Compactibility: AI and the Idea-Expression Inversion

Anirban Mukherjee
Hannah Hanwen Chang

February 21, 2026

Anirban Mukherjee (anirban@avyayamholdings.com) is Principal at Avyayam Holdings. Hannah H. Chang (hannahchang@smu.edu.sg; corresponding author) is Associate Professor of Marketing at the Lee Kong Chian School of Business, Singapore Management University. This research was supported by the Ministry of Education (MOE), Singapore, under its Academic Research Fund (AcRF) Tier 2 Grant, No. MOE-T2EP40124-0005.

Abstract

This Essay identifies a novel challenge to the idea-expression dichotomy: *compactibility*, the capacity of AI systems to extract the functional essence of a copyrighted work and re-express it in an entirely new form, tailored to a specific purpose. Unlike traditional summarization, which yields a static condensation, compaction exploits generative AI's ability to produce alternative expressions and to dynamically select among them. We argue that compactibility poses three escalating threats: the statistical fragility of similarity-based copying inferences when AI output has effectively infinite range; the ability to tune AI systems to evade substantial similarity metrics while retaining market-substituting value; and, most fundamentally, the power of compaction to separate ideas from expression with a precision that renders the original expression disposable. These capabilities invert the economic assumptions underlying copyright: expression becomes cheap and infinitely reproducible while the ideas it conveys become the extractable resource. We conclude that both the administrability of the idea-expression boundary and the normative case for drawing it where we do require reassessment.

Keywords: Idea-Expression Dichotomy, Substantial Similarity, Copyright Law, Artificial Intelligence, Generative AI.

JEL codes: K11, O34, L86.

Table of Contents

Introduction

Copyright law draws its central line between protectable expression and unprotectable ideas—a boundary observed in at least 166 countries and foundational to every major intellectual property regime.[1] In U.S. law, the distinction traces to *Baker v. Selden*, where the Supreme Court held that copyright in a book did not extend to the accounting system it described: the author's specific language was protectable, but the underlying method was not.[2] This distinction rests on an implicit assumption: that expression is a stable, identifiable artifact—the scarce resource that merits protection—distinct from the idea it conveys. For over a century, the dichotomy has served

[1] Agreement on Trade-Related Aspects of Intellectual Property Rights art. 9(2), Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299 [hereinafter TRIPS Agreement] ("Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.").

[2] 101 U.S. 99, 102 (1879) ("The copyright of a book on book-keeping cannot secure the exclusive right to make, sell, and use account-books prepared upon the plan set forth in such book.").

as the primary mechanism for adjudicating the boundary between lawful inspiration and unlawful appropriation.

This Essay argues that the idea-expression dichotomy faces a fundamental challenge from what we term *compactibility*—the capacity of AI systems to extract the functional essence of a copyrighted work and re-express it in an entirely new form, tailored to a specific purpose.[3] Unlike traditional summarization, which produces a static condensation of a document, compaction is dynamic and goal-directed: an AI agent reads a copyrighted work, identifies the elements most relevant to a user's request, discards the original expression, and generates a new expression of the extracted content.[4] A textbook can be compacted into a course module; a news article into a market briefing; a novel into a podcast script—each output purpose-built for the user, each retaining the value of the original, and each bearing no mimetic similarity to it. Compaction does not merely blur the line between idea and expression; it treats expression as a disposable interface: something to be read through, extracted from, and regenerated at will.

Compactibility builds on two properties of generative AI that independently threaten similarity-based copyright enforcement. When the output space is effectively infinite, overlap between any two works becomes a statistical inevitability rather than evidence of derivation, rendering similarity-based inference of copying unreliable.[5] And AI systems can be tuned to evade the specific metrics copyright law uses to establish infringement while retaining market-substituting value, converting a legal standard into a technical optimization target.[6] But compaction poses the

[3] We refer to the exercise of this capability on a particular work as *compaction*.

[4] The mechanism underlying compaction is what Mukherjee and Chang term "fluid agency"—the stochastic, dynamic, and adaptive capability of AI systems that enables them to modulate their means while pursuing specified ends. *See* Anirban Mukherjee & Hannah H. Chang, *Fluid Agency in AI Systems: A Case for Functional Equivalence in Copyright, Patent, and Tort*, 21 Wash. J. L. Tech. & Arts 1 (2026), https://digitalcommons.law.uw.edu/wjlta/vol21/iss1/3 [https://perma.cc/4LNK-FLSG] [hereinafter Mukherjee & Chang, Fluid Agency]. For the broader doctrinal framework, *see* Anirban Mukherjee & Hannah Hanwen Chang, *From 'Custom' to 'Code': A Doctrine of Functional Norms for Emergent Practices in Agentic Markets* (Feb. 9, 2026) (unpublished manuscript), https://ssrn.com/abstract=5391955 [https://perma.cc/BF4N-J9QX] [hereinafter Mukherjee & Chang, Functional Norms].

[5] *See* Anirban Mukherjee & Hannah Hanwen Chang, *Beyond Pairwise Comparisons: A Distributional Test of Distinctiveness for Machine-Generated Works in Intellectual Property Law* (Jan. 26, 2026) (unpublished manuscript), https://arxiv.org/abs/2601.18156 [https://perma.cc/HXH7-SM8D] (demonstrating that AI outputs are distributionally distinct from training data, complicating overlap-based infringement analysis) [hereinafter Mukherjee & Chang, Distributional Distinctiveness].

[6] *See* Anirban Mukherjee & Hannah Hanwen Chang, *Engineering Rivalry: A 'Duty to Disrupt' Algorithmic Convergence*

deeper challenge: it does not merely complicate the measurement of similarity—it eliminates the similarity altogether, extracting ideas from their expressive form and regenerating expression *de novo*.

In this Essay, we describe the mechanics of compaction and the malleability of expression that makes it possible (Part I). We then show that compaction severs the link between access and mimetic similarity on which copyright's enforcement framework depends (Part II). Finally, we argue that compaction inverts the economic assumptions underlying the idea-expression dichotomy—expression becomes cheap and infinitely reproducible while ideas become the extractable resource—and that both administering this boundary and justifying it require reassessment (Part III).

