

# Public Interest, the True Soul: Copyright’s Fair Use Doctrine and the Use of Copyrighted Works to Train Generative AI Tools

Elizabeth Spica, Ph.D., J.D.\*

## Abstract

*Generative Artificial Intelligence (AI) tools like ChatGPT, Gemini, and Stable Diffusion are trained using datasets that often include entire copyrighted works. The use of copyrighted works to train these increasingly prominent tools has proven so controversial that a senior executive from one of the most well-known AI firms recently resigned, stating it was “exploitative” for AI developers to scrape and use mass amounts of internet data to train their systems without first gaining consent from copyright holders.<sup>1</sup> In this article, I examine these concerns and ultimately conclude that the ingestion of entire copyrighted works for the purpose of training Generative AI tools likely constitutes a transformative use under U.S. copyright’s Fair Use Doctrine. In arriving at this conclusion, I first provide a brief overview of current litigation, then introduce case law instructive to the application of fair use to challenges presented by Generative AI tools. Next, I apply that case law to the ingestion of copyrighted works for training Generative AI tools, arguing that such use constitutes a transformative, fair use under U.S. copyright law. Finally, the article ends with optimism, exploring implications for the future of copyright law and outlining the numerous avenues that creators still have to uphold their exclusive rights.*

## Table of Contents

I.	Introduction .....	68
II.	The U.S. Copyright Fair Use Doctrine .....	72
	A. Factor 1. “[T]he purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes.”	73
	B. Factor 2. “[T]he nature of the copyrighted work.” .....	75
	C. Factor 3. “[T]he amount and substantiality of the portion used in relation to the copyrighted work as a whole.” .....	76

---

\* J.D., University of Tennessee College of Law; Ph.D., Higher Education Administration, University of Tennessee Department of Educational Leadership and Policy Studies. The author would like to thank Professor Lucy Jewel, whose insight and expertise helped shape this article.

<sup>1</sup> Zoe Kleinman, *AI Chief Quits over ‘Exploitative’ Copyright Row*, BBC.COM (Nov. 17, 2023), <https://www.bbc.com/news/technology-67446000.amp>.

D.	Factor 4. “[T]he effect of the use upon the potential market for or value of the copyrighted work.” .....	78
III.	Application of Fair Use Case Law to Generative Artificial Intelligence	
(AI)	Tools .....	80
A.	Factor one weighs in favor of fair use, because the purpose and character of the use of works ingested into Generative AI fundamentally differ from the use of the Generative AI tools themselves. ....	80
B.	Factor two also weighs in favor of fair use, even when expressive works are ingested, because of the transformative nature of Generative AI tools .....	82
C.	Factor three weighs in favor of fair use, because copying entire works is likely necessary to achieve Generative AI tools’ transformative purpose(s). ....	83
D.	Factor four also weighs in favor of fair use because a market that preferences the personalized output of Generative AI would likely create a separate market from that of readers who prefer works from individual creators. ....	84
IV.	Implications & Future Directions .....	88
V.	Conclusion .....	90

## I. Introduction

While perhaps novel to the general public, Generative Artificial Intelligence (AI) tools are certainly nothing new. In the 1960s, mathematician Alan Turing posited that computers could impersonate humans.<sup>2</sup> In that same decade, computer scientist Joseph Weizenbaum’s ELIZA program engaged users in real-time, plausible human conversation.<sup>3</sup> Now in the 2020s, Generative AI tools are commonplace, generally referring to computer systems designed to imitate human cognitive functions, including generating text content in response to user inputs or prompts.<sup>4</sup>

Generative AI systems are trained through content “ingested” or supplied to the system via large datasets. Commonly used Generative AI tools like OpenAI’s ChatGPT rely upon Large Language Models (“LLMs”), which leverage predictive algorithms based on statistical patterns and parameters derived from large text-based datasets.<sup>5</sup> Other tools, such as Google’s Gemini, are instead trained on Large

<sup>2</sup> Corydon Ireland, *Alan Turing at 100*, THE HARVARD GAZETTE (Sept. 13, 2012), <https://news.harvard.edu/gazette/story/2012/09/alan-turing-at-100>.

<sup>3</sup> *Id.*

<sup>4</sup> Class Action Complaint at 51–52, *Authors Guild v. OpenAI*, No. 1:23-cv-08292 (S.D.N.Y. 2023).

