WHY THE FAIR USE DOCTRINE SHOULD NOT FAVOR AI

ABSTRACT

In recent years, the development of artificial intelligence (AI) has caused tension in the area of copyright law. Copyright holders worry about the potential degradation of their rights and infringement by AI. Specifically, various authors have filed lawsuits in district courts across the country concerning the use of copyrighted works in training AI systems. In response to this tension, AI companies have countered that downloading complete copyrighted works into databases for AI systems to learn from is fair use, an affirmative defense to copyright infringement. When this issue first arose, courts were largely deferential to AI companies because as a newly developing industry, courts wanted to afford it the opportunity to grow. In more recent cases, however, there is a notable trend in rulings being more favorable to the copyright holders as knowledge and proliferation of AI becomes more widely known and controversial.

This Note will assert that fair use should not be an applicable defense to copyright infringement when AI companies use copyrighted works to train their AI systems. When analyzing the four elements of fair use in cases of downloading copyrighted works into AI databases, the elements weigh against the defendants in nearly all cases. Removing the defense would prevent companies from using copyrighted works in AI training and reduce the possibility of copyright infringement in AI's responses.

I. INTRODUCTION 552
II. LEGAL BACKGROUND 553
A. OVERVIEW OF COPYRIGHT LAW AND AI 553
B. INFRINGEMENT 555
C. FAIR USE 556
III: AI SYSTEMS DOWNLOADING COPYRIGHTED MATERIAL FOR LEARNING 559
A. MACHINE LEARNING OVERVIEW 559
B. ARGUMENTS OF COPYRIGHT HOLDERS AND AI COMPANIES 560
C. WHAT COURTS HAVE RULED 561
IV. USE OF COPYRIGHTED MATERIAL IN AI TRAINING IS NOT FAIR USE 562
A. FAIR USE ELEMENTS APPLIED TO AI TRAINING 562
B. RECENT TRENDS IN COURTS 565
C. GREATER PROTECTIONS FOR COPYRIGHT HOLDERS' RIGHTS 567
V. CONCLUSION 568

I. INTRODUCTION

The interaction of artificial intelligence (AI) with the law has been a fascinating discussion point in recent years.[1] As AI technologies continue to improve and solve increasingly complex problems, their relationship with the law becomes an evolving conundrum. There is no better example of this complexity than the tension between AI technologies and intellectual property fields, specifically that of copyright.[2]

Copyright law exists to ensure that a person's creative works are protected, which in turn incentivizes people to create new works.[3] With the

1. *See* Matthew M. Beier, *The AI Revolution in Law: There's No Turning Back*, 97 WIS. LAW. 41 (2024) (stating "[g]enerative AI has recently been the subject of much of the legal tech revolution discussion").

2. *See generally* Danielle S. Van Lier, *Artificial Intelligence*, *in* 3 ENTERTAINMENT LAW 3D: LEGAL CONCEPTS AND BUSINESS PRACTICES § 24:14 (Thomas D. Selz et al. eds., 2025) (highlighting the interaction between copyright law and AI and the complexities that arise, most notably regarding AI's ability to infringe copyrighted works).

3. *See* Noah Ehrenpreis, *Intellectual Property Law—Frustrating the Purpose of the Copyright Clause: First Circuit Restrictively Interprets Copyright Registration Requirements*—Torres-Negrón

introduction of AI, copyright holders worry about the possibility of AI infringing on their intellectual property rights.[4] For example, many copyright holders have taken issue with the use of their copyrighted materials in AI training, the process used to teach AI pattern recognition.[5] Some AI companies have responded to lawsuits from copyright holders claiming fair use, an affirmative defense to copyright infringement claims.[6] Currently, authorities are unsure how to address the growing issue of copyright infringement by AI, and have left the issues largely to the courts.[7] This Note will outline the issues surrounding copyright and fair use as related to AI, then discuss the current state of these issues, and will conclude by arguing that the fair use defense should not be applicable to the use of copyrighted material in AI training.

II. LEGAL BACKGROUND

A. OVERVIEW OF COPYRIGHT LAW AND AI

The United States Constitution vests Congress with the authority to establish copyright.[8] The outlined purpose is "[t]o promote the Progress of Science and useful Arts by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[9] Copyright protections are owed to "original works of authorship fixed in any tangible medium of expression."[10] Categories of works protected include literary works, musical works, and dramatic works, as well as accompaniments to those works, including lyrics.[11] Once an author creates an original work and obtains a valid copyright, they are granted exclusive

v. J N Records, LLC, 504 F.3d 151 (1st Cir. 2007), 42 SUFFOLK U. L. REV. 389, 391–92, 392 n.23 (2009).

4. *See generally* Van Lier, *supra* note 2 (noting there were "over 50 lawsuits . . . relating to the unauthorized use of copyrighted content" at the time of publishing).

5. *See generally id.* (explaining the frustrations copyright holders have outlined in recent lawsuits concerning AI training).

6. *See* Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 391, 397 (D. Del. 2025).

7. *See generally* CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW (2025) (articulating that although the Copyright Office has provided guidance, judicial interpretations have largely governed how copyright law applies to generative AI).

8. *See* U.S. CONST. art. I, § 8, cl. 8.

9. *Id.*

10. 17 U.S.C. § 102.