I. The Mechanics of Compaction

A. From Summarization to Compaction

AI systems interact with copyrighted works at three analytically distinct stages: (1) *training-time ingestion*, in which works are copied into datasets to build the model's parameters; (2) *inference-time access*, in which the system retrieves or loads copyrighted sources into working memory to answer a specific query; and (3) *output generation*, in which the system produces user-facing text that may range from verbatim excerpt to high-level paraphrase.[7]

Questions relating to the first stage—whether training on copyrighted works is itself infringement—are currently the focus of most AI copyright litigation.[8] This stage, however, is not the focus of this Essay as these questions are conceptually distinct from those arising

in Two-Sided Agentic AI Markets (Feb. 9, 2026) (unpublished manuscript), https://ssrn.com/abstract=5382575 [https://perma.cc/PNT5-9XSZ] (arguing that AI systems can be architected to achieve specific market outcomes through design choices, and that intent-based legal frameworks collapse when outcomes are emergent properties of design) [hereinafter Mukherjee & Chang, Duty to Disrupt].

[7] *See* Katherine Lee, A. Feder Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, 72 J. Copyright Soc'y U.S.A. 251 (2025) (developing a supply-chain taxonomy of the generative-AI pipeline and analyzing the distinct copyright questions each stage raises).

[8] *See, e.g.*, N.Y. Times Co. v. Microsoft Corp., No. 23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023) (alleging that defendants used copyrighted articles to train large language models).

from compaction, which relates to the second and third stages: the retrieval, processing, and re-expression of copyrighted material at inference time. Even an AI system trained entirely on non-copyrighted data can perform compaction if it has access to copyrighted sources at query time.

Compaction is distinct from traditional summarization. Summarization is a *source-determined* process. Given a document, a summarizer produces a shorter version that captures its principal content. The output is static in a critical sense: it does not change depending on who reads it or why. A summary of a judicial opinion will contain the same key holdings whether it is prepared for a lawyer, a journalist, or a student. The relationship between the original and the summary is fixed, and the summary is evaluated against the original for fidelity.

Compaction, by contrast, is *query-determined*. The same source document produces different outputs depending on the purpose for which information is being extracted. When an AI system is asked to "explain the holding of this case for a patent attorney," it extracts different elements than when asked to "identify the policy implications for consumer protection." The compacted output is not a condensed version of the document; it is a purpose-built extraction of whatever elements are useful for a specific downstream task, with everything else discarded.[9]

The result is a cascade of transformations. The original copyrighted expression is read by the AI system, compacted into a query-relevant extract, and regenerated as new text for the user. At no

[9]The technical architecture that underlies compaction at scale is retrieval-augmented generation (RAG) and its variants. *See* Patrick Lewis et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, *in* 33 Advances in Neural Info. Processing Sys. 9459 (H. Larochelle et al. eds., 2020). In a standard RAG workflow, a primary AI system receives a user's query, identifies relevant source documents, retrieves them, and generates a response that synthesizes the retrieved information with the model's own capabilities. Recent extensions of this architecture—including cache-augmented generation and agentic retrieval—delegate the retrieval and processing steps to secondary AI agents, which are themselves capable of reasoning about what information is relevant to the primary system's task. They do not merely retrieve chunks of text by keyword match; they read, comprehend, and selectively extract—performing, in effect, a context-specific editorial judgment about what matters and what does not. The technical literature describes these capabilities variously as query-focused summarization, contextual compression, and agentic retrieval. Recent work on recursive language models takes this further: a root model delegates source processing to recursively spawned sub-models, each of which extracts and compresses before passing results upward, such that the original expression is progressively stripped through multiple layers of abstraction. *See Recursive Language Models*, Prime Intellect (Jan. 1, 2026), https://www.primeintellect.ai/blog/rlm [https://perma.cc/L4AX-PSKZ]. Our use of the term "compaction" is legal rather than technical: it denotes the downstream consequence of these capabilities for copyright—the extraction of market-substituting value from copyrighted expression without reproducing that expression.

point in this chain does the original expression survive intact. What survives is the informational content—the "idea" in copyright's terminology—stripped from its original expressive form and re-clothed in new language, new structure, and often a new medium.

B. The Malleability of Expression

The cascade described above is possible because generative AI renders expression *malleable*—and this malleability is not contingent on any particular system's design or behavior. It is a foundational property of generative modeling itself.

Generative models, whether they produce text, images, audio, or video, learn to represent their training data at a level of abstraction that separates *what is said* from *how it is said*. In the learned representational space, semantically similar options reside as neighbors: synonymous phrasings, stylistic variations, alternative visual renderings of the same scene.[10] Once this structure has been learned, generating different surface realizations of the same underlying content is native to the technology. A text-to-image model prompted to paint a sunset over the ocean will produce a thousand distinct images—a thousand distinct expressions of a single idea—without exercising anything resembling adaptive or goal-directed behavior. It is simply sampling from a learned space of possibilities. The same is true of a language model completing a sentence: given a message to convey, it can choose any words.[11]

---

[10]This property reflects how generative models represent data as high-dimensional vectors—commonly known as embeddings—where spatial proximity corresponds to semantic similarity. The foundational demonstration was Word2Vec, in which vector arithmetic captures semantic relationships: the vector for "king" minus "man" plus "woman" yields the vector closest to "queen." *See* Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado & Jeff Dean, *Distributed Representations of Words and Phrases and their Compositionality*, *in* 26 Advances in Neural Info. Processing Sys. 3111 (2013). Modern generative AI extends this principle from individual words to entire documents and cross-modal concepts, such that texts with similar meanings or images with similar visual content cluster in the learned space, enabling interpolation between representations and generation of novel variations from the same underlying semantic content. For a comprehensive overview of vector semantics and embeddings, *see* Daniel Jurafsky & James H. Martin, Speech and Language Processing ch. 5 (3d ed. draft 2026), https://web.stanford.edu/~jurafsky/slp3/ [https://perma.cc/29DJ-34Q8].

