Pro Forma Copyright: AI Compaction and the Idea-Expression Impasse

Anirban Mukherjee
Hannah Hanwen Chang

Draft — April 18, 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

AI systems access copyrighted works in response to a specific query, extract the ideas embedded in them, discard the original expression, and dynamically reassemble individually unprotectable units—words, pixels, notes—into new expression that reproduces no protectable expression from any single source. This Article introduces this process and terms it *compaction*.

Copyright polices similarity, not derivation. Compaction makes copyright *pro forma*: formally present, functionally empty; compacted outputs are causally derived from copyrighted works but share no protectable expression with them. Every relevant doctrinal test—substantial similarity, derivative work claims, intermediate copying, volitional conduct attribution—functions as designed and yet finds nothing actionable.

By generating functionally equivalent yet dissimilar expression at negligible cost and on demand, compaction erodes the economic value of copyright's exclusive rights over expression. Ideas provide no fallback: they are constitutionally unprotected to prevent monopolies over facts and methods, and the same technology that defeats expression-protection defeats idea-protection even more thoroughly.

The result is a structural impasse, not a patchable doctrinal gap. The intended state of the idea-expression dichotomy was protected expression and unprotected ideas; the emerging state is unprotected expression and unprotected ideas.

Keywords: Idea-Expression Dichotomy, Fair Use, Copyright Law, Artificial Intelligence.
JEL codes: K11, O34, L86.

Table of Contents

Introduction

A lawyer preparing for oral argument issues a query to an AI research assistant, such as Google's Deep Research, OpenAI's equivalent, or Perplexity Pro. The system retrieves 127 sources—paywalled law review articles, newspaper investigations, government reports, copyrighted textbooks—reads each one, identifies the elements relevant to the query, and produces a three-thousand-word research memo.[1] The memo is excellent: well-organized, analytically precise, responsive to the lawyer's specific need. It synthesizes doctrinal arguments from a dozen scholarly articles, integrates empirical findings from three proprietary market reports, and frames the analysis in the register of appellate advocacy. It reproduces no protectable expression from

[1]Current AI research products routinely synthesize dozens to hundreds of sources per query. *See, e.g., Introducing Deep Research*, Google DeepMind (Dec. 11, 2024), https://blog.google/products/gemini/google-gemini-deep-research/ [https://perma.cc/4HLC-EWC9]; *Introducing Deep Research*, OpenAI (Feb. 2, 2025), https://openai.com/index/introducing-deep-research/ [https://perma.cc/B33M-XH4K]. Perplexity AI's Pro Search product similarly synthesizes information from multiple copyrighted web sources. *See* 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].

any single source. No reader could identify which works contributed, how much each contributed, or that any particular source was used at all.

The authors of the 127 works receive nothing—no payment, no attribution, no notification. Each contributes some fragment of analysis, some data point, some conceptual framework. None contributes enough to be identifiable in the output. The memo is, in a precise sense, a *collage*: an output assembled word by word from the statistical residue of many works, shaped not by any single source but by the query itself.[2]

The collage substitutes for each of those works in the one dimension that matters to the lawyer: it answers her question. She will not read any of the 127 originals. She will not subscribe to the journals that published them. She will not purchase the textbooks from which their data was drawn. The collage extracts what she needs from all of them and leaves behind what copyright protects—the expression.[3]

These systems are not merely hypothetical. They are already deployed at scale, and their implications have already provoked legal confrontation—including in domains where the value at stake lies not in facts or analysis but in visual and performative expression itself.[4] ByteDance's Seedance 2.0 can generate footage of figures who look like Tom Cruise but are not Tom Cruise: new performances, new visual compositions, produced on demand and tunable along a continuous gradient of resemblance.[5] The motion picture industry, which has driven major copyright

[2] *See infra* Part II (defining compaction and collage).

[3] This pattern is the subject of active litigation. *See* Complaint, N.Y. Times Co. v. Perplexity AI, Inc., No. 25-cv-10106 (S.D.N.Y. filed Dec. 5, 2025) (alleging that Perplexity's AI search product synthesizes copyrighted news content from paywalled sources and delivers it to users who never visit the original articles).

[4] *See, e.g.,* Getty Images (US), Inc. v. Stability AI, Inc., No. 23-cv-00135 (D. Del. filed Feb. 3, 2023) (alleging that AI image generation infringes copyrights in photographs); Andersen v. Stability AI Ltd., No. 23-cv-00201 (N.D. Cal. filed Jan. 13, 2023) (visual artists challenging AI image generation systems). For the industry response to AI video generation, *see infra* note 5.

[5] Within days of Seedance 2.0's launch in February 2026, every major Hollywood studio sent cease-and-desist letters to ByteDance alleging copyright infringement. *See* Gene Maddaus, *After AI Video of 'Tom Cruise' Fighting 'Brad Pitt' Goes Viral, Motion Picture Association Denounces 'Massive' Infringement on Seedance 2.0*, Variety (Feb. 12, 2026), https://variety.com/2026/film/news/motion-picture-association-ai-seedance-bytedance-tom-cruise-1236661753/ [https://perma.cc/4CMD-5RP5]; Todd Spangler, *Paramount Sends ByteDance Cease-and-Desist Letter Over Seedance AI Videos*, Variety (Feb. 14, 2026), https://variety.com/2026/film/news/paramount-disney-bytedance-cease-and-desist-seedance-ai-infringement-ip-1236663663/ [https://perma.cc/YD6J-MHG7]. Disney, Paramount, Warner Bros., Netflix, and Sony each sent separate cease-and-desist letters within one week. *See also* Press Release, Motion Picture Ass'n, Statement on Seedance 2.0 (Feb. 13, 2026) ("unauthorized use of U.S. copyrighted works on a massive scale"); Press

legislative pushes since the VCR, sees the threat most acutely because film is expression-heavy: its market value lies in specific visual, performative, and narrative choices, precisely the elements copyright is best equipped to protect. If even these works are becoming vulnerable to decomposition and regeneration, the implications for works whose value lies primarily in ideas (journalism, scholarship, legal analysis, market research) are far more severe. For a growing class of those works, the protective function is already being defeated. The question is fundamental:[6] *What does it mean to protect expression when expression can be dynamically disassembled and reassembled into a collage drawn from hundreds of sources—or thousands?*

The process at work is what this Article terms *compaction*: the ability of AI systems to access copyrighted works at inference time, learn the ideas embedded in them, discard the original expression, and generate new expression from the learned ideas, directed by a specific query.[7] The output is a collage of words, shaped word by word, drawn in part from training and in part from what is recovered at inference. Each fundamental unit of expression—a word, a pixel, a note—is individually unprotectable.[8] Expression has always been aggregation: a specific arrangement of these units into a coherent whole. What AI does is disaggregate existing expressions and

---

Release, SAG-AFTRA, SAG-AFTRA Statement on Seedance 2.0 (Feb. 13, 2026), https://www.sagaftra.org/sag-aftra-statement-seedance-20 [https://perma.cc/K6EA-R6TZ] ("blatant infringement… includes the unauthorized use of our members' voices and likenesses"). Senators Blackburn and Welch subsequently demanded ByteDance shut down Seedance entirely. *See* Letter from Sens. Marsha Blackburn & Peter Welch to Liang Rubo, CEO, ByteDance (Mar. 16, 2026). The studios' claims center on training-time ingestion of copyrighted works; this Article addresses the distinct inference-time mechanism described in Part II. The likeness dimension raises right-of-publicity questions distinct from the copyright analysis this Article develops. The copyright point is narrower: the visual expression itself—camera angles, lighting, composition, editing, performance—is dynamically regenerable.

[6] This question is distinct from those currently dominating AI copyright litigation—whether training on copyrighted works constitutes infringement, *see* N.Y. Times Co. v. Microsoft Corp., No. 23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023), or whether AI-generated outputs should receive copyright protection, *see* Thaler v. Vidal, 43 F.4th 1207 (Fed. Cir. 2022); U.S. Copyright Office, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16,190 (Mar. 16, 2023).

[7] We refer to the exercise of this capability as *compaction*, and to the susceptibility of a work to this process as its *compactibility*. The technical architecture that most commonly underlies compaction at scale is retrieval-augmented generation (RAG). *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). For the broader concept of query-directed adaptive processing in AI systems, *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]. Part II develops the full taxonomy.

[8] *See infra* Part III (developing the fundamental-units argument and its consequences for the boundaries of expression).

reaggregate them to specification. The result treats expression as a disposable interface—something to be read through, extracted from, and regenerated at will.[9]

When expression can be learned from and recomposed at the level of its fundamental units, copyright protection becomes *pro forma*—formally present, functionally empty.[10] The new expressions are causally derived from the originals but share no protectable expression with them—and the law protects against expressive similarity, not mere causal derivation.[11] Consequently, every relevant doctrinal test—substantial similarity, derivative work claims, intermediate copying, volitional conduct attribution—functions as designed and yet finds nothing actionable: the tests provide the right[12] answers that collectively yield the wrong outcome.[13] The boundaries of what counts as protectable expression become fluid, dissolving under a technology that operates at the sub-expressive level.[14] And because functionally equivalent expression can be generated at negligible cost, the economic value of copyright's monopoly over any specific expression collapses.[15] Ideas provide no fallback: they are constitutionally unprotected to prevent monopolies over facts and methods, and the same technology that defeats expression-protection defeats idea-protection even more thoroughly.[16] The result is a structural impasse, not a patchable doctrinal gap.[17]

---

[9] *See infra* Part III (showing that collage generalizes the logic of *Sega v. Accolade* and *Google v. Oracle* to expressive works generally).

