

TRANSFORMING TRAINING GENERATIVE ARTIFICIAL INTELLIGENCE

FARAH JONI[†]

I. INTRODUCTION	676
II. BACKGROUND	678
<i>A. How are These Machines so Smart?</i>	678
<i>B. History and Purpose Behind Copyright Law and Fair Use</i>	680
<i>C. The Origin of Transformative Use</i>	682
<i>D. The Evolution of Transformative Use</i>	683
<i>E. Intermediate Copying</i>	686
<i>F. Supreme Court Developments</i>	687
III. ANALYSIS	690
<i>A. Introduction to GenAI Transformative Use</i>	690
<i>B. The Role of Intermediate Copying in GenAI Training</i>	691
<i>C. GenAI: Transforming Through Distinct Purpose and Utility</i>	693
<i>D. Transformative Training</i>	696
<i>E. Policy</i>	698
IV. CONCLUSION.....	699

[†] B.S.B.A, 2020, *summa cum laude*, Wayne State University; J.D., expected 2025, Wayne State University Law School.

I. INTRODUCTION

Generative artificial intelligence (GenAI) has completely transformed the landscape of content creation.¹ In this context, GenAI does not refer to *Terminator*-like machines that come to life for the destruction of humanity.² Instead, GenAI refers to deep learning (DL) models that allow computers to accomplish human-like tasks.³ In artificial intelligence training, DL is a method where programmers teach computers how to complete tasks and process data akin to the human brain.⁴ Typical GenAI models are structured like chatbots and respond to human-entered prompts to produce output based on their training.⁵

Despite these generative models' success within several industries,⁶ content creators have criticized GenAI.⁷ Creators are not fearful of an apocalyptic rise of machines taking over the world, or even their jobs, but instead accuse GenAI developers of a lesser evil: copyright infringement.⁸ Creators have raised concerns about software developers using their copyrighted content in training GenAI models and have brought copyright infringement suits against GenAI developers.⁹ The outcome of these suits

1. *Is Generative AI a Game Changer?*, J.P. MORGAN RSCH. (Mar. 20, 2023), <https://www.jpmorgan.com/insights/global-research/artificial-intelligence/generative-ai#:~:text=What%20are%20the%20impacts%20of,new%20business%20models%20and%20applications> [https://perma.cc/EM5B-8QM6] [hereinafter *Generative AI*].

2. Michael Goldberg, *What the Terminator Teaches Us About AI and the Need for Better Data*, AUTHORITY (Apr. 28, 2020), <https://aithority.com/guest-authors/what-the-terminator-teaches-us-about-ai-and-the-need-for-better-data/> [https://perma.cc/YB7S-YPGD].

3. Jim Holdsworth & Mark Scapicchio, *What is Deep Learning?*, IBM (Jun. 17, 2024), <https://www.ibm.com/topics/deep-learning> [https://perma.cc/24W3-HHQ3] (defining deep learning as "a subset of machine learning that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain.").

4. *What is Deep Learning?*, AMAZON WEB SERVS., <https://aws.amazon.com/what-is/deep-learning/#:~:text=Deep%20learning%20is%20a%20method,produce%20accurate%20insights%20and%20prediction> [https://perma.cc/SN3Y-67PR].

5. See *infra* Part II.A.

6. See, e.g., *Generative AI*, *supra* note 1, (describing how different industries use GenAI: the financial sector utilizes it to analyze market data and health care industries use GenAI generate data to assist with pharmaceutical testing).

7. Dylan Walsh, *The Legal Issues Presented by Generative AI*, MIT SLOAN (Aug. 28, 2023), <https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai> [https://perma.cc/N5EE-TMXE].

8. *Id.*

9. See, e.g., *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223, 2024 U.S. Dist. LEXIS 24618 (N.D. Cal. Feb. 12, 2024); see also Complaint, *Chabon v. OpenAI, Inc.*, 3:23-cv-04625 (court Sept. 8, 2023); Complaint, *Silverman v. OpenAI, Inc.*, 3:23-cv-03416 (court July 7, 2023); Complaint, *Getty Images, Inc. v. Stability AI, Inc.*, 23-cv-00135 (court Feb. 3, 2023); Joe Panettieri, *Generative AI Lawsuits Timeline: Legal Cases vs. OpenAI*,

could potentially reshape the GenAI landscape by limiting lawful GenAI systems to those solely trained on public domain data or under licenses.¹⁰

GenAI developers' best strategy for defending against copyright infringement lawsuits is to invoke the fair use defense.¹¹ The fair use doctrine is an affirmative defense to copyright infringement claims.¹² It permits an individual to use a copyrighted work without infringing on the owner's copyrights in limited circumstances, including reproduction in copies or phonorecords for purposes like criticism, commentary, news reporting, teaching with multiple copies for classroom use, scholarship, or research.¹³

The fair use doctrine is codified in the Copyright Act and lays out four factors that courts use to determine whether a party accused of copyright infringement has a valid fair use defense: (1) the purpose and character of the use, including commercial nature, also known as transformative use; (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.¹⁴ Courts balance these factors and apply them on a case-by-case basis to determine whether a copier used an author's work fairly, while favoring works that add something new with a different purpose or character than the original work.¹⁵

Professor Edward Lee categorizes copyright use in technology into three categories: creational, operational, and output.¹⁶ Creational uses involve using copyrighted works exclusively in the development phase of technology, typically through reverse engineering.¹⁷ This type of use focuses on copying *functional* aspects of the software in order to create new technology.¹⁸ Operational uses are also functional, but unlike creational uses, they take place after the technology has already been created.¹⁹ An example of an operational use is internal processing or

Microsoft, Anthropic and More, SUSTAINABLE TECH. PARTNER (Jan. 26, 2024), <https://sustainabletechpartner.com/topics/ai/generative-ai-lawsuit-timeline/> [<https://perma.cc/ZUH8-3G98>] (giving a timeline of generative artificial intelligence lawsuits filed).

10. Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 745 (2021).

11. 17 U.S.C. § 107.

12. *Id.*

13. *Id.*

14. *Id.*

15. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994); *see, e.g., id.* at 579; *see also Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1187 (2021) (explaining that the fair use doctrine is flexible, and its application varies depending on the context).

16. Edward Lee, *Technological Fair Use*, 83 S. CAL. L. REV. 797, 842 (2010).

17. *Id.*

18. *Id.*

19. *Id.* at 843.

temporary copying for transmission purposes.²⁰ Finally, an output use refers to the distribution, display, or performance of the copyrighted work to the public, such as through photocopying or broadcasting.²¹ Due to GenAI's unique capabilities, developers training it do not neatly fit into any one of these categories. GenAI developers engage in a combination of creational use and output use in training GenAI systems.

This Note will focus on the first factor of the fair use defense, the purpose and character of use, to show that a developer's use of unlicensed data in training GenAI is inherently transformative. Next, it will explore traditional applications of transformative use and what it means for programmers using copyrighted works to train GenAI in the current legal landscape. Finally, this Note will conclude with suggested legislative measures to protect and codify GenAI's training data use.

