

PLAGIARISM, COPYRIGHT, AND AI

Mark A. Lemley* & Lisa Larrimore Ouellette†

Critics of generative AI often describe it as a “plagiarism machine.” They may be right, though not in the sense they mean. With rare exceptions, generative AI doesn’t just copy someone else’s creative expression, producing outputs that infringe copyright. But it does get its ideas from somewhere. And it’s quite bad at identifying the source of those ideas. That means that students (and professors, and lawyers, and journalists) who use AI to produce their work generally aren’t engaged in copyright infringement. But they are often passing someone else’s work off as their own, whether or not they know it. While plagiarism is a problem in academic work generally, AI makes it much worse because authors who use AI may be unknowingly taking the ideas and words of someone else.

Disclosing that the authors used AI isn’t a sufficient solution to the problem because the people whose ideas are being used don’t get credit for those ideas. Whether or not a declaration that “AI came up with my ideas” is plagiarism, failing to make a good-faith effort to find the underlying sources is a bad academic practice.

We argue that AI plagiarism isn’t—and shouldn’t be—illegal. But it is still a problem in many contexts, particularly academic work, where proper credit is an essential part of the ecosystem. We suggest best practices to align academic and other writing with good scholarly norms in the AI environment.

Introduction

Imagine a student who asks ChatGPT for a “novel thesis for a law school seminar paper on a fresh approach to calculating patent damages.” Seconds later the screen offers: “Corporations often transfer patents to foreign subsidiaries in tax havens, and experts should rely on these transfer prices when valuing patents for damages purposes.” The student writes a paper based on this thesis, verifies their citations, and feels proud of their creative work. They fail to discover, however, that Jennifer Blouin and Melissa Wasserman made precisely that argument in a [2018 symposium essay](#), which was in ChatGPT’s training dataset.³ Blouin and Wasserman might be justifiably annoyed

* William H. Neukom Professor of Law, Stanford Law School; partner, Lex Lumina LLP.

† Deane F. Johnson Professor of Law, Stanford Law School. Thanks to Brian Frye, James Grimmelmann, Rose Hagan, Matthew Sag, Pam Samuelson, and Jessica Silbey for comments on an earlier draft. © 2025 Mark A. Lemley and Lisa Larrimore Ouellette.

³ Blouin and Wasserman build on (and cite) earlier work, including [Andrew Blair-Stanek](#).

at this unattributed copying of their idea, particularly if the student publishes their paper as a Note. Should they have any recourse?

[Thousands](#) of pages of judicial pleadings and law review commentary have been devoted to whether generative AI systems infringe authors' copyrights in the works that are bulk ingested during model training. But in our hypothetical, the copyright answer is clear: the student used only Blouin and Wasserman's *idea*, not their way of expressing it. And [ideas aren't copyrightable](#).

There is a problem, however, with copying ideas from AI output. What the student has committed is *plagiarism*—the appropriation of another's work and insights without acknowledgement. Courts and litigants sometimes conflate copyright infringement and plagiarism, perhaps because they often occur together, such as when an extensive passage is paraphrased without a footnote. But they are distinct problems. Excerpt an entire chapter with attribution and you've infringed but not plagiarized; set forth someone's idea as your own and you've plagiarized but not infringed. This distinction is important. Copyright infringement is a [legal wrong](#) that can make you liable for enormous [damages](#), [attorneys' fees](#), and even [jail time](#). Plagiarism, by contrast, isn't illegal, though it can lead to academic punishment, and its reputational effects might cost you [your job](#) or [delay](#) your becoming President of the United States by decades. Despite this distinction, the legal literature has barely considered how AI raises plagiarism concerns that are distinct from copyright.⁴

Carefully thinking about AI-facilitated plagiarism is important because many different institutions are struggling with how to regulate AI-based copying. [Legislatures and courts](#) are considering what kinds of generative AI uses should trigger a legal cause of action. Universities and journals are scrambling to [rewrite](#) honor codes and policies regarding AI use. Law schools and law reviews have [yet to settle](#) on standard norms for writing with AI. And yet AI use for academic work by both professors and students is now [pervasive](#).

In this Essay, we provide a roadmap for how these new governance schemes should address plagiarism concerns, disentangling the distinct harms at issue. Some generative AI

⁴ On plagiarism with generative AI in the courts, see Amy B. Cyphert, *Generative AI, Plagiarism, and Copyright Infringement in Legal Documents*, 25 MINN. J.L. SCI. & TECH. 49, 56–59 (2024). Some law reviews have begun to consider AI policies but have not recognized this plagiarism concern. See Nachman Gutowski, *Disclosing the Machine: Trends, Policies, and Considerations of Artificial Intelligence Use in Law Review Authorship*, JACKSONVILLE U. L. REV. (forthcoming 2025).

practices relate to copyright’s legal protection of economic incentives for creating new works—and the fair-use exception that supports valuable follow-on creations. An overlapping but distinct set of practices relate to plagiarism’s nonlegal protection of academic integrity, which centers on honesty and transparency about the origins of the material in a new work. A third category of practices concerns scholarly norms of quality, such as verifying assertions or familiarizing oneself with the relevant literature. Each of these three categories of harm—copyright infringement, plagiarism, and bad scholarly practices—serves distinct normative ends. Any regulatory framework for AI-assisted authorship should thus address each category on its own terms rather than assuming they rise and fall together.

It might be tempting to expand copyright (or other areas of substantive law) to reach AI-facilitated plagiarism, perhaps because copying material without attribution feels instinctively wrong. But intellectual property (IP) law isn’t designed to punish every act of free-riding, and there are [sound policy reasons](#) for copyright law’s limits. Instead, the growing problem of AI-facilitated plagiarism should be addressed through the extralegal norms and academic sanctions that have long governed these kinds of concerns. Plagiarism is a problem, but it is not—and should not be—a *legal* problem.

