AI technology.³⁶⁰

In a recent registration decision, the Copyright Office implemented this policy by registering as a compilation a novel by Elisa Shupe created with AI assistance.³⁶¹ Shupe, a retired US Army veteran, authored *AI Machinations: Tangled Webs and Typed Words* solely using text generated by ChatGPT. Shupe's initial copyright registration application, submitted in October 2023, was denied by the Copyright Office.³⁶² However, after appealing, Shupe received copyright registration for the "selection, coordination, and arrangement" of the AI-generated text in her book, retroactively effective from the original application date.³⁶³ Note that the Copyright Office did not grant protection to the actual AI-generated text.

³⁵⁸ U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence* 88 Fed. Reg. 16, 190 (March 15, 2023), https://www.copyright.gov/ai/ai_policy_guidance.pdf.

³⁵⁹ U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence* 88 Fed. Reg. 16, 190 (March 15, 2023), https://www.copyright.gov/ai/ai_policy_guidance.pdf.

³⁶⁰ U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence* 88 Fed. Reg. 16, 190 (March 15, 2023), https://www.copyright.gov/ai/ai_policy_guidance.pdf.

³⁶¹ Kate Knibbs, *How One Author Pushed the Limits of AI Copyright*, WIRED (Apr. 17, 2024), <https://www.wired.com/story/the-us-copyright-office-loosens-up-a-little-on-ai/>.

³⁶² *Id.*

³⁶³ *Id.*

Instead, Shupe’s copyright covers only her compilation and organization of the AI-generated content.³⁶⁴

Following the same line of reasoning, the Copyright Office has recognized copyright protection in Kashtanova’s *Zarya of the Dawn* for the non-AI generated elements, and Keirse’s *A Single Piece of American Cheese*, which explicitly acknowledged the use of artificial intelligence in the application to the Copyright Office.³⁶⁵ While *Zarya of the Dawn* consisted of an arrangement of AI images in the form of a comic book, *A Single Piece of American Cheese* was created using in-painting tools so that each individual AI-generated element was combined seamlessly together to form a single image.³⁶⁶ In the Copyright Office’s ruling for *A Single Piece of American Cheese*, they agreed with the author that the creative process behind the work was similar to that of a collage artist.³⁶⁷ These cases demonstrate the different and unique ways in which AI-work can be selected and arranged, and the level of complexity that can be achieved in the creative process when working with AI tools.

Under the Copyright Act, a compilation is eligible for copyright protection as a work derived from an author’s process of selecting, coordinating, and organizing pre-existing materials.³⁶⁸ Compilations can be: (1) fact compilations, or (2) collective works. A fact compilation involves the collection of publicly available data, such as names, addresses, and phone numbers, and organizing it in a manner that demonstrates minimal creative effort. Examples of fact compilations include databases, directories, and lists.³⁶⁹ A collective work, meanwhile, embodies a compilation that assembles a number of separate and independent contributions, each a work in its own right, into a cohesive whole, such as a magazine, anthology, or encyclopedia.³⁷⁰

In my view, the Copyright Office’s approach to protecting AI-assisted works as compilations is misguided, as such works may not be suitable for copyright protection for three reasons. First, in terms of the means of creating works, human authors who create works with the assistance of AI are likely to engage in more than just the act of compiling content. According to the Copyright Act, compiling refers to an author’s selection, coordination, or arrangement of factual data or pre-existing works, which

³⁶⁴ This “thin copyright” prohibits others from reproducing the work in its entirety but does not restrict the reuse or rearrangement of the AI-generated text itself.

³⁶⁵ Invoke, *How We Received the First Copyright for a Single Image Created Entirely with AI-Generated Materials*, (2025) <https://44037860.fs1.hubspotusercontent-na1.net/hubfs/44037860/Invoke-First-Copyright-Image-AI-Generated-Material-Report.pdf>.

³⁶⁶ *Id.*

³⁶⁷ *Id.*

³⁶⁸ 17 U.S.C. § 101.

³⁶⁹ *Lists, Directories, and Databases Under Copyright Law*, JUSTIA, <https://www.justia.com/intellectual-property/copyright/lists-directories-and-databases/>.

³⁷⁰ 17 U.S.C. § 101.

ultimately assembles them into a new, coherent work as a whole.³⁷¹ In essence, the act of compiling involves two steps. Initially, a compiler selects data or works from available sources, which involves exercising judgment in choosing the type of data or works and the specific sources from which this information is obtained.³⁷² Based upon the selected data and works, the compiler coordinates or arranges them in a manner that ensures the resulting work is original.³⁷³ This second process involves ordering or grouping the selected data and works in a manner that goes beyond traditional methods, such as “the alphabetical, chronological, or sequential listings of data.”³⁷⁴

In fact, human authors can develop new, original works based on AI-generated output through means more dynamic and creative than the act of compiling.³⁷⁵ Some of those means include:

- **Editing:** By revising and refining AI-generated content, authors can enhance the work’s coherence, readability, or artistic quality, often adding their own creative touch.
- **Adaptation:** Transforming AI-generated content into different formats or mediums, such as turning a text into a screenplay or converting a digital artwork into a sculpture, demonstrates the human author’s creativity in reimagining the original material.
- **Expansion:** Authors may enrich AI-generated works by adding creative elements like illustrations, unique design features, or multimedia components, increasing the work’s overall appeal and distinctiveness.

³⁷¹ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 350–51 (ruling that (“[a] factual compilation is eligible for copyright if it features an original selection or arrangement of facts, but the copyright is limited to the particular selection or arrangement. In no event may copyright extend to the facts themselves.”)).