[10] This Article brackets training-time ingestion and examines only what happens at inference time—when an AI system accesses copyrighted works in response to a specific query. Major pending training-time cases include *New York Times Co. v. Microsoft Corp.*, No. 23-cv-11195 (S.D.N.Y. filed Dec. 27, 2023); *Doe 1 v. GitHub, Inc.*, No. 22-cv-06823 (N.D. Cal. filed Nov. 3, 2022); *Andersen v. Stability AI Ltd.*, No. 23-cv-00201 (N.D. Cal. filed Jan. 13, 2023); Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc., 529 F. Supp. 3d 1261 (M.D. Fla. 2021). These cases raise important but analytically distinct questions.

[11] *See infra* Part III.

[12] "Right" here means consistent with the doctrinal tests' intended design—not normatively endorsed. The Article does not contest copyright's purpose of protecting original expression; it argues only that current doctrine is structurally ineffective against compaction. *See infra* Parts III, V.

[13] *See infra* Part III.

[14] *See infra* Part III.

[15] *See infra* Part IV.

[16] *See infra* Conclusion.

[17] *See infra* Parts III–IV and the Conclusion. *Cf.* Micaela Mantegna, *ARTificial: Why Copyright Is Not the Right Policy Tool to Deal with Generative AI*, 134 Yale L.J.F. 1126, 1128 (2024) (arguing that copyright is "not the right policy tool" for generative AI). Mantegna's analysis, like most of the prior literature on AI and copyright, addresses generative models that produce content from training data without accessing copyrighted works at inference time. *See id.* at 1131–35 (surveying GANs, diffusion models, and large language models as the relevant technical architectures).

The Article proceeds in four Parts. Part I establishes the economic premise on which copyright's protective function has historically rested: expression was expensive to produce, and protecting it was sufficient to incentivize the production of the ideas embedded within it. Part II defines compaction—the mechanism that breaks this premise—and distinguishes it from every prior condensation technology. Part III, the Article's most detailed section, demonstrates that compaction makes the boundaries of expression fluid: it walks through each doctrinal pathway—substantial similarity, derivative works, intermediate copying, volitional conduct, and the compounding barriers created by multi-source collage—and shows that each operates as designed and finds nothing actionable. Part IV traces the economic consequences: the cost structure that sustained the idea-expression dichotomy has inverted, and the institutional mechanisms that depended on it—deterrence, licensing, investment signals—are collapsing. The Conclusion synthesizes the diagnosis into a structural impasse: expression cannot be effectively protected against a technology that operates at the sub-expressive level, and ideas cannot and should not be protected either. The authors of those 127 works have a right that is formally complete and functionally empty.[18] This Article explains why.

I. The Doctrine and Its Implicit Premise

Copyright does not protect ideas. This is black-letter law: section 102(b) of the Copyright Act provides that "[i]n 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."[19] *Baker v. Selden* established the principle that copyright in a book on accounting did not extend to the accounting system the book described.[20] *Feist* confirmed that copyright protects only

This Article addresses a different technological phenomenon: modern AI systems equipped with search, retrieval, and agentic capabilities that access copyrighted works in real time to answer specific queries. *See supra* note 7. Compaction is a property of these full-stack AI systems, not of generative AI alone. The arguments here supplement rather than displace the prior literature.

[18] *See infra* Conclusion.

[19] 17 U.S.C. § 102(b) (2024).

[20] 101 U.S. 99, 103 (1879).

original expression, not facts or ideas.[21] Yet authors whose works are valuable primarily for their ideas—journalists, scholars, analysts—have been adequately compensated through copyright for two centuries.[22] If copyright protects only the wrapping and not the contents, how has the system worked for works whose contents are what consumers want?

The resolution is that ideas and expression are consumed as a bundle.[23] A reader who wants to learn what a book teaches must read the book. A student who wants a textbook's analysis must purchase the textbook. A journalist's scoop reaches its audience only through the article. The expression is the delivery vehicle; there is no way to extract the cargo without boarding the vehicle.[24] This creates what economists call a tied good: the idea (unprotectable) is tied to the expression (protectable). Because copyright gives the author control over the expression, the author captures economic value from both components of the bundle—even though copyright formally protects only one.[25] The expression is not the thing the consumer ultimately values. But it is the *only road* to the thing the consumer values. The road is the toll.[26]

The economic justification for copyright, formalized by Landes and Posner, is that creative works are public goods—costly to produce, cheap to copy.[27] Copyright solves this by granting a limited monopoly over expression, enabling authors to charge a price that covers their fixed costs.[28] But this framework contains an assumption so foundational it is rarely stated: that the

[21] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991) (holding that copyright protects only original expression, requiring "independent creation plus a modicum of creativity"—a standard that attaches to the selection and arrangement of elements, not to the elements themselves).

[22] *See* William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. Legal Stud. 325, 326–27 (1989) (modeling copyright's economic function as enabling authors to recoup fixed costs of production through control over copies).

[23] *See* Landes & Posner, *supra* note 22, at 326 (observing that "an author's fixed costs of creating a work are not recovered unless the work is sold"—implicitly assuming that ideas reach consumers only through the copyrighted expression that embeds them).

[24] *See supra* notes 19–20 and accompanying text.

[25] Copyright has always tolerated the extraction of unprotectable elements from copyrighted works—ideas, facts, methods—even when extraction requires copying. *See* Molly Shaffer Van Houweling, *The Freedom to Extract in Copyright Law*, 103 N.C. L. Rev. 445, 448–52 (2025) (theorizing "extractive use" as a doctrinal category and arguing that AI is a scale transformation of an activity copyright has always permitted).

[26] *See* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 558 (1985) (describing copyright as an "engine of free expression" that "supplies the economic incentive to create and disseminate ideas").

[27] *See* Landes & Posner, *supra* note 22, at 325.

[28] *Id.* at 326–29.

consumer's payment for expression *also* captures the value of the ideas, because the two are consumed together.[29] Landes and Posner did not need to state this assumption because, in 1989, no technology could systematically unbundle the two. The assumption was invisible because it was universally true.

That assumption can now be stated precisely. For any copyrighted work, the value to the consumer can be decomposed into two components: V_i, the value of the ideas, facts, and analysis the work contains; and V_e, the value of the expression itself—the aesthetic pleasure, the narrative craft, the distinctive voice.[30] For many categories of works—journalism, textbooks, legal analysis, market research—V_i vastly exceeds V_e.[31] A Bloomberg terminal subscriber does not pay for elegant prose; she pays for market intelligence. A law student does not buy a casebook for literary style; she buys it for doctrinal analysis. Yet copyright protects only V_e. The author's ability to capture V_i depends entirely on the bundling of ideas with expression. Expression is not valuable primarily for its own sake; it is valuable because it is the *only road* to the ideas. Copyright's control over the road gives the author pricing power over the destination. Expression is, in economic terms, a tollbooth.[32]

This framework has worked for the entire history of copyright because expression was expensive to produce.[33] Producing a new aggregation of words, notes, or pixels—one that conveyed the same ideas in a different form—required substantial human creative labor: choosing words, composing sentences, structuring arguments, arranging notes into melodies.[34] The monopoly over a specific expression had economic value precisely because producing an equivalent alternative

---

[29] *Id.* at 326.

[30] *See supra* notes 20–21 and accompanying text (establishing that copyright does not extend to ideas or systems).

[31] Part IV develops this asymmetry through the concept of the *investment-to-expression ratio*: the ratio between the investment required to produce a work's underlying ideas and the investment required to produce its specific expression. *See infra* Part IV.

[32] *See Harper & Row*, 471 U.S. at 558.

[33] *See generally* Landes & Posner, *supra* note 22, at 326 (modeling the incentive effects of changes in copying costs on creative production).

[34] *See* Feist, 499 U.S. at 345 ("[T]he requisite level of creativity is extremely low; even a slight amount will suffice." The low threshold reflects the assumption that even minimal creative effort in arranging expression is worth protecting because the alternative—recreating equivalent expression—requires comparable effort).

was costly. A competing textbook required its own research and its own prose. A competing news article required its own reporters. The dichotomy's protective function—protecting expression as a proxy for incentivizing the production of ideas—depended on this implicit premise: that a monopoly over one expression effectively controlled access to the ideas within it, because generating an alternative expression of the same ideas required comparable investment.[35]

If this premise fails—if equivalent expression can be generated at negligible cost—then the monopoly over any specific expression loses its economic value. The right is formally intact, but it protects one arrangement among an effectively infinite number of alternatives, each equally capable of conveying the same ideas. The protection becomes pro forma. Part II describes the technology that breaks the premise.