AI does, however, present the plagiarism and scholarly practice problems in a new light. While some scholars—and many university committees—have begun to think about the proper rules for using and disclosing the use of AI in academic writing, merely disclosing the use of AI is unlikely to satisfy those whose ideas the AI copied and returned in response to a prompt. And a disclosure rule is insufficient to distinguish between uses of AI almost everyone would consider acceptable, such as correcting grammar and typos, and ones that are much more problematic, such as turning in a paper whose thesis and text were largely composed by AI. This isn’t a copyright problem. It isn’t even always a plagiarism problem. But it may be a problem of academic integrity.

We explain the problem in Part I. In Part II, we explain why it isn’t—and shouldn’t be—a legal problem. In Part III, we suggest best practices for teachers, students, and scholars using AI, and offer some thoughts as to how those best practices might be adapted outside the academic context.

I. How Generative AI Breaks Citation Chains

ChatGPT and other generative AI systems are now widely used for academic writing. Indeed, [surveys](#) suggest that an overwhelming

majority of students are using AI for tasks including suggesting research ideas and generating drafts. For example, in a spring 2024 [survey of Harvard undergraduates](#), almost 90% reported using generative AI, and over 50% used these tools for writing assignments, including “coming up with ideas.”

But generative AI is fundamentally different from research tools that link to original sources. A [large language model \(LLM\)](#) like ChatGPT produces text by [predicting](#) likely word sequences based on patterns in its training data, rather than by retrieving and crediting specific prior works. The result is a form of [probabilistic](#) generation of new content that makes it difficult or impossible to figure out why any given output is generated, much less which training material contributed to it or whether the output’s core “idea” is similar to the idea in a training source. AI output thus may implicitly contain facts or ideas from training data, but without attribution. In short, it breaks traditional citation chains.

Even when AI tries to attribute things to sources, it frequently gets things wrong. That shouldn’t be surprising. While some have grown accustomed to treating it like a search engine, it’s called *generative* AI for a reason—it makes things up on the fly in response to prompts based on word and token relationships it has [learned during training](#). If you ask an LLM to give you citations, it will, but quite often it will make those up too. More than 150 [lawyers](#) (and even a couple of [judges](#)) who have been caught [citing cases that don’t exist](#) or introducing other AI errors can testify to that. And while new techniques like “[retrieval-augmented generation](#)” and “[deep research](#)” target some of the problems of entirely invented citations by having the LLM focus on particular sources from the real world, LLMs using those methods often [get things wrong](#) even if the sources they cite really exist—sometimes as much as [90%](#) of the time. And some of those [hallucinations](#) can be life threatening, as when Google’s Bard answered a question asking what to do in case of a seizure by taking text from a reputable medical site describing what *not* to do.

The problem of hallucinations is well documented, though it keeps happening, probably because it is intrinsic to software that generates new content on the fly in response to user queries. Because it is so well known, people are learning (albeit slowly and inconsistently) that if an LLM cites a source, they need to check the source to see whether it exists and what it says.

But even if AI companies can somehow get the hallucination problem under control, or if users universally learn to verify sources before trusting them, hallucinating sources obscures a deeper problem: the source of ideas and text the AI itself generates.

Generative AI gives answers in response to prompts created by users. From the perspective of an AI user, this can create an [illusion of originality](#), where users mistake ideas that ultimately [originated in training material](#) for their own insights. While there are some circumstances in which asking the right question is the most important part of an idea, that isn't true of much scholarship. Working through the answer to a question, or coming up with an idea no one had thought of before, isn't always or even usually just a function of [asking the right question](#). If a student (or a professor!) asks ChatGPT to write a paper on a particular topic, the ideas in that paper come from somewhere—perhaps a single source in the training data, perhaps a combination of sources, or perhaps an original idea from the AI itself. (While courts have rejected the idea that AI can [author](#) or [invent](#) things, AI can generate content that would qualify for IP protection if a human had originated it.)

The problem with presenting these ideas as their own would be evident to most people if they weren't using AI. If I ask someone else to write my paper for me, it would be obvious that I was doing something inappropriate if I tried to pass the paper off as my own work. Similarly, if I read someone else's article and incorporated its ideas into my paper, it would be obvious that I was doing something inappropriate if I didn't credit the source of those ideas. Copying ideas without credit is plagiarism. But authors are less likely to recognize that copying ideas from AI-generated output that is based on human-authored training data raises the same issues as copying ideas directly from a human-authored source.

The issue isn't just that AI may not credit (or even be able to identify) the part of its training that produced the idea, or that users often misunderstand how generative AI works. It's also that technological writing assistance makes it more cognitively challenging to recognize where ideas came from. Authors may be [reluctant](#) to view a computer as the [originator of ideas](#). And as summarized by a [recent review](#) of the psychological literature in the generative AI context, "research suggests that AI users are at risk of failing to correctly monitor the extent of their own contribution when being assisted by an AI." For example, in [one study](#), participants were more likely to attribute ownership of a ghostwritten postcard to themselves when they were told it was produced by an AI than when they were told it was produced by a human ghostwriter. This is likely an example of "[cognitive externalization](#)," in which AI is viewed simply as a [tool](#) for offloading portions of a writer's work. There is also evidence that people are [more likely to cheat](#) when delegating tasks to AI.

The result is that students turn in papers—and professors write articles—that include ideas and concepts (and perhaps even text) first put forward by others but that don’t cite those others. This is a sort of unintended plagiarism, in which the writer doesn’t credit the originator of the idea because they may not even know that there is such an originator.