[11]We have elsewhere termed the adaptive, goal-directed quality of modern AI systems "fluid agency"—the capacity to modulate means while pursuing specified ends. *See* Mukherjee & Chang, Fluid Agency, *supra* note 4. Fluid agency amplifies the architectural malleability described here by making the selection among expressive possibilities query-directed and purpose-driven—the difference between randomly sampling from the space of possible sunsets and selecting the one that best serves a user's specific request. The legal significance of this adaptive quality for the attribution of volitional conduct is explored in Section II.B.

Consequently, given any idea, such as a legal argument, a melody, or a narrative arc, the system can generate an effectively infinite number of distinct expressions of it. It can translate between languages, convert between media, shift between registers, and interpolate between styles. A single compacted representation of a novel's plot can yield a screenplay, an audio dramatization, a children's adaptation, and a critical essay, none of which bear textual resemblance to the original or to each other.

This is not a hypothetical capability. Google's NotebookLM converts uploaded documents—including copyrighted books and articles—into conversational podcast scripts in the voices of AI-generated hosts. Perplexity AI reads web sources and generates original-language answers that synthesize information from copyrighted articles the user never sees—a practice that has already prompted copyright litigation.[12] AI-powered educational platforms compact textbooks into interactive lesson modules. In each case, the original expression is consumed, its informational content extracted, and new expression generated—expression that is, as we have demonstrated elsewhere, distributionally distinct from the source material even when the underlying ideas are directly derived from it.[13]

This has a specific and consequential implication for copyright. Copyright law has long assumed that expression is the scarce, costly artifact—the creative labor that produces a specific sequence of words, notes, or brushstrokes. Ideas are abstract and common; expression is concrete and individual. Protection attaches to the latter precisely because it is the product of authorial effort.[14]

---

[12] *See NotebookLM Now Lets You Listen to a Conversation About Your Sources*, Google (Sept. 11, 2024), https://blog.google/technology/ai/notebooklm-audio-overviews/ [https://perma.cc/Q9VF-APQX]; Beatrice Nolan, *Perplexity, the $18 Billion AI 'Answer Machine,' Wants to Make Peace with Publishers. They're Not Buying It.*, Fortune (Aug. 26, 2025), https://fortune.com/2025/08/26/perplexity-lawsuits-publishers-ai-search-nikkei-news-corp/ [https://perma.cc/EP8N-JRBC]. The New York Times has sued Perplexity AI for precisely this form of content extraction. Complaint, N.Y. Times Co. v. Perplexity AI, Inc., No. 25-cv-10106 (S.D.N.Y. filed Dec. 5, 2025). For a discussion of the copyright implications of AI-generated search answers, *see generally* James Grimmelmann, *Copyright for Literate Robots*, 101 Iowa L. Rev. 657 (2016) (discussing the copyright implications of machines that read and process copyrighted text).

[13] *See* Mukherjee & Chang, Distributional Distinctiveness, *supra* note 5 (finding that AI-generated outputs are statistically distinguishable from training data across multiple generative architectures, and that this distinctiveness increases with model capability).

[14] *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991) (holding that copyright protects only original

But for a generative AI system, expression is not scarce. Instead, expression can be generated infinitely from the compacted representation of ideas, inverting the foundational assumption of the idea-expression dichotomy. The "idea" becomes the valuable, extractable resource. The "expression" becomes disposable packaging, regenerated on demand for each new use. Compaction is the process by which this inversion is achieved. And once expression is malleable, a system can read a work and produce a substitute that preserves its value while shedding any resemblance to it—rendering the access-plus-similarity framework that copyright relies on structurally incapable of response.

II. The Collapse of Substantial Similarity

A. Access Without Similarity

Copyright infringement analysis proceeds in two steps. The plaintiff must first show that the defendant had *access* to the copyrighted work, and then that the defendant's work is *substantially similar* to protectable elements of the original.[15] Courts have developed two principal methods for evaluating substantial similarity. The Ninth Circuit applies a two-part *Krofft* test: an "extrinsic" prong that objectively compares specific protectable elements such as plot, theme, dialogue, mood, and sequence, and an "intrinsic" prong that asks whether an ordinary observer would perceive the works as substantially similar in their "total concept and feel."[16] The Second Circuit's *Altai* framework uses "abstraction-filtration-comparison": the court abstracts the work into layers from the most concrete (literal text) to the most abstract (general idea), filters out unprotectable elements (ideas, facts, scènes à faire), and compares what remains.[17] Both approaches share a

expression, not facts or ideas, and that originality requires "independent creation plus a modicum of creativity").

[15] *See* Arnstein v. Porter, 154 F.2d 464, 468 (2d Cir. 1946) (establishing the access-plus-similarity framework for copyright infringement analysis).

[16] Sid & Marty Krofft Television Prods., Inc. v. McDonald's Corp., 562 F.2d 1157, 1164 (9th Cir. 1977); *see also* Cavalier v. Random House, Inc., 297 F.3d 815, 822 (9th Cir. 2002) (refining the extrinsic test to focus on specific protectable elements).

[17] Computer Assocs. Int'l, Inc. v. Altai, Inc., 982 F.2d 693, 706–11 (2d Cir. 1992).

common structural premise: that infringement is established by demonstrating *similarity* between the original and the allegedly infringing work.