II. BACKGROUND

A. How are These Machines so Smart?

Understanding how GenAI developers process training data is pivotal in determining whether GenAI uses unlicensed copyrighted works lawfully. GenAI owes its success to DL and neural networks.²² Neural networks are complex computing processes that analyze patterns and arrangements in large data sets to predict new output based on a given prompt.²³ For example, if developers feed a GenAI model vast amounts of stories, the model will eventually be capable of replicating key story elements.²⁴ These elements include the structure of the plot, the nature of the characters, underlying themes, and various narrative techniques.²⁵ Over time, the model becomes proficient in recognizing and recreating the storytelling.²⁶ The same concept applies to non-text GenAI platforms, which include image and music GenAI models.²⁷

20. *Id.*

21. *Id.*

22. Kim Martineau, *What is Generative AI?*, IBM: RESEARCH (Apr. 20, 2023), <https://research.ibm.com/blog/what-is-generative-AI> [<https://perma.cc/6XN5-TH73>].

23. Owen Hughes, *Generative AI Defined: How it Works, Benefits and Dangers*, TECHRUPUBLIC (Aug. 7, 2023), <https://www.techrepublic.com/article/what-is-generative-ai/> [<https://perma.cc/R38E-FHTM>] (explaining how generative AI is trained).

24. *Id.*

25. *Id.*

26. *Id.*

27. James Vincent, *All These Images were Generated by Google's Latest Text-to-Image AI*, THE VERGE (May 24, 2022), <https://www.theverge.com/2022/5/24/23139297/google-imagen-text-to-image-ai-system-examples-paper> [<https://perma.cc/59ME-5ZEF>].

Both sophisticated industries and laypeople can access GenAI.²⁸ For instance, consumer industries use GenAI as a marketing tool to personalize experiences, content, and product recommendations.²⁹ Laypeople use GenAI for just about *anything*. Perhaps one wonders what it would look like for a cat dressed in a Frankenstein costume to sip on an espresso. By entering “cat sipping on espresso dressed in a Frankenstein costume” in an image-generating GenAI chat box, like DALL-E, the intelligence model will generate a state-of-the-art photograph that turns the silly thought into reality within mere seconds.³⁰ These text-to-image GenAI models are capable of producing images with different styles, limited only by a user’s imagination.³¹

A single author’s work may not majorly impact a GenAI program because these DL models require millions of data points for training.³² Further, the data goes through many layers of training to effectively produce output: the input layer, the hidden layer, and the output layer.³³ The first layer is the input layer, where a developer feeds the original data into the GenAI system.³⁴ The second is the hidden layer, where the DL model processes and analyzes data.³⁵ This is where the real magic happens. In the hidden layer, GenAI systems learn to recognize patterns and adapt their behavior accordingly.³⁶ The hidden layer process can be conceptualized by thinking about the human brain.³⁷ If a person looks at a picture of an animal they have never seen before and needs to identify it, they compare the unknown animal’s features with those of animals they already know.³⁸ One might look at its eyes, ears, size, number of legs, and fur pattern to compare it to other animals.³⁹ If one sees hooves, one might think it is a cow or a deer.⁴⁰ If the eyes look like a cat’s, one might identify

28. Michael Chui et al., *The Economic Potential of Generative AI: The Next Productivity Frontier*, MCKINSEY DIGITAL (Jun. 14, 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> [https://perma.cc/YP96-9XVR].

29. *Generative AI*, BOSTON CONSULTING GROUP, <https://www.bcg.com/capabilities/artificial-intelligence/generative-ai> [https://perma.cc/ZLS9-YTBL] (last visited Jan. 2, 2024) (defining AI and explaining how it is trained and utilized).

30. Vincent, *supra* note 27.

31. *Id.*

32. *Generative AI*, *supra* note 1.

33. *What is Deep Learning?*, *supra* note 4.

34. *Id.*

35. *Id.*

36. *Id.*

37. *Id.*

38. *Id.*

39. *What is Deep Learning?*, *supra* note 4.

40. *Id.*

it as a type of wildcat.⁴¹ DL neural networks use hidden layers similarly.⁴² When a DL algorithm identifies an animal image, each hidden layer processes a distinct animal feature and attempts to categorize it accurately.⁴³ Once the algorithm categorizes the animal, it generates a response through the output layer.⁴⁴

As fascinating as this learning process is, it will often require developers to use copyrighted data to train, or teach, the GenAI model. The unauthorized use of copyrighted training data raises several copyright infringement issues.⁴⁵ First, because the training process involves making digital copies of original works as an intermediate step, these copies may infringe on the copyright holders' exclusive right to reproduce their work.⁴⁶ Moreover, there are concerns as to whether the copyrighted training data will reappear in the GenAI output.⁴⁷ Fair use principles and legislative intervention can provide solutions to these concerns.

B. History and Purpose Behind Copyright Law and Fair Use

The framers of the Constitution established copyright protection to advance innovation.⁴⁸ Copyright protection, enumerated in Article I, Section 8, Clause 8 of the U.S. Constitution, actively promotes the advancement of science and useful arts while protecting an original creator's work.⁴⁹ It recognizes that intellectual creativity is vital for society's growth.⁵⁰ In exchange, the law incentivizes creators by permitting them to profit from their work and affords six exclusive rights currently codified in 17 U.S.C §106.⁵¹ These exclusive rights include the ability to reproduce, distribute, perform, display, and create derivative works from their original creation and for sound recordings, and the ability to perform the copyrighted work publicly through digital audio transmission.⁵²

41. *Id.*

42. *Id.*

43. *Id.*

44. *Id.*

45. Christopher T. Zirpoli, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3 (2023), <https://crsreports.congress.gov/product/pdf/LSB/LSB10922> [<https://perma.cc/U2SZ-S9R4>].

46. *See Id.*; 17 U.S.C. § 106.

47. Zirpoli, *supra* note 45.

48. U.S. CONST. art. I, § 8, cl. 8.

49. *Id.*

50. Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1109 (1990) (defining the goals of copyright).

51. *Id.* at 1107; 17 U.S.C. § 106.

52. 17 U.S.C. § 106.