II. Why Copyright Shouldn’t Expand to Police Plagiarism

U.S. copyright law prevents the copying of creative expression. It does not, however, prohibit copying another’s [ideas](#), with or without credit. Copying the literal text of an article is infringement, at least unless excused by fair use or some other defense. Copying the [expressive heart of a work](#) (like a well-defined character from a novel) may also be infringement even if the defendant uses different words. But copying [facts](#) and [ideas](#) is [not](#) something copyright law forbids. To the contrary, the point of copyright is to [encourage](#) the dissemination of ideas by allowing different people to express those ideas in different ways.

A. Copyright Lawsuits Conflate Compensation and Credit

Generative AI trains on large datasets of copyrighted content. Many authors of the works on which AI models are trained are upset. There have been more than fifty U.S. copyright [lawsuits](#) against generative AI companies.⁵ Those cases have led to [two](#) early [decisions](#) finding that [training AI](#) is [fair use](#) because it produces a [transformative](#) result. There are [plenty](#) of [interesting issues](#) here that have produced numerous [scholarly articles](#).⁶ But they aren’t the focus of this paper.

⁵ One of us (Lemley) represents or has represented the defendants in some of these suits.

⁶ For arguments that training should be fair use, see, for example, Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 748 (2021) (“In this Article, we argue that ML systems should generally be able to use databases for training, whether or not the contents of that database are copyrighted.”); Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem*, 93 WASH. L. REV. 579, 622–23 (2018) (arguing that using copyrighted works for AI training is “highly transformative”); Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887, 1914 (2024). *But see* Robert Brauneis, *Copyright and the Training of Human Authors and Generative Machines*, 48 COLUM. J.L. & ARTS 1, 59 (2024). For early discussion of the issue, see generally Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J.

Our focus is on the *output* of generative AI. LLM output generally doesn't infringe copyright because it isn't [substantially similar](#) to the protectable expression of any of the inputs on which the AI is trained. There are exceptions, often traceable to [duplications](#) in the training dataset or to [deliberate efforts](#) by the user to prompt infringement. And particular models seem to “[memorize](#)” certain works for reasons scholars don't fully understand. But for the most part, if you ask generative AI to give you a paper on a topic, it won't give you anything much like a particular prior paper. From a copyright perspective, that should be the end of the question.

But it isn't. Complaints in these lawsuits (and [related press](#)) often raise the [concern](#) that the authors aren't only uncompensated—they also [aren't getting credit](#) for the use of their work. And when they do, they often turn to the language of plagiarism. Content creators are [fond of referring to](#) generative AI as “[nothing more than a plagiarism machine](#).” (That's [not true](#).) By invoking plagiarism, these commentators generally do *not* mean AI is something that should be legal but ought to earn moral disapproval in certain limited contexts. Rather, they see the argument as either adding moral opprobrium to a claim of copyright infringement or as a reason to expand copyright infringement to reach conduct that, because it is plagiarism, ought therefore to be illegal. Others who don't use the term “plagiarism” are explicit in proposing that copyright broaden to reach credit or [personality-based harms](#), among [other injuries](#).⁷

It isn't just laypeople or trade groups that sometimes conflate plagiarism and copyright infringement. The Supreme Court has mistakenly described [copyright infringement as plagiarism](#). So has Judge Learned Hand, perhaps the best-known copyright jurist of all time. [Nichols v. Universal Pictures](#) (1930)—the main teaching case for the test for copyright infringement—repeatedly refers to an alleged infringer as a “plagiarist.” [Hand's assertion](#) that “no plagiarist can excuse the wrong by showing how much of his work he did not pirate” has been quoted in over one hundred other copyright cases, including by [the Supreme Court](#).⁸ And even prominent judges who understand

COPYRIGHT SOC'Y U.S.A. 291 (2019); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607 (2009).

⁷ But see Dennis Crouch, *Using Intellectual Property to Regulate Artificial Intelligence*, 89 MO. L. REV. 781, 843–44 (2024) (arguing against using IP to solve non-IP problems relating to AI); see also Oren Bracha, *The Work of Copyright in the Age of Machine Production*, 38 HARV. J.L. & TECH. 171, 215 (2024) (same).

⁸ For other statistics, search Westlaw for Judge Hand's quotation.

the difference, like Judge Richard Posner, use the fact of plagiarism to change the copyright law to treat the plagiarist as a wrongdoer when they would [otherwise be entitled to](#) a fair use defense.⁹

B. Distinguishing Copyright Infringement, Plagiarism, and Bad Scholarly Practices

But plagiarism isn't—and shouldn't be—conflated with copyright infringement. It is crucial to maintain the conceptual boundary between the two. Copyright infringement is a violation of a legal right: It occurs when someone copies protectable expression from a work without permission (outside of fair use or other defenses). By contrast, plagiarism isn't a legal cause of action at all—it is an ethical or academic offense. And then there is a third category: what we might call “bad scholarly practices” or substandard research habits, which might not rise to plagiarism but still violate disciplinary norms of rigor. The three often get tangled in practice, but they are analytically distinct. Let's unpack them:

Copyright infringement is an economic right that requires copying of protectable expression. Any [expressive work](#)—a book, poem, song, or even computer program—with a “[modicum of creativity](#)” receives protection as soon as it is “fixed” (e.g., written down or typed). But copyright has limits. Facts and ideas aren't themselves copyrightable, though the particular expression of a fact or idea is. Copyright also has a (somewhat) [limited term](#), after which works enter the public domain and are free for anyone to use.

If you write a paper based on others' copyrighted works, you have infringed their copyrights if your paper has “[substantial similarity](#)” to their protectable expression, unless your borrowing constitutes “[fair use](#).” Infringement doesn't require that you pass off the new material as your own; even with full attribution, copying substantial protected expression can infringe. For example, if you publish a full chapter of an in-copyright book and credit the original author, you've still infringed—you just haven't plagiarized.