11. *Id.*

rights to exercise on their work.[12] These rights include the right to reproduce the work, the right to prepare derivative works, the right to distribute reproductions of the work, and the right to perform, display, or transmit the work publicly.[13] These rights allow for the creator to see profitability of the work while also deterring would-be copiers from trying to profit off of the creator's work.[14]

The subject matter of copyright is where the intersection with AI emerges.[15] In recent years, the Copyright Office, the governing body that issues copyrights, has generally maintained that AI-generated works cannot be copyrighted unless there is "sufficiently creative" human contribution to the work.[16] The Copyright Office and courts have placed a large emphasis on the importance of human authorship in works.[17] As such, in many rulings, these bodies cite to the use of technology as a tool for creation rather than the creator itself.[18] However, AI works differently from a typical computer program as it does not produce a specific outcome designed by a programmer.[19] Instead, it pulls data from a database and assembles it in an order based on recognition of patterns over time.[20] This process is known as "machine learning," and is the basis for many of the generative AI models in use today.[21]

The Copyright Office has maintained its stance that AI-generated works with little human input cannot be copyrighted.[22] However, an issue largely unaddressed by the Copyright Office is whether the initial use of copyrighted

12. *See* Jessica L. Gillotte, Note, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2666 (2020) (citing 17 U.S.C. § 106 (2019)).

13. *See* 17 U.S.C. § 106.

14. *See* Gillotte, *supra* note 12, at 2666.

15. *See generally id.* at 2659.

16. *See* ZIRPOLI, *supra* note 7 (showing several examples of failed copyright petitions that failed to distinguish the human contribution from the AI contribution).

17. *See generally id.* (discussing the role of human authorship in determining copyright protection for AI-generated content); Gillotte, *supra* note 12, at 2658–59 (stating "AI-generated artwork may also undermine copyright law's purpose of incentivizing *human* authors to create new, original works, rather than merely incentivizing the *creation* of new, original works"(emphasis in original)).

18. *See* Gillotte, *supra* note 12, at 2667 (stating that computer programmers or users can hold copyright for works made by computers, not the computer program itself because it has no "legal personhood" to own property).

19. *See* Ran He, Jie Cao & Tieniu Tan, *Generative Artificial Intelligence: A Historical Perspective*, 12 NAT'L SCI. REV. 1, 4–5 (2025), https://pmc.ncbi.nlm.nih.gov/articles/PMC11970245/pdf/nwaf050.pdf [https://perma.cc/V739-PA36].

20. *See id.*

21. *See id.*

22. *See* ZIRPOLI, *supra* note 7.

material in AI training is a violation of copyright law.[23] Litigation on this issue has sprung up across the country, with over fifty lawsuits pending at the federal level.[24] Most of these lawsuits claim that the use of copyrighted material in AI training is an infringement upon the rights of the copyright owner.[25] To address the arguments of both sides and argue a position, this Note will first briefly explain copyright infringement and the defense of fair use.

B. INFRINGEMENT

Copyright infringement occurs when a person who is not the copyright holder uses copyrighted material.[26] To prove copyright infringement, a plaintiff must show "(1) that [they] own[] a valid copyright . . . and (2) that [the defendant] copied 'constituent elements of the copyrighted work.'"[27] In most situations, proof of direct copying is difficult to find, so courts rely on other factors to determine if a defendant copied the work.[28] To establish copying, the plaintiff must show evidence that the defendant had access to the original work and that the defendant's work is substantially similar to the original.[29] Access can be determined by direct access, which means the plaintiff has proof that the defendant directly received a work from them.[30] If there is no proof of direct access, plaintiffs may still show there is an inference that the defendant had access to the work if there is "a reasonable possibility of access," such as if the work was widely and publicly disseminated.[31]

---

23. *See generally* Van Lier, *supra* note 2 (noting that the Copyright Office released a pre-published report that did not take a definitive stance on infringement, rather it suggested solutions that could avoid litigation such as pre-approved licensing).

24. *See id.*

25. *See id.*

26. 17 U.S.C. § 501 ("Anyone who violates any of the exclusive rights of the copyright owner as provided by sections 106 through 122 or of the author as provided in section 106A(a), or who imports copies or phonorecords into the United States in violation of section 602, is an infringer of the copyright or right of the author, as the case may be.").

27. Calhoun v. Lillenas Publ'g, 298 F.3d 1228, 1232 (11th Cir. 2002) (citing Feist Publ'ns, Inc., v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991)).

28. *See* Original Appalachian Artworks, Inc. v. Toy Loft, Inc., 684 F.2d 821, 829 (11th Cir. 1982).

29. *See Calhoun*, 298 F.3d at 1232 (citing *Original Appalachian Artworks, Inc.*, 684 F.2d at 829 (citation omitted)).

30. *See* Selle v. Gibb, 741 F.2d 896, 901 (7th Cir. 1984).

31. *See id.* (citing Abkco Music, Inc. v. Harrisongs Music, Ltd., 722 F.2d 988, 998 (2d Cir. 1983)).