Although preserving creative work is important, granting creators excessively broad protection hinders copyright law's objective of promoting creativity and authorship in society.⁵³ If the law prohibits any kind of unauthorized use of a copyrighted work, the original author would have a "copyright monopoly."⁵⁴ Moreover, allowing unlimited exclusive rights to copyright protection is impractical because all creative work builds on previous ideas—nothing is entirely new.⁵⁵ Innovators take existing knowledge and ideas to create something that might seem new but is actually a development of a past work.⁵⁶ Disciplines like philosophy, literary criticism, history, and natural sciences continually revisit and build upon past ideas and theories.⁵⁷

To address this concern, Congress limited the scope of the potential "copyright monopoly"⁵⁸ through the fair use defense to copyright infringement.⁵⁹ The defense allows individuals to use copyrighted material without the copyright owner's permission if it benefits society in certain ways.⁶⁰ Such benefits include using parts of a copyrighted work for education, commentary, or research.⁶¹ Fair use is thus an essential part of copyright law because it encourages the spread of knowledge and innovation without significantly harming the copyright owner's ability to profit from their work.⁶² Judge Pierre Leval suggested that when courts consider whether use of a copyrighted work is fair use, they should always look back to the fundamental goal of copyright law: stimulating creativity and learning.⁶³

In applying the fair use doctrine, courts consider four factors to evaluate whether the unlicensed use of a copyrighted work is fair use: (1) the purpose and character of the use, including commercial nature, also known as transformative use; (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 use's effect on the copyrighted work's potential market or value.⁶⁴ Courts have typically balanced the four statutory fair use factors and applied them on a case-by-case basis.⁶⁵

53. Leval, *supra* note 50, at 1109.

54. *Id.*

55. *Id.*

56. *Id.*

57. *Id.*

58. *Id.* at 1110.

59. 17 U.S.C. § 107.

60. *Id.*

61. *Id.*

62. Leval, *supra* note 50, at 1109.

63. *Id.*

64. 17 U.S.C. § 107.

65. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994).

C. The Origin of Transformative Use

Judge Leval famously coined the term “transformative use” in copyright law to describe a way of using copyrighted material to add something new, with a different purpose or character, when compared to an original work.⁶⁶ For a use to be transformative, it must alter the original work with new insights or meaning.⁶⁷ Case law has further elaborated on this idea.⁶⁸

In *Campbell v. Acuff-Rose Music, Inc.*⁶⁹, the Supreme Court first applied Judge Leval’s interpretation of transformative use, holding that a commercial parody of a song was fair use under copyright law, even if it was for profit, because it transformed the original work in a significant way.⁷⁰ Acuff-Rose Music, Inc., sued the rap group 2 Live Crew and its record company for allegedly infringing Acuff-Rose Music’s copyright of Roy Orbison’s song “Oh, Pretty Woman” with a version titled “Pretty Woman.”⁷¹ Roy Orbison’s original song is a romantic rock ballad celebrating the beauty of a woman.⁷² In contrast, 2 Live Crew’s version is a rap parody that uses a similar tune, but changes the lyrics to include mocking language criticizing the original’s naive romanticism by contrasting it with the stark realities of urban street life.⁷³ This difference in message and tone, along with the purpose of parody and commentary, was central to the Court’s finding of fair use.⁷⁴ Furthermore, the Court emphasized that parodies inevitably mimic an original work to make their point, so they may have some claim to use of the original work.⁷⁵

The Court reasoned that when applying the fair use defense to a parody, the key question is whether one can *reasonably perceive* the character of use as parodic.⁷⁶ The Court held that one could reasonably perceive 2 Live Crew’s song as commenting on or criticizing the original song by substituting predictable lyrics with shocking ones.⁷⁷ The group made fun of how dull and ordinary the original song seemed to them.⁷⁸

66. Leval, *supra* note 50, at 1111.

67. *Id.*

68. See *infra* notes 69–83 and accompanying text; see discussion *infra* Sections II.C, II.E, and II.F.

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

70. *Id.* at 572.

71. *Id.* at 571.

72. *Id.* at 583.

73. *Id.*

74. *Id.*

75. *Campbell*, 510 U.S. at 581.

76. *Id.* at 582.

77. *Id.* at 574.

78. *Id.* at 583.

The Court clarified that transformative works lie at the heart of fair use not only by giving “breathing space within the confines of copyright,” but ultimately, by furthering the sciences and the useful arts.⁷⁹

The Court justified the parody’s transformative value and benefit to society because the parody highlighted aspects of an earlier work while creating a new work in the process.⁸⁰ The Court emphasized that the more transformative a new work is, the less significance it will afford to other factors that weigh against fair use in the first prong, like commercialism.⁸¹ In other words, *even if* a work has commercial purposes, it can still be fair use if it adds something new with a transformative purpose, like a parody critiquing the original work.⁸² Other courts have followed the *Campbell* precedent by finding parodies to be transformative when their purpose is to ridicule an original work for comedic purposes.⁸³

After the Supreme Court’s decision in *Campbell*, courts must analyze whether a new work is transformative by considering how much the current use differs in purpose or character from the original—the greater the difference between the original work and the new work, the more it supports transformative use.⁸⁴ In a transformative use analysis, courts do not solely weigh physical changes to distinguish an original work from new work, but focus on the use’s purpose as well.⁸⁵

D. The Evolution of Transformative Use

Courts have applied *Campbell* and found lawful transformative use in a wide variety of cases alleging copyright infringement, including textual

79. *Id.* at 579.

80. *Id.*

81. *Campbell*, 510 U.S. at 579.

82. *Id.*

83. *See e.g.*, *Lombardo v. Dr. Seuss Enterprises, L.P.*, 729 Fed. Appx. 131 (2d Cir. 2018) (holding that a parody of “How the Grinch Stole Christmas” is transformative because it was mocking the Grinch for comedic effect by imitating the style of the Grinch and mocking “the naïve, happy world of the Whos.”).

84. James L. Buchwalter J.D., Annotation, *Transformative Use of Copyrighted Material Under 17 U.S.C.A. § 107*, 84 A.L.R. FED. 3D ART. 3 (2023).

85. *Id.*

materials⁸⁶ and visual arts.⁸⁷ In internet search engine cases, technology developers rightfully relied on precedent to justify fair use.⁸⁸ For example, in *Authors Guild v. Google, Inc.*, the Second Circuit applied *Campbell* and decided that Google's scanning of copyrighted books was fair use.⁸⁹ Google created a digital database of copyrighted books by scanning them and making the text searchable.⁹⁰ Users could see snippets of the books in the search results but not the whole book.⁹¹ Authors Guild argued that Google copied its work without permission, while Google claimed the use was fair under § 107 because the database was a useful tool and did not harm book sales.⁹²

The court held that Google's book scanning project was a transformative use because it did not *only* reproduce the texts of books, but instead created a new and beneficial tool for users to search and locate information in books, significantly transforming the way users interact with the text.⁹³ The Second Circuit based its ruling, in part, on the idea that search results, including small snippets from books, serve a different purpose and provide a different meaning than the original works.⁹⁴ The tool changed the way people used books from just reading them to being able to search through them like a database, which helped users research key terms without replacing the original books' use.⁹⁵ Snippets give enough context to help users evaluate the book's relevance without

86. *Compare* Newport-Mesa Unified Sch. Dist. v. State of California Dep't of Educ., 371 F. Supp. 2d 1170 (C.D. Cal. 2005) (holding that a school distributing test protocols to special education students' parents was fair use because such distribution was for a nonprofit educational use and the test protocols with answers are transformative. Had the school distributed a blank test, it would not have been transformative), *with* Bell v. Moawad Grp., LLC, 326 F. Supp. 3d 918 (D.Ariz. 2018) (holding that consulting firms use of a sports psychology authors book is not transformative because the firm posted an excerpt from the authors book as "an inspirational quote in the form of an image." The excerpt was not significantly altered and conveyed the same message the author developed.).