Plagiarism is typically [defined](#) as the use of someone else's language, ideas, or work without sufficient credit. It is about being honest and transparent about where material in your paper comes from, partly to give credit to the people whose work you relied on, but also to help readers really understand and evaluate your scholarship. Plagiarism is an ethical offense in academia (and journalism, etc.), enforceable by social sanction or institutional discipline but not by

⁹ See RICHARD A. POSNER, THE LITTLE BOOK OF PLAGIARISM 16–17 (2007).

lawsuits. It is also contextual, with both norms and sanctions defined by the relevant community. For example, legal scholarship has a norm of [more extensive citation](#) than most other disciplines.

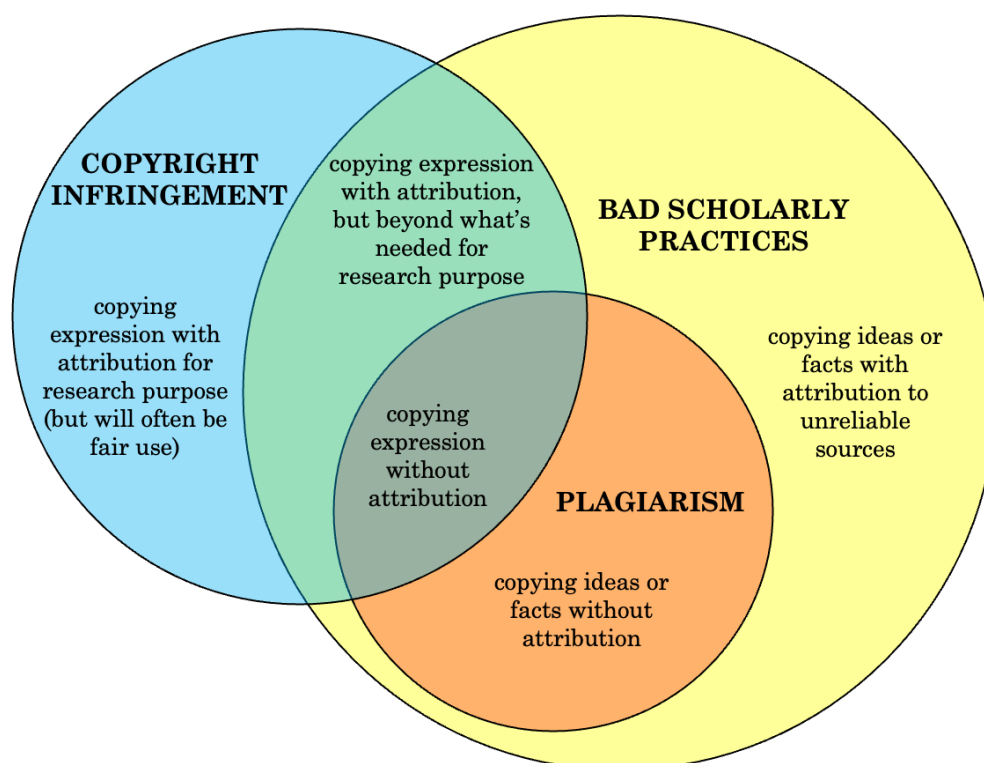
One can plagiarize uncopyrightable material. If a historian articulates a novel theory (an idea) based on newly discovered facts and you use the idea and facts in an article without credit, you've plagiarized, even though you haven't [infringed any copyright](#). Similarly, publishing your former advisor's mathematical model as your own is plagiarism but [not copyright infringement](#). The distinction between copyright infringement and plagiarism has been aptly summarized by [Brian Frye](#):

[C]opyright infringement and plagiarism overlap, but are not co-extensive. Copyright law prohibits certain unauthorized uses of copyrighted works, irrespective of attribution, and plagiarism norms prohibit unattributed copying of certain expressions, facts, and ideas, irrespective of copyright protection. Using an original element of a copyrighted work with attribution may be copyright infringement, but cannot be plagiarism, and copying a fact or idea without attribution may be plagiarism, but cannot be copyright infringement.

Bad scholarly practices form a third category that refers to breaches of disciplinary norms of rigorous research and writing. These norms are more contested and reflect a more subjective judgment of how good your scholarship is rather than an academic integrity issue.

Consider a researcher who sees a claim on Twitter and inserts it in their paper. If what they use is uncopyrightable—like an idea or fact—then it's not copyright infringement. If they honestly cite the tweet where they learned it, then it's not plagiarism. But other scholars might still criticize this as sloppy scholarship because the researcher should have verified the fact against a more reliable source or taken the time to figure out whether the idea had already been written about in the literature.

These three categories of copyright infringement, plagiarism, and bad scholarly practices are distinct but overlapping. Figure 1 shows a Venn diagram to illustrate their relationships.

Figure 1. Copyright Infringement vs. Plagiarism vs. Bad Scholarly Practices

All plagiarism is a bad scholarly practice because honesty about sources is foundational to scholarship. But there are plenty of bad scholarly practices that aren't plagiarism, such as accurate citation to unreliable sources or citing a derivative source for an idea without citing the originator of the idea. And there are instances of plagiarism or bad scholarship that aren't copyright infringement—copying ideas, facts, or public domain text without credit.

Copyright infringement could be plagiarism, or just bad scholarship, or good scholarship. For example, if you extensively copy protected expression without attribution, that's plagiarism (and a bad scholarly practice) as well as copyright infringement. If you extensively copy protected expression *with* attribution, such as by using someone's text for the same purpose while citing them, then it's no longer plagiarism, but it is copyright infringement, and it is still a bad scholarly practice because you are substituting someone else's work for your own. Crediting the original creator doesn't get you out of infringement (despite what many [YouTubers](#) seem to think).