II. Compaction

A. What Compaction Is

Compaction is the process by which an AI system reads a copyrighted work, learns the ideas embedded in it, discards the original expression, and generates a new expression of the learned content tailored to a specific query.[36] The output is purpose-built: it serves the user's specific need, in the user's preferred format, retaining no protectable expression from the source.[37]

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

[35] *See* Landes & Posner, *supra* note 22, at 327–28 (showing that when copying costs approach zero, the author's ability to recoup fixed costs collapses—a model that implicitly assumes producing *new equivalent expression*, not just copying, requires investment).

[36] *See supra* note 7 (defining compaction and compactibility).

[37] The term denotes the downstream legal consequence of these capabilities—the extraction of market-substituting value from copyrighted expression without reproducing that expression.

user-facing text that may draw on retrieved sources.[38] Most current litigation focuses on stage one—whether training on copyrighted works is itself infringement.[39] This Article focuses on stages two and three: the retrieval, processing, and re-expression of copyrighted material at inference time. Even a system trained entirely on non-copyrighted data can perform compaction if it has access to copyrighted sources at query time.[40]

Compaction has four defining elements. First, *retrieval*: the AI accesses one or more copyrighted works at inference time in response to a user's query. Second, *extraction*: it identifies the elements of each work that are relevant to the query—typically ideas, facts, methods, or analytical frameworks rather than protectable expression. Third, *discarding*: it strips the original expression, retaining only the extracted content in compressed, internal representation. Fourth, *re-expression*: it generates new text (or other media) that conveys the extracted content in a form tailored to the user's purpose.[41] The output is *query-determined*: the same work yields different compacted outputs depending on who asks and why.[42]

The critical distinction is between *learning* and *memorizing*.[43] Prior AI systems that raised copyright concerns were primarily memorizing and regurgitating—reproducing training data verbatim.[44] Compaction is categorically different. The AI *learns the idea* from the expression

[38] *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).

[39] *See supra* note 10 and accompanying text.

[40] This is because compaction operates at inference time, not training time. The technical architecture is retrieval-augmented generation. *See supra* note 7.

[41] The technical architecture that most commonly underlies compaction at scale is retrieval-augmented generation (RAG) and its variants. *See* Mukherjee & Chang, Fluid Agency, *supra* note 7. Recent work on recursive language models extends this architecture 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].

[42] *See supra* note 7 and accompanying text (defining compaction as query-directed extraction and re-expression).

[43] We use *learn*, *encode*, and cognate terms throughout as functional descriptions of model behavior—shorthand for the extraction and re-encoding of semantic content—not as claims about machine cognition. Nothing in this Article's argument turns on the resolution of that debate. *See* Emily M. Bender & Alexander Koller, *Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data*, *in* Proceedings of the 58th Annual Meeting of the Ass'n for Computational Linguistics 5185 (2020) (arguing that language models trained on form alone cannot achieve genuine understanding).

[44] *See infra* Part III (discussing NYT Exhibit J as evidence that depends on the AI *failing* to transform).

and then *rebuilds* expression from the learned idea, directed by a query. Between extraction and re-expression, there is a representational step: the AI has encoded the idea. The original expression is not shuffled or paraphrased—it is not retained as expression. New expression is generated from what was learned, not from what was read.[45]

A concrete illustration: a user asks an AI system, "Explain the holding of [a copyrighted Supreme Court opinion analysis] for a patent attorney." The system retrieves the article, identifies the elements relevant to patent practice, discards the original prose, and generates a memo in the register of patent advisory practice.[46] The memo conveys the article's analytical framework in language suited to its new audience. It serves a different purpose (practice, not scholarship), addresses a different audience (patent attorneys, not academics), and is written in a different register (advisory, not analytical). Access is unambiguous—the AI read the article in full. But the resulting memo reproduces no protectable expression from the original in text, structure, or organization. The author's expression has been consumed and discarded. Her ideas have been extracted and repackaged.

Not all works are equally susceptible to this process. This Article uses the term *compactibility* to denote the degree to which a work's market-relevant value can be extracted and re-expressed without retaining protectable expression.[47] A news article reporting factual findings is highly compactible: its value lies overwhelmingly in facts and analysis that can be fully extracted. A novel with a distinctive narrative voice is less compactible: its value lies partially in the specific expression. This gradient matters because the works most vulnerable to compaction are the works whose value lies in ideas—and those are the works the copyright system was designed to incentivize producing.[48]

---

[45] *See* Landes & Posner, *supra* note 22, at 326 (modeling the economics of expression production; the model assumes that producing new expression from the same ideas requires investment, an assumption that compaction defeats).

[46] This scenario is a straightforward application of the agentic RAG architecture. *See supra* note 41 and accompanying text. For real-world examples, *see supra* note 1.

[47] *See infra* Part IV (developing the investment-to-expression ratio and the categories of vulnerability).

[48] *See supra* Part I (establishing the implicit premise that protecting expression was sufficient to incentivize idea production); *see infra* Part IV (developing the economic consequences of breaking that premise).

B. From Compaction to Collage

The single-source scenario, while illustrative, understates the problem. In practice, AI systems do not compact one work at a time.[49] When a user issues a deep research query—"What is the current state of fair use doctrine as applied to generative AI?"—the AI system does not retrieve one source. It retrieves dozens, sometimes hundreds. Each source contributes some fragment: a case holding from one article, a policy argument from another, a statistical finding from a third, a historical observation from a fourth. The system extracts from each, discards each source's expression, and produces a single output—a research memo, a market briefing, a lesson module—that synthesizes all of them. The output is not derived from any single work. It is a *collage*: an assemblage of fragments drawn from many sources, reassembled into a new whole that serves the user's specific purpose.[50]

The scale of this assemblage is already large and growing. Google's AI Overviews, launched in 2024, synthesize information from multiple web sources into a single answer panel displayed above organic search results; by late 2024, Google reported that AI Overviews had reached over one billion users globally.[51] Perplexity AI reads web sources—including paywalled articles—and generates original-language answers that synthesize copyrighted material the user never sees, a practice that has prompted copyright litigation.[52] Google's NotebookLM converts uploaded documents—including copyrighted books and articles—into conversational podcast scripts in the voices of AI-generated hosts.[53] A single deep research query today may involve one hundred sources.[54] Tomorrow, a thousand or ten thousand.

---

[49] *See supra* note 1 and accompanying text.

[50] Some of what shapes the output comes from the model's training; some comes from what is recovered at inference. The provenance of any given sentence is, in practice, unknowable.

[51] *See* Hema Budaraju, *AI Overviews: About Last Week*, Google (May 30, 2024), https://blog.google/products/search/ai-overviews-update-may-2024/ [https://perma.cc/YEK5-CDCX]; Sundar Pichai, Alphabet Inc. Q3 2024 Earnings Call (Oct. 29, 2024) (stating that AI Overviews reached over one billion users globally).

[52] *See* Complaint, N.Y. Times Co. v. Perplexity AI, Inc., No. 25-cv-10106 (S.D.N.Y. filed Dec. 5, 2025); Nolan, *supra* note 1.

[53] *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].

[54] *See supra* note 1.

A collage is not a creation from nothing. It is an *interpolation* within the space of existing expressions. All frontier AI systems are foundationally statistical machines: they model the probability distribution over possible sequences of words (or pixels, or notes), learned from training data and conditioned on retrieved sources.[55] The output is a point in this learned distribution—a weighted combination of patterns the system has encountered. The AI assembles its output from fragments of learned patterns, much as a physical collage assembles an image from fragments of existing photographs. The difference is that the AI's fragments are sub-lexical—smaller than words, smaller than phrases—making the assemblage invisible to ordinary observation.[56]

C. What Compaction Is Not

To understand the specificity of this phenomenon, it helps to distinguish compaction from four superficially similar processes.

First, compaction is not summarization. Summarization is source-determined: given a document, a summarizer produces a shorter version that captures its principal content.[57] Compaction is query-determined: the same source produces different outputs depending on the user's purpose.[58] A summary of a judicial opinion contains the same holdings for any reader; a compacted output extracts different elements for a patent attorney than for a consumer protection advocate.

Second, compaction is not paraphrase. Paraphrase restates the same content in different words—the goal is fidelity to the original's meaning and scope. Compaction selectively extracts and

---

[55] *See* Tomas Mikolov et al., *Distributed Representations of Words and Phrases and their Compositionality*, *in* 26 Advances in Neural Info. Processing Sys. 3111 (2013) (demonstrating that meaning can be represented as high-dimensional vectors, where spatial proximity corresponds to semantic similarity); 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] (comprehensive overview of vector semantics and embeddings). For empirical evidence that generative models function as semantic interpolators within a learned latent space, producing outputs that are semantically human-like yet stochastically distinct from training sources, see Anirban Mukherjee & Hannah Hanwen Chang, Beyond Pairwise Comparisons: A Distributional Test of Distinctiveness for Machine-Generated Works in Intellectual Property Law (Jan. 2026) (unpublished manuscript), https://arxiv.org/abs/2601.18156 [https://perma.cc/HXH7-SM8D] [hereinafter Mukherjee & Chang, Distributional Distinctiveness].