87. *See, e.g.*, Blanch v. Koons, 467 F.3d 244 (2d Cir. 2006) (holding that the use of a copyrighted photograph from a magazine in a collage was transformative because the new use of the photograph was for a commentary on social consequences of the mass media, differing entirely from the original authors goals in creating the image, which was to exhibit German art-gallery space).

88. *See* Jenny Quang, *Does Training AI Violate Copyright Law?*, 36 BERKELEY TECH. L.J. 1407, 1417 (2021); *see also* Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015); Field v. Google Inc., 412 F. Supp. 2d 1106 (D.Nev. 2006); Perfect 10, Inc. v. Yandex N.V., 962 F. Supp. 2d 1146 (N.D. Cal. 2013), *as amended* (Sept. 6, 2013).

89. Authors Guild v. Google, Inc., 804 F.3d 202, 208 (2d Cir. 2015).

90. *Id.* at 207.

91. *Id.* at 209.

92. *Id.* at 208, 211.

93. *Id.* at 217.

94. *Id.* at 218.

95. *Authors Guild*, 804 F.3d at 217.

harming the author's copyright interests.⁹⁶ If users decide that the book is relevant to their interest area, the tool compels the users to buy the book; the tool only provides a portion small enough to give context, but not the entire book.⁹⁷

Courts stretched the confines of fair use even more when they ruled that fair use may apply when an unauthorized user makes minimal physical alterations to an original creation, provided the context of the new work's usage differs significantly from the original's.⁹⁸ In *Kelly v. Arriba Soft Corp.*, the Ninth Circuit held that when an image search engine's developer used another author's exact images as small thumbnails, the new use constituted transformative use.⁹⁹

In *Kelly*, the purpose and character of the work differed from the original because professional photographer, Kelly, intended his images to be aesthetically pleasing to viewers, while search engine developer, Arriba Soft Corp., intended the thumbnails to serve "as a tool to index and improve access to images."¹⁰⁰ The Ninth Circuit reasoned that thumbnails are more transformative than reproductions that *can* be used for the same purpose as the original because the thumbnails at issue here repurpose the images from their original entertainment or aesthetic intent into guides leading users to information sources.¹⁰¹

Similarly, in *Nunez v. Caribbean Inter. News Corp.*, the First Circuit held that a news outlet's verbatim reproduction of a photograph was transformative use because the original author's purpose was different from the copier's.¹⁰² The original photographer took photos of a model so they could appear in a modeling catalog, not a newspaper.¹⁰³ The factor weighed in favor of fair use because the newscaster's use of the original photograph transformed the model's photos into news.¹⁰⁴

Cases like *Kelly* and *Nunez* support the notion that a court may find fair use even when someone directly copies an original work, given that the new author's purpose differed from the original's.¹⁰⁵

96. *Id.*

97. *Id.* at 217–18.

98. See *Faulkner Literary Rts., LLC v. Sony Pictures Classics Inc.*, 953 F. Supp. 2d 701 (N.D. Miss. 2013) (finding that although a film used a single line from a novel, due to its vastly different context and purpose, it qualified as fair use); see also *In re DMCA Subpoena to Reddit, Inc.*, 441 F. Supp. 3d 875 (N.D. Cal. 2020).

99. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003).

100. *Id.*

101. *Id.* at 819.

102. *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 22 (1st Cir. 2000).

103. *Id.* at 23.

104. *Id.*

105. See generally, *Nunez*, 235 F.3d 18; *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).

E. Intermediate Copying

Not only have Courts found transformative use in cases involving physical and purposive transformation, but Courts have also found transformative use in fact patterns that completely differ in context from the original use; intermediate copying is one of those examples.¹⁰⁶ Generally, in copyright infringement cases involving intermediate copying, defendants make unlicensed and non-public copies of an original work and use them as a preliminary step in achieving a different product.¹⁰⁷ Courts have permitted intermediate copying when individuals make copies for reverse-engineering purposes and courts have held that these copies are transformative and non-infringing.¹⁰⁸

In *Sega Enterprises Ltd v. Accolade, Inc.*, for example, the Ninth Circuit notably addressed intermediate copying, holding as fair use Accolade's copying of Sega's software code as a preliminary step to make its games compatible with Sega's videogame system, even though Accolade's new product was competing with some of Sega's products.¹⁰⁹

In *Sega*, instead of paying Sega a licensing fee, Accolade reverse-engineered Sega's games to understand the technical details of how they communicated with Sega's Genesis console.¹¹⁰ Accolade's intermediate copying of Sega's software code constituted fair use, as the copying was *necessary* for achieving compatibility with Sega's Genesis console.¹¹¹ This use is protected under 17 U.S.C. §102(b) of the Copyright Act because the ideas and functional concepts embodied in the code are not copyrightable.¹¹² The Ninth Circuit's decision confirmed that even the direct copying of an author's work when the author does not make the original works available to the public may constitute fair use if the copier uses the original works only as a means to a completely different end.¹¹³

Nearly a decade later, in *Sony Comput. Entm't, Inc. v. Connectix Corp.*, the Ninth Circuit reaffirmed its position regarding transformative use in intermediate copying.¹¹⁴ In *Sony*, Connectix reverse-engineered and

106. Jiarui Liu, *An Empirical Study of Transformative Use in Copyright Law*, 22 STAN. TECH. L. REV. 163, 171 (2019).

107. *Id.* at 214; *See Sony Comput. Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000) (holding that the defendant's intermediate copying was permissible because it facilitated access to unprotected ideas and functional elements in the PlayStation's BIOS).

108. *See Liu, supra* note 106, at 215; *Kelly*, 336 F.3d at 822; *Authors Guild v. Google, Inc.*, 804 F.3d 202, 229 (2d Cir. 2015).

109. *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522 (9th Cir. 1992).

110. *Id.* at 1514.

111. *Id.* at 1520, 1522.

112. *Id.* at 1519, 1526.

113. *Id.* at 1526.