To determine if there is substantial similarity, courts use an ordinary observer test.[32] If "an average lay observer would recognize the alleged copy as having been appropriated from the original work," then the similarity would be considered substantial.[33] Other courts have employed a "more discerning ordinary observer test," which looks at not just the copyrightable elements of the works, but also noncopyrightable "similarities in such aspects as the total concept and feel, theme, characters, . . . and setting" separately.[34] If the similarities in the two works only concern those that are noncopyrightable, then it is not enough to establish substantial similarity, and the plaintiff cannot prove copying occurred.[35]

Once a plaintiff has shown they own a valid copyright and the defendant has copied constituent elements of their copyrighted work, they have a prima facie claim for copyright infringement.[36] This creates a rebuttable presumption that the defendant has copied the work.[37] If a defendant cannot produce evidence of independent creation to rebut, then they are liable for copyright infringement.[38] However, there are other defenses available for copyright infringement, namely the affirmative defense of fair use.[39]

C. FAIR USE

Fair use is a defense to copyright infringement designed to protect individuals who use copyrighted material for a legitimate purpose, including "criticism, comment[ary], news reporting, teaching (including multiple copies for classroom use), scholarship, or research."[40] The prohibition against copyright infringement is intended to protect creators from others profiting off of their work; therefore, the fair use defense exists to protect those who are not seeking to profit.[41] There are four factors that are relevant in determining if the use of a copyrighted work is fair.[42] Those factors are:

---

32. *See Calhoun*, 298 F.3d at 1232.

33. *Id.* (quoting *Original Appalachian Artworks, Inc.*, 684 F.2d at 829 (citation omitted)).

34. Lewinson v. Henry Holt & Co., 659 F. Supp. 2d 547, 564–65 (S.D.N.Y. 2009) (quoting Boisson v. Banian, Ltd., 273 F.3d 262, 272–73 (2d. Cir. 2001)).

35. *See id.* (citing Gal. v. Viacom Int'l, Inc., 518 F. Supp. 2d 526, 543 (S.D.N.Y. 2007)).

36. *Calhoun*, 298 F.2d at 1232.

37. *See Original Appalachian Artworks, Inc.*, 684 F.2d at 829.

38. *See id.* (first citing Miller v. Universal City Studios, Inc., 650 F.2d 1365, 1375 (5th Cir. 1981); and then citing Novelty Textile Mills, Inc. v. Joan Fabrics Corp., 558 F.2d 1090, 1092 n.2 (2d Cir. 1977)).

39. *See* 17 U.S.C. § 107; Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 590 (1994).

40. *See* 17 U.S.C. § 107.

41. *See* Gillotte, *supra* note 12, at 2666 (stating that "[t]hese rights protect authors against copyists who seek to profit from an author's work without expending the resources, time, and energy that the author did in creating her work").

42. *See* 17 U.S.C. § 107.

> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.[43]

The purpose and character factor is meant to "consider[] the reasons for . . . the copier's use of an original work."[44] If the work is copied for a commercial nature, this factor will likely weigh against defendants, but it is not dispositive.[45] The court will also look at whether the alleged copy was transformative of the original material.[46] To be transformative, the secondary work must have a "further purpose or different character" than the original work.[47] For example, a criticism must often refer back to the original work it is critiquing.[48] However, the critique is not using the material for the same purpose, but rather, using material to add context to the overall argument the secondary author is making.[49] Courts weigh both the commercial and transformative nature of the secondary work to conclude the first factor.[50]

The second factor rests on the nature of the original work, hinging on the overall creativity of the work itself with "[m]ore creative works [receiving] more protection."[51] Because emphasis is placed on creativity, this factor weighs more heavily in favor of fictitious works, like novels and fine arts.[52] While this factor exists in text, it "rarely play[s] a significant role in the determination of a fair use dispute" in practice.[53]

---

43. *Id.*

44. Warhol Found. Visual Arts v. Goldsmith, 598 U.S. 508, 527–28 (2023).

45. *See id.* at 531 (citing Google LLC v. Oracle America, Inc., 593 U.S. 1, 32 (2021) (highlighting that use of the work in a commercial nature weighs against a defendant because the user is using others' work for their profit, which is anathema to the purpose of copyright law, incentivizing creativity)).

46. Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 397 (D. Del. 2025) (citing *Warhol*, 598 U.S. at 529–31).

47. *See Warhol*, 598 U.S. at 531–33.

48. *See id.* at 532 (stating "other commentary or criticism that targets an original work may have compelling reason to 'conjure up' the original by borrowing from it").

49. *See id.* (explaining that "a use may be justified because copying is reasonably necessary to achieve the user's new purpose"); 17 U.S.C. § 107 (extrapolating the statutorily included example with justification as to why it is a commonly accepted instance of fair use).

50. *See Thomson Reuters*, 765 F. Supp. 3d at 397.

51. *See id.* at 399 (quoting 4 NIMMER ON COPYRIGHT § 13F.06[A]).

52. *See id.*

53. Authors Guild v. Google Inc., 804 F.3d 202, 220 (2d Cir. 2015) (quoting WILLIAM F. PATRY, PATRY ON FAIR USE § 4.1 (2015)).

The third factor is "the amount and substantiality of the portion used in relation to the copyrighted work as a whole."[54] For this factor, courts consider both "the quantity and value of the materials used."[55] Key language that courts use to explain this factor is that the interpolated work must not "take the heart" of the original copyrighted work.[56] For example, in *Harper & Row Publishers, Inc. v. Nation Enterprises*, the defendant used quotes from the plaintiff's book for their article.[57] Though the quotes were only a meager part of the article in its quantity, the article was "structured around the quoted excerpts which serve[d] as its dramatic focal points."[58] This was seen as "taking the heart" of the quoted books and the third factor ultimately weighed in favor of the plaintiffs.[59]

The fourth factor is "the effect of the use upon the potential market for or value of the copyrighted work."[60] Courts have historically given this factor the most weight in determining whether use is fair.[61] The facts used to determine this factor can largely be the same as those used for the third factor.[62] If a substantial quantity of copyrighted work is used in another work, then it affects the market value of the original work, since there is a substitute elsewhere.[63] Courts should consider both the extent of the market harm directly and "'whether unrestricted and widespread conduct of the sort . . . would result in a substantially adverse impact on the potential market' for the original."[64]

Ultimately, "[f]air use is a mixed question of law and fact."[65] Courts should weigh these four factors when analyzing an assertion of the fair use defense.[66] Factor two should impact the decision the least, while factor four should be the most influential factor.[67] If a court finds there is enough

54. 17 U.S.C. § 107(3).

55. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 586 (1994) (quoting Folsom v. Marsh, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841)).

56. *See Thomson Reuters*, 765 F. Supp. 3d at 399 (citing *Campbell*, 510 U.S. at 587).

57. *See* Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 565–66 (1985).

58. *Id.* at 566.

59. *See id.*

60. 17 U.S.C. § 107(4).

61. *See Thomson Reuters*, 765 F. Supp. 3d at 400 (citing *Harper & Row*, 471 U.S. at 566 (stating "[f]actor 4 'is undoubtedly the single most important element of fair use'")).

62. *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 587 (1994).

63. *See generally id.* (likening "the degree to which the parody may serve as a market substitute" based on the similarity to the original work as evidence of it affecting the market for the work).

64. *Id.* at 590 (citations omitted).

65. *Harper & Row*, 471 U.S. at 560 (citing Pac. & S. Co. v. Duncan, 744 F.2d 1490, 1495, n.8 (11th Cir. 1984)).

66. *See generally Thomson Reuters*, 765 F. Supp. 3d at 401 (showing that the judge is using these factors to decide the case).

67. *Id.*

evidence to support fair use on the facts proven by the defendants, then the court should grant summary judgment for the defendants.[68]

III: AI SYSTEMS DOWNLOADING COPYRIGHTED MATERIAL FOR LEARNING

A. MACHINE LEARNING OVERVIEW

Understanding how AI operates is central to assessing whether its training processes infringe upon copyright protections. AI is a general term for computer programs meant to emulate human thinking and decision-making.[69] This type of thinking is quantified as being able to perform "both technical functions, such as finding an optimal math solution, and functions traditionally associated only with the human brain, such as processing natural language."[70] AI programs are commonly model-based algorithms.[71] These algorithms use data collected via external stimuli to gradually learn how to generate responses.[72] Instead of following explicit instructions from the programmers, the model is given a set of data to collect inferences from and give responses via those inferences.[73] This is also known as machine learning.[74] In essence, the model teaches the chatbot to respond based on the data it has access to by using that data to predict the word that comes next.[75]

Machine learning is the process of the AI program gradually understanding how to recognize patterns received in its data set.[76] The program gradually gets feedback from the user and trains itself to better recognize the pattern.[77] An AI program can learn in both supervised and unsupervised formats.[78] Supervised learning is a more manual type of learning where a user inputs a label to a piece of data and the program can recall that label and associate it with the item later.[79] Unsupervised learning,

68. *See id.* at 397, 401.

69. *See generally* Gillotte, *supra* note 12, at 2660 (citing STUART J. RUSSEL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 1 (3d ed. 2010)).

70. *Id.* (citing RUSSEL & NORVIG, *supra* note 69, at 1–2).

71. *See* He, Cao & Tan, *supra* note 19, at 3–4.

72. *See id.* at 4.

73. *See id.*

74. *See id.*

75. *See id.*

76. *See* Gillotte, *supra* note 12, at 2660 (citing M. Tim Jones, *A Beginner's Guide to Artificial Intelligence, Machine Learning, and Cognitive Computing*, IBM DEVELOPER (June 1, 2017), https://developer.ibm.com/articles/cc-beginner-guide-machine-learning-ai-cognitive/#machine-learning [https://perma.cc/TTQ8-UHN4]).

77. *See id.* (citing RUSSEL & NORVIG, *supra* note 69, at 693).

78. *See id.*

79. *See id.*

however, is when the program is left to take its own initiative and "observes patterns in the input data without explicit feedback or labels and instead refines its algorithms iteratively by comparing its performances over time."[80]

The issue of copyright arises not from the process of machine learning, but in the data sets compiled for AI programs to draw from.[81] Authors have accused AI companies of downloading their original works into these databases and have filed suits claiming the downloading is infringement.[82] To explore whether their claims have validity, the next section of this Note will highlight the arguments made by copyright holders, as well as the counterarguments and defenses of the AI trainers.

B. ARGUMENTS OF COPYRIGHT HOLDERS AND AI COMPANIES

For the first fair use factor, copyright holders have argued in court that the AI training does not only train the AI to recognize patterns, but also "memorize their works' creative elements."[83] A central theme in many copyright holders' arguments is that machines "should not be allowed to do what people do."[84] On the third and fourth factors, authors argue that the AI companies copy more than necessary, and the extensive copying in turn dilutes the market value of their work.[85]

Defendants counter that the use is fair because it is largely transformative.[86] The purpose being transformative helps defendants establish factors one, three, and four of their defense because it refrains from "taking the heart" of the work and therefore does not conflict with the original source's market.[87]

---

80. *See id.* (citing RUSSEL & NORVIG, *supra* note 69, at 694–95).

81. *See generally* Van Lier, *supra* note 2 ("This training or 'input' phase . . . raises critical copyright questions. During training, copies of these works are . . . encoded as part of the AI model's internal functioning.").