[56] *See infra* Part III (developing the fundamental-units argument and the consequent dissolution of expression's boundaries).

[57] The output does not change depending on who reads it or why.

[58] *See supra* note 7 and accompanying text (defining compaction as query-directed extraction and re-expression).

recombines: it takes some elements, discards others, and may combine the extracted elements with material from other sources.[59] The output is not a restatement of the original; it is a purpose-built extraction from it.

Third, compaction is not a search snippet. A search snippet is a pointer: it routes the user *to* the source, preserving the economic link between access and payment.[60] Compaction routes the user *around* the source, delivering the value of the copyrighted work without referral. The search engine preserves the gateway function of expression; compaction erodes it.[61]

Fourth, compaction is not a derivative work adaptation. A film adaptation of a novel incorporates protectable expression from the original—characters, dialogue, plot structure.[62] A compacted output does not incorporate protectable expression; it strips the expression and retains only unprotectable ideas, facts, and methods.[63] Derivative work doctrine requires incorporation of expression. Compaction is defined by its elimination.

D. Why This Time Is Different

The impulse to extract and condense copyrighted material is as old as copyright itself. Abridgments—authorized condensations of books—were among the earliest copyright controversies in English law.[64] Digests and headnotes condensed judicial opinions into navigable indexes.[65]

---

[59] *See supra* Part II.A (describing compaction's four elements: retrieval, extraction, discarding, and re-expression).

[60] *See* Authors Guild, Inc. v. Google, Inc., 804 F.3d 202, 216–18 (2d Cir. 2015) (holding that search-engine snippets serve a transformative purpose by allowing users to identify relevant works); Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007) (finding that thumbnail images in search results serve as pointers to the original).

[61] *See supra* Part I (developing the gateway function concept and the economic premise on which it rests).

[62] *See* 17 U.S.C. §§ 101, 106(2) (2024) (defining "derivative work" as a work "based upon one or more preexisting works").

[63] *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); 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 applies only when "specific expression that is traceable to a particular work" is incorporated into the generated output).

[64] *See* Gyles v. Wilcox (1740) 26 Eng. Rep. 489 (Ch.) (addressing whether an abridgment of a copyrighted book constituted infringement—often cited as a precursor to the fair use doctrine).

[65] *See* West Publ'g Co. v. Mead Data Cent., Inc., 799 F.2d 1219, 1222–23 (8th Cir. 1986) (describing the key-number system as an organizational framework for classifying legal principles).

Search engines cached and snippeted copyrighted web pages.[66] Five technological disruptions—the printing press, the photocopier, the internet, file-sharing, search engines—each lowered the cost of reproducing expression.[67] None severed the bundling of ideas with expression.[68] In every case, the extraction operated within an economic system that channeled revenue—directly or indirectly—to the producers of the original works. The gateway was bent but never broken. Expression was always *static*: one fixed arrangement compared to another fixed arrangement. The enforcement architecture was built for this world.[69]

The closest historical analog to AI compaction is the human research assistant. A law firm hires an associate to read cases and prepare memos. A consulting firm hires an analyst to read reports and synthesize findings. In economic terms, these professionals do exactly what AI does: they read expression, learn the ideas, and re-express them for a specific purpose.[70] Yet the gateway survives. The firm buys the Westlaw subscription. The analyst reads the Bloomberg terminal. Someone paid for access to the expression, and the payment funds the production of the ideas embedded within it. The toll was collected.

Human beings have always read copyrighted works, extracted ideas, and produced new works.[71] This is not only legal—it is constitutionally essential. The idea-expression dichotomy exists precisely to permit it.[72] But the objection that "humans have always done this" confuses the

---

[66] *See supra* note 60 and accompanying text.

[67] *See generally* Landes & Posner, *supra* note 22, at 326 (modeling the incentive effects of changes in copying costs on creative production).

[68] The printing press lowered production cost but required purchasing the copy—the gateway held. The photocopier reproduced expression, triggering copyright—the gateway held. The internet made distribution free, but reading still required accessing expression—the gateway held. File-sharing copied expression—the MP3 *was* the song—and copyright responded. *See* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U.S. 913 (2005). Search engines indexed expression but routed users *to* the source—the gateway held. *See supra* note 60.

[69] *See generally* 1 Melville B. Nimmer & David Nimmer, Nimmer on Copyright § 1.01[B] (rev. ed. 2024) (tracing the expansion of copyright from books to maps, charts, photographs, motion pictures, sound recordings, and digital works).

[70] *See* Landes & Posner, *supra* note 22, at 332–33 (modeling how the price of copies funds the author's fixed costs of creation).

[71] *See Authors Guild v. Google, Inc.*, 804 F.3d 202, 212–13 (2d Cir. 2015) (observing that readers have always extracted ideas and information from copyrighted works).

[72] *See supra* notes 20, 19 and accompanying text.

legality of an activity with its economic consequences.[73] Human extraction was tolerable because human constraints preserved the economic equilibrium on which copyright's incentive structure depended. Four dimensions of difference transform the same activity into a structurally different phenomenon. *Scale*: a single AI system can compact every work in a corpus simultaneously, serving millions of users—millions of works processed simultaneously, not dozens per week.[74] *Residual demand*: the user who receives the compacted output has no reason to access the original—the economic transaction that funded idea production disappears.[75] *Gateway bypass*: the entire value proposition of AI answer engines is that the user does not need to visit the source.[76] *Cost*: negligible at the margin, making extraction accessible to everyone rather than only well-resourced institutions.[77] Each is a quantitative difference. Together, they produce a qualitative transformation: the equilibrium that made human extraction harmless depended on human constraints, and AI removes the constraints.

A technology that operates at this scale, eliminates residual demand for the original, bypasses the economic gateway entirely, and costs nothing at the margin does not merely accelerate what research assistants have always done. It erodes the economic equilibrium that made the idea-expression dichotomy a workable framework for compensating idea producers.[78] What makes compaction categorically different from every prior condensation technology is not the kind of activity but three features that no predecessor shared: *learning* (not copying—the AI processes and reconstructs rather than shuffling expression), *dynamic output* (not static—each output is one of millions, tailored to a specific query), and *negligible cost of re-expression.*[79] This is the first condensation technology that breaks the premise on which the dichotomy's protective function

[73] *See* Landes & Posner, *supra* note 22, at 326 (copyright's incentive structure depends on authors' ability to recoup fixed costs through the expression market).

[74] *See supra* note 51 and accompanying text (describing the scale of AI-generated summaries).

[75] *See supra* Part I (describing the bundling of ideas with expression).

[76] *See supra* notes 52–53 and accompanying text.

[77] *Cf.* Landes & Posner, *supra* note 22, at 327–28 (showing that when copying costs approach zero, the author's ability to recoup fixed costs collapses).

[78] *See supra* Part I (establishing the implicit premise that protecting expression was sufficient to incentivize idea production only when expression was expensive to produce).

[79] *See supra* Part II.A (defining the four elements of compaction).

rested—because it operates below the level at which copyright has purchase, at the level of ideas rather than expression.

III. Fluid Boundaries

Any expression, decomposed to its most fundamental units, is always unprotectable. A word is not copyrightable. An individual musical note is not copyrightable. A single pixel is not copyrightable. This is not controversial; it is axiomatic.[80] The building blocks of expression—the units from which all creative works are assembled—belong to the commons. At the level of fundamental units, there is nothing to protect.

Expression has always meant something more than its units. It has meant *aggregation*: a specific arrangement of words into sentences, sentences into paragraphs, paragraphs into arguments; a particular sequence of notes into melody, melody into composition; a particular composition of pixels into image, images into visual narrative.[81] Copyright protects the aggregation, not the units. A novel is not a pile of words; it is a structure. A symphony is not a collection of notes; it is an architecture. Copyright's promise is that an author's specific aggregation—the particular way she arranged the building blocks—is hers.[82] The aggregation is what makes expression protectable; without it, there is nothing but commons.

AI operates precisely at the aggregation level. It can disaggregate any existing expression into its fundamental units and reaggregate them into a new expression that retains the functional content of the original while constituting an entirely different arrangement.[83] Every word in an article can be swapped for a synonym, every sentence restructured, every paragraph reordered. The result is a new aggregation—formally a new "expression"—that conveys the same ideas, the

[80] *See Feist*, 499 U.S. at 349–50 (copyright does not protect "building blocks" of expression); 17 U.S.C. § 102(b).

[81] *See Feist*, 499 U.S. at 348 (copyright protects originality in "selection, coordination, [and] arrangement"); Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 58 (1884) (copyright protects the author's "original intellectual conception" expressed through creative choices).