³⁷² *See, e.g., Key Publ’ns, Inc. v. Chinatown Today Pub. Enter., Inc.*, 945 F.2d 509, 513 (2d Cir. 1991) (“Selection implies the exercise of judgment in choosing which facts from a given body of data to include in a compilation.”); *Want Ad Dig., Inc. v. Display Advert., Inc.*, 653 F. Supp. 2d 171, 179 (N.D.N.Y. 2009) (“Want Ad does not satisfy this definition, as it does not exercise any judgment in deciding which classifieds to include in the *Want Ad Digest*.”); William F. Patry, 2 PATRY ON COPYRIGHT § 3:66 (2024).

³⁷³ *Key Publ’ns*, 945 F.2d at 512.

³⁷⁴ 2 PATRY, *supra* note 372, § 3:66. (“‘Arrangement’ or ‘coordination’ refers to the ordering or grouping of data into lists or categories that go beyond the mere mechanical grouping of data as such, for example, the alphabetical, chronological, or sequential listings of data.”).

³⁷⁵ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (proposed Mar. 16, 2023) (to be codified at 37 C.F.R. pt. 202).

- **Remixing:** By altering, rearranging, or combining parts of AI-generated content with other materials, authors create new works that reflect their individual creative vision. For example, an artist may blend AI-generated music with original compositions, or a filmmaker could use AI-generated scripts as the foundation for a unique and original screenplay.

These diverse approaches to incorporating human creativity into AI-generated content are much more complex and dynamic than mere compilation. As such, original, expressive works that involve many methods of creation may warrant the same copyright protection that is afforded to types of works other than compilation.

Second, the requirement that a copyrightable compilation should contain a large amount of data or number of works may render this protection unsuitable for works created by human authors with the engagement of AI technologies. As a rule of thumb, the more extensive the pool of material available for selection, coordination, or arrangement, the greater the likelihood that a compilation will qualify for copyright protection. Conversely, when there is limited material to choose from, the chances of a compilation being protected are reduced. This is because minimal efforts, or *de minimis* efforts, do not meet the criteria for copyright protection.³⁷⁶ Following the legislative intent of the Copyright Act,³⁷⁷ the Copyright Office “generally will not register a compilation containing only two or three elements, because the selection is necessarily *de minimis*.”³⁷⁸

This approach may not adequately address the needs of human authors who create works in collaboration with AI technologies, as these works may not always involve extensive data or a large number

³⁷⁶ See, e.g., *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, 499 U.S. 340, 362, 111 S. Ct. 1282, 113 L. Ed. 2d 358 (1991) (ruling that “the selection and arrangement of facts cannot be so mechanical or routine as to require no creativity whatsoever.”); *Silverstein v. Penguin Putnam, Inc.*, 368 F.3d 77 (2d Cir. 2004); *Mid America Title Co. v. Kirk*, 59 F.3d 719, 722 (7th Cir. 1995) (selection used failed to meet Feist’s standards because “[selecting which facts to include . . . was not a matter of discretion based on . . . personal judgment or taste, but instead it was a matter of convention and strict industry’ standards”); *Sem-Torq, Inc. v. K Mart Corp.*, 936 F.2d 851 (6th Cir. 1991) (five separate double-sided signs containing phrases such as “For Rent/For Sale” not sufficiently original); *Financial Information, Inc. v. Moody’s Investors Service, Inc.*, 751 F.2d 501 (2d Cir. 1984) (daily bond cards; no room for selectivity, choice, or judgment in reading of financial publications, cutting out bond redemption notices, extracting names of bond issuers, description of the bond issue, redemption price, date, agent, and serial number of the bond being called).

³⁷⁷ H.R. Rep. No. 1476, 94th Cong., 2d Sess. 47, at 122 (1976) (observing that a work “where relatively few separate elements have been brought together,” such as “a composition consisting of words and music, a work published with illustrations or front matter, or three one-act plays,” would not be considered a collective work).

³⁷⁸ Compendium of U.S. Copyright Office Practices, § 312.2.

of pre-existing works. Many AI-assisted creations can involve a human author's substantial creative input, even if the number of elements included in the work is limited. In such cases, the focus should be on the human author's creative contribution rather than the mere quantity of data or works involved.

Further, the compilation category of copyright law may not provide sufficient protection for works that involve significant human creativity but do not include a large number of elements. For instance, a human author might use Photoshop to add original elements to an AI-generated image. The Copyright Office has already acknowledged that this single image could qualify as a copyrightable work. However, if we strictly apply the compilation criteria, this image would not receive copyright protection due to the limited number of elements involved.

Third, the authorial intent of new works that humans create by drawing upon AI-generated content differs from that of compilations. For the former, human authors typically aim to integrate creative elements from their own minds with those generated by AI systems, resulting in a unified work that embodies both types of elements.³⁷⁹ This single, unified work could be a novel, painting, song, or video that has been edited, refined, combined, or remixed with AI-generated content. In contrast, the intent behind a compilation is to present an unconventional method for people to access various individual pieces or sets of data or works through the compilation itself. The compiler intends to allow others to use these different pieces or sets separately.³⁸⁰

The Copyright Office maintains that a unified work, such as a novel with multiple chapters, does not qualify as a compilation due to the intentional cohesion behind its creation.³⁸¹ Yet, its decisions to register AI-assisted works, such as *AI Machinations: Tangled Webs and Typed Words*, as compilations of AI-generated content contradict this stance, as the author deliberately crafted a unified work by editing and integrating AI-generated works.³⁸² Moreover, these rulings overlook the

³⁷⁹ Katherine O'Toole & Emőke-Ágnes Horvát, *Extending Human Creativity with AI*, 34 J. of Creativity (Aug. 2024), <https://www.sciencedirect.com/science/article/pii/S2713374524000062>.