82. *See* ZIRPOLI, *supra* note 7 ("Copyright owners have filed several dozen lawsuits claiming that creating these digital copies without permission to train AI systems infringes their exclusive right to make reproductions (or copies).").

83. *See* Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007, 1021–22 (N.D. Cal. 2025).

84. *See id.* at 1022 (referring to arguments made by plaintiffs referencing Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc. 765 F. Supp. 3d 382, 398 (D. Del. 2025)).

85. *Id.* at 1030–32 (contending that the use in AI training will result in "an explosion of works competing with their works").

86. *See id.* at 1021.

87. *See generally Thomson Reuters*, 765 F. Supp. 3d at 398–400 (noting that this decision was influenced by the fact that the copying was not transformative and thus invaded the potential market of the original).

C. WHAT COURTS HAVE RULED

Recently, courts have addressed whether AI training constitutes fair use based on arguments made by both plaintiffs and defendants.

The first case ruling on fair use and AI was *Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc*.[88] In this case, Thomson Reuters, the company responsible for the legal database Westlaw, accused Ross of copying Westlaw data to train their own fledgling legal AI program.[89] This data included the headnotes of Westlaw, which are original editorializations.[90] Thomson Reuters filed for summary judgment on copyright infringement and fair use, while Ross filed a motion for summary judgment in their favor for copyright infringement and fair use.[91] The court ruled in favor of Thomson Reuters by granting summary judgment on fair use.[92] It reasoned that Thomson Reuters prevailed on fair use factors one and four, while Ross prevailed on factors two and three.[93] Factor four weighs more importantly than the others, while factor two weighs less than the others, so after the court balanced those factors, it ruled that Ross's use was not fair.[94]

The next case concerning fair use and AI was *Bartz v. Anthropic PBC*.[95] In this case, AI firm Anthropic downloaded millions of copyrighted books, some legally purchased and some illegally pirated, to train its AI system.[96] The plaintiffs were authors whose works were put into the AI system.[97] The plaintiffs sued for copyright infringement while the defendants claimed fair use.[98] The court held that the use of the non-pirated works was fair.[99] The court reasoned that the work being taken from books to simply give the AI knowledge to understand word patterns was a transformative use, which also went to fair use factors three and four.[100]

---

88. *See generally id.* (holding that AI training using copyrighted works was not fair use).

89. *See id.* at 391.

90. *See id.* at 390–91.

91. *See id.* at 391.

92. *Id.* at 401 (granting partial summary judgment on copyright infringement to remand for remaining factual questions and granting full summary judgment against Ross's pled defenses, including fair use).

93. *Id.*

94. *See supra* notes 53, 61 and accompanying text.

95. *See generally* Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (holding that training AI on lawfully acquired work was fair use).

96. *See id.* at 1014.

97. *Id.*

98. *Id.* at 1014, 1018.

99. *Id.* at 1034.

100. *See id.* at 1022, 1029–30, 1032.

Shortly thereafter, the court in *Kadrey v. Meta Platforms, Inc.* made a similar decision.[101] Meta Platforms has an AI system called Llama that used the works of Kadrey in its training.[102] Here, the court ruled that the use was fair, but left open for interpretation the fourth factor, which could sway future rulings.[103] The court said it is possible that with the possession of the original works, there remains the potential for the system to create derivative works which could dilute the market for the original.[104] This would make the fourth factor weigh against fair use and could sway future cases, but the plaintiffs failed to argue on that point.[105]

IV. USE OF COPYRIGHTED MATERIAL IN AI TRAINING IS NOT FAIR USE

A. FAIR USE ELEMENTS APPLIED TO AI TRAINING

The Supreme Court has described fair use as a "flexible" doctrine that should be approached on a case-by-case basis.[106] However, in the case of AI training, leaving the analysis to different judges who hold different understandings of AI can lead to rulings at odds with each other.[107] In a controversial recent executive order, the President of the United States articulated that the United States should have a unified approach concerning the industry of AI and dissuaded states from making their own laws and decisions about AI.[108] To be in line with this unified approach, the issue of AI training should become uniform instead of taking a case-by-case approach. This subsection will analyze each fair use factor as related to AI training and explain what courts have gotten wrong so far.

For the first factor, the purpose and character of the use must be analyzed.[109] Courts have determined that the sub-elements of this factor are

101. *See generally* Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026 (N.D. Cal. 2025) (holding that Meta's AI training was fair use).

102. *See id.* at 1042.

103. *See generally id.* at 1055 (pointing to an argument outlined in dicta from the ruling that the judge would find convincing in this and other cases concerning AI and fair use).

104. *See id.* (stating that the ability of AI to create derivative works from the use of copyrighted material at a much faster rate than the original author could have massive implications on the potential market of the work).

105. *See id.*

106. *See* ZIRPOLI, *supra* note 7.

107. *See generally Kadrey*, 788 F. Supp. 3d 1026 (showing a case where the judge paused to consider the implications and potential arguments for the fourth factor); Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (showing a case where the judge dismissed similar arguments that the judge in *Kadrey* opined about).

108. *See* Exec. Order No. 14365, 90 Fed. Reg. 58499 (Dec. 11, 2025) (likening the national strategy of lawmaking concerning AI to the courts and how that approach should follow).