Compaction renders this premise structurally unsound. Consider a paradigmatic scenario: a user asks an AI system to "explain the key arguments in [a copyrighted law review article] and draft a memo applying them to my client's case." The AI retrieves the article via agentic RAG, extracts its analytical framework, discards the original text, and generates a memo in the user's preferred format. Access is unambiguous—the AI read the copyrighted article in full. But the resulting memo bears no resemblance to the original's text, structure, or organization. The *Krofft* extrinsic test finds no shared protectable elements in the literal expression. The intrinsic test finds no similarity in "total concept and feel" because the works are in entirely different forms and registers. The *Altai* filtration step strips away the unprotectable ideas and methods, and what remains of the original's protectable expression is absent from the memo entirely.[18]

The most prominent AI copyright case to date illustrates the limitation. In *The New York Times Co. v. Microsoft Corp.*,[19] the Times's strongest *similarity-based* evidence of infringement consisted of "Exhibit J"—a compilation of instances where ChatGPT reproduced near-verbatim passages from Times articles.[20] This red-marked, word-for-word overlap was the evidentiary backbone of the plaintiff's case. But this strategy depends entirely on the AI system *failing* to transform—on the model regurgitating memorized training data rather than generating new expression. Compaction, by definition, does not regurgitate. It extracts and re-expresses. An AI system performing compaction on the same Times articles would produce outputs conveying the same information in entirely different words, and the Exhibit J strategy would yield nothing.

---

[18] To be precise, the objective tests—*Krofft*'s extrinsic prong and *Altai*'s comparison step—would fail because no protectable expression is reproduced. The subjective intrinsic test—whether an ordinary observer would perceive the works as similar—might in theory detect a shared "feel" in some cases, but our empirical work demonstrates that human observers are unable to reliably distinguish AI-generated works from their source materials even in controlled settings, performing at approximately chance level. *See* Mukherjee & Chang, Distributional Distinctiveness, *supra* note 5.

[19] *See supra* note 8.

[20] *See* Mukherjee & Chang, Distributional Distinctiveness, *supra* note 5 (discussing the reliance on verbatim overlap metrics in current AI copyright litigation and observing that "the law currently lacks a tool to measure distinctiveness absent exact replication").

This is not a marginal evidentiary problem that can be patched through doctrinal refinement. It is a recurring fact pattern—access plus non-mimetic substitution—that similarity doctrine was not built to police. The substantial similarity framework was built to catch *copying*—the reproduction of expression. Courts can and do find infringement based on nonliteral copying—the taking of protectable structure, sequence, or organization.[21] But compaction can be calibrated to avoid even structural similarity, retaining only the unprotectable ideas and recasting them in an entirely new organizational framework.[22] The framework thus confronts a category of appropriation it was never designed to detect: works that are *derived* from a specific source but bear no *similarity* to it.

One might object that even if the reproduction right fails, compaction resembles the preparation of a derivative work—a "recasting" or "adaptation" under 17 U.S.C. § 106(2) (2024). But derivative work infringement, like reproduction infringement, requires that the defendant's work incorporate protectable expression from the original.[23] A compacted output that retains only unprotectable ideas, facts, and methods does not incorporate protectable expression in any legally cognizable sense. The derivative work right thus fails for the same structural reason: both the reproduction and derivative work rights presuppose that the defendant's work contains some trace of the plaintiff's protectable expression, and compaction is defined by the elimination of precisely that trace.

A remaining possibility is that compaction involves actionable intermediate reproduction—the

---

[21] *See Altai*, 982 F.2d at 706–07; *see also* Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (recognizing that infringement may extend beyond literal copying to the taking of protected patterns).

[22] This structural mismatch is most acute for informational, analytic, and instructional works—news articles, textbooks, research reports, legal analyses—where the market value resides primarily in unprotectable ideas, facts, and methods rather than in distinctive expressive choices. For works whose market depends on protectable expression itself—novels with distinctive characters, films with protectable visual sequences, music with original melodic structure—compaction may be less able to strip all similarity, because the protectable elements are more tightly fused with the content a user seeks. *See* Benjamin L.W. Sobel, *Elements of Style: Copyright, Similarity, and Generative AI*, 38 Harv. J.L. & Tech. 49, 74–75, 101–02 (2024) (arguing that "style"—the cumulative effect of individually unprotectable expressive choices—may itself constitute protectable expression, and that generative AI is specifically designed to reproduce these higher-order patterns of similarity).

[23] *See* Litchfield v. Spielberg, 736 F.2d 1352, 1357 (9th Cir. 1984) (holding that a work is not an infringing derivative work unless it incorporates protectable expression from the preexisting work); *see also Altai*, 982 F.2d at 701 (applying filtration analysis to derivative work claims); Oren Bracha, *Generating Derivatives: AI and Copyright's Most Troublesome Right*, 25 N.C. J.L. & Tech. 345, 384 (2024) (arguing that the derivative work right "can only apply when specific expression that is traceable to a particular work, rather than unprotectable metainformation, is incorporated into the generated output").

copying of a copyrighted work into system memory during retrieval and processing.[24] But this theory confronts a double bind. The first barrier is procedural: expression similarity is not merely the legal *test* for copyright infringement—it is the *evidentiary predicate* that initiates the litigation process. Without some similarity in the output pointing back to a specific copyrighted work, a plaintiff cannot plausibly allege that a system compacted *their* work during a particular user's session. Demanding session-specific retrieval logs without such evidence is the kind of speculative fishing expedition that courts will not authorize at the pleading stage.[25] Intermediate copying is thus not merely "hard to prove"—it is a claim that cannot be *initiated* without the very output similarity that compaction eliminates.

The second barrier is substantive, and we develop it in Section III.B: even if the procedural hurdle were cleared—through leaked logs, whistleblowers, or admissions—the intermediate copy may itself be fair use. Recent decisions treating the transformative processing of lawfully obtained works as protected activity suggest that inference-time compaction may enjoy even stronger fair use protection than training-time ingestion. The result is a doctrine that works precisely as designed—no output similarity means no viable claim; transformative intermediate processing means fair use—yet permits systematic market substitution. This is not merely another doctrinal gap. It is the letter of copyright law operating as written while defeating the law's animating purpose.

B. The Volitional Conduct Gap

Even if a court were willing to stretch substantial similarity to capture compacted outputs, or to pursue an intermediate copying claim despite the barriers described above, plaintiffs face a second, equally fatal barrier: identifying a human who performed the volitional act of copying. Copyright's exclusive rights are implicated only by the volitional act of a person who "causes

[24] *See* MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511, 518 (9th Cir. 1993) (holding that loading software into RAM creates a "copy" under the Copyright Act).