³⁸⁰ Put differently, the protectible authorship in such a collective work stems from the original selection, coordination, and arrangement of the independent works included in the collective work. *See, e.g., American Dental Ass'n v. Delta Dental Plans Ass'n*, 126 F.3d 977, 979 (7th Cir. 1997) ("Classification is a creative endeavor. Butterflies may be grouped by their color, or the shape of their wings, or their feeding or breeding habits, or habitats, or the attributes of their caterpillars, or the sequence of their DNA; each scheme of classification could be expressed in multiple ways.").

³⁸¹ *Collective Works*, COPYRIGHT.GOV, <https://copyright.gov/eco/help-collective-work.html> (last visited Nov. 9, 2025).

³⁸² Lauren Cury & Anna Shaw, *Windows Into the Future: New Horizons & Implications for the Copyrightability of AI-Generated Works*, HOGAN LOVELLS (May 9, 2024), <https://www.engage.hoganlovells.com/knowledgeservices/news/windows-into-the-future-new-horizons-implications-for-the-copyrightability-of-ai-generated-works> ("Schupe also noted that ChatGPT

fact that authors can produce original works by creatively transforming AI-generated content in ways that go well beyond simple compilation.

2. Explanations about Human Creativity

Requiring authors to properly explain how they have modified AI-generated content would protect not only compilations, but also original works created through the various methods discussed above. The U.S. Copyright Office could enforce this by incorporating a mandatory field into copyright application forms, where applicants describe their specific contributions to AI-assisted outputs. For example, authors might submit a concise statement (100–200 words) outlining the scope of their alterations, thereby giving examiners a clear view of the human role and reducing ambiguity. These disclosures would empower examiners to evaluate whether the human input satisfies the “modicum of creativity”³⁸³ and human authorship thresholds essential for copyright eligibility,³⁸⁴ rather than relying on vague claims of authorship. Ultimately, such a mandate would promote transparency, expedite the review process, and discourage assertions over minimally altered AI-generated material.³⁸⁵ Below are sample explanations that the Copyright Office could use to guide authors in fulfilling this requirement.

Example 1: A writer uses a language model like GPT to generate a draft chapter and seeks copyright for the final novel.

Sample Explanation:

“I used ChatGPT 4.0 to generate a rough draft of a sci-fi novel chapter by providing the prompt: ‘A 500-word scene set on a Martian colony where a detective uncovers a conspiracy.’ I generated three drafts and selected one for its narrative structure. I then revised the draft extensively, rewriting 70% of the text to enhance character development, add dialogue reflecting my unique voice, and incorporate a subplot about environmental collapse that I created independently. I also structured the chapter within a 12-chapter novel, where all other chapters were written entirely by me. My revisions and original contributions dominate the final work.”

did not do all of the work. Her appeal described the long hours she spent on her draft, her tailored prompting and the edits she made to the bot’s responses. According to Shupe, nearly every sentence was modified by her.”).

³⁸³ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 346 (1991) (“[T]he requisite level of creativity is extremely low; even a slight amount will suffice. The vast majority of works make the grade quite easily, as they possess some creative spark, ‘no matter how crude, humble or obvious’ it might be.”).

³⁸⁴ Copyright Office AI Guidance, 88 Fed. Reg. 16190.

³⁸⁵ *Id.*

This explanation highlights the writer's iterative prompt design, selection, and substantial revisions, showing that the AI output was a starting point rather than the final product. It allows the Copyright Office to protect the human-authored text and structure while potentially excluding the raw AI draft, ensuring transparency about the creative process.

Example 2: An artist uses Midjourney to generate a base image of a futuristic cityscape and seeks copyright protection for the final artwork.

Sample Explanation:

"I used Midjourney to generate a base image by crafting a detailed prompt: 'A neon-lit futuristic city at dusk with towering skyscrapers, flying cars, and a cyberpunk aesthetic, in the style of Blade Runner.' I iterated through 15 outputs, selecting one that best matched my vision. I then imported the image into Adobe Photoshop, where I spent 12 hours adding original elements, including a central character in a unique outfit I designed, hand-painted graffiti on buildings, and adjusted lighting to emphasize a moody atmosphere. These additions and edits reflect my creative vision and artistic skill, transforming the AI output into a distinct work."

This explanation elucidates the artist's role in prompt design, curation, and substantial post-processing.³⁸⁶ It establishes creative control that extends beyond simply accepting an AI output, thereby justifying copyright eligibility for the human-authored components (e.g., character design, graffiti, and lighting adjustments). The Copyright Office can deny protection to the raw AI-generated image, as seen in the *Zarya of the Dawn* case (2023), while affording copyright to works of human authorship.³⁸⁷

These examples illustrate that, to qualify for protection, a human author must introduce sufficient modifications to the AI-generated content.³⁸⁸ Minor tweaks or the addition of inconsequential elements will not suffice to render the content copyrightable.³⁸⁹ Aligning with

³⁸⁶ Mackenzie Caldwell, *What Is an "Author"?—Copyright Authorship of AI Art Through a Philosophical Lens*, 61 Hous. L. Rev. 367, 406 (2023).

³⁸⁷ See *supra* Part I.B.

³⁸⁸ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. at 16192–93 ("In other cases, however, a work containing AI-generated material will also contain sufficient human authorship to support a claim to copyright. For example, a human may select or arrange AI-generated material in a sufficiently creative way that 'the resulting work as a whole constitutes an original work of authorship.' Or an artist may modify material originally generated by AI to such a degree that the modifications meet the standard for copyright protection.").