109. *See* 17 U.S.C. § 107(1).

whether the use is commercial and whether it is transformative.[110] With AI being the burgeoning industry it is, there is no question as to whether the purpose is commercial, as having AI with more data to pull from leads to better responses and better market value.[111] Thus, this factor depends on whether AI training is a transformative use.[112] Different courts have arrived at different conclusions.[113] In *Reuters*, the court ruled that Ross's use of Thomas Reuters's Headnotes was not transformative because it was building its own competing legal database, which is not a "further purpose or different character" than the original use.[114] Whereas the court in *Bartz* held that the use of books was "quintessentially transformative" because, just like humans can, the AI program was simply reading and extrapolating patterns to learn.[115] On this point, it seems the first factor favors fair use by the AI companies, as the catalogs the companies downloaded were so vast that it cannot be said that each individual work had the same purpose as the AI program.[116]

For the second factor, courts must consider the nature of the work and measure its creativity and originality as a whole.[117] Though courts can debate the degree of creativity in each work, it remains true that copyrightable material contains at least some degree of creativity.[118] Since the works downloadable for AI training can span from novels to fact books, and published and unpublished works, this factor inherently favors original

110. *See* Warhol Found. Visual Arts v. Goldsmith, 598 U.S. 508, 527–29 (2023).

111. *See generally* NJENGA KARIUKI, CHAPTER 4: ECONOMY, ARTIFICIAL INTELLIGENCE INDEX REPORT 2025, STAN. U. HAI (2025), hai_ai-index-report-2025_chapter4_final.pdf [https://perma.cc/X33K-4FYK] (addressing the commercial advantage and strategy AI training has as weight in favor of commercial, transformative use).

112. *See supra* notes 46–49 and accompanying text.

113. *See generally* Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382 (D. Del. 2025); Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (showing different outcomes in the ruling of fair use).

114. *Thomson Reuters*, 765 F. Supp. 3d at 397–98.

115. *Bartz*, 787 F. Supp. 3d at 1021–22.

116. *See id.* at 1022–23.

117. 17 U.S.C. § 107(2); *Thomson Reuters*, 765 F. Supp. 3d at 399 (citing 4 NIMMER ON COPYRIGHT § 13F.06).

118. *See Bartz*, 787 F. Supp. 3d at 1029 (stating that "all of Authors' books – all published, whether non-fiction or fiction – contained expressive elements").

works.[119] Though it is the least compelling factor, this factor still weighs against fair use.[120]

The third factor of amount and substantiality asks courts to weigh how much the interpolated work "takes the heart" of the original.[121] However, that is not necessary in most cases of AI training, as the AI companies are downloading entire works for their training.[122] Court rulings on this factor seem to look too broadly at the overarching AI training.[123] Though the overall AI program is minimally made up of each individual work, the program has taken the entirety of the copyrighted work.[124] Ruling that this factor weighs in favor of AI companies misses the forest for the trees, because each individual work's heart was taken to create the larger heart of the AI program's learning.[125] This factor should weigh against fair use in AI training.

Finally, for the fourth factor, courts look to whether the alleged copy affects the market for the original work.[126] This factor is the crux of whether use is fair because it concerns the profitability of the original work, which infringement law is supposed to protect.[127] As such, this factor is weighed the most heavily in decisions about fair use.[128] This factor is also largely linked with the third factor because how substantially the work is available affects its individual market.[129] Additionally, this factor considers not only

---

119. *See id.* (explaining this factor is partially meant to set up assessment for the factors that follow, such as highlighting the difference between the nature of the original work and the nature of the secondary work for comparison of substantiality taken).

120. *See Thomson Reuters*, 765 F. Supp. 3d at 399 (citing Authors Guild v. Google, Inc., 804 F.3d 202, 220 (2d Cir. 2015) (stating this factor rarely plays a significant role in the outcome of the case)).

121. *See generally* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 587 (1994) ("We conclude that taking the heart of the original and making it the heart of a new work was to purloin a substantial portion of the essence of the original.").

122. *See Bartz*, 787 F. Supp. 3d at 1014 ("An artificial intelligence firm downloaded . . . millions of copyrighted books . . . .").

123. *See id.* at 1029–31 (ruling that the use of copyrighted works was fair for AI training because the works were "reasonable in relation to the purpose of the copying"); *Campbell*, 510 U.S. at 586.

124. *See generally Bartz*, 787 F. Supp. 3d at 1030 (expressing that "entire works were copied").

125. *Contra id.* at 1029–31 (opposing the view of the court that the small heart of each work in comparison to the millions downloaded means it should not be substantial).

126. *See* 17 U.S.C. § 107(4).

127. *See* Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985) (emphasizing the central purpose of copyright law is to incentivize creativity with profit); Gillotte, *supra* note 12, at 2666.

128. *See Harper & Row*, 471 U.S. at 566.

129. *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 587 (1994) ("The facts bearing on this [third] factor will also tend to address the fourth, by revealing the degree to which the [work] may serve as a market substitute . . . .").

the original work's actual market, but also potential markets that could occur from derivative work by the author.[130]

Though copyright questions in outputs of AI are an entirely separate subject, it is necessary to consider when analyzing the potential market effect of AI training.[131] Since AI training stores the entire original work in its database, it can pull details from that work that can affect its market value.[132] Among the complainants in the *Authors Guild v. OpenAI, Inc*. class action lawsuit is *Game of Thrones* author George R. R. Martin, who alleges that the AI program ChatGPT wrote a sequel to his books using his copyrighted intellectual property.[133] Martin alleges that ChatGPT was able to write this sequel in his style because his books had been downloaded without his permission to train the AI.[134] Even though this case is pending, it shows there is a reasonable belief that LLMs (language learning models) downloading complete works can have the potential to create derivative works, thereby affecting the potential market for the original works.[135] This factor should weigh against fair use.[136] In sum, factors two, three, and four weigh against a finding of fair use, while only factor one favors it, suggesting that the use of copyrighted materials in AI training does not qualify as fair use.[137]

B. RECENT TRENDS IN COURTS

Judicial treatment of AI training indicates a trend towards outcomes becoming increasingly unfavorable to AI under the fair use doctrine. Initially, courts and scholars were deferential to the newly expanding AI industry, in part due to unfamiliarity with the complexity of AI legal issues.[138] In fact, several cases at the federal district court level saw dismissals at the summary

130. *See id.* at 592.