114. *Sony Computer Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 607 (9th Cir. 2000).

copied Sony's copyrighted basic input-output system (BIOS) in the process of creating its software program, Virtual Game Station (VGS).¹¹⁵ The district court ruled that VGS simply replaced the Sony PlayStation console and was not transformative because computer screens and television screens are interchangeable.¹¹⁶

However, the Ninth Circuit disagreed, stating that VGS was transformative because it enabled users to play PlayStation games on personal computers.¹¹⁷ This opened up new gameplay opportunities in places where a PlayStation console and television might not be available.¹¹⁸ Further, the court noted that VGS "is a wholly new product, notwithstanding the similarity of uses and functions between the Sony PlayStation and the Virtual Game Station."¹¹⁹ The court dismissed Sony's reliance on decisions emphasizing that a format change is not a transformation because VGS is not merely a format change, it also included expressive elements.¹²⁰ Therefore, the court declared that the intermediate copies that Connectix made in the process of creating the software were transformative and non-infringing.¹²¹

F. Supreme Court Developments

In 2021, in *Google LLC v. Oracle Am., Inc.*, the Supreme Court clarified the balance between copyright protection and innovation in the context of fair use in software development.¹²² Oracle America sued Google for copyright infringement when Google copied 11,500 lines of code from Oracle's Application Programming Interface (Sun Java API)¹²³ to build a software platform for mobile devices.¹²⁴

The Court categorized the Sun Java API into three essential parts: 1) implementing code, which tells the computer how to do specific tasks; 2) the "method call," in which special commands in the API trigger these

115. *Id.* at 598.

116. *Id.* at 607.

117. *Id.* at 606.

118. *Id.*

119. *Id.*

120. *Sony Computer*, 203 F.3d at 607 (citing *Infinity Broadcast Corp. v. Kirkwood*, 150 F.3d 104, 108 (2d Cir.1998) "change of format, though useful, is not technically a transformation.").

121. *Id.* at 607–08.

122. *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021).

123. *Oracle America, Inc. v. Google, Inc.*, 750 F.3d 1339, 1349 (Fed. Cir. 2014). (describing it as a tool that "allow[s] programmers to use ... prewritten code to build certain functions into their own programs, rather than write their own code to perform those functions from scratch.").

124. *Google*, 141 S. Ct. at 1186.

tasks; and 3) “declaring code,” which organizes and labels these tasks in a neat system.¹²⁵ Oracle alleged that Google copied the declaring code without authorization and that the Java API packages were creative works, which copyright law protects because of the way Oracle designed and organized them, making them more than just a functional element of the Java programming language.¹²⁶

The Supreme Court held that Google’s copying of the code was fair use and did not violate copyright law.¹²⁷ The Court explained that the original developer created the code so that users could use it in desktops and computers, whereas Google copied the code to develop its Android-based smartphones.¹²⁸ In other words, Google copied the code for a different purpose from Oracle’s purpose in creating the code.¹²⁹ Google copied only those parts of the API necessary for smartphone program functionality, allowing programmers to use these tasks without needing to learn a new programming language.¹³⁰

The Supreme Court acknowledged that Google directly copied parts of the API to allow the programmers access to programs that perform specific tasks, much like Sun’s original purpose in creating the computer programs.¹³¹ Still, the Court held that limiting fair use based on this rationale alone would overly restrict functional unauthorized use of a copyrighted computer program, such as for teaching or research purposes.¹³² The Court determined that Google’s use of the program to develop a new platform ultimately aligns with the fundamental constitutional goal of copyright, which is to foster creative progress.¹³³

After *Oracle*, the next major development to fair use in the Supreme Court’s 2023 decision *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, where the Court heard a challenge that redefined the judiciary’s role in evaluating whether a work is transformative.¹³⁴ In *Warhol*, Vanity Fair sought to license photographer Lynn Goldsmith’s photograph of the musician Prince for use as an “artistic reference.”¹³⁵ Goldsmith agreed to this use, on the condition that Vanity Fair use her

125. *Id.* at 1201.

126. *Id.* at 1201, 1202.

127. *Id.* at 1190.

128. *Id.* at 1203.

129. *Google*, 141 S. Ct. at 1203.

130. *Id.*

131. *Id.*

132. *Id.*

133. *Id.*

134. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

135. *Id.* at 515.

photo only one time.¹³⁶ Vanity Fair licensed pop artist Andy Warhol to create a silkscreen portrait of the photo of the musician, titled *Orange Prince*, for the purpose of illustrating a magazine story about Prince.¹³⁷ Vanity Fair published Warhol's Prince silkscreen and credited Goldsmith for the photograph in addition to paying her \$400.¹³⁸

Despite Vanity Fair granting Warhol only a limited license of use, Warhol derived 15 additional works from Goldsmith's original photograph.¹³⁹ The Andy Warhol Foundation for the Visual Arts, Inc. ("AWF") licensed one of the silkscreens to another magazine, Condé Nast, for the same purpose as the Vanity Fair publication.¹⁴⁰ Condé Nast paid AWF \$10,000 in compensation for this deal but gave Goldsmith nothing.¹⁴¹ Goldsmith accused AWF of copyright infringement for using her photograph, prompting AWF to file a lawsuit against her.¹⁴² The Supreme Court reviewed the case to address whether the first prong of the fair use analysis favored AWF or Goldsmith.¹⁴³

In evaluating the first statutory factor of the fair use analysis, the Court emphasized that *Campbell* requires the Court to consider *to what extent* the use is different.¹⁴⁴ A court is more likely to find fair use when the original work is very different from the new work.¹⁴⁵ This difference need not only be physical; photographers often license their works to magazines to illustrate stories about musicians, as Goldsmith did with her photographs of Prince in Newsweek, Vanity Fair, and People magazine.¹⁴⁶ Additionally, photographers may license their photographs so artists can use them for artistic reference, as Goldsmith did with Vanity Fair and Andy Warhol.¹⁴⁷ When AWF licensed the silkscreen portrait *Orange Prince* to Condé Nast, it was doing so to illustrate a story about a musician, like Goldsmith *could* have done with her photograph.¹⁴⁸ The Court reversed the district court's analysis, holding that judges should not "assume the role of art critic" and try to interpret the artistic intent or a work's meaning.¹⁴⁹

136. *Id.*

137. *Id.*

138. *Id.*

139. *Id.*

140. *Andy Warhol*, 598 U.S. at 515.

141. *Id.*

142. *Id.*

143. *Id.*

144. *Id.* at 529.

145. *Id.*

146. *Andy Warhol*, 598 U.S. at 534–35.

147. *Id.* at 535.

148. *Id.*

149. *Id.* at 544.