³⁸⁹ See Jane C. Ginsburg, *The Concept of Authorship in Comparative Copyright Law*, 52 DEPAUL L. REV. 1063, 1078–80 (2003) (noting that originality is synonymous with copyright

the proposed watermarking mandate, every AI-generated work would bear an embedded watermark which would facilitate identification of human-made alterations. The following are examples of three degrees of modification, each with varying copyrightability outcomes for the works concerned:

- *De minimis* modification: An author uses AI-generated text to create a story and only makes minor grammatical corrections or adjustments to punctuation. In this case, the author's contribution would be considered minimal, and the resulting work would likely not be eligible for copyright protection.
- Sufficient amount of modification: An author significantly restructures a passage of AI-generated text, adds new characters, and alters the narrative to create a cohesive and unique story. The author's contribution in this case goes beyond mere editing and demonstrates a sufficient level of originality, making the work potentially eligible for copyright protection.
- Substantial amount of modification: An author uses AI-generated text as a starting point but extensively rewrites the content, develops a completely new plot, and introduces original themes and characters. The resulting work is predominantly the author's own creation, with only minimal reliance on the AI-generated content. In this case, the work may qualify for copyright protection under Section 102(a), as the author's contribution is substantial and dominant.

Detail before Photoshop



Detail after Photoshop



Image Example 1: A work containing only a *de minimis* change (not copyrightable).³⁹⁰

In this image of Zarya from *Zarya of the Dawn*, the version on the left was generated by Midjourney following Kashtanova’s prompt. For inclusion in her book, Kashtanova simply edited the distortion on the lips using Adobe Photoshop. This minor alteration, addressing only the blemish on the lips, should be considered a *de minimis* change.



Image Example 2: A work containing a sufficient amount of changes (to be protected as a work with AI creativity).³⁹¹

Here, we will assume that the main image on the left is an AI-generated image. As to the four images on the right, the author has altered the background by adding new color schemes, bird features, and dynamic

³⁹⁰ Letter from Robert Kasunic, Assoc. Reg. of Copyrights and Dir. of Registration Pol’y and Prac., U.S. Copyright Off., to Van Lindberg, Taylor English Duma LLP 12 (Nov. 21, 2022), <https://www.copyright.gov/doc/zarya-of-the-dawn.pdf><https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>, p. 12.12.12.12.12.

³⁹¹ *Background Diffusion*, CUTOUT.PRO, <https://www.cutout.pro/photo-editing-background>.

visual effects. These alterations represent sufficient changes which would make the four images eligible for protection as works with AI creativity.

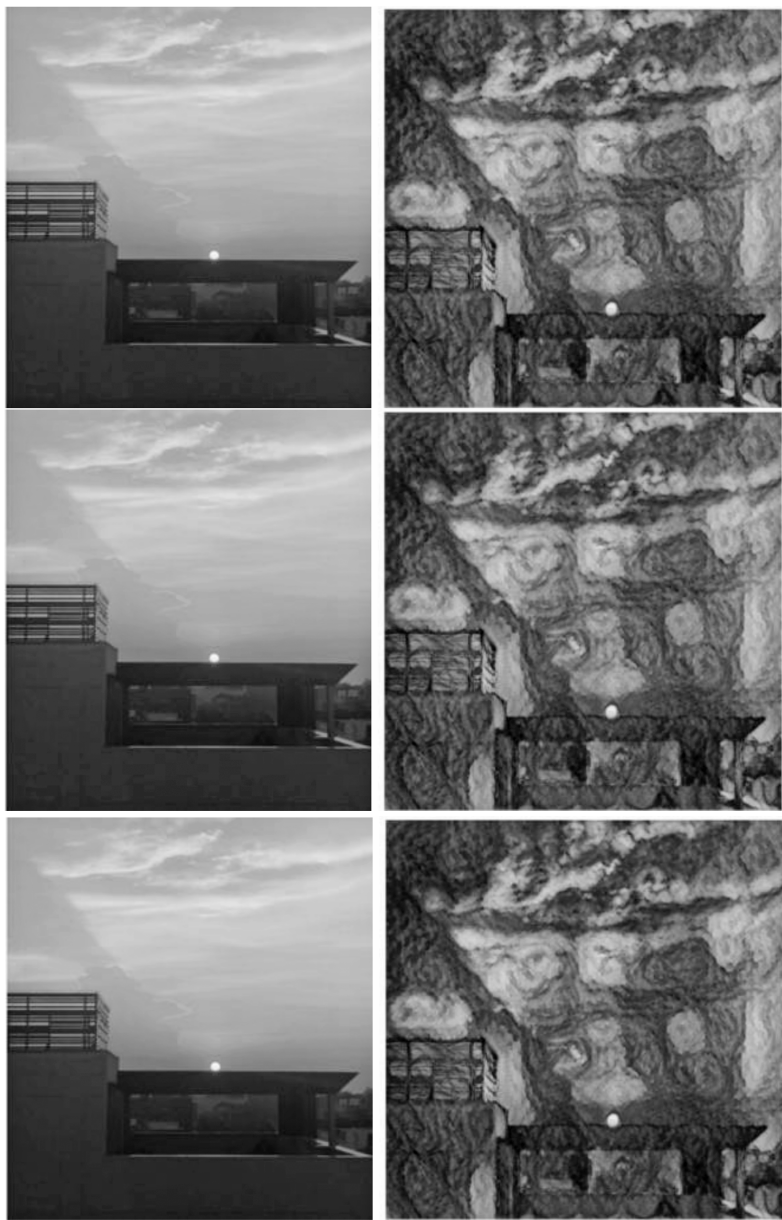


Image Example 3: A work containing substantial changes (to be protected as a Section 102(a)-protected human work).³⁹²