131. *See* Class Action Complaint & Jury Trial Demanded ¶ 2, Authors Guild v. OpenAI, Inc., No. 1:23-cv-8292, (S.D.N.Y. Sep. 19, 2023) (stating that "Defendants' LLMs can spit out derivative works: material that is based on, mimics, summarizes, or paraphrases Plaintiffs' works, and harms the market for them").

132. *See id.* ¶ 243 (referencing a study done that measured the "memorization" capability of ChatGPT).

133. *See id.* ¶ 242.

134. *See id.* ¶¶ 241–45.

135. *See id.* ¶¶ 2–3.

136. *See id.* ¶ 4.

137. *See generally* Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382, 401 (D. Del. 2025) (showing that courts should balance the factors after deciding each one to determine if the fair use defense stands).

138. *See generally id.* at 391 (asserting that this judge was previously quick to dismiss the case, but as he learned more about AI, the case grew more worthy of consideration).

judgment stage in favor of defendants.[139] However, courts became more open to the possibility of infringement by AI as more suits were filed.[140] The year of 2025 was when courts began to consider and make rulings on the possibility of fair use.[141]

The *Reuters* case specifically shows a marked change in how courts perceive the possibility of AI.[142] In this case, the court initially denied the plaintiff's motions for summary judgment on infringement and fair use.[143] However, as the case proceeded to trial, the court learned more about AI training and the often pernicious practices of AI companies.[144] After doing so, the court reconsidered its own decision and partially reversed it, granting summary judgment for Thomson Reuters on their infringement and fair use motions.[145]

In *Bartz v. Anthropic PBC*, the defendants were able to partially prevail on their defense of fair use.[146] In this case, the court was willing to hold that the use of legally obtained copies was fair, but ruled against their use of pirated works.[147] In *Kadrey v. Meta Platforms, Inc.*, the court ruled that the use was fair, but left open the possibility for future findings against fair use.[148] The court here said that there is an argument for the fourth fair use factor to weigh against similar cases, as AI programs are capable of "gener[ating] literally millions of secondary works . . . . [n]o other use . . . has anything near the potential to flood the market with competing works the way [AI] training does."[149] The court was unable to act upon that thought as

139. *See id.* (initial incorrect ruling of partial summary judgment, decided to go farther); Andersen v. Stability AI Ltd., 700 F. Supp. 3d 853, 868 (N.D. Cal. 2023) (dismissing the infringement claim against DeviantArt but retaining the infringement claim against Stability).

140. *See, e.g.*, Tremblay v. OpenAI, Inc., 716 F. Supp. 3d 772 (N.D. Cal. 2024); N.Y. Times Co. v. Microsoft Corp., 777 F. Supp. 3d 283 (S.D.N.Y. 2025) (showing examples of cases where the issues were heard and considered).

141. *See, e.g.*, *Thomson Reuters*, 765 F. Supp. 3d 382 (ruling against fair use for the use of editorial headnotes); Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025) (ruling in favor of fair use in AI training); Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026 (N.D. Cal. 2025) (ruling in favor of fair use in AI training, but leaving open the possibility of future rulings against).

142. *See generally Thomson Reuters*, 765 F. Supp. 3d 382 (showing a direct change in thinking from a judge who had previously ruled on fair use to alter his decision).

143. *Id.* at 391.

144. *See id.*; *see also Bartz*, 787 F. Supp. 3d at 1019 (asserting the factual basis that some of the copyrighted works were pirated illegally, referencing the pernicious practice).

145. *Thomson Reuters*, 765 F. Supp. 3d at 401.

146. *See Bartz*, 787 F. Supp. 3d at 1034.

147. *See id.*

148. *See generally* Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1055 (N.D. Cal. 2025) (addressing the dicta used by the judge to outline how he thought plaintiffs should have argued this factor).

149. *Id.*

it was not an argument brought by the plaintiffs, but this shows the potential position courts might have if someone were to argue that position.[150]

Even though these cases had varying outcomes and language used by the courts, they all point toward a growing sense of skepticism toward AI training as it is being performed today.[151] Whether it is the *Bartz* court's admonition of the pirated copies, the *Kadrey* court's pondering of arguments against the fourth fair use factor, or the *Reuters* court revising itself, these all show that courts are gradually becoming more open to the idea that AI companies are violating copyright law.[152] As more cases get decided and more suits filed, as well as with President Trump's executive order demanding national unity concerning AI strategy, it will become paramount for courts to grasp a unified approach to these increasing alleged transgressions.[153]

C. GREATER PROTECTIONS FOR COPYRIGHT HOLDERS' RIGHTS

Finally, the last question to ponder is why this issue is so contentious and why it matters. Fair use is a decades-old doctrine meant to protect people who use copyrighted material for legitimate purposes.[154] If this is the purpose of fair use, why should AI receive more scrutiny under the fair use doctrine? The answer surrounds the fact that copyright law is meant to protect and promote *human* creation.[155] There is a broad public interest in supporting the creation of arts and literature, as it enhances culture and provides leisure.[156] Previous courts have held that computers acting at the direction of programmers cannot hold copyright.[157] AI is a different form of technology,

---

150. *See generally id.* (connecting the judge's words to the logical step of pursuing his suggested argument).