It is less common for something to be good scholarly practice but still be copyright infringement, since most good scholarly uses of protected expression should be [fair use](#). Nonetheless, there are examples. Some stem from the fact that courts can get the fair use

question wrong. Adolf Hitler's publisher [successfully sued](#) American politician and journalist Alan Cranston in the 1930s for translating the full version of *Mein Kampf* into English to inform Americans that the official English-language version released by the Nazis downplayed some of the worst parts of the book. Probably that should have been fair use because it was in the public interest and Cranston disclaimed any profit. But even if the court was right to condemn it because he translated the whole text, Cranston's act was still a public service that we wouldn't morally condemn. Something similar may be true of those who copy videos of police or ICE misconduct in order to publicize the wrong. That may be [copyright infringement](#), but it is hard to fault the practice as a moral or social matter.

Academic norms may also permit closer copying for purposes of accurate disclosure than copyright law does. [Salinger v. Random House](#) (1987) found against an academic author who too-closely paraphrased author J.D. Salinger's letters, but if anything, the best scholarly practice is probably direct quotation and not paraphrasing, something that is even more likely to be infringement. Copyright law might also condemn reusing your own prior figures that some journal made you assign them the copyright in, but academic practice would surely permit that reuse with attribution. The strict liability nature of copyright infringement also offers some possibilities. In [Lipton v. Nature Co.](#) (1995), the defendants did everything right in an effort to get permission and give attribution, but unbeknownst to them, the licensor had gotten the work from another licensor—who it turns out had stolen it decades before. That was infringement, but it wasn't something people would treat as a moral wrong.

C. Should the Law Protect Attribution?

Attribution matters to [creators](#). But U.S. law offers [no general attribution right](#).¹⁰ As discussed above, some uncredited copying will constitute copyright infringement; in those cases, the copyright owner can require [credit](#) as a [condition](#) of use. And the Visual Artists Rights Act grants an [attribution right](#) for limited-edition works of [visual art](#). In general, however, failing to credit someone is not—standing alone—a legal wrong. Authors may desire credit for many uses of unprotected elements of their works, or use of works that have entered the public domain, or uses that qualify as fair use, but they have no such claim under copyright laws.

¹⁰ Mark A. Lemley, *Rights of Attribution and Integrity in Online Communications*, in REAL LAW @ VIRTUAL SPACE: COMMUNICATION REGULATION IN CYBERSPACE 251, 251–67 (Susan J. Drucker & Gary Gumpert eds., 1999).

Nor can authors rely on other IP laws to circumvent these limits. In [*Dastar v. Twentieth Century Fox*](#) (2003), the Supreme Court rejected an attempt to use the [Lanham Act](#)—the federal trademark statute—as a plagiarism remedy. In that case, a company that repackaged a public-domain video series without crediting the original producer was sued for misrepresenting the “origin” of the work. The Supreme Court rejected this effort to read the Lanham Act “as creating a cause of action for, in effect, plagiarism—the use of otherwise unprotected works and inventions without attribution.” Allowing this kind of end run around copyright’s limits would upset what the Court described as the “carefully crafted bargain” of IP law.

[Some scholars](#) have proposed creating a new legal right of attribution to fill this gap, following jurisdictions such as Europe with stronger protection of moral rights. But we agree with [Rebecca Tushnet’s skepticism](#): “Legitimate claims for credit are simply too varied and contextual, and copyright law already too complex and reticulated,” to support the creation of a new legal cause of action, absent any independent legal harm. A rule broad enough to cover all uncredited uses of others’ ideas or words seems likely to conflict with many of the limits built into IP law’s grant of economic rights and to impose unmanageable line-drawing problems.

To be clear, our position is not that attribution doesn’t matter; in the following section, we describe the reputational and epistemic harms that plagiarism imposes. But not every harm is a *legal* wrong. And where the harm falls within academia and other knowledge-producing communities, the appropriate remedy lies there too.

D. Should We Care About Plagiarism at All?

If plagiarism isn’t copyright infringement, or any other form of legal wrong, what’s the problem? We think there are at least two reasons why academic norms properly discourage plagiarism: protecting academic authors’ reputational interests and protecting readers’ ability to evaluate academic work.

First, citations are the currency of academic writing. Academics rarely get paid directly for their work, which they distribute for free or even [pay to have published](#). Their payment is in the form of scholarly reputation. And that reputation depends on recognition of the intellectual contributions they make. As [David Nimmer put it](#), “the entire incentive for [the] creation [of academic articles] is (from the celestial perspective) to advance the frontiers of human knowledge and (from the earthly vantage) to win their authors recognition,” such that attribution “is not an afterthought”—“it IS the incentive (or a large

part of it) which therefore must enjoy protection for the enterprise to continue sensibly. ‘Citing is paying’ in this environment.”

Second, writing or publishing a paper under one’s own name implicitly represents that the author has contributed the ideas and words contained therein. Of course, not everything everyone writes is entirely original; we all stand on the shoulders of giants.¹¹ But that is precisely the role citation plays. If someone else said it better, it’s fine to quote them, but the quotation marks make it clear what you’ve added. Similarly, it’s fine to get ideas from elsewhere, but citing the source of those ideas communicates that fact to the reader. It also allows the reader to properly attribute the words or ideas to their originator, not only fulfilling the reputational currency goal but also allowing the reader to distinguish the author’s contributions from borrowed material. That enables the reader to evaluate the quality of the paper, whether for purposes of grading a student or making a reputational judgment about the contribution of a scholarly work.

Some commentators have downplayed these concerns. Brian Frye provocatively argues that plagiarism is not only [harmless](#) but can also be [desirable](#). And in work with Megan Boyd, he has argued that law schools should be [teaching](#) students how to plagiarize, including with [generative AI](#). We disagree. Much of Frye’s attack on plagiarism norms is really an attack on efforts to treat plagiarism as a [legal wrong](#)—a form of copyright infringement, trademark infringement, or fraud. As noted above, we agree with him on this point. He directs other arguments [against](#) visceral dislike of plagiarism as an inherent wrong, independent of instrumental concerns; again, we have no objection.