<sup>5</sup> George Lawton, *What is Generative AI? Everything you Need to Know*, TECHTARGET (last updated Jan. 15, 2024), <https://www.techtargget.com/searchenterpriseai/definition/generative-AI> (last visited Mar. 15, 2024). While this article references the most widely recognized AI tools, numerous others exist. For example, Jasper, AI-Writer and Lex also generate text; Dall-E, Midjourney and Stable Diffusion generate images; Amper, Dadabots and MuseNet generate music; CodeStarter, Codex,

Multimodal Models (“LMMs”) that allow input based on multiple formats including text, video, and audio.<sup>6</sup> Regardless of input format, the computations enabled by these datasets allow the system to recognize user input and generate human-like output in response.<sup>7</sup>

Contrary to popular belief, however, Generative AI processes are far less than automated. Humans play a key role in training, namely, in cleaning and organizing the datasets and in defining and fine-tuning model parameters to further hone and polish the output.<sup>8</sup> In general, the higher quality and larger the dataset, the more humanlike the output generated.

The looming question with these technological innovations is whether the ingestion of copyrighted works without prior permission from the copyright holder constitutes infringement. The answer will determine the outcome of a torrent of pending lawsuits alleging that Generative AI tools have infringed upon authors’ exclusive rights. For example, *Andersen v. Stability AI Ltd.*, alleged the infringement of multiple artists’ visual works in training tools including Stability AI, MidJourney, and DeviantArt.<sup>9</sup> Similarly, in February 2024, the Southern District of New York consolidated complaints alleging copyright infringement based on the use of “a broad array of works of fiction” in training OpenAI’s ChatGPT system.<sup>10</sup> Another class action suit, *Leovy, et al. v. Google LLC*, alleges that Google infringed plaintiffs’ exclusive derivative rights by training its Bard (now Gemini) AI tool using plaintiffs’ copyrighted works.<sup>11</sup> Yet another suit in the Middle District of Tennessee, *Concord Music Group, Inc. v. Anthropic PBC*, involves the alleged infringement of publicly available but nonetheless copyrighted musical lyrics and compositions when Anthropic created its Claude chatbot, “a next generation AI assistant built for work and trained to be safe, accurate, and secure.”<sup>12</sup>

The lawsuits seem unending, with *New York Times v. Microsoft* targeting Microsoft and OpenAI for allegedly infringing “millions of The Times’ copyrighted news articles, in-depth investigations, opinion pieces, reviews, how-to guides, and

---

GitHub Copilot and Tabnine generate code; and, tools like Descript, Listnr and Podcast.ai synthesize voices.

<sup>6</sup> Cem Dilmegani, *Large Multimodal Models (LMMs) vs Large Language Models (LLMs)*, AIMULTIPLE (last updated Jan. 10, 2024), <https://research.aimultiple.com/large-multimodal-models> (last visited Mar. 17, 2024).

<sup>7</sup> See Class Action Complaint at 51–60, *Authors Guild*, No. 1:23-cv-08292.

<sup>8</sup> See Asiri Piyajanaka, *The Role of Human Programmers in Training, Guiding, and Evaluating Generative AI Models*, MEDIUM (Aug. 6, 2023), <https://medium.com/@asiripiyajanaka/the-role-of-human-programmers-in-training-guiding-and-evaluating-generative-ai-models-89ec56320c2f>.

<sup>9</sup> *Andersen v. Stability AI Ltd.*, No. 23-cv-00201-WHO, 2024 U.S. Dist. LEXIS 22837 (N.D. Cal. Feb. 8, 2024), at \*1, \*3.

<sup>10</sup> Class Action Complaint at 1, *Authors Guild*, No. 1:23-cv-08292.

<sup>11</sup> Class Action Complaint at 11, *Leovy v. Google LLC*, No. 3:23-cv-03440-AMO (N.D. Cal. 2024).

<sup>12</sup> Complaint at 4, *Concord Music Group, Inc. v. Anthropic PBC*, No. 3:23-cv-01092 (M.D. Tenn. 2023). See also <https://claude.ai/> (accessed Mar. 16, 2024).

more.”<sup>13</sup> Even the world of legal technology has been unable to escape the reach of Generative AI. *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc.*, for example, concerns in part the alleged infringement of material from the Westlaw legal research platform.<sup>14</sup> Here, Thomson alleged that ROSS infringed Westlaw’s proprietary Headnotes and Key Number System in developing its own AI-based legal search platform, which allows users to search by “language you would use with a colleague.”<sup>15</sup> In developing this competing system, ROSS engaged LegalEase, a third party licensed to access Westlaw content.<sup>16</sup> LegalEase allegedly engaged with Westlaw’s content at a rate 40 times its historical usage pattern (and five times greater than the average monthly usage of Westlaw’s top 100 law firms).<sup>17</sup>

As a threshold matter, each use above likely constitutes prima facie copyright infringement, as “anyone who violates any of the exclusive rights of the copyright owner. . . is an infringer of the copyright[.]”<sup>18</sup> Plaintiffs in each of these cases would likely have little difficulty showing the elements required to prove infringement, specifically (1) the copying of a copyrighted work, (2) that the infringer had access to that work, and (3) a substantial similarity between the copyrighted work and the work copied for ingestion into the Generative AI tool.<sup>19</sup> Infringement alone, however, is insufficient to establish liability when defendants assert a valid affirmative defense.