[82] *See Feist*, 499 U.S. at 345 ("[T]he requisite level of creativity is extremely low; even a slight amount will suffice.").

[83] *See supra* note 55 and accompanying text.

same analysis, the same informational content, in a different arrangement of units. And the AI can do this not once but infinitely, producing as many alternative aggregations as desired, each differing from the original and from every other, each serving as a formally distinct "expression" of the same underlying ideas.[84]

This is not creation de novo. It is interpolation within the latent space—the mathematical representation of all expressions the model has learned.[85] The output space is continuous. The distance between any two expressions is measurable, and the AI can produce any expression that falls between them. A collaged output is a point in this continuous space, shaped by the query, influenced by training, conditioned on retrieval. It is not a leap into uncharted territory; it is a step to an adjacent point in a space already mapped by existing expressions.

The result is a shift from *static* to *dynamic* expression. Copyright's enforcement architecture was built for static-against-static comparison: one fixed expression (plaintiff's work) compared to another fixed expression (defendant's work).[86] With compaction, one side of the comparison has become dynamic—a process that generates an unbounded number of outputs, each a different point in a continuous space, each tailored to a different query.[87] A given output may be one of millions produced in a day. The comparison itself becomes intractable: what is the court comparing? Which variation? How similar is "too similar" when the space is continuous and the target is a moving process rather than a fixed object?[88]

The consequence for copyright is severe. "Protecting expression" now means protecting a specific aggregation when infinite alternative aggregations are available at negligible cost. This is the legal equivalent of holding an exclusive right to a particular grain of sand on an infinite beach.

---

[84] *See supra* notes 55 and accompanying text (describing the representational architecture that enables AI systems to separate content from form and generate multiple distinct expressions of the same underlying semantic content).

[85] *See supra* note 55 (describing how generative models represent data as high-dimensional vectors where spatial proximity corresponds to semantic similarity).

[86] *See supra* Part I (establishing that expression was always static—one fixed arrangement compared to another).

[87] *See supra* Part II.B (describing the scale of AI-generated collages—over a billion queries per day through Google AI Overviews alone).

[88] *Cf. Perfect 10*, 508 F.3d at 1165–68. In *Perfect 10*, static thumbnails had determinable provenance—each thumbnail corresponded to one photograph. Compaction eliminates this one-to-one mapping.

The right is real. It is enforceable. No one will ever need to infringe it—because adjacent grains, indistinguishable in function, are free for the taking. The protection is formally complete and functionally empty.[89] An author's copyright in her specific arrangement of words is undiminished. But an AI system that can produce a thousand alternative arrangements conveying the same ideas has no reason to use hers. The exclusive right remains; the exclusivity is illusory.

This reveals what has happened to the idea-expression boundary. Hand's abstractions test assumed that idea and expression occupied different cognitive layers—that there were, in the metaphor's own terms, layers to peel.[90] At the most concrete level, a work consists of its specific words; at a higher level, its structure; at a still higher level, its themes; and at the most abstract level, a general idea.[91] The court's task was to identify the boundary between protectable expression (the lower layers) and unprotectable idea (the upper layers). This framework presupposed that expression was a stable stratum—a layer that could be identified, compared, and protected. The fundamental-units argument reveals that expression is not a layer. It is a point in a continuous space of possible aggregations. AI can move to any other point in that space—shifting words, restructuring sentences, reordering arguments—and produce a new expression that is formally distinct but functionally identical. The boundary between "idea" and "expression" is not merely hard to draw, as Hand conceded. It dissolves when the technology operates at the level of individual units that are themselves unprotectable. The boundary has no purchase below the level of aggregation. And AI operates precisely at that level.

The outputs of compaction are thus *causally derived without protectable similarity*—and the law protects against expressive similarity, not mere causal derivation. This structural mismatch between the doctrinal test and the phenomenon manifests in every pathway to legal relief.[92] The

[89] *Cf.* Landes & Posner, *supra* note 22, at 326 (arguing that copyright's economic value depends on the cost of producing substitutes; when substitutes can be produced at negligible cost, the economic value of the exclusive right approaches zero even though the legal right remains intact).

[90] *See* Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930) (Hand, J.); *see also* Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960) (Hand, J.) (acknowledging that the test "inevitably" involves ad hoc judgments).

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

[92] *See* Mark A. Lemley, *How Generative AI Turns Copyright Upside Down*, 25 Colum. Sci. & Tech. L. Rev. 21, 39–44 (2024) (concurrent articulation: similarity fails as a proxy for copying when AI outputs vary widely from their

doctrine confronts a category of appropriation it was never designed to detect: works that are derived from specific sources but share no protectable expression with any of them.[93]

To show this concretely, consider a single recurring scenario. An AI answer engine with agentic RAG receives a user query: "What are the leading arguments for and against treating AI-generated inventions as patentable subject matter?" The system autonomously retrieves 84 sources, including a paywalled law review article by Professor X published in the *Yale Law Journal*. It reads Professor X's article in full, extracts her analytical framework—a four-part typology of patent eligibility theories—combines it with material from the other 83 sources, and generates a 4,000-word research memo.[94] The user never sees any of the 84 sources. The memo shares no protectable expression with any single one. Professor X's four-part typology is present in the memo's analysis but is expressed in different words, in a different structure, combined with frameworks from other sources, and attributed to no one.

The doctrine does not malfunction. It functions exactly as designed—and what it produces, applying each test, is the conclusion that the collage is not actionable.

Substantial similarity fails.[95] Under the *Krofft* test, the extrinsic prong compares specific protectable elements: Professor X's article and the AI memo share none—different structure, different organization, different prose, different register.[96] They share *ideas*—the four-part typology—but the extrinsic test, properly applied, filters ideas out. The intrinsic prong finds no similarity in total concept and feel.[97] Under the *Altai* framework, abstraction decomposes the article into

---

sources).

[93] This structural mismatch is most acute for high-IER works—journalism, textbooks, legal analysis—where the market value resides primarily in unprotectable ideas. For works whose market depends on the expression itself, collage may be less able to strip all similarity. *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).

[94] This scenario is a straightforward application of the agentic RAG architecture described in Part II. *See supra* note 41 and accompanying text. For real-world examples, *see supra* note 1.

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

[96] 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).

[97] *See Krofft*, 562 F.2d at 1164.

layers; filtration strips ideas, facts, and scènes à faire; comparison finds nothing remaining.[98] The test works. The result is that Professor X's analytical framework—the product of months of research—passes into the memo without legal consequence.[99] The strongest available evidence strategy—the kind of verbatim overlap demonstrated in *NYT v. Microsoft*'s "Exhibit J"—depends on the AI *failing* to transform.[100] Collage, by definition, does not regurgitate.[101]

The derivative work right fails.[102] Section 106(2) requires incorporated expression. *Litchfield* established that a work is not an infringing derivative unless it "incorporates a portion of the copyrighted work in some form."[103] Professor X's memo contains her typology (an idea), case holdings (facts), and policy arguments (ideas). None of her expression appears.[104] Reading Section 106(2) broadly enough to reach causal derivation even without incorporated expression would extend copyright to ideas—what Section 102(b) forbids and *Baker v. Selden* was designed to prevent.[105]

The intermediate copying double bind forecloses both procedural and substantive paths.[106] The AI unambiguously copies Professor X's article when it retrieves it. But proving this requires discovery, which requires surviving a motion to dismiss under *Iqbal*.[107] Without output similarity pointing back to her work, Professor X cannot plausibly allege that the system compacted *her* article during *that* session. The claim cannot be *initiated* without the very similarity that collage eliminates.[108] And even if the procedural barrier is cleared, the intermediate copy may be fair use.

---

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

[99] *See also Nichols*, 45 F.2d at 121 (recognizing that as a work is abstracted to higher levels of generality, protection eventually ceases).

[100] *See* Mukherjee & Chang, Distributional Distinctiveness, *supra* note 55 (discussing the reliance on verbatim overlap metrics in current AI copyright litigation).

[101] *See supra* note 6.

[102] 17 U.S.C. §§ 101, 106(2) (2024).

[103] Litchfield v. Spielberg, 736 F.2d 1352, 1357 (9th Cir. 1984).

[104] *See Altai*, 982 F.2d at 701; *see also* Bracha, *supra* note 63, at 384 (arguing that the derivative work right applies only when "specific expression that is traceable to a particular work" is incorporated).

[105] *See supra* notes 19, 20 and accompanying text; *see also Feist*, 499 U.S. at 349–50.

[106] *See* MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511, 518 (9th Cir. 1993).

[107] Ashcroft v. Iqbal, 556 U.S. 662, 678 (2009); *see also* Bell Atl. Corp. v. Twombly, 550 U.S. 544, 570 (2007).

[108] Expression similarity is not merely the legal test for infringement—it is the evidentiary predicate that initiates the litigation process.