Frye’s primary objection to anti-plagiarism norms focuses on the first justification we introduced above, related to the incentive value of attribution for academics. Frye argues that these norms can be [justified](#) only if they give us new scholarship we wouldn’t otherwise get (the economic justification for copyright) and that there is [no evidence](#) that they do so.¹² But he merely asserts that. We agree data would be

¹¹ The most familiar version of this statement comes from Isaac Newton: “If I have seen further [than others,] it is by standing on [the] shoulders of Giants.” ROBERT K. MERTON, ON THE SHOULDERS OF GIANTS 31 (1965). Fittingly, Newton borrowed (plagiarized?) the aphorism from earlier writers, with the first known use dating to Bernard of Chartres around 1126. *See id.* at 273–74.

¹² Relatedly, he argues that anti-plagiarism norms favor incumbents. But it seems more plausible to us that a world of rampant scholarly plagiarism would favor incumbents more because if two scholars put forth the

desirable; this is ultimately an empirical question.¹³ But given the reputation economy he concedes exists in the academic world, the existing norms against plagiarism, and the [evidence](#) that creators care a great deal about attribution, it would be surprising if attribution of work had no incentive effect. Certainly, scholars think it matters; they [fight](#) over who is [listed as the author](#) on [papers and patents](#) and [in what order](#), and we can think of few if any examples of people submitting academic work anonymously. [Tenure](#) and [promotion](#) committees also act as if it matters. Frye's [claims](#) that attribution is swamped by other incentives for academics to write—"to publish articles, find a teaching job, attend conferences, land speaking engagements, get tenure, be promoted, receive research funding, move to a different school"—ignores the fact that all of those incentives depend on the identification of the author as the person who has come up with the ideas and written the articles that drive all those things.

Frye [responds](#) that those aren't necessary features of a system; academics could build our norms differently, and perhaps doing so would [improve](#) social welfare. We're skeptical, but for our purposes it doesn't matter. In the system we have, reputation and academic contribution matter, and plagiarism breaks the mechanism we have to evaluate those things.

More importantly, Frye's focus on whether attribution norms provide optimal incentives for producing scholarship—and his [related conclusion](#) that there is no harm from plagiarism by students who aren't creating public-facing work—fails to grapple with the second key argument against plagiarism. As we explained above, anti-plagiarism norms aren't just about protecting the reputational interests of scholars you cite; they are also about being honest and transparent about your scholarly methodology, including the sources of your words and ideas and what you actually contributed to the project. These values apply to any academic project, whether the audience is the broad readership of a scholarly publication or only the single instructor for a university class.

same idea—one plagiarizing from the other—people will be more likely to associate the idea with the more prominent scholar.

¹³ Frye says in [his response](#) to us that the burden should be on us to prove the incentive effects of plagiarism. But that strikes us as backwards. It is Frye, not we, who proposes changing what Frye himself says has been a norm "for at least 2000 years." The burden of proof ought to lie with those who want to change the norm.

III. Preserving Attribution Norms in the AI Era

There's been lots of attention to the hallucination concern, particularly for hallucinated citations. But commentators have paid very little attention to the plagiarism concern, which needs to be better recognized.

If plagiarism is the taking of someone else's ideas or words without attribution, there would seem to be a simple solution to the problem of AI plagiarism: disclose that you used AI in writing the paper. We do indeed think disclosure is a necessary response to the use of AI in academic writing. But it isn't a sufficient response. "Using AI" can mean anything from having it correct your grammar and spelling (surely not plagiarism), to having it fill in citations for you (not a good idea, as we've noted above), to giving AI an idea and having it write a paper for you to having AI come up with the idea itself. If the AI fills in citations, those citations may be wrong—entirely made up or real citations that don't support the cited proposition. In either case, the citations don't reflect proper attribution. If the AI contributed significant writing or ideas, even disclosing that fact may hide the true source of those ideas. And the nature of the writing matters, because the norms around copying and attribution differ in student, academic, and professional writing.

We need different norms for different circumstances. In this Part, we suggest some best practices for plagiarism and AI.

A. Education

In the education context—including law schools—we recommend a focus on clear policies, granular disclosure requirements, and pedagogical design that encourages genuine student learning. The most salient plagiarism concern in student writing (at least unpublished student writing) is the accurate representation of what the student did and what they got from somewhere else so that instructors can fairly evaluate the student's work. Practical guidelines will allow instructors to uphold academic standards without categorically banning useful AI tools.

First, schools and instructors should [set clear policies and expectations](#). Teachers should have significant discretion in whether and to what extent they allow AI to be used in their classes. Professors and schools should be clear up front in syllabi and school policies, so that students understand that using AI to generate ideas without proper attribution is a form of plagiarism and will be treated as such. For example, Stanford Law School's generative AI policy—which we helped draft—specifies that instructors may not “authorize students to contravene standard academic norms concerning plagiarism” and that

“[p]lagiarism includes using an idea obtained from AI without attribution or submitting AI-generated text verbatim without quotation marks.”¹⁴ Where AI use is permitted in limited ways, the rules should be precise; e.g., “AI may be used for proofreading, but not for drafting substantive text,” or “you may use AI to brainstorm topics, but you must cite the AI for any significant ideas that you incorporate from its suggestions,” or “before using AI for any part of drafting your paper, check with me so that I can guide you through the often tricky academic integrity issues that result from incorporating AI text.”