³⁹² Letter from U.S. Copyright Off. Rev. Board to Alex Garens, Day Pitney LLP 6 (Dec. 11, 2023), <https://www.copyright.gov/rulings-filings/review-board/docs/SURYAST.pdf>

In this scenario, let us assume that the image on the left is AI-generated. The artist has then extensively edited the image in Photoshop and painted over it to produce a new image that has a similar composition but an entirely different style and color palette. These substantial modifications transform the images into two distinct graphic works. Under Section 102(a), these images, due to their substantial alterations, are eligible for copyright protection.³⁹³

It should be clarified that in all three use cases, the components of the work that are generated by the AI will not be protected by copyright, and thus even though the image as a whole is copyrightable, the artist will have no infringement claim against those who copy only the AI-generated parts.³⁹⁴ In many cases, such as in the case of *A Single Piece of American Cheese* where inpainting was used extensively, the value of the copyright in the piece will greatly outweigh the value of the AI-generated parts individually.³⁹⁵ Consequently, the exclusion of AI-generated content from copyright protection will likely not undermine protection over the core of the artist's work.

By mandating a proper description of actions such as crafting prompts, selecting or curating outputs, or making substantive edits, the Copyright Office gains insight into the human's creative influence. For example, an applicant might explain: "I provided a detailed prompt describing a surrealist painting, iterated through 20 AI-generated outputs, selected one, and manually adjusted colors and added original figures." This clarifies that the final work reflects human decision-making, not just AI generation. Without such an explanation, works can only be defined in binary terms, either with or without AI generation, obscuring the human's role.³⁹⁶

Hence, the explanation requirement would help clarify the role of human creativity in altering AI-generated Works. AI systems, such as DALL-E, Midjourney, or GPT models, can produce sophisticated outputs that appear original, but their generative processes are often opaque, relying on complex algorithms and training data that are not fully disclosed.³⁹⁷ This opacity makes it difficult to discern whether a work is primarily the

³⁹³ *Id.* at 6–7.

³⁹⁴ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. at 16192–93; *see also* Ginsburg, *supra* note 389, at 1074 (noting the proportional relationship between increased machine use in the creation of a work and author's higher burden of proving human determination of work's form and content).

³⁹⁵ Alec Winshel, *U.S. Copyright Office Grants Registration to AI-Generated Artwork*, JSEL (Mar. 24, 2025), <https://journals.law.harvard.edu/jsel/2025/03/u-s-copyright-office-grants-registration-to-ai-generated-artwork/>.

³⁹⁶ *See* Lin Kyi et al., Governance of Generative AI in Creative Work: Consent, Credit, Compensation, and Beyond, in PROCEEDINGS OF THE 2025 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYS., Art. 197, at 9.

³⁹⁷ Pasquale & Sun, *supra* note 50, at 209 ("To create and improve their AI models, large technology firms have undermined authors' proprietary control over their works by using these works as training data, without consent and often through opaque processes.").

product of human creativity or AI automation. Requiring an explanation of the human author's contributions forces applicants to articulate their specific role in shaping the AI-generated output, thereby shedding light on the creative process.

By substantially promoting output transparency by illuminating human creativity, the explanation requirement delineates authors' contributions from AI inputs, enhancing regulators' institutional capacity to preserve the integrity of copyright system. As proposed earlier, the Copyright Office would mandate that authors specify the scope of their modifications to AI-generated content, thereby establishing a clear demarcation between watermarked AI outputs and human-introduced creative elements. Designed and implemented properly, this requirement would ensure that the Copyright Office have a clearer understanding of how authors create new works by adapting from AI-generated content, thereby helping Office develop a more rigorous practice of examining creativity and ensure that works registered for copyright protection are sufficiently original. Hence, the output transparency that the requirement promotes aligns with the goals of copyright protection while embracing the transformative potential of AI in creative industries.

This requirement makes the creative process more visible, ensuring that the Copyright Office and the public (if the application is publicly accessible) understand how the work was produced. It demystifies the "black box" of AI generation by documenting the human's active participation, such as designing prompts or post-processing outputs, which might otherwise be assumed to be purely machine-driven.³⁹⁸

By standardizing the disclosure of human contributions, the explanation requirement also ensures that the creative process is documented in a way that is accessible to examiners as well as the public, given that copyright registrations are public records. This transparency will help demystify works adapted from AI-generated contents, making it clear how they were produced and why they qualify for protection. It also sets a precedent for consistent evaluation, reducing the risk of granting copyrights to works with minimal human input, or denying copyright for sufficiently creative works.

Beyond bolstering the Copyright Office's regulatory framework, the explanation requirement would enhance public awareness of the creative process, particularly how human authors engage with AI outputs. This mandate exposes the workflow, facilitating traceability of a work's origins. In cases where an AI-generated output is challenged for resembling a copyrighted piece, the applicant's explanation can serve as evidence of

³⁹⁸ Mark A. Lemley & Bryan Casey, Fair Learning, 99 Tex. L. Rev. 743, 767 (2021); Katherine Lee, A. Feder Cooper & James Grimmelman, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, 71 J. COPYRIGHT SOC'Y U.S.A. 159, 159–202 (2024).

original contributions, thereby heightening accountability. For instance, if a dispute arises over a song's similarity to an existing track, the applicant's description of their prompt design and post-processing can demonstrate their independent creative effort, fostering trust in the copyright system.³⁹⁹

More broadly, the explanation requirement would propel cultural and social progress in multiple interconnected ways.⁴⁰⁰ Authors' disclosures can unveil the interdependent essence of creativity in the AI era, where humans leverage existing cultural artifacts (others' works) and algorithmic creations (AI outputs). By promoting transparency in creation, this practice deepens comprehension of how new works derive from predecessors, whether human or machine-made. Greater transparency would also enrich the public documentation of creativity. Disclosures of modifications to AI content contribute vital data to the cultural archive, rendering processes visible and assisting audiences, scholars, and future creators in better tracing evolutions of ideas, styles, and techniques.