The Court also weighed AWF's commercial use of Goldsmith's photo against fair use, noting that "although a use's transformativeness may outweigh its commercial character, here, both elements point in the same direction."¹⁵⁰ The Supreme Court held that AWF's use of Goldsmith's photograph for Warhol's Prince Series and its commercial purposes did not support the fair use defense.¹⁵¹ The Supreme Court's decision in *Warhol* narrowed the application of transformative use by suggesting that aesthetic changes are not enough and purpose of use extends to commercial uses.¹⁵²

III. ANALYSIS

A. Introduction to GenAI Transformative Use

The legal landscape regarding transformative use has significantly evolved since Judge Leval first coined the term.¹⁵³ Starting with the Supreme Court's decision in *Campbell*¹⁵⁴ where the Court recognized parody as transformative due to its critical nature, the journey has continued through *Authors Guild*,¹⁵⁵ where the Second Circuit highlighted transformative use in terms of utility, and culminated in *Oracle*,¹⁵⁶ where the Supreme Court permitted copying of a software code's creative elements due to its contribution to scientific advancement.¹⁵⁷ Notably, cases like *Sega*¹⁵⁸ and *Kelly*¹⁵⁹ further demonstrate courts' willingness to permit copying for functional purposes or when repurposing creative elements.¹⁶⁰

However, the Supreme Court's ruling in *Andy Warhol v. Goldsmith* added a layer of complexity by suggesting that a court assessing a copied work's transformative nature under copyright law requires considering more than just aesthetic changes—it requires evaluating a shift in the original work's meaning or purpose, including a market purpose.¹⁶¹

Courts have not decided whether GenAI's training data use is transformative. These cases, then, must become a lens through which

150. *Id.* at 537–538.

151. *Id.* at 551.

152. *Andy Warhol*, 598 U.S. at 549–50.

153. *See supra* Part II.

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

155. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 217 (2d Cir. 2015).

156. *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1187 (2021).

157. *See supra* Part II.F.

158. *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

159. *Kelly v. Arriba Soft Corp* 336 F.3d 811 (9th Cir. 2003).

160. *Id.*

161. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

courts will examine GenAI's use of copyrighted work in training data.¹⁶² Judicial trends should lean towards courts viewing GenAI's use of training data as transformative because of the following factors: (1) GenAI developer's training data use is a form of intermediate copying, and (2) GenAI's utility and purpose are distinct from the original authors.

B. The Role of Intermediate Copying in GenAI Training

Training DL models on copyrighted works will almost always involve the reproduction of entire works.¹⁶³ In the learning process, AI algorithms ingest copyrighted works to learn, which ultimately requires the AI algorithm to make copies and then learn from the patterns in those works.¹⁶⁴ This use can potentially violate the copyright owner's reproduction right under 17 U.S.C. § 106(1).¹⁶⁵ However, transformative use, specifically intermediate copying principles, permit this type of reproduction.¹⁶⁶ The process of inputting training data into GenAI systems is fundamentally transformative because it involves teaching the deep learning system to recognize patterns, understand contexts, and generate new content based on those learned concepts; the process does not permit the DL model to replicate the input data verbatim and reproduce it.¹⁶⁷ Ultimately, it gives the work an entirely different purpose and character, as permitted by fair use principles.

The complex DL training process requires GenAI to process and learn patterns in data, similar to how the human brain learns a new concept.¹⁶⁸ When learning a new concept, both humans and machines copy an original author's work to learn from the ideas it conveys, not the artistic style.¹⁶⁹ Students in history class learn from a textbook to understand the history conveyed and apply it to the future, not to copy the textbook author's stylistic tone. Similarly, a DL model scans written text from a fiction book to learn word sequencing and sentence structure, not the author's uniqueness or expressiveness.¹⁷⁰ In the realm of GenAI, software reverse engineering cases give direction on the application of fair use in relation

162. *See supra* notes 154–161.

163. U.S. PATENT AND TRADEMARK OFFICE, *Public Views on Artificial Intelligence and Intellectual Property Policy*, 23 (Oct. 2020), https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf [<https://perma.cc/U9NA-JCNS>].

164. *Id.* at 24.

165. Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 61 (2017).

166. *See supra* Part II.D.

167. Hughes, *supra* note 23.

168. Lemley, *supra* note 10, at 749.

169. *Id.*

170. *Id.*

to copyrighted training data because developers train these models by copying software code, and in some instances, the developers directly copy as a preliminary step. Although it is still copying, courts have decided this type of copying is transformative because the original only serves a functional purpose in creating a new work.¹⁷¹

Whether intermediate copying supports a finding of fair use becomes a gray area when developers copy the original work for its non-expressive purpose, but the GenAI system instead learns patterns that are inherently expressive. Professor Benjamin Sobel distinguished *Sega* from the GenAI developer's use of training data because the accessed content in *Sega* was for a non-expressive purpose, and he further argued that it is possible for machines to learn from creative elements because "it learns from vast corpora of input data without the guidance of human-programmed rules, rather than by applying narrower sets of rules and facts that are predetermined by its programmers."¹⁷² Sobel warned that because these programs lack human control, they may ingest creative elements as well.¹⁷³ Although GenAI may access creative elements of the training data, it does so only *incidentally* while accessing unprotectable elements; the process of training GenAI is to teach the GenAI system to recognize patterns, not to replicate a single authors work.¹⁷⁴ It is learning from the elements, just like how Accolade copied *Sega*'s games to understand the technical details of how they communicated with *Sega*'s Genesis console as well as how Connectix learned Sony's code to make it compatible with the Sony PlayStation console.¹⁷⁵ As previously mentioned, programmers train GenAI programs to learn from the content it is being trained on, not to replace the original creators.¹⁷⁶ Programmers teach GenAI systems to

171. See, e.g., *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); *Ticketmaster Corp. v. Tickets.Com, Inc.*, No. CV997654HLHVBKX, 2003 WL 21406289, at *5 (C.D. Cal. Mar. 7, 2003).

172. Sobel, *supra* note 176, at 58.

173. *Id.*

174. See generally *supra* notes 33–44 and accompanying text; see Hughes, *supra* note 23.

175. *Sony Computer Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 606 (9th Cir. 2000).

176. See *supra* Part II.A.; see also George Lawton, *What is Generative AI? Everything You Need to Know*, TECH TARGET, <https://www.techtarget.com/searchenterpriseai/definition/generative-AI> [https://perma.cc/E9FZ-V97Z] (last visited Mar. 17, 2024) (“[T]o generate text, various natural language processing techniques transform raw characters (e.g., letters, punctuation and words) into sentences, parts of speech, entities and actions, which are represented as vectors using multiple encoding techniques. Similarly, images are transformed into various visual elements, also expressed as vectors. One caution is that these techniques can also encode the biases, racism, deception and puffery contained in the training data.”).

recognize patterns in sentence structure, word sequence, and visual elements for image-generating GenAI programs.¹⁷⁷

Even if Sobel is correct and GenAI training captures and learns the creative elements of the original author's work, the Supreme Court's decision in *Oracle* may permit this copying.¹⁷⁸ In *Oracle*, Google used the SunJava API's declaring code, which contained creative elements.¹⁷⁹ Yet, the Supreme Court *still* found that this type of copying was a transformative use and emphasized that the purpose behind the copying was to create new products that expand the usefulness of Android-based smartphones.¹⁸⁰ The declaring code that Google copied had structural and organizational elements that reflected the original author's creativity.¹⁸¹ Likewise, the training data that programmers feed into GenAI programs has creative elements, but GenAI developers use this training data for a different purpose, for a different end-product, and essentially for a different market than the original author.