Second, while disclosure is an important first step to avoiding AI plagiarism, it must be specific disclosure. It isn’t enough to say: “I used AI to help me with this paper.” For a teacher to evaluate a student’s contribution, the student must be clear about what part of their work came from AI and what part was their own. Teachers should err on the side of requiring not just general disclosure but specific identification of text and ideas that came from AI, ideally with a link to the relevant AI prompt and output. It should be a norm in academic writing to quote AI contributions just as one [would quote any other publication](#).¹⁵ Even if what is used isn’t an exact quote, a concept that comes from AI or words that are rephrased from AI still deserve citation so the teacher can understand what is original to the student. For students writing legal scholarship, the latest edition of *The Bluebook*—the standard citation guide—includes [instructions](#) for citing AI-generated content, including that authors “save a screenshot capture of that output as a PDF.”

¹⁴ *Use of Generative AI Technology*, STAN. L. SCH., <https://law.stanford.edu/office-of-student-affairs/use-of-generative-ai-technology> (on file with authors).

¹⁵ For some AI uses that instructors may exercise their discretion to allow, it won’t be feasible or desirable to use quotation marks. If an author whose first language isn’t English asks AI to improve the grammar of an essay, for example, the result may be small changes throughout multiple sentences. In that case we think it is fine to disclose how the author (student or academic) used the AI without specific quotation.

Quotation also shouldn’t be necessary where the author uses AI to summarize their own work in an abstract or conclusion. As long as the summary is taken from the author’s own words in the paper, a generalized disclosure of how AI was used should suffice to avoid plagiarism concerns. As with all plagiarism norms, resolving these kinds of evolving edge cases will be contextual and should be guided by the underlying concerns of protecting reputational interests and being transparent with readers about scholarly methodology.

Third, citing the AI should not be the end of the story. Citing the AI avoids plagiarism, but we should also teach students to use AI suggestions as the starting place for research, not the end point. Students should go beyond “AI said so” to verify facts or find the actual source of an idea, in the same way that students have learned to follow links from Wikipedia to more authoritative sources. AI tools like retrieval-augmented generation and deep research may help identify underlying sources. This is good scholarly practice, as noted above. But even if the student will never be a scholar and the paper will never see the light of day, doing so is pedagogically useful for another reason: Going through the exercise will help students understand that AI regularly makes mistakes and that both the ideas and the sources it cites need to be critically examined.

Fourth, educators should structure assignments to reduce the temptation of AI ghostwriting. There are several tools for this. One is staged assignments, where students submit topic proposals, outlines, and lists of sources as the work progresses. Requiring all assertions to be supported with citations to credible sources can encourage good research practices, limit AI’s effectiveness, and make it easier for teachers to spot AI use through hallucinated citations. Additionally, where feasible, instructors can incorporate oral discussions to test students’ understanding of the work they submitted.

Finally, institutions should back up these norms with sanctions. If a student misrepresents AI-generated work as their own, they should face the same potential consequences as in any other case of plagiarism, including a failing grade, academic probation, or other university-imposed discipline. As in other plagiarism inquiries, the sanction should be proportional to the offense, with attention to whether the student intentionally deceived or merely misunderstood the rules. Of course, detecting undisclosed AI usage poses a challenge, though hallucinated citations should create a presumption of AI use. This is why our emphasis is on transparency, disclosure, and prevention, not after-the-fact detection.

B. Academic Publishing

The importance of proper citation in academic publishing is well established, but norms for proper citation of AI-generated ideas are still contested. For example, a [2025 Nature survey](#) of over five thousand science academics found that 13% thought it was appropriate to use AI to draft a paper without disclosure, another 52% thought it was appropriate with disclosure, and the remaining 35% thought it was never appropriate.

In setting AI norms in academic publishing, it is important to recognize that the stakes of AI-facilitated plagiarism are even higher than in the classroom context. A scholarly article that appropriates others' ideas without credit raises not only the concern that readers will misunderstand the author's contribution, but also the concern that those who actually did come up with the ideas won't be credited. Accordingly, while the same specific disclosure and quotation norms we suggest for student work should apply to scholarship, we would go further. It isn't acceptable in published academic work to rely on AI as the source of an idea without verification that the idea really does owe its origin to AI. Any idea that comes from AI must be [investigated](#) so the source of the idea can be cited.

We therefore disagree with [claims](#) that disclosure of AI use should be voluntary. For example, in their argument against disclosure norms, [Kevin Frazier and Alan Rozenshtein write](#) that "treating AI use as a form of plagiarism" is "conceptually flawed" because the idea that AI output "is a product of appropriation from the authors whose works were used in its training . . . raises vital questions about intellectual property and fair use" but not "traditional academic ethics." But as we have explained, AI output can contain ideas drawn from the training data without attribution, and representing these ideas as your own can be plagiarism, even though it is rarely copyright infringement.

These norms should be tempered with a requirement of reasonableness in the AI world, just as in the rest of academia. Not every citation needs to turn into a history expedition.¹⁶ But scholars owe each other a good faith effort to find core sources.¹⁷ And relying on a statement from AI as evidence that the AI itself is the source of the idea doesn't satisfy that standard of reasonableness. It is possible the idea really does owe its origin to AI, but in those rare cases, it is worth verifying that fact. Nor is it sufficient to rely on whatever sources the AI cites without checking them. Those sources [may not exist](#), or they may not say what the AI [claims](#) they say. Checking the sources might show that they originated the idea, in which case academics should cite

¹⁶ Brian Frye thus [mischaracterizes our proposal](#) in suggesting that we would demand attribution of "[a]nything and everything" that "has already been said by someone, somewhere, sometime, or might as well have been said."

¹⁷ Similarly, a scholar who independently comes up with an idea can write about that idea without being a plagiarist, but they owe a duty of good faith to other scholars to invest some effort in determining whether the idea has already appeared in the literature.

them. Or they might cite other papers for the source of the idea, and the scholar can follow those citations.

Academic norms aren't definitively recorded anywhere. But journals can establish policies and expectations about how AI will be cited and used. Such [policies](#) are already common among [science journals](#). Law reviews have been [slower](#) to adopt AI policies, but [a few](#) have [done so](#), and others should catch up. Universities can also establish norms in [written policies](#) governing tenure and promotion or (more problematic) when plagiarism accusations surface.