C. Addressing Problems

1. Potential Problems with AI Watermarking

The proposal to implement AI watermarking is technologically feasible in practice, though it remains an imperfect solution. Critics may contend that watermarking should be avoided due to the relative ease with which watermarks can be removed or altered. This objection, however, lacks persuasiveness; the existence of malicious actors does not provide a compelling reason to forgo preventive measures altogether.

Watermarks have been utilized for decades in images and videos, and despite their vulnerability to removal, they have proven effective in accrediting creators and curbing casual copyright infringement among typical users.⁴⁰¹ For example, watermarking serves as a form of copyright management information, enabling identification of copyrighted works and their creators. Under the Digital Millennium Copyright Act (DMCA),⁴⁰² unauthorized removal or tampering with watermarks is prohibited, safeguarding the integrity of ownership claims.⁴⁰³ Malicious actors intent on stripping watermarks should be viewed as exceptions, as most users lack either the technical expertise or the motivation to undertake such efforts—even if the process is relatively straightforward. Consequently,

³⁹⁹ Julianna Simpson, *Auctor Ex Machina: Artificial Intelligence and Authorship in Zarya of the Dawn*, 20 J. BUS. & TECH. L. 1, 26 (2025).

⁴⁰⁰ Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1177 (2007).

⁴⁰¹ See JULIE E. COHEN, *CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE* 161 (2012) (exploring watermarking as a socio-technical copyright strategy for digital content like images and videos).

⁴⁰² Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998).

⁴⁰³ *Id.* at 17 U.S.C. § 1202(b).

watermarking can still yield significant benefits in the majority of cases, even if it falls short of being a perfect fix.⁴⁰⁴

Lawmakers should complement this approach by enacting penalties for those who intentionally remove watermarks or supply tools to facilitate such actions, further deterring malicious interference and reinforcing the efficacy of the watermarking solution. Implementing laws and regulations that protect watermarked content and penalize unauthorized removal, or tampering would discourage such activities and hold violators accountable. Legislators can learn from the DMCA, which penalizes the circumvention of technological measures (including digital watermarks) that control access to copyrighted works. The DMCA sets forth both civil remedies and criminal penalties for those who breach these protective technological measures.⁴⁰⁵

Furthermore, when weighed against privacy rights such as the right to anonymity or the right to pseudo-anonymity, greater transparency through watermarking should be prioritized. These privacy interests do not justify concealing AI-generated content as the concealment of AI is more prone to facilitate malicious acts than legitimate ones. It is, however, important to consider what kind of information will be recorded for the sake of watermarking. For example, metadata can reveal a surprising amount of information about a file's author. Currently, most social media platforms automatically remove metadata from uploaded files to protect people's location information and to save storage.⁴⁰⁶ Likewise, retrieval-based methods will require user consent to store all model outputs indefinitely for detection purposes.⁴⁰⁷ Privacy considerations will not prevent AI-watermarking, but we should be wary of what kinds of information we reveal when implementing watermarks and whether they are entirely necessary.⁴⁰⁸

The effect of watermarking obligation as a hindrance to technological advancement should also be considered. The implementation costs will raise the barrier for entry into the AI sector and make it difficult for

⁴⁰⁴ Nihal Krishan, *AI Watermarking Could Be Exploited by Bad Actors to Spread Misinformation. But Experts Say the Tech Still Must Be Adopted Quickly*, FedScoop, (Jan. 3, 2024) <https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/>.

⁴⁰⁵ 17 U.S.C. §§ 1203–1204.

⁴⁰⁶ Jacob Hoffman-Andres, *AI Watermarking Won't Curb Disinformation*, EFF, (Jan. 5, 2024), <https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation>.

⁴⁰⁷ Shayne Longpre et al., *Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?*, MIT Generative AI PubPub (Apr. 24, 2024), <https://mit-genai.pubpub.org/pub/uk7op8zs/release/2>; Siddarth Srinivasan, *Detecting AI Fingerprints: A Guide To Watermarking And Beyond*, Brookings (Jan. 4, 2024), <https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond>.

⁴⁰⁸ *Watermarking Standards and Policy for AI-Generated Media*, CyberPeace Inst. (May 10, 2024), <https://www.cyberpeace.org/resources/blogs/watermarking-standards-and-policy-for-ai-generated-media>.

smaller start-ups to operate.⁴⁰⁹ Additionally, requiring companies to adopt watermarking methods that trade off output accuracy or quality will stifle progress in AI.⁴¹⁰ Making the detection tool publicly available may also risk revealing proprietary information about the model, which can undermine an AI developer's competitive advantage.⁴¹¹ These problems do not provide persuasive reasons to abandon watermarking altogether, but they should be taken seriously to ensure that regulations do not impose excessive burdens that are counter-productive. Treating AI companies fairly will be important in maintaining a good relationship that encourages co-operation between regulators and AI developers.