[25] *See* Ashcroft v. Iqbal, 556 U.S. 662, 678 (2009) (requiring "factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged").

in some meaningful way an infringement."[26] Automated processes that copy without human direction do not, standing alone, give rise to direct liability.[27]

With compaction, the chain of volitional conduct can be fatally attenuated. A user may issue a prompt—a question or instruction such as "summarize the current debate over Chevron deference" or "explain recent developments in fair use doctrine"—that does not specify which copyrighted works to access or how to process them. The user has issued a facially noninfringing prompt: that prompt names no copyrighted work and requests no reproduction.[28] Yet the AI system may autonomously retrieve a paywalled law review article, a copyrighted treatise, or a news report, compacting each at inference time to assemble its response.[29] The user may never even learn which works were accessed, let alone have selected them or directed their use. The AI agent, for its part, is not a legal person capable of volitional conduct in the sense the doctrine requires.[30]

---

[26] CoStar Grp., Inc. v. LoopNet, Inc., 373 F.3d 544, 550 (4th Cir. 2004). *See also* Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1370 (N.D. Cal. 1995) (holding that an ISP's automatic storage and transmission of copyrighted material did not constitute direct infringement because the copying was not volitional). For analysis of the volitional conduct doctrine as applied to generative AI, *see* Aleksander J. Goranin, *A Deep Look at Copyright's Volitional Conduct Doctrine and Generative Artificial Intelligence*, 74 Emory L.J. 1127 (2025).

[27] *See* Cartoon Network LP v. CSC Holdings, Inc., 536 F.3d 121, 131 (2d Cir. 2008) (holding that an automated remote-storage DVR system did not directly infringe because it was the customer, not the cable company, who initiated each recording).

[28] The analysis differs where a user specifically directs the system to process a particular copyrighted work—for example, by uploading a PDF and requesting its summary. In that scenario, the user's volitional conduct is clear, and the doctrinal difficulty described here does not arise.

[29] This issue has a structural antecedent in web search. Search engines also crawl, index, and cache copyrighted webpages, and at query time retrieve and display snippets of them in response to facially noninfringing queries. Courts have generally tolerated this processing in part because the snippet functions as a pointer that routes the user to the copyrighted source, preserving the economic link between access and payment. *See* Authors Guild, Inc. v. Google, Inc., 804 F.3d 202, 216–18 (2d Cir. 2015); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007); Field v. Google, Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006). However, compaction inverts this dynamic: where a search snippet sends the user *to* the source, a compacted output sends the user *around* it—delivering the value of the copyrighted work without referral.

[30] *See* Thaler v. Vidal, 43 F.4th 1207, 1213 (Fed. Cir. 2022) (holding that an AI system cannot be an "inventor" under the Patent Act because it is not a "natural person"); U.S. Copyright Office, Compendium of U.S. Copyright Office Practices § 306 (3d ed. 2021) (requiring human authorship as a condition of copyright registration). For an analysis of the conceptual foundations of volition in copyright and the challenges posed by automated systems, *see* Mala Chatterjee & Jeanne C. Fromer, *Minds, Machines, and the Law: The Case of Volition in Copyright Law*, 119 Colum. L. Rev. 1887, 1896–1904 (2019) (arguing that machines may possess a functional analog of volition sufficient for legal purposes). For a practical framework for tracing culpability through AI systems despite this doctrinal gap, *see* Anirban Mukherjee & Hannah Hanwen Chang, *Operational Agency: A Permeable Legal Fiction for Tracing Culpability in AI Systems*, SMU Sci. & Tech. L. Rev. (forthcoming 2026), https://ssrn.com/abstract=5680063 [https://perma.cc/V6LG-D685] [hereinafter Mukherjee & Chang, Operational Agency].

Secondary liability doctrines offer no clear alternative. Traditional contributory infringement requires knowledge of *specific* infringing activity and a material contribution to it,[31] but the operator of a general-purpose AI system has no knowledge of which copyrighted works any particular user's query will cause the system to access. Vicarious liability requires both a direct financial interest in the infringing activity and the right and ability to supervise it,[32] but the autonomous nature of agentic AI—the very fluid agency that enables compaction—makes real-time supervision of each query's interaction with copyrighted material impracticable. And the *Sony* defense—that a technology capable of "substantial noninfringing uses" should not give rise to contributory liability merely because it can also be used to infringe—applies with considerable force to general-purpose AI systems, whose compaction capabilities are one function among many.[33]

The convergence of these failures—the inability to establish similarity, the inapplicability of the derivative work right, the double bind facing intermediate copying claims, and the inability to attribute volitional conduct—means that for the paradigmatic class of compaction-based appropriation—query-time extraction that yields non-mimetic, purpose-built substitutes—existing doctrine offers no reliable enforcement pathway. The question, then, is not merely whether the current legal tools need updating. It is whether the conceptual foundation on which those tools rest—the idea-expression dichotomy itself—remains sound, whether it can still do the work we ask of it. We turn to that question next.

[31] *See* Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc., 443 F.2d 1159, 1162 (2d Cir. 1971); *see also* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 930 (2005).

[32] *See* Grokster, 545 U.S. at 930 n.9.

[33] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442 (1984).

III. The Idea-Expression Dichotomy Revisited

A. The Original Bargain and Its Inversion

The idea-expression dichotomy rests on an economic bargain. Copyright grants authors a limited monopoly over their *expression*—their specific creative choices in language, structure, and form—in order to incentivize the production of creative works. But it withholds protection from the *ideas* those works contain, because ideas are the building blocks of cumulative knowledge and artistic progress.[34] As the Supreme Court has explained, copyright serves as an "engine of free expression" by "establishing a marketable right to the use of one's expression" that "supplies the economic incentive to create and disseminate ideas."[35]

This bargain rests on two assumptions. The first is that expression is the *scarce, costly* element—the creative labor that merits protection. Ideas, by contrast, are assumed to be abundant and freely discoverable. The second is that the boundary between idea and expression, while difficult to draw in practice, is *in principle* a real and adjudicable distinction. Judge Learned Hand acknowledged the difficulty in *Nichols v. Universal Pictures Corp.*, observing that "[n]obody has ever been able to fix that boundary, and nobody ever can."[36] But Hand's observation was a confession of *practical difficulty*, not a declaration of *conceptual impossibility*. The assumption was that the boundary existed, even if its precise location was contestable in every case.