The Fair Use Doctrine, a judicial common law doctrine codified in the 1976 Copyright Act, offers such a defense.<sup>20</sup> An equitable rule, the Fair Use Doctrine seeks to balance the time-limited monopoly awarded for the “fruits of [creators’] intellectual labor” against the First Amendment right and public interest in “encourag[ing] free expression and creativity[.]”<sup>21</sup> As such, copyright’s Fair Use Doctrine establishes a four-criteria, case-by-case approach to examine the “endless variety of situations and combinations of circumstances that can rise in particular cases.”<sup>22</sup>

---

<sup>13</sup> Complaint at 1–2, *The New York Times Company v. Microsoft Corporation*, No. 1:23-cv-11195 (S.D.N.Y. 2023). *See also* Kyle Wiggers, *OpenAI Claims New York Times Copyright Lawsuit is Without Merit*, TECHCRUNCH (Jan. 8, 2024), <https://techcrunch.com/2024/01/08/openai-claims-ny-times-copyright-lawsuit-is-without-merit/> (covering OpenAI’s response to suit).

<sup>14</sup> *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc.*, 529 F. Supp. 3d 303, 307 (D. Del. 2021).

<sup>15</sup> ROSS, <https://www.rossintelligence.com/features> (last visited Mar. 17, 2024).

<sup>16</sup> *ROSS Intelligence Inc.*, 529 F. Supp. 3d at 309.

<sup>17</sup> *Id.*

<sup>18</sup> 17 U.S.C. § 501(a).

<sup>19</sup> 17 U.S.C. § 501; *See Three Boys Music Corp. v. Bolton*, 212 F.3d 477, 486 (“By establishing reasonable access and substantial similarity, a copyright plaintiff creates a presumption of copying”).

<sup>20</sup> 17 U.S.C. § 107.

<sup>21</sup> *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 432 (1984); *Comedy III Prods., Inc. v. Gary Saderup, Inc.*, 21 P.3d 797, 808 (2001).

<sup>22</sup> *Harper & Row, Publs. v. Nation Enters.*, 471 U.S. 539, 560 (1985) (noting that, while juries may initially decide fair use, judges review appeals de novo as questions of law rather than fact); *see also* H.R. REP. NO. 94-1476 at 66 (1976) (hesitating to freeze the fair use doctrine in time due to the need to evolve and adapt to “rapid technological change”); 17 U.S.C. § 107 (carving out an exception that allows copying “for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research”).

A major barrier to courts' ability to keep pace with the exponential surge of AI-related lawsuits is a lack of guidance as to which side of fair use the datasets used to train Generative AI tools fall.<sup>23</sup> Legal scholars have discussed the fair use status of artificial intelligence training materials in light of societal costs and benefits, as well as in light of recent Supreme Court decisions.<sup>24</sup> Other scholars have examined misperceptions related to Generative AI processes themselves.<sup>25</sup> None, however, have undertaken a holistic fair use analysis against case law precedent directly focused on the transformative aspects of new works that incorporate entire copyrighted works. To address this gap, I search for guidance in case law precedent including *Authors Guild v. Google, Inc.*, *Authors Guild, Inc. v. HathiTrust*, *Google, LLC v. Oracle America, Inc.*, and, most recently, *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*. In examining the issue through these precedents, I conclude that the ingestion of full copyrighted works for the purposes of creating Generative AI tools qualifies as a fair, transformative use.

Clarifying the fair use status of ingesting entire copyrighted works into Generative AI is a foundational matter. Until courts provide guidance, creators remain under the threat of being deprived of work and employment, a precarious limbo striking dangerously close to the heart of human rights. The Universal Declaration of Human Rights emphasizes that all human beings are “born free and equal in dignity and rights[,]” including the right “to work . . . and to protection against unemployment.”<sup>26</sup> If the fruits of creative labors simply become fodder for the AI machine, used without limit, compensation, or consequence, creators arguably become less free, less dignified, and retain fewer rights. Court guidance is essential to determine how society moves forward.