The challenges outlined above have two potential solutions. The first would require all generative AI service providers to implement mandatory watermarking, embedding visible or invisible identifiers in all AI-generated content. The second allows for an opt-in model, permitting AI companies below a certain scale, for example, those with fewer than one million users, to decide whether to apply watermarks to synthetic content. This opt-in approach offers a significant advantage: it promotes output transparency as a regulatory goal while providing flexibility to explore alternative methods for authenticating AI-generated content. Additionally, it could alleviate financial pressures on small and medium-sized AI companies by sparing them the high costs of developing proprietary watermarking technologies.

2. Potential Problems with AI Explanations

Requiring authors to provide proper explanations of their contributions to AI-related works, while beneficial for transparency and compliance with U.S. copyright law, could introduce several challenges. These problems span administrative, legal, practical, and ethical dimensions, potentially complicating the application process for creators, examiners, and the U.S. Copyright Office.

First, mandating a contribution explanation adds a layer of complexity to the copyright application process, increasing the workload for both applicants and Copyright Office examiners. Creators, especially those unfamiliar with AI tools or copyright law, may find it challenging to articulate their contributions in a clear, concise manner.⁴¹² Small-scale creators, such as independent artists or writers, might lack the resources

⁴⁰⁹ Siddarth Srinivasan, *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*, Brookings (Jan. 4, 2024), <https://www.brookings.edu/articles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/>.

⁴¹⁰ Haiqa Yousaf, *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*, MEDIUM (Jan. 9, 2024), <https://medium.com/@haiqajabeen888/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond-b03e7142ef7c>

⁴¹¹ Srinivasan, *supra* note 409.

⁴¹² *Developments in the Law—Artificial Intelligence*, 138 HARV. L. REV. 1554 (2025), <https://harvardlawreview.org/print/vol-138/developments-in-the-law-artificial-intelligence/>.

or expertise to describe technical processes like prompt design or post-processing, potentially leading to incomplete or poorly drafted explanations. Additionally, the creative process behind a work can be quite complicated and long. Having to continuously keep track of one's interaction and iterations for the sake of keeping records for a proper explanation can hinder the author's creativity.⁴¹³

To alleviate this, the Copyright Office could furnish templates or exemplars on its website tailored to prevalent AI scenarios (e.g., visual art, music, literature), enabling applicants to delineate contributions efficiently. A visual arts template, for instance, might prompt: "I utilized [AI tool] with prompt [X], selected output [Y], and incorporated [specific edits]." Furthermore, AI-driven tools could flag inadequate or ambiguous explanations, curtailing manual oversight; natural language processing might verify coverage of core elements like prompt design, curation, and post-processing.⁴¹⁴

Moreover, the explanation requirement relies on applicants to provide accurate and truthful explanations, but applicants may exaggerate or misrepresent contributions, and verifying claims is challenging. For example, an applicant might claim significant creative input (e.g., "I extensively modified the AI-generated image") when they only made minimal changes, such as adjusting brightness.⁴¹⁵ Verifying these claims is challenging, as the Copyright Office typically does not investigate the veracity of applications beyond reviewing submitted materials.⁴¹⁶ But the availability of a paper trail of documentation behind the creative process will still be of great importance.⁴¹⁷

Unlike plagiarism or infringement, which can sometimes be detected through comparison, assessing the accuracy of contribution explanations at the registration level requires technical expertise or access to original AI outputs, which may not be feasible. For example, an applicant using DALL-E might not retain all generated outputs to prove their curation

⁴¹³ Katherine B. Forrest, *The Ethics and Challenges of Legal Personhood for AI*, 133 YALE L.J.F. 1175, 1192 (2024), <https://www.yalelawjournal.org/forum/the-ethics-and-challenges-of-legal-personhood-for-ai>.

⁴¹⁴ Ayesha Zafar, *Balancing the Scale: Navigating Ethical and Practical Challenges of Artificial Intelligence (AI) Integration in Legal Practices*, 4 DISCOV. ARTIF. INTELL. 27, 36 (2024), <https://doi.org/10.1007/s44163-024-00121-8>.

⁴¹⁵ Christopher T. Zirpoli, *Generative Artificial Intelligence and Copyright Law*, Crs Report No. LSB10922 (Congressional Research Service, 2025), <https://www.congress.gov/crs-product/LSB10922>.

⁴¹⁶ U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices* § 309.2 (3d ed.2021), <https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf> ("Ordinarily, the Office will not conduct its own factual investigation to confirm the truth of the statements made in the application. However, the Office may take administrative notice of facts or matters that are known by the Office or the general public, and may communicate with the applicant if the application appears to contain inaccurate or erroneous information.").

⁴¹⁷ *Id.*

process. Dishonest or vague explanations could lead to improper grants of copyright for works with minimal human input, undermining the integrity of the copyright system and potentially flooding it with low-effort AI-generated registrations.

When such issues reach the court, however, applicants' lack of documentation will make it difficult for them to substantiate their claim, especially if the defendant challenges the validity of their copyright.⁴¹⁸ Furthermore, any dishonesty can be challenged during cross-examination and experts can be called to assist the court with technical matters. Nonetheless, registration establishes *prima facie* evidence of valid copyright in the facts stated in the certificate where the work is registered within 5 years of publication.⁴¹⁹ Furthermore, the high costs of litigation will mean that in many cases, it will be difficult to challenge claims from registered copyright owners.

One solution could be implementing a stringent declaration of accuracy mechanism for copyright registration. Applicants would be required to sign a declaration affirming the truthfulness of their explanation, with penalties for willful misrepresentation.⁴²⁰ Including such a declaration in the application may not dissuade malicious actors, but it can at least prompt ordinary applicants to consciously think about and verify the accuracy of their explanations.