151. *See generally* Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382 (D. Del. 2025) (revising the earlier decision upon closer examination of the issues with AI); *Bartz*, 787 F. Supp. 3d 1007 (refusing to treat pirated works as training copies); *Kadrey*, 788 F. Supp. 3d 1026 (opining about arguments the plaintiffs should have made on factor four of fair use).

152. *Compare Kadrey*, 788 F. Supp. 3d 1026; *with Bartz*, 787 F. Supp. 3d 1007; *and Thomson Reuters*, 765 F. Supp. 3d 382 (showing all three courts' collective shifts against AI companies).

153. *See, e.g.*, Class Action Complaint & Jury Trial Demanded, Authors Guild v. OpenAI, Inc., No. 1:23-cv-8292, (S.D.N.Y. Sep. 19, 2023) (expressing the ongoing state of the issue that is still actively being litigated).

154. *See* 17 U.S.C. § 107 (citing Google LLC v. Oracle Am., Inc., 593 U.S. 1, 20 (2021) (outlining the stated purpose for the fair use defense)).

155. *See* Gillotte, *supra* note 12, at 2658–59 (emphasis in original).

156. *See generally* Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) (positing that "the ultimate aim [of copyright law] is . . . to stimulate artistic creativity for the general public good").

157. *See* Gillotte, *supra* note 12, at 2667 (stating that computer programmers or users can hold copyright for works made by computers, not the computer itself because it has no "legal personhood") (first citing Annemarie Bridy, *Coding Creativity: Copyright and the Artificially*

however, as it simulates the ability to think for itself and is trained to mimic original human writing.[158] For the first time in history, humans have competition in creativity, and AI can do it with a fraction of the time and effort.[159] This competitor could severely depress the incentive for creativity and dilute the market with AI-generated works based on truly original works.[160] Fair use's inherent flexibility should be skewed toward protecting human creators, and a way to do that is by holding that fair use should weigh against AI training.[161] Another potential policy solution could be requiring agreements, not simply valid purchases, between AI trainers and copyright holders for use in AI training.[162]

V. CONCLUSION

The interaction between copyright law and AI has caused discussion and contention in recent years.[163] Creators and copyright holders worry about the depressed value of their creativity in an era where AI can recreate their works in seconds.[164] The complex issue of whether AI training is fair use has led to conflicting opinions in the federal court system.[165] With over fifty active lawsuits, intellectual property lawyers, judges, and lawmakers should familiarize themselves with these issues and work to find solutions that protect and promote human creativity.[166]

Intelligent Author, 2012 STAN. TECH. L. REV. 5, 21; and then citing U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 306 (3d ed. 2017)).

158. *See id.* at 2660.

159. *See* Kadrey v. Meta Platforms, Inc., 788 F. Supp. 3d 1026, 1055 (N.D. Cal. 2025) ("This case, . . . involves a technology that can generate literally millions of secondary works, with a miniscule fraction of the time and creativity used to create the original works it was trained on.").

160. *See id.*

161. *See generally id.* (stating that "[i]ndeed, it seems likely that market dilution will often cause plaintiffs to decisively win the fourth factor—and thus win the fair use question overall").

162. *See* Van Lier, *supra* note 2 ("Licensing is frequently discussed as a potential remedy for copyright issues triggered by GenAI's training stage. In creative industries, licensing would necessarily precede the training stage as the first step in a model's life cycle. This would give copyright owners the opportunity to opt-out of any potentially unwanted uses of their copyrighted works, assuming adequate notice. To date, licensing arrangements between creators and GenAI companies have been negotiated on a purely voluntary, bilateral basis—no court has ordered such licensing, nor has Congress or any regulatory agency mandated it as a prerequisite to training.").

163. *See generally id.* (outlining that there are over fifty active lawsuits between various authors' groups and AI companies).

164. *See generally Kadrey*, 788 F. Supp. 3d at 1055 (reiterating that AI can recreate and derive from original works with a fraction of the time and effort it takes a human author to write).

165. *Compare* Thomson Reuters Enter. Ctr. GMBH v. Ross Intel. Inc., 765 F. Supp. 3d 382 (D. Del. 2025), *with* Bartz v. Anthropic PBC, 787 F. Supp. 3d 1007 (N.D. Cal. 2025), *and Kadrey*, 788 F. Supp. 3d. 1026 (looking at the three central cases on AI and fair use to compare their reasoning, attitude, and outcome).

166. *See* Van Lier, *supra* note 2; *see also supra* Section IV.C (emphasizing the need for greater protections of copyright in the age of AI).

Gabriel Almlie[*]

[*] 2027 J.D. Candidate at the University of North Dakota School of Law. Thank you to Editors and Members of the NORTH DAKOTA LAW REVIEW for their tireless work in getting this piece ready for publication. Thank you to all my music and writing teachers for fostering a passion for creativity and expression in me, which heavily influenced my decision to write about this issue. Thank you to my friends and family for their continued support. Finally, a special thank you to Professor Nikola Datzov for serving as a subject matter consultant and helping me solidify a narrow, interesting, and relevant topic to explore.