Sega v. Accolade held that intermediate copying to extract unprotectable functional specifications was fair use.[109] *Google v. Oracle* held that copying API declarations—expression serving as an "interface to ideas"—was fair use.[110] Both treated expression as instrumental: valuable not in itself but as a pathway to something else. Collage generalizes this logic: it treats *all* copyrighted expression as an interface to the ideas, analysis, and information the work contains.[111] In *Bartz v. Anthropic*, the court held that training an AI model on lawfully obtained books was "spectacularly transformative."[112] If training-time ingestion is transformative, inference-time collage—where a work is re-expressed in an entirely different form—is arguably more so.[113] Without similarity, the claim cannot start. With evidence of copying, the copying may be fair use. Both paths are foreclosed.

The volitional conduct gap removes the final avenue.[114] The user issued a generic query naming no copyrighted work. The AI autonomously retrieved Professor X's article among 83 others. The user never learned which works were accessed. The AI is not a legal person capable of volitional conduct.[115] The chain of volition is fatally attenuated.[116] Secondary liability fares no better: contributory infringement requires knowledge of *specific* infringing activity;[117] the *Sony*

[109] Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1527–28 (9th Cir. 1992).

[110] Google LLC v. Oracle Am., Inc., 593 U.S. 1, 25–27, 35 (2021).

[111] *Cf.* Jacqueline C. Charlesworth, *Generative AI's Illusory Case for Fair Use*, 26 Vand. J. Ent. & Tech. L. 323 (2025) (arguing that intermediate-copying precedents are inapplicable to generative AI because such systems exploit intrinsic expressive value).

[112] Bartz v. Anthropic, PBC, No. 24-cv-05417, slip op. at 11 (N.D. Cal. June 23, 2025).

[113] The *Kadrey* court's recognition that AI outputs could inflict cognizable "dilution" harm suggests a countervailing principle—but even that court found for the defendant on fair use. *See* Kadrey v. Meta Platforms, Inc., No. 23-cv-03417, slip op. at 29 (N.D. Cal. June 25, 2025).

[114] Copyright's exclusive rights are implicated only by the "volitional act" of a person who "causes in some meaningful way an infringement." CoStar Grp., Inc. v. LoopNet, Inc., 373 F.3d 544, 550 (4th Cir. 2004).

[115] *See supra* note 6; *see also* Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1370 (N.D. Cal. 1995); Cartoon Network LP v. CSC Holdings, Inc., 536 F.3d 121, 131 (2d Cir. 2008).

[116] *See CoStar*, 373 F.3d at 550; *cf. Netcom*, 907 F. Supp. at 1370. For the most comprehensive scholarly treatment of this gap, *see* Aleksander J. Goranin, *A Deep Look at Copyright's Volitional Conduct Doctrine and Generative Artificial Intelligence*, 74 Emory L.J. 1127, 1145–62 (2025) (tracing the doctrine from its RAM-copy origins through proximate causation and evaluating classic analogies against generative AI). *But see* Mala Chatterjee & Jeanne C. Fromer, *Minds, Machines, and the Law: The Case of Volition in Copyright Law*, 119 Colum. L. Rev. 1887, 1920–32 (2019) (arguing that machines' functional behavior may satisfy the volition requirement and that the law may treat system operation as the firm's volitional conduct).

[117] *See* Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc., 443 F.2d 1159, 1162 (2d Cir. 1971).

defense—that a technology capable of "substantial noninfringing uses" should not give rise to contributory liability—applies with considerable force; *Grokster*'s inducement exception requires affirmative steps to foster infringement that the generic-query scenario does not present.[118]

The multi-source structure creates one further problem. When 84 sources each contribute fragments to a single output, the collage is a *hydra*: it has many heads and no body.[119] Each head is individually non-infringing. The hydra as a whole substitutes for the market value of all 84 sources combined. Copyright, structured as a system of individual rights, has no mechanism for addressing it.[120] The evidentiary hydra: no single source's expression is identifiable in the result—provenance is practically unknowable, because the output is a statistical interpolation across all sources.[121] The *de minimis* hydra: each extraction falls below the threshold—a single analytical framework from one of 84 sources, re-expressed in new words, is not a substantial taking.[122] The *de minimis* threshold exists for good reason—it prevents copyright from monopolizing building blocks. Collage ensures that every taking falls below it. The attribution hydra: the causal chain from any individual work to any measurable market harm is hopelessly attenuated—diffuse across 84 sources per query, cumulative across thousands of queries, and practically unattributable.[123] Fair use's fourth factor—"the effect of the use upon the potential market for or value of the copyrighted work"[124]—is structurally inadequate: the harm is diffuse, cumulative, unprovable, and offset by findings of transformative purpose.[125] A structural problem cannot be solved by retail adjudication.

The fair use framework's treatment of "different markets" deepens the problem. In *Cariou*

[118] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 442 (1984); *see also Grokster*, 545 U.S. at 936–37 (requiring evidence that the distributor acted with the object of promoting infringement); *supra* note 68.

[119] *See supra* Part II.B (introducing the collage phenomenon).

[120] *See* 17 U.S.C. § 501(a) (2024) (defining infringement as violation of the rights of "the copyright owner").

[121] *Cf.* Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165–68 (9th Cir. 2007). In *Perfect 10*, each thumbnail image corresponded to one source image; provenance was trivial to establish. Collage eliminates this one-to-one mapping.

[122] *See* Ringgold v. Black Entm't Television, Inc., 126 F.3d 70, 74–75 (2d Cir. 1997).

[123] *See Perfect 10*, 508 F.3d at 1168 (finding that "potential harm to [the plaintiff's] market remains hypothetical" even when source-level provenance was clear).

[124] 17 U.S.C. § 107(4) (2024); *see Harper & Row*, 471 U.S. at 566 (describing the market effect as "undoubtedly the single most important element of fair use").

[125] *See supra* Part III (demonstrating the Cariou-Warhol tension in the transformative use framework).

v. Prince, the Second Circuit held that appropriation art was transformative fair use because the works had a "different character" and served a "different market."[126] Applied to collage, Cariou's logic is extraordinarily favorable to AI systems:

Original Work	Collaged Output	Different Market?
Law review article	Client memo	Yes: scholarship vs. practice
Newspaper article	Market briefing	Yes: news vs. finance
Textbook	Lesson module	Yes: reference vs. pedagogy
Investigative report	Policy summary	Yes: journalism vs. government
Novel	Podcast script	Yes: literary fiction vs. audio

Compacted outputs are *designed* to serve different functions—that is the definition of query-directed extraction.[127] Cariou's test is satisfied structurally, not incidentally. Before AI, the escape hatch was narrow: one artist, limited physical collages. AI removes the constraint entirely—unlimited outputs, unlimited formats, unlimited markets, at negligible cost per output.[128] The escape hatch that was a crack in the doctrine becomes, at scale, the doctrine's dominant feature.[129] Cariou was decided against the backdrop of positive marginal cost of expression. When expression costs nothing, the "different market" finding is trivially achievable for every output.

Cariou reveals something deeper than a doctrinal loophole. It shows that the idea-expression dichotomy is not ontological but market-contingent. The same intellectual content—an analytical framework, a factual finding, a legal argument—counts as "protected expression" when embodied in its original market (the law review article) but becomes an extractable "idea" when re-expressed for a different market (the client memo).[130] If the idea-expression distinction is market-contingent,

[126] Cariou v. Prince, 714 F.3d 694, 706–07 (2d Cir. 2013). The transformative use inquiry descends from *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).

[127] *See supra* Part II.A (defining compaction as query-directed extraction and re-expression).

[128] *See supra* Part I (describing the economic premise that made expression expensive to produce).

[129] *Cf. Perfect 10*, 508 F.3d at 1165 (finding transformative purpose in an information retrieval context).

[130] *Cf. Baker v. Selden*, 101 U.S. at 104–05 (holding that the "use" of a system described in a copyrighted work is

and market-specific outputs can be generated on demand, then the distinction itself becomes manipulable—a function of the AI system's output parameters rather than a property of the underlying work.[131]

The Supreme Court recognized this problem, in part, in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*.[132] Warhol narrowed Cariou: when a secondary work serves "substantially the same purpose" as the original, the first fair use factor weighs against the defendant regardless of expressive alteration.[133] But applying the "same purpose" test to collage requires identifying the "ideas" the output shares with its sources—the very operation Hand conceded "nobody has ever been able" to perform.[134] Warhol's correction is necessary but insufficient. It identifies the right concern—market substitution despite formal transformation—but provides no stable test for resolving it in the collage context, where multi-source synthesis has no single original against which to measure "same purpose."[135]

The barriers are *sequentially compounding*: failure at any stage is individually dispositive, and Professor X must clear each one to reach the next.[136] Each test is a gate. Each gate is locked. And each lock engages as intended—not stuck, not broken, not in need of repair. The doctrine finds that there is no substantial similarity. The doctrine finds that no protectable expression is incorporated. The doctrine finds that intermediate copying cannot be proved without output similarity. The doctrine finds that no human performed the volitional act. The doctrine finds that

---

free, though the "explanation" is protected—a distinction that presupposes stable boundaries between use and explanation).

[131] *See supra* Part II.B (describing how collage generates novel outputs for novel markets from existing source material).