Another solution could be to create a spot-checking mechanism. This would implement random audits of AI-altered work applications, requiring applicants to submit supporting materials (e.g., original AI outputs or editing files) in select cases.⁴²¹ This deters exaggeration without burdening all applicants. In practice, however, this will still mean that most applicants will still have to retain supporting evidence for the creative process behind a work to satisfy the audit requirements if they are called upon. A reliable method for recording and storing this information will be necessary to streamline the process and avoid requiring AI users to record their activity manually.

⁴¹⁸ *Statement of Policy on Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 88 Fed. Reg. 16190 (Mar. 16, 2023) (“a court may disregard a registration in an infringement action pursuant to section 411 (b) of the Copyright Act if it concludes that the applicant knowingly provided the Office with inaccurate information, and the accurate information would have resulted in the refusal of the registration”); *Unicolors, Inc. v. H&M Hennes & Mauritz, L.P.*, 142 S. Ct. 941, 948 (2022) (requiring that the applicant “was actually aware of, or willfully blind to” the inaccurate information).

⁴¹⁹ U.S Copyright Office, *Copyright Basics*, Circular 1, (2021), at 5, <https://www.copyright.gov/circs/circ01.pdf#page=7>.

⁴²⁰ A fine of up to \$2,500 is permitted under 17 U.S.C. § 506(e) for false representations in a copyright application, and registration cancellation may occur through judicial processes or Copyright Office actions if inaccuracies are identified.

⁴²¹ For the idea of AI auditing, see generally Edwin A. Farley & Christian R. Lansang, *AI Auditing: First Steps Towards the Effective Regulation of Artificial Intelligence Systems*, 38 HARV. J.L. & TECH. 1 (2025).

Additionally, professional AI service providers could be deployed to collaborate with the Copyright Office to design tools to assist users in satisfying these requirements for copyright registration. For example, service providers may produce official and watermarked records for AI users to help validate their claims and make fraud more difficult. These records can show the full editing history of a work as well as what kind of prompts or parameters were inputted by the user.⁴²² Although this may create data privacy concerns surrounding the increased monitoring and recording of user activity, it is an acceptable trade-off for greater transparency over the creative process. Furthermore, users unconcerned with copyright protection should still be able to opt out from such data collection.

CONCLUSION

The rise of AI has brought forth a complex array of “black box” challenges, particularly obscuring transparency in AI content creation and poised to similarly affect AI-generated products and robots in the near future.⁴²³ Much attention has centered on input opacity, the inscrutable nature of AI algorithms and datasets. Yet output opacity, arising from the deliberate concealment of AI’s role in content creation, warrants equal scrutiny due to the profound harms it inflicts on individuals and society alike.⁴²⁴

As AI-generated content becomes increasingly indistinguishable from human-authored works, devising mechanisms to promote output transparency—while fostering dynamic collaboration between human and AI creativity—has emerged as an urgent priority. The theory of differentiated creativity, as this Article proposes, illuminates this imperative: human creativity unfolds as a vibrant, purposeful process, ignited by intentionality, infused with emotions drawn from lived experiences, and

⁴²² Arian F. Jabbary, Alexander J. Liederman & Nikhil T. Pradhan, *Clarifying the Copyrightability of AI-Assisted Works*, Nat’l L. Rev. (Apr. 10, 2024), <https://natlawreview.com/article/clarifying-copyrightability-ai-assisted-works> (“Maintain detailed records of human prompts and modifications, such as arranging, adapting, or refining AI outputs.”); Edward Lee, *Prompting Progress: Authorship in the Age of AI*, 76 FLA. L. REV. 1445, 1559 (2024) (“Prompt engineering encourages people to experiment with generative AI and to engage in an iterative process of creation, meaning the creator edits and refines the initial work through a series of successive prompts to produce the work the creator envisions.”).

⁴²³ See FRANK PASQUALE, FRANK PASQUALE, *NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI* 7 (2021) (“Robotic systems and AI should not counterfeit humanity. . . . There is a critical difference between humanizing technology and the counterfeiting of distinctively human characteristics.”).

⁴²⁴ See Ian Ayres & Jack M. Balkin, *The Law of AI is the Law of Risky Agents Without Intentions*, U. CHI. L. REV. ONLINE 2 (2024), https://lawreview.uchicago.edu/sites/default/files/2024-11/Ayres_Balkin_Law%20of%20Risky%20Agents.pdf (“When people use AI, they should be responsible for the harms that occur when the risks of using the technology are realized.”).

steered by autonomous agency over the final outcome. In stark contrast, AI creativity proceeds mechanistically, propelled by predefined models and data inputs, bereft of genuine intent, emotional depth, or autonomous control.

Anchored in this theory of differentiated creativity, the legal standard of output transparency advanced in this Article provides a framework for demarcating AI and human contributions. Beyond guiding policy deliberations and legal reforms, this standard boasts a key practical edge: its enforcement is technically and juridically simpler than that of algorithmic or data transparency. Output transparency sidesteps the complexities of AI models and vast datasets, presenting a feasible path to effective regulation of AI-generated content.

To navigate the shifting landscape of creative production, regulators must move swiftly to mandate AI output transparency. In particular, they should legislate requirements for embedding watermarking in AI-generated outputs and obligate creators to disclose modifications that transform such content into original human-authored works. These steps, as this Article has demonstrated, will empower the legal system to protect and cultivate the creative synergies between humans and AI, heralding a more innovative and equitable society in the AI age.