Compaction inverts both assumptions.

---

[34] 17 U.S.C. § 102(b) (2024) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work."). *See also* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 349–50 (1991) (reaffirming that copyright protects original expression but "does not extend to the facts themselves").

[35] Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 558 (1985).

[36] *Nichols*, 45 F.2d at 121.

1. *The Inversion of Scarcity*

On the first: expression is no longer scarce. An AI system capable of compaction can generate an effectively infinite number of distinct expressions from a single compacted representation of the underlying ideas. The "creative labor" of expression—choosing words, structuring arguments, crafting prose—is no longer the exclusive product of time-intensive human effort; it is the output of a probabilistic process that can produce as many variations as desired, in any medium, register, or style.[37]

But the deeper problem is not that expression has become cheap. It is that *ideas* have become extractable. In the pre-AI world, the idea-expression dichotomy posed a limited practical problem because there was no efficient way to extract the "idea" from a work without also encountering—and potentially copying—its expression. To learn what a book teaches, you had to *read* the book. The expression was, in effect, the delivery mechanism for the idea, and copyright's control over the expression gave the author indirect control over access to the idea. Compaction breaks this link. An AI agent can read the book, extract its ideas, and discard the expression entirely—all in seconds, at scale, for any user who asks. The author's expression no longer serves as a gatekeeper.

The distinction from a human research assistant is instructive. A researcher who reads a book and prepares a summary for a client has also extracted ideas from expression. But the researcher operates at human speed: one person, one book, one summary. More importantly, the researcher *accessed* the work—she read the book, and typically the book was purchased or borrowed. The author's expression served as the necessary gateway through which the ideas passed, and the act of access preserved the economic link between value and expression. Compaction eliminates this gateway function at scale. An AI system can read a work once and serve its extracted ideas to an unlimited number of users, none of whom need access the original. The economic logic of copyright—that controlling expression gives authors indirect control over the value of their ideas—dissolves when expression is no longer the bottleneck through which ideas must pass.

The economic value of the work thus shifts from the expression to the ideas, but the ideas are

[37] *See supra* Section I.B (discussing the malleability of expression as a foundational property of generative modeling).

precisely what copyright does not—and constitutionally cannot[38]—protect.

2. The Unmeasurable Boundary

The second assumption—that the idea-expression boundary is in principle adjudicable—is equally destabilized, but for a reason more fundamental than the practical difficulties compaction creates. Hand's abstractions test[39] proceeds by successive layers of generalization: at the most concrete level, a work consists of its specific words; at a higher level, its plot; at a still higher level, its themes; and at the most abstract level, a general idea that is unprotectable. The court's task is to identify the level at which protectable expression gives way to unprotectable ideas.

For human works, this test rests on an implicit but foundational premise: that the author's cognition is itself *layered*. A novelist holds a concept—the corrupting influence of power—as one kind of mental representation, and renders it through plot, character, and prose that constitute a different kind. The concept and its expression are distinguishable not because they sit at different points on a single continuum, but because they are different in *kind*: one is abstract and conceptual, the other concrete and linguistic. Hand's layers of abstraction work because they track layers of cognition. When a court peels back expression to locate the underlying idea, it is retracing—however imperfectly—the path by which an author moved from concept to words. The test is administrable because there are, in fact, different layers to peel.

AI systems do not necessarily generate through layers that correspond to the idea-expression distinction Hand's test implicitly tracks. A diffusion model may develop internal representations of recognizable concepts, but generation proceeds without an intervening stage of ideation—these are mathematical regularities across training data, not concepts the system holds and then independently chooses to express. A language model may engage in intermediate reasoning that more closely approximates ideation, but the depth of such reasoning is continuous and tunable—a design parameter, not a stable cognitive architecture.

---

[38] *See* U.S. Const. art. I, § 8, cl. 8 (empowering Congress to secure rights for "limited Times" to "promote the Progress of Science and useful Arts"); *Baker*, 101 U.S. 99, 103.

[39] *See Nichols*, 45 F.2d 119, 121.

The representations that drive generation are thus not necessarily "ideas" in the sense the copyright regime contemplates. They are compressed mathematical relationships *across* expressions—not a separate cognitive layer from which expression is produced, but the residue of expression itself, distilled through the network's parameters.[40] Moreover, these representations can generate outputs at *any* level of abstraction, and the AI can position its output at whatever point on the gradient is most useful for the user—or most evasive of copyright liability.[41]

This matters because it determines what kind of problem the idea-expression boundary presents. For human works, the boundary is a *locational* problem: the layers exist, and the difficulty is finding where one ends and another begins. For AI outputs, the boundary is an *administrability* problem: the generative process may not track the layered cognitive architecture that Hand's test assumes, and the system can position its output at any point on the abstraction continuum—making the boundary not merely hard to locate but impossible to stabilize.

That is, the idea-expression dichotomy assumes a *boundary*, however fuzzy, between layers that are different in kind. Compaction need not reveal such layers—only a *continuum* that systems can traverse at will. Whether an AI system's architecture replicates the cognitive separation between concept and expression is system-dependent and, from the output alone, unknowable—making the "ideas" a reader identifies potentially nothing more than post-hoc abstractions projected onto text generated without any such separation. Ultimately, ideas matter only to machines that think, and AI systems, despite the moniker of intelligence, may not think very much or even at all. In sum, Hand's test asks courts to peel back layers. But there may be no consistent layers to peel—only a

[40] To be clear, some AI systems do engage in intermediate reasoning—chain-of-thought processes that approximate human ideation. But this reasoning is contingent (it can be enabled or disabled), variable in depth (from none to extensive), and opaque (hidden within the network's computations, where it may not faithfully represent the model's actual computational process). The existence of optional, variable, and unfaithful reasoning processes makes the idea-expression boundary not merely difficult to draw but system-dependent and unverifiable—a further reason the dichotomy cannot serve as a stable legal standard for AI-generated content. *See* Jason Wei et al., *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*, *in* 35 Advances in Neural Info. Processing Sys. 24824 (2022) (showing that chain-of-thought reasoning is elicited through prompting rather than being intrinsic to the model); Miles Turpin et al., *Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting*, *in* 36 Advances in Neural Info. Processing Sys. (2023) (finding that chain-of-thought explanations can be systematically unfaithful to the model's actual computational process).