[132] Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 598 U.S. 508, 541–42 (2023).

[133] *See Warhol*, 598 U.S. at 532–33.

[134] *Nichols*, 45 F.2d at 121; *see Warhol*, 598 U.S. at 551–54 (Kagan, J., dissenting) (arguing that the majority's "purpose" analysis is indeterminate at the applicable level of generality).

[135] Warhol assumed a bilateral relationship—one secondary work, one original. Collage is radically multilateral. *See* Jane C. Ginsburg, *Fair Use in the US Redux: Reformed or Still Deformed?*, [2024] Sing. J.L.S. 52, 64–71 (documenting the *Oracle*/*Warhol* doctrinal whiplash and arguing that transaction-cost justifications should not override market-substitution concerns).

[136] To prevail, Professor X must: (1) demonstrate substantial similarity—she cannot; (2) show incorporated expression as a derivative work—it is absent; (3) survive a motion to dismiss on intermediate copying without output similarity—she cannot plead it; (4) identify a human who performed the volitional act—no one did; and (5) overcome the hydra's evidentiary, *de minimis*, and attribution barriers—each is independently fatal.

each individual extraction is *de minimis*.[137] Each test asks its intended question and returns its intended answer. The collective result: systematic market substitution without triggering any enforcement mechanism the law provides. The formal right is intact. The practical protection is empty.[138]

This is not a gap to be patched by doctrinal refinement. It is the letter of copyright law operating as written while defeating the law's animating purpose.[139] The fluid boundaries described in this Part are the doctrinal manifestation of a deeper economic transformation—to which we now turn.

IV. The Economics of Change

Copyright's implicit bargain rested on a specific economic structure.[140] Expression was the scarce, costly artifact—the product of human creative labor—and ideas were the abundant commons on which future creators could freely build.[141] Protecting expression while leaving ideas free struck a reasonable balance because expression was the *gateway* to ideas.[142] To access the ideas in a newspaper article, you had to read the article. To access the analysis in a treatise, you had to buy the treatise.[143] Copyright controlled the expression; the author captured value from both the expression and the ideas because the bundle was indivisible. This was not an accident of technology. It was a structural feature of the information economy that persisted from Gutenberg

[137] Each finding aligns with the doctrine's intent, yet each is contestable. On substantial similarity: Professor X's four-part typology might constitute protectable selection and arrangement under *Feist*, 499 U.S. at 348, though filtration would likely strip it as an unprotectable analytical method. On incorporated expression: the claim assumes successful compaction; where AI systems fail to transform—as in the verbatim overlaps at issue in *NYT v. Microsoft*—protectable expression does appear in the output. *See supra* note 6. On intermediate copying: emerging disclosure regimes, *see* Regulation (EU) 2024/1689, art. 53(1)(d) (AI Act transparency obligations), may eventually bypass the output-similarity predicate, though none currently does so routinely. On volitional conduct: *see supra* note 116 (arguing that functional system behavior may satisfy the volition requirement); this Article assumes current doctrine. On *de minimis*: *see supra* note 122; the threshold can be crossed even for brief uses, but multi-source collage structurally ensures each individual taking falls below it.

[138] *See infra* Conclusion (developing the impasse).

[139] U.S. Const. art. I, § 8, cl. 8.

[140] *See* Landes & Posner, *supra* note 22, at 326–27.

[141] *Id.* at 326.

[142] *See supra* Part I.

[143] *See Harper & Row*, 471 U.S. at 558.

through Google.[144]

Collage inverts the scarcity. Ideas are now the expensive element—journalism costs money to produce, research requires years of work, legal analysis demands expertise.[145] Expression is the free element—AI regenerates it at negligible cost from any combination of fundamental units, in any format, for any purpose.[146] Copyright protects the cheap thing—any particular aggregation of words, notes, or pixels—and leaves the expensive thing—the ideas that required real investment to produce—unprotected. The law has not changed. The economics have. And the law's effectiveness depended on the economics.[147]

Three conditions, all now satisfied, produce pro forma protection. *First, the malleability condition*: expression can be regenerated at will from fundamental units.[148] Because generative AI can produce an effectively infinite number of distinct expressions from the same underlying ideas, no one needs to copy the author's expression. *Second, the extraction condition*: ideas can be extracted at inference time from copyrighted sources.[149] The AI system retrieves the work when a user issues a query, processes it, and extracts the relevant ideas in real time. *Third, the enforcement condition*: similarity-based enforcement cannot detect the extraction.[150] The compacted output reproduces none of the original's protectable expression. Similarity avoidance is not a side effect; it is the operational definition of successful collage.

The market for information-rich works operates through a production chain with four links. First, *investment*: publishers, universities, and research institutions fund the production of ideas—paying reporters, researchers, analysts. Second, *embedding*: the ideas are embedded in copyrighted expression. Third, *distribution*: the expression is distributed through markets—subscriptions, book sales, database access. Fourth, *revenue*: revenue from distribution flows back to fund further

[144] *See supra* Part II.D (tracing the gateway function through five prior technological disruptions).

[145] *See supra* Part I (establishing the V_i/V_e decomposition).

[146] *See supra* Part III (demonstrating that infinite alternative aggregations are available at negligible cost).

[147] *Cf.* Landes & Posner, *supra* note 22, at 326 (arguing that copyright's economic value depends on the cost of producing substitutes).

[148] *See supra* Part III.

[149] *See supra* note 41 and accompanying text.

[150] *See supra* Part III.

investment.[151] Collage short-circuits this chain at the third link. If ideas can be extracted from expression and recombined into collages without going through the distribution market, revenue does not flow back to fund investment. The chain breaks not because anyone violates copyright but because the economic link through expression has been severed.[152]

Many information markets depend on cross-subsidization: the economic value of ideas is captured through bundled access to expression.[153] A Bloomberg terminal subscription funds not just data delivery but the analytical journalism that gives the data context. A Westlaw subscription funds not just access to case law but the editorial enhancements that make the law navigable. Collage enables selective extraction of the ideas without paying for the bundle. An AI system can read the Bloomberg analysis, extract the market-moving insight, and deliver it to a user with no Bloomberg subscription.[154]

The vulnerability depends on where the author's investment resides—in the ideas or in the expression. Four categories illustrate the high-IER landscape.[155] *Investigative journalism*: the investment is months or years of reporting; the expression is professionally competent but fungible—an AI can deliver the findings in a purpose-built briefing. *Academic textbooks*: the investment is years of scholarship; an AI can compact a textbook into a customized study guide. *Legal analysis and treatises*: the investment is deep legal research; an AI can read a treatise and produce a memo applying its analysis to a specific question. *Market research*: the investment is data collection and quantitative analysis; AI systems are already extracting and re-expressing market intelligence from paywalled research.[156]

This creates a deep irony. Copyright provides the *least* functional protection to the works that need it *most*—high-IER works: journalism, textbooks, legal analysis, expert research—and the

[151] *See* Landes & Posner, *supra* note 22, at 326–29 (modeling the feedback loop between distribution revenue and investment in creative production).

[152] *See supra* Part I (establishing that the bundling of ideas with expression is the mechanism through which copyright incentivizes idea production).

[153] *See supra* Part I.

[154] *See supra* note 1 and accompanying text.

[155] *See supra* note 31 and accompanying text.

[156] *See supra* Part II (defining compaction and its four elements).

most functional protection to the works that need it *least*—works whose expression is harder to substitute.[157] The Copyright Clause empowers Congress to "promote the Progress of Science and useful Arts."[158] The works most vulnerable to collage—journalism, scholarship, research, expert analysis—are the "Science" works the Clause was designed to promote.[159] Copyright's formal architecture fulfills the constitutional command. The functional result defeats it.

The institutional consequences are already visible. Copyright deters unauthorized copying through the threat of statutory damages, injunctive relief, and attorneys' fees.[160] This deterrence presupposes that unauthorized use is *detectable*. When extraction is undetectable—when the collaged output reproduces no protectable expression from any source—the deterrence mechanism fails.[161] The formal penalties deter nothing, because they cannot be triggered. Deterrence without detection is an empty threat.

Copyright enables voluntary licensing markets. These licenses depend on the licensor's ability to withhold access—to make the expression a scarce good whose distribution the licensor controls.[162] Collage undermines this leverage. If an AI system can retrieve a work at inference time, extract its ideas, and deliver them in regenerated expression—all without authorization—the inference-time licensing market collapses.[163] The licensor's product (access to expression) is rendered worthless by a technology that bypasses the expression entirely.

Copyright also functions as an investment signal: it tells publishers and creators that their intellectual investments will be protectable.[164] When that signal weakens, investment decisions shift at the margin. Publishers invest less in expensive idea production and more in compaction-resistant content: entertainment, personality-driven work, experiential offerings—content whose value *is*

---

[157] *See Harper & Row*, 471 U.S. at 558 (copyright as "engine of free expression"); *see also supra* Part I.

[158] *See supra* note 139.

[159] *See supra* Part I.