Finally, just as instructors serve as a check in the educational context, peer reviewers can serve as a check for academic publications by surfacing uncited ideas that come from AI. We note, however, that peer reviewers too are increasingly using AI (and some paper authors are responding with [hidden prompts](#) to affect AI reviewers). Peer reviewers, like academic authors, should not put excessive reliance on AI to the exclusion of their own ideas and judgment. And they too should disclose their use of AI.

We can't fail academics like we can students. But we can deny or delay tenure and promotion, and we can require authors to retract papers that do not cite sources or disclose their use of AI. But the ultimate sanction for AI plagiarism in the academic world is also the ultimate currency in that world—professional reputation.

C. Legal Practice

Things are very different in legal practice. Law relies heavily on precedent, and fidelity to existing language is an important part of that. The result is that plagiarism hits different in legal writing. Courts regularly copy ideas and text from parties' briefs, often without citation or quotation. And the parties are generally fine with that; their arguments prevailed, after all. Lawyers drafting contracts regularly copy provisions from prior contracts. Doing so [may](#) help [standardize](#) contract text and therefore make understanding and interpreting contracts easier.¹⁸ Lawyers often copy ideas (and sometimes exact text) from the briefs and complaints of others. There is a whole industry of "copycat" complaints in which class action lawyers file follow-on lawsuits after governments file enforcement actions or even copying other class action suits. These copies are probably copyright infringement in a [strict sense](#), but enforcement is very rare.

¹⁸ Cf. Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CALIF. L. REV. 479 (1998) (questioning the extent of the network effects in standard contract terms).

Plagiarism seems to matter less in law than elsewhere because we care more about the ideas themselves and less about their source. If a criminal defendant can get a charge dismissed by relying on the text of someone else's successful motion to suppress evidence, we want them to do so rather than making them write a different (and possibly less persuasive) argument. And while cases should be cited accurately, there is less value to citing non-binding documents like prior [legal briefs](#) in different cases.

AI doesn't change this dynamic. What it does is raise the risk of hallucination—either false statements of law or inaccurate citations to authority. That risk is heightened in the legal context because appeal to authority is a central feature of how we decide cases. As noted above, there are hundreds of cases in which lawyers have cited nonexistent cases hallucinated by AI. In response, an increasing number of courts now [require disclosure](#) of the use of AI in briefs. [The ABA](#) has now issued ethics guidance on AI use. And courts are [punishing](#) the lawyers who cite hallucinated cases—at least the ones they catch—with [sanctions](#), [bar discipline](#), and potentially even [harsher penalties](#).¹⁹

Disclosure of AI use is a possible means of identifying potential hallucinations. But disclosing that AI wrote your brief isn't sufficient to identify hallucinations. At most it puts people on notice to check the citations in the brief carefully. But both the filing lawyer and the lawyer on the other side should be doing that already. AI creates a new problem; lawyers for the most part weren't making up citations before AI. But it isn't a plagiarism problem. Plagiarism is likely to matter only when lawyers copy another lawyer's brief that in turn was generated by AI. But even if our norms allow that copying, the lawyers doing it ought to be checking the cases they cite before they cite them, AI or no.

D. The Downside of Disclosure?

Disclosure might sound like an unambiguous good. [Jacob Noti-Victor](#) argues that hidden AI authorship of creative works like movies is normatively problematic and discusses private and public ways to have more disclosure. It seems reasonable to want people to have more information about the source of the works they encounter.

¹⁹ See, e.g., Order at *10–13, *Mavy v. Comm'r of Soc. Sec. Admin.*, 2025 WL 2355222 (D. Ariz. Aug. 14, 2025) (No. 2:25-cv-00689) (removing lawyer from case, striking the brief, and ordering the lawyer to notify every other judge before whom they have appeared that they cited fake cases).

But disclosure of AI may have a downside. A number of studies (though not all) have found an “AI penalty.” The [same works](#) are viewed [less favorably](#) when users are told they were generated by AI [than when told they were generated by people](#).²⁰ If people read AI-generated arguments and content more negatively simply because it came from AI, they may unfairly discount ideas offered by students, scholars, or lawyers because those ideas are accurately identified as coming from AI.

The existence of this AI penalty is [contested](#), however, and may not apply in all contexts. Further, it may well be an artifact of the novelty of the technology that is unlikely to persist. But if it does, it is a downside to disclosure. That doesn’t mean disclosure isn’t worth it. But it does mean we should consider strategies to mitigate this bias in evaluating papers.

Conclusion

Generative AI rarely outputs protected expression, but it routinely regurgitates ideas, often without accurate attribution. That isn’t a copyright problem. But in some domains, particularly academia, it is a plagiarism problem. The norms governing AI use in student writing or scholarship are still developing, and the risk of AI-facilitated plagiarism has yet to be widely recognized. But this risk is real, and it should be governed like other plagiarism problems: by setting clear and enforceable standards through the academic institutions where plagiarism’s harms are most salient. Schools should adopt rules for student work requiring specific disclosure of the ideas and text generated by AI so the reader can know where the ideas and text in papers came from. Scholars should do more, investigating the ideas and claims AI makes to find their actual source.

AI is here to stay. So are the scholarly values of honesty, transparency, and credit that make cumulative knowledge possible. We do not need new legal causes of action to preserve those values. Rather, the members of expert, disciplinary communities should insist—through thoughtful pedagogy, clear guidelines, editorial practice, and professional discipline—that AI doesn’t relieve authors of these responsibilities.

* * *

²⁰ Joseph J. Avery, Camilla Hrdy & W. Michael Schuster, The AI Penalty in Trade Secret Law (2025) (unpublished manuscript).