C. GenAI: Transforming Through Distinct Purpose and Utility

Generally, the works that programmers use to train GenAI models serve a different purpose in the training process than their original aesthetic and entertainment purposes.¹⁸² In *Authors Guild*, the Second Circuit recognized Google's book scanning's transformative nature where Google's purpose was to create a searchable index rather than to replicate books for reading.¹⁸³ In *Kelly*, the First Circuit applied a similar rationale, deeming the *direct copying* of thumbnails fair use because Arriba Soft Corp. used the thumbnails as an indexing tool, not for their artistic expression.¹⁸⁴

Likewise, when programmers feed GenAI training data, the goal is not direct reproduction, but rather enabling the program to learn and, thereby, create something distinct and original. Just as the search engine in *Authors Guild* transforms books into searchable databases without replacing their original use, GenAI transforms data inputs from creative works into a tool used to train the GenAI program.

177. *Id.*

178. *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021).

179. *Id.* at 1190, 1201.

180. *Id.* at 1203.

181. *Id.*

182. HOWARD B. ABRAMS & TYLER T. OCHOA, 2 THE LAW OF COPYRIGHT § 15:153 (2024) (ebook).

183. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 217 (2d Cir. 2015).

184. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003).

Furthermore, programmers are not using GenAI to displace the original training data's market or audience. For instance, OpenAI rightfully argues that the original works targeting human audiences for entertainment value support the training AI systems' transformative nature.¹⁸⁵ GenAI's goal is not to replace the authors but rather to create a functional and useful GenAI tool.

To address the transformative use of the input data, it is helpful to look at the output to illustrate the distinctive purpose of use for training data. If one entered the following prompt into ChatGPT-4: "Write a short paragraph describing the molecular breakdown of a plant using Shakespearean English," the deep learning model might generate the following paragraph:

In the realm where green tendrils and leaves doth unfold, a mystical transformation lies within. The noble Chlorophyll, that painter of verdure, captures the sun's golden rays, weaving them with air's own breath, carbon dioxide, and water drawn from Gaia's bosom. This alchemy, named Photosynthesis, doth create the sweet elixir of glucose and the breath of life, oxygen. Strong cellulose builds its fortress, whilst starches, like hidden treasure, store the sun's bounty. Thus, in this microscopic tapestry, plants transmute celestial fire into earthly sustenance, a marvel of Nature's own hand.¹⁸⁶

OpenAI utilizes thousands of data points to teach its language model, ChatGPT, about topics such as plant anatomy, which could have included sourced copyrighted works or public domain works.

Originally, the authors of these works intended others to use the works for human enjoyment and education, but OpenAI repurposes them to train its AI model, enabling it to generate unique content, such as descriptions of plant anatomy in Shakespearean English which is a use that the original creators did not foresee. Whether from Shakespeare, a scientist, or any other creator, OpenAI uses their copyrighted works to fulfill a different purpose. The used input and new expression do not substitute the original work because programmers feed GenAI programs millions of data points to yield a response and "nobody looking to read a specific webpage

185. OpenAI, Comment Letter on Request for Comments on Intellectual Property Protections for Artificial Intelligence Inventions (Nov. 27, 2019), https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf#page=5 [<https://perma.cc/KPG9-2M7Y>].

186. OpenAI, AI Generated Description, CHATGPT (Mar. 2, 2024), <https://chat.openai.com/> (click to 'message ChatGPT' and input "Write a short paragraph describing the molecular breakdown of a plant using Shakespearean English").

contained in the corpus used to train an AI system can do so by studying the AI system or its outputs.”¹⁸⁷

Jessica Gillotte notes that in the world of artwork-generating GenAI, “an engineer could similarly digitize copyrighted artworks to assemble a corpus of training data for her AI program.”¹⁸⁸ This use is significantly transformative because it is different from the original author’s goals, which are conveying artistic expression, making a living, or honing their craft.¹⁸⁹ Unlike the original author, the engineer uses these works as training data to train the AI model to recognize patterns.¹⁹⁰ The millions of data points allow the GenAI program to enhance its algorithm by increasing the differences between its produced works and the output that the training data refines.¹⁹¹ The engineer is essentially changing the purpose of these artworks because digitizing and ultimately training GenAI on copyrighted images is different than the original artist’s creation.¹⁹²

The Supreme Court’s decision in *Warhol*¹⁹³ does not change this notion. An overly broad reading of *Warhol* suggests that the Court departed from *Campbell* by equating transformative and commercial uses. Perhaps this argument would apply if hypothetically, a user abuses a for-profit GenAI model and directs the program to generate a specific author’s work and the algorithm succeeds in doing that. Based on *Warhol*, a court should then use commercial use as a *factor* in the transformative use analysis.¹⁹⁴ In *Warhol*, the Court only reemphasized that purpose of use takes many forms, including commercial forms, regardless of whether the copier physically altered the original work.¹⁹⁵ A GenAI program’s commercial aspect does not change the fact that the programs used these works for a *completely different* purpose than the original authors intended that others use the works. Although commercial use is a factor to be considered in the transformative use analysis, distinct utility and purpose could still favor a finding for transformative use.

The Supreme Court ruled in favor of Goldsmith, not simply because AWF profited off the reproductions, but because Goldsmith sometimes

187. See *supra* note 185, at 5.

188. Jessica L. Gillotte, *Copyright Infringement in AI-Generated Artworks*, 53 U.C. DAVIS L. REV. 2655, 2683–84 (2020).

189. *Id.* at 2684.

190. *Id.*

191. *Id.*

192. *Id.*

193. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

194. *Id.* at 526–39.

195. *Id.*

also profited from the *same use*.¹⁹⁶ When AWF licensed the silkscreen portrait entitled *Orange Prince* to Condé Nast, it was doing so to illustrate a story about a musician, like Goldsmith *could* have done with her photograph.¹⁹⁷ On the other hand, GenAI training does not serve a similar purpose as the thousands of data points that GenAI used to study the breakdown of plants and the thousands of data points that GenAI used to learn the language of Shakespearean English.

This transformative purpose alone supports a finding of fair use, but future courts should consider other factors in the transformative use analysis given GenAI's unique capabilities. In evaluating the transformativeness of GenAI, future courts should not aim to further complicate what constitutes transformative use but instead should leave it up to the legislature to adopt a clear law regarding transformations for AI.