[41] This is the adversarial tuning problem identified in the Introduction, now operating not merely on a model's loss function but on the full idea-expression continuum. *See supra* notes 5–6 and accompanying text.

continuous gradient of latent similarity.

B. The Administrability Crisis

This creates a dual administrability crisis. On one side, the tools for *detecting* infringement fail: substantial similarity cannot identify compaction-based appropriation, and volitional conduct cannot be attributed to any person in the processing chain. On the other side, the tools for *defining* the boundary fail: if expression can be regenerated at any point on the abstraction continuum, there is no stable level at which a court can declare "this is idea" and "that is expression." The dichotomy's two components, the enforcement problem and the definition problem, collapse simultaneously.

The result is a policy dilemma with no easy resolution. We cannot extend copyright to protect ideas—that would create monopolies over facts, methods, and concepts, precisely the outcome *Baker v. Selden* was designed to prevent.[42] But we cannot ignore compaction, because it enables the systematic extraction of value from copyrighted works without any infringement under existing law, potentially undermining the economic incentives for creation that copyright exists to preserve.[43]

Some may argue that fair use doctrine, with its sensitivity to market harm, can fill this gap.[44] But fair use faces its own difficulties in the compaction context. On one hand, compaction is *purposefully different* from the original—the user receives a query-specific extract, not a condensed version of the work—which might support a finding of transformative use. On the other hand, compaction directly *substitutes for the market* for the original, because the user obtains the value

[42] *See Baker*, 101 U.S. 99, 101–04.

[43] The challenge extends beyond U.S. law. The European Union has addressed AI's interaction with copyrighted material through text-and-data-mining exceptions, Directive 2019/790, arts. 3–4, 2019 O.J. (L 130) 92, and AI Act obligations regarding copyright-compliance policies and public summaries of training content, Regulation (EU) 2024/1689, art. 53(1)(c)–(d), 2024 O.J. (L 2024/1689). In February 2026, the European Publishers Council filed an EU antitrust complaint against Google over AI-generated search summaries, framing the issue in terms of content extraction and compensation rather than expressive similarity—a reframing that itself illustrates the doctrinal shift this Essay describes. *See* Press Release, Eur. Publishers Council, Formal Antitrust Complaint Against Google over AI Overviews and AI Mode (Feb. 10, 2026), https://www.epceurope.eu/our-ai-competition-complaint. The comparative dimensions of compaction across jurisdictions merit separate treatment.

[44] 17 U.S.C. § 107 (2024).

of the work without purchasing or accessing it. The Supreme Court's recent *Warhol* decision sharpened this concern, holding that when a secondary work serves the same commercial purpose as the original, the first fair use factor weighs against the defendant even when expression is substantially altered—importing market-substitution logic into the very analysis of transformativeness.[45] The doctrine thus sends mixed signals: compaction alters expression radically, which should weigh in favor of transformative use, but substitutes for the original's market.

Recent decisions suggest that where the procedural barriers described in Section II.A can be cleared, the intermediate copying at the heart of compaction may itself qualify as fair use—completing the double bind. In *Bartz v. Anthropic*, the court held that training an AI model on lawfully obtained books was "spectacularly transformative," reasoning that the model's outputs served entirely different purposes than the originals.[46] Yet just two days later, a different judge in the same district recognized the converse: AI-generated outputs could inflict cognizable market harm through "dilution," and had the plaintiffs presented evidence of such harm, the fourth factor would have gone to a jury.[47] If training-time ingestion is transformative, inference-time compaction—where a work is re-expressed in an entirely different form for a specific user's purpose—is arguably more so. *Sega Enterprises Ltd. v. Accolade, Inc.* similarly held that intermediate copying of copyrighted software to extract unprotectable functional specifications was fair use.[48] But *Sega* involved software—a highly functional work. Compaction involves literary and analytical works where the copyrighted expression is thick. The critical insight is that compaction treats expressive works *as if* they were functional data, mining them for their ideas while discarding their expression. The intermediate copying precedents thus paradoxically extend fair use protection to

[45] *See* Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 541–42 (2023); *see also* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985) (describing the effect on the market for the original as "undoubtedly the single most important element of fair use").

[46] Bartz v. Anthropic, PBC, No. 24-cv-05417, slip op. at 11 (N.D. Cal. June 23, 2025). The court distinguished lawfully obtained copies from pirated ones, holding that only the former qualified for fair use protection.

[47] *See* Kadrey v. Meta Platforms, Inc., No. 23-cv-03417, slip op. at 29 (N.D. Cal. June 25, 2025) (ruling for Meta on fair use due to insufficient evidence of market harm, but describing market dilution as "so important in this context" that evidence of it would have precluded summary judgment).

[48] 977 F.2d 1510, 1527–28 (9th Cir. 1992).

the extraction of expressive works by treating them as functional inputs.[49]

Google LLC v. Oracle America, Inc. reinforces this pattern.[50] The Court held that copying Java API declarations to build a new platform was fair use, reasoning in part that the copying was necessary to allow programmers to access underlying ideas through a familiar interface. The distinction from compaction is real: API declarations carry thin, functional copyright, while many compacted works—journalism, analysis, textbooks—carry thick copyright; and Google built *upon* the Java ecosystem to create a new platform, whereas compaction *substitutes for* the original. But the parallel is equally telling: the Court treated expression as an interface through which users accessed underlying functionality—precisely how AI treats copyrighted expression during compaction. If expression can be characterized as an interface to ideas, fair use doctrine may protect systematic value extraction.