[160] *See* 17 U.S.C. §§ 502–505 (2024).

[161] *See supra* Part III (demonstrating the enforcement gap).

[162] *See Harper & Row*, 471 U.S. at 546–47.

[163] *See supra* Part I.

[164] *See* Landes & Posner, *supra* note 22, at 326; *see also supra* note 139 and accompanying text.

its expression rather than its ideas.[165] The market, responding rationally to the new incentive structure, produces less of what is socially most valuable and more of what is commercially most defensible.

The objection that this is merely creative destruction—that every major information technology has displaced incumbent business models—conflates *competition* with *extraction*.[166] Competition produces new value: a competing textbook offers a different pedagogical approach; a competing news outlet covers the same events with different reporters. Collage does not produce new ideas. It extracts existing ideas without compensating their producers and re-expresses them at negligible cost.[167] The encyclopedia analogy fails because the encyclopedia invested in editorial synthesis. The search engine analogy fails because the search engine preserved the gateway.[168] AI collage does neither. It takes the encyclopedia's synthesis role without the encyclopedia's investment, and the search engine's access to sources without the search engine's referral function. It is a mechanism that captures the value of idea production while externalizing its cost.

The institutional response is already visible, and it is telling that it has not come through copyright. Publishers have turned to competition law: the European Publishers Council's February 2026 antitrust complaint against Google frames AI-generated search summaries as an abuse of market dominance, not a copyright infringement—an implicit concession that copyright's similarity-based architecture cannot do this work.[169] Collective licensing proposals seek compensation at the retrieval layer rather than the output layer. Retrieval-layer regulation—requiring AI systems to log, license, or compensate for copyrighted sources accessed at inference time—targets the point of extraction rather than the point of output.[170] Each of these responses implicitly acknowledges the pro forma condition: copyright's output-policing architecture has been bypassed, and the

---

[165] *See supra* Part I (identifying the investment-to-expression ratio as the measure of vulnerability to compaction).

[166] *See* Landes & Posner, *supra* note 22, at 332–33 (distinguishing between competitive creation, which increases the supply of ideas, and free riding, which decreases the incentive to produce them).

[167] *See supra* Part II (defining compaction as extraction and re-expression of ideas from copyrighted expression).

[168] *See supra* note 60.

[169] *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 [https://perma.cc/YA7Y-9XSG].

[170] *See supra* note 41 and accompanying text.

institutional response has moved elsewhere.

Conclusion

Copyright was designed for a world in which expression was a stable, scarce, costly artifact—an aggregation of individually unprotectable elements (words, notes, pixels) into arrangements that only human labor could produce.[171] AI systems access dozens or even hundreds of copyrighted works at inference time, learn the ideas embedded in them, and dynamically reconstruct those ideas in new expression without reproducing protectable expression. These outputs are causally derived without protectable similarity, and the law polices similarity, not derivation.

The formal architecture of copyright—its doctrines, its tests, its remedies—is complete and operates as intended. And yet the result, for the works that most need protection, is systematic market substitution that copyright law is structurally incapable of reaching: the law protects what is now cheap and leaves unprotected what is now valuable.[172] When AI can disaggregate existing expression into its fundamental units and reaggregate them at negligible cost, "expression" ceases to demarcate a stable category that the law can protect against recombination.[173]

This Article uses the term *pro forma* to describe this condition: a legal right that exists formally in the statute books, is formally invocable in litigation, and is constitutionally valid—but provides no functional protection for the interest it was designed to serve.[174] The author retains the exclusive right to control her expression. No one has copied her expression into the output. No derivative work incorporates it. Her copyright is fully intact. And yet the economic value of her work—the ideas, the analysis, the investigation—has been extracted, re-expressed, and delivered to consumers who never encountered her expression and never will. The right is real. It is enforceable. No one will ever need to infringe it.

---

[171] *See supra* Part I.

[172] *See supra* Part IV.

[173] *See supra* Part III.

[174] *See* 17 U.S.C. § 106 (2024).

Not all works are equally vulnerable. Works valued primarily for their expression—novels, poetry, music—retain substantial protection because their value depends on specific expressive choices that cannot be fully extracted through compaction.[175] The works for which copyright becomes pro forma are those whose value lies primarily in ideas embedded within fungible expression: journalism, scholarship, legal analysis, market research. If AI outputs reproduce protectable expression—verbatim passages, closely paraphrased structures—existing doctrine reaches those cases.[176] But these are failure modes of immature compaction; as the technology matures, the zone of unreachable extraction grows while the zone of actionable reproduction shrinks.[177]

The pro forma thesis is distinct from the claim that copyright has a *gap* that can be patched. A gap implies that the existing framework is structurally sound and merely needs supplementation—a new statutory provision, a revised fair use factor, a broader definition of derivative work. The pro forma thesis holds that the problem is not a gap but a structural condition. The doctrinal tests are not producing wrong answers that could be corrected by amending a statutory provision.[178] They are producing *intended* answers that collectively yield the wrong outcome. The distinction matters because the response to a gap is a patch, while the response to a structural condition requires rethinking which element of creative works the law protects—a far more fundamental inquiry.

Nor is the thesis that copyright is *obsolete* and should be replaced. Copyright continues to function effectively for traditional forms of appropriation—verbatim copying, unauthorized reproduction, piracy, close paraphrase.[179] The pro forma claim is narrower: copyright fails *specifically* where collage-based extraction is involved—where the extraction is of ideas rather than expression, where the output serves a different purpose in a different market, and where no

[175] *See supra* Part II (defining compactibility as the gradient of vulnerability).

[176] *See supra* Part III (discussing NYT Exhibit J and the verbatim-overlap evidence strategy).

[177] *See supra* Part III.

[178] *See supra* Parts III–IV.

[179] *See supra* Part II (distinguishing compaction from paraphrase and derivative work adaptation).

expressive similarity connects source to output.[180]

The structural inversion might seem to suggest that copyright should be extended to protect ideas. But this is foreclosed—and rightly so—by the deepest principles of the copyright system. *Baker v. Selden* established that methods and systems cannot be monopolized through copyright.[181] Section 102(b) codifies the exclusion.[182] The freedom to use ideas is essential to cumulative progress—the very "Progress of Science and useful Arts" the Copyright Clause invokes.[183] Protecting ideas would create monopolies over facts, methods, and concepts.[184] Ideas must remain free. Any regime that attempted to protect them would chill the cumulative progress that copyright is meant to promote.[185]

But the foreclosure goes deeper than policy. If copyright's enforcement machinery cannot hold onto expression—a relatively concrete, identifiable, bounded legal category—it certainly cannot hold onto ideas, which are definitionally more amorphous, more fluid, more resistant to legal boundaries.[186] The same technology that defeated expression-protection would defeat idea-protection even more thoroughly. Ideas have even fewer stable boundaries than expression. Every argument in Part III—that AI can disaggregate and reaggregate at will, that the output space is continuous, that similarity-based enforcement cannot track a dynamic target—applies with greater force to ideas than to expression.[187]

The result is a genuine impasse. The intended state of the idea-expression dichotomy was protected expression and unprotected ideas. The emerging state is unprotected expression and unprotected ideas. For works vulnerable to compaction, everything is unprotected.

---

[180] *See supra* Parts III–IV.

[181] *See supra* note 20.

[182] *See supra* note 19.

[183] *See supra* note 139.

[184] *See Baker*, 101 U.S. at 103; *see also Feist*, 499 U.S. at 349–50.

[185] *See supra* notes 20, 19.

[186] *See supra* Part III (demonstrating that the boundaries of expression dissolve when the technology operates at the level of fundamental units).

[187] *See supra* Part III.

The dichotomy has not broken down as a principle. The world has changed around it.[188] Copyright has become the wrong answer to a new problem while remaining the right answer to an old problem—the static copying of expression rather than the dynamic reconstruction of expression from ideas.

The impasse is not a rhetorical choice to avoid prescription. It is a logical consequence of the diagnosis. Expression cannot be effectively protected against a technology that operates at the sub-expressive level.[189] Ideas cannot and should not be protected.[190] And yet, as the cost of translating ideas into expression falls, ideas become relatively more valuable—they are increasingly what we would want to incentivize.[191] The refusal to protect ideas becomes a functional refusal to protect *anything*—for the category of works whose production requires the greatest investment.

This Article does not propose to resolve this impasse. The diagnosis is itself the contribution. To mistake a structural condition for a patchable gap is to invite reforms that address symptoms while leaving the underlying inversion intact—broader derivative work definitions, enhanced statutory damages, mandatory licensing—none of which addresses the core problem that the element copyright protects (expression) is no longer the element that carries economic value (ideas).[192] Before the law can respond, it must understand what it is responding to.

---

[188] *See supra* Part I (establishing the implicit economic premise on which the dichotomy's protective function rested).

[189] *See supra* Part III.

[190] *See supra* notes 20, 19 and accompanying text.

[191] *See supra* Part IV (tracing the economic inversion and the investment signal failure).

[192] *See supra* Parts III–IV.