D. Transformative Training

The fair use defense requires a fact-specific analysis, under which courts balance the four factors in relation to each other.¹⁹⁸ The problem with courts relying on the fair use analysis to determine whether GenAI is a transformative use is that it is highly unpredictable.¹⁹⁹ Jenny Quang advocated for abandoning the fair use doctrine as a whole and creating a “safe harbor” provision for data mining within U.S. copyright law.²⁰⁰ This proposal would provide clarity on allowed uses of training data, treat commercial and non-commercial uses alike, and limit data mining to non-expressive uses.²⁰¹

While this strategy appears to support the use of training data for GenAI, not distinguishing between commercial and non-commercial purposes of use paradoxically risks obstructing the core aim of copyright law, which is to encourage innovation. Although GenAI training is inherently transformative for the reasons discussed earlier, a proposal to not distinguish between commercial and non-commercial use fails to consider some genuine risks. Specifically, there is a risk that GenAI could generate outputs that resemble an original creator's work, and if the program charges a fee for its service, it could ultimately discourage creators from creating. Professor Grimmelman agrees and even goes as far

196. *Id.* at 559 (Kagan, J., dissenting).

197. *Id.* at 519–20.

198. *See supra* notes 14–15 and accompanying text.

199. Lemley, *supra* note 10, at 763.

200. Quang, *supra* note 88, at 1430–35.

201. *Id.* at 1433.

to suggest that the current state of copyright law favors robots over humans and will lead to the fall of humanity.²⁰²

Professor Mark Lemley suggests a concept called “Fair Learning,” centered on the idea that if an AI learns from the non-copyrightable parts of a copyrighted work—such as facts or patterns—courts should generally consider this use fair, even if other aspects of the fair use analysis might not support such a finding.²⁰³ The proposal also distinguishes between using copyrighted material for direct competition and using copyrighted material for educational or innovative purposes, suggesting that courts should generally allow the latter.²⁰⁴

Currently, copyright laws lack specific provisions that navigate the complexities of training GenAI systems. Current copyright laws and the fair use doctrine rely on principles that courts established before the advent of advanced technologies.²⁰⁵ Instead of complicating transformative use principles further and muddying a doctrine that is already unpredictable, this Note suggests legislative measures that require amending the Copyright Act to include provisions that explicitly address GenAI training programs. These provisions would provide clear definitions of what constitutes “generative artificial intelligence” and clearly distinguish between “learning” and “copying” in the realm of GenAI. The statute should codify the principles established in the intermediate copying cases,²⁰⁶ which will provide clarity on the type of copying that the courts allow. This proposal will ultimately guide GenAI developers to train responsibly, so that GenAI programs are unable to generate copies of the original authors’ work and will carefully balance the promotion of innovation and technological advancement.

The legislative provision should adopt Professor Lemley’s “Fair Learning” criteria²⁰⁷ and set forth conditions under which courts consider such copying non-infringing due to its necessity for the learning process. Further, the proposed legislation should create safe harbor provisions for developers who use copyrighted materials in ways that meet established

202. James Grimmelman, *Copyright for Literate Robots*, 101 IOWA L. REV. 657, 676–677 (2016) (“We are teaching robots to write like us and read like us-- sometimes for our own edification or entertainment, sometimes as a side effect of the global struggle of algorithm against algorithm for aggregated slivers of human attention. Already, computers can compose music and write news stories. What if there comes a day when they have no further need of our creative facilities at all, when robots are superintelligent, surpassing human cognitive abilities as we surpass banana slugs? Superintelligent computers would pose an existential risk to humanity.”)

203. Lemley, *supra* note 10, at 782.

204. *Id.* at 779–80.

205. *See supra* Part II.

206. *See supra* Part II.B; *see also supra* notes 114–31.

207. Lemley, *supra* note 10, at 776–79.

transformative use criteria, protecting them from infringement claims. Due to the complexity and uncertainty in training GenAI, the statute should include a provision requiring transparency and documentation requirements in training data, ultimately protecting copyright holders' interests. Congress should also adopt these provisions for public policy reasons.

E. Policy

Copyright law aims to foster the progress of science and the arts by protecting the rights of creators for their original works.²⁰⁸ However, too broad of a protection could stifle creativity and innovation, as all creators build on past ideas.²⁰⁹

Despite concerns that GenAI will surpass human intelligence, ultimately leading to the fall of humanity,²¹⁰ GenAI is not all that bad. In fact, GenAI can enhance human capabilities, stimulate employment growth, and complement human work and intellect similar to how mobile technology and the internet have transformed lives over the past 25 years.²¹¹ GenAI could boost the shift to green energy by helping develop sustainable products and improve energy systems, ultimately creating millions of clean energy jobs.²¹² In education, GenAI could make learning more accessible worldwide by offering training in various languages and tailoring education to individual needs.²¹³ In agriculture, using GenAI alongside sensors and drones could lead to better crop management and new careers in high-technology farming.²¹⁴ Although the healthcare industry is still implementing GenAI, its future is promising in improving diagnoses, managing disease outbreaks, discovering new drugs, and offering customized treatments.²¹⁵

To maximize the potential of GenAI while safeguarding creator's rights, lawmakers should amend the Copyright Act to include specific provisions that govern how developers train GenAI systems. This

208. *See supra* Part II.

209. *See supra* Part II.

210. *See supra* note 208 and accompanying text.

211. Sudeep Kesh, Shankar Krishnamurthy, Nick Patience, *Can Generative AI Create a Productivity Boom?*, S&P GLOBAL (Jan. 10, 2024), [https://www.spglobal.com/en/research-insights/featured/special-editorial/look-forward/can-generative-ai-create-a-productivity-boom#:~:text=This%20technology%20has%20profound%20implications,global%20economy%20and%20equitable%20development.&text=Generative%20AI%20\(GenAI\)%20can%20accelerate,over%20the%20last%2025%20years](https://www.spglobal.com/en/research-insights/featured/special-editorial/look-forward/can-generative-ai-create-a-productivity-boom#:~:text=This%20technology%20has%20profound%20implications,global%20economy%20and%20equitable%20development.&text=Generative%20AI%20(GenAI)%20can%20accelerate,over%20the%20last%2025%20years) [https://perma.cc/VJ8F-XZP9].

212. *Id.*

213. *Id.*

214. *Id.*

215. *Id.*

approach will ensure that innovative technology continues to enhance societal progress without compromising the intellectual property rights that fuel creativity.

IV. CONCLUSION

Over the past century, the courts evolved copyright law and redefined transformative use principles to fit different modes of expression and different contexts of use.²¹⁶ Implementing current case law to GenAI training may show that GenAI training is inherently transformative, but relying on the fair use defense alone is risky because this doctrine is highly unpredictable. Thus, the legislature should amend the Copyright Act and address GenAI directly, providing clearer guidelines on permissible copying. The complexities of GenAI training require a new way of looking at copyright law and clearer guidelines for addressing these complex issues. If courts apply a rigid and inflexible understanding of copyright laws that Congress passed long before technological advancements, courts would severely hinder technological and societal progress, enabling a “copyright monopoly.”²¹⁷

Congress should actively encourage GenAI training by passing laws to ensure that the law remains relevant and effective in promoting creation and innovation while protecting creators’ rights. Supporting technological development is key for growth; if not for computing advancements, society might still rely on carrier pigeons for communication.

216. *See supra* Part II.

217. Leval, *supra* note 50, at 1110.