The result is a law that matches its own formal requirements while defeating its own animating purpose. Copyright protects expression because expression was historically the scarce, costly artifact—the product of creative labor that merited exclusive rights. Ideas were left free because they were the building blocks of cumulative progress. But compaction inverts this scarcity: the idea is now what is expensive to produce—the journalism, the analysis, the research—while expression is what is cheap to generate. Current law thus protects the cheap thing and leaves the expensive thing unprotected. The fair use framework, far from filling the gap, may affirmatively *widen* it—classifying the extraction pipeline as transformative and the intermediate copying as protected, even as the end result is systematic market substitution.

Moreover, the doctrine's case-by-case application provides neither the predictability that AI developers need nor the protection that authors seek. This unpredictability is particularly corrosive because both sides require *ex ante* guidance. AI developers must make design choices—whether to enable retrieval of copyrighted sources, how to handle paraphrasing, whether to attribute—before

[49] For the argument that these intermediate-copying precedents are inapplicable to generative AI because such systems exploit intrinsic expressive value rather than merely extracting unprotectable functional elements, *see* Jacqueline C. Charlesworth, *Generative AI's Illusory Case for Fair Use*, 26 Vand. J. Ent. & Tech. L. 323 (2025).

[50] 593 U.S. 1 (2021).

any particular query is issued, and authors must decide whether to invest in works whose value may be extracted before any particular AI system processes them. A doctrine that resolves these questions only *ex post*, case by case, provides no design-time signal to either party.

We do not propose a specific doctrinal solution here—the problem is too fundamental for a quick fix, and the risk of overcorrection (chilling beneficial AI innovation) is as real as the risk of undercorrection (allowing systematic free-riding on creative labor). We have argued elsewhere that AI governance may require shifting from intent-based liability to design-based duties—engineering systems to respect legal norms by construction rather than policing violations after the fact.[51] Given the structural unreliability of output-stage enforcement, the retrieval stage—where copyrighted works are actually accessed and processed—may represent a more natural point of intervention. Whether these approaches can be adapted to copyright—requiring AI systems to be designed in ways that preserve the value of expression, even when they are technically capable of circumventing its protection—is a question that merits urgent attention.

What is clear is that the status quo is untenable. The idea-expression dichotomy was designed for a world in which expression was the stable, costly output of human creativity and ideas were the abstract commons on which future creators could freely build. In that world, protecting expression while leaving ideas free struck a reasonable balance. In a world where AI can extract ideas from expression instantly, generate new expression trivially, and traverse the continuum between the two at will, that balance no longer holds. Both sides of the dichotomy—what we protect and where we draw the line—require re-examination.

Conclusion

This Essay has argued that compaction—the capacity of AI systems to extract the functional essence of a copyrighted work and re-express it in an entirely new form—poses a fundamental

[51] *See* Mukherjee & Chang, Duty to Disrupt, *supra* note 6 (proposing a legally enforceable duty to engineer procompetitive AI behavior through system design); Mukherjee & Chang, Operational Agency, *supra* note 30 (developing a framework for tracing legal consequences through AI systems to the human design choices that shaped them).

challenge to copyright's most basic boundary: the idea-expression dichotomy.

The challenge operates at every level of copyright's enforcement architecture. Substantial similarity fails because compaction produces works derived from a source but bearing no resemblance to it. The derivative work right fails because compaction strips the protectable expression that infringement requires. Intermediate copying claims face a double bind: without output similarity they cannot be initiated, and even if that barrier is cleared, the processing may qualify as fair use. Volitional conduct cannot be attributed to any person in the chain. And the fair use framework, far from filling the gap, may affirmatively protect the extraction pipeline. The result is a doctrine that works precisely as designed while permitting systematic market substitution—a disjunction between copyright's letter and its animating purpose.

These failures are not accidents of doctrinal design. They are consequences of a foundational assumption that no longer holds: that expression is a stable, scarce, and identifiable artifact, distinguishable from the ideas it conveys. Compaction reveals expression to be malleable, infinitely reproducible, and traversable along a continuum that AI can navigate at will. The dichotomy that copyright law draws between idea and expression—a line that Judge Hand famously conceded "nobody has ever been able to fix"[52]—has not merely blurred. It has been rendered unmeasurable by a technology that treats both sides of the line as raw material for recombination.

The path forward is unlikely to run through output policing alone. If expression can always be regenerated to defeat similarity metrics, measuring infringement at the output stage is a losing strategy. More promising avenues may include retrieval-layer governance—requiring AI systems to log, license, or compensate for copyrighted sources accessed at inference time—or the design-based duties we have explored elsewhere, which would make rights-respecting behavior a system requirement rather than an afterthought.[53] The institutional response is already underway: publishers have begun challenging AI answer engines not through copyright's similarity framework but through competition law, seeking compensation for the value extracted

[52] Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930).

[53] *See* Mukherjee & Chang, Duty to Disrupt, *supra* note 6; Mukherjee & Chang, Operational Agency, *supra* note 30.

rather than the expression reproduced.

We do not claim to know where the law should ultimately settle. The reasons for refusing to protect ideas remain compelling, and any expansion of copyright's reach risks stifling the cumulative creativity that the system exists to promote. But continuing to anchor the entire copyright edifice on the protection of expression alone—when expression has become the most fungible element in the creative ecosystem, and when the distinction between idea and expression presupposes a cognitive architecture that a crucial mode of content generation does not possess—requires a candor that current doctrine has not yet achieved. Compactibility does not merely reveal a gap in copyright's defenses; it exposes a structural inversion in which the law's formal architecture—protecting expression, freeing ideas—operates to protect what is now cheap and leave unprotected what is now valuable. The question is no longer where to draw the line between idea and expression. It is whether a line can be drawn at all—and if not, what replaces it.