When courts clarify the fair use status of datasets used to train Generative AI, they also further contour the ways in which copyright ultimately promotes “the Progress of Science and useful Arts.”<sup>27</sup> To balance the desire and presumed need to incentivize creativity against the broader public interest to freely use that creativity, copyright law has granted creators a time-limited monopoly and exclusive rights to reproduce, prepare derivative works, distribute copies, and in some cases, perform,

---

<sup>23</sup> Kyle Wiggers, *The Current Legal Cases against Generative AI are Just the Beginning*, TECHCRUNCH (Jan. 27, 2023), <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/>.

<sup>24</sup> See Andrew Torrance & Bill Tomlinson, *Training is Everything: Artificial Intelligence, Copyright, and “Fair Training”*, 128 DICK. L. REV. 233, 255 (2023) (weighing fair use arguments against the “societal costs . . . and benefits . . . of allowing AI to make easy use of copyrighted works”); see also Gary Myers, *Artificial Intelligence and Transformative Use After Warhol*, 81 WASH. & LEE L. REV. ONLINE 1, 19 (2023).

<sup>25</sup> See Michael Murray, *Generative AI Art: Copyright Infringement and Fair Use*, 26 SMU SCI. & TECH. L. REV. 259, 259 (2023) (clarifying “the actual science and technology that goes into the creation and operation of a contemporary visual generative AI system”); see also Van Lindberg, *Building and Using Generative Models Under US Copyright Law*, 18 RUTGERS BUS. L. REV. 1, 3 (2023) (explaining machine learning models to conclude that “building and using generative ML models is either outside the scope of copyright or is a fair use”).

<sup>26</sup> G.A. Res. 217A (III), Universal Declaration of Human Rights (Dec. 10, 1948).

<sup>27</sup> U.S. Const. art. I, § 8, cl. 8.

display, and/or transmit the work publicly.<sup>28</sup> On the most catastrophic level, if the ingestion of copyrighted works is deemed a fair use, a chilling effect will bring creativity and innovation to a grinding halt. More realistically, however, creators wishing not to participate will in turn develop ways to prevent AI systems from accessing their works.<sup>29</sup> Either way, court guidance is needed in this pivotal moment, and this article seeks to inform that guidance.

Parties in the aforementioned suits likely agree that the copyright issues surrounding Generative AI remain fast-paced and ever-evolving. The sooner that courts contour the issues, the sooner resolution will occur, which means the onslaught of claims will be less likely to drown the American litigation system. Once courts clarify the infringement status of Generative AI datasets, creators will also receive much-needed clarification on the remedies available to protect their exclusive rights. To examine this issue, I first outline key Fair Use Doctrine principles with respect to relevant case law, including *Authors Guild v. Google, Inc.*, *Authors Guild, Inc. v. HathiTrust*, *Google, LLC v. Oracle America, Inc.*, and, most recently, *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, each of which informs the fair use status of ingesting entire copyrighted works into Generative AI.<sup>30</sup> In applying this insight, I conclude that the ingestion of copyrighted works to train Generative AI constitutes a fair, transformative use. In the final section, I consider implications and directions for future research, concluding with optimism that, regardless of how courts decide this issue, creators still have several ways to reserve their exclusive rights under U.S. copyright law.

## II. The U.S. Copyright Fair Use Doctrine

Copyright exists “[t]o promote the Progress of Science and useful Arts” by providing authors a time-limited monopoly over the various exclusive rights discussed in the previous section.<sup>31</sup> By incentivizing creativity through the granting of these exclusive rights, copyright law hopes to “stimulate activity and progress in the arts for the intellectual enrichment of the public.”<sup>32</sup> In turn, to achieve this balance, the Fair Use Doctrine provides an exception that allows anyone to use copyrighted works without infringement and without first gaining permission from the copyright

---

<sup>28</sup> 17 U.S.C. § 106.

<sup>29</sup> Ariel Bogle, *New York Times, CNN and Australia’s ABC Block OpenAI’s GPTBot Web Crawler from Accessing Content*, THE GUARDIAN (Aug. 24, 2023, 8:31 PM), <https://www.theguardian.com/technology/2023/aug/25/new-york-times-cnn-and-abc-block-openai-gptbot-web-crawler-from-scraping-content> (reporting how several news outlets have blocked chatbots like GPTBot to help prevent A.I. tools from scraping their content); see also Julia Angwin, *The Internet is About to Get Much Worse*, N.Y. TIMES (Sept. 23, 2023), <https://www.nytimes.com/2023/09/23/opinion/ai-internet-lawsuit.html>.

<sup>30</sup> *Authors Guild v. Google, Inc.*, 578 U.S. 941 (2016); *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014); *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021); *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

<sup>31</sup> U.S. Const. art. I, § 8, cl. 8; see also *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578 (1994); *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 154 (1975).

<sup>32</sup> Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1107 (1990); see also 17 U.S.C. §§ 102, 106, 302–05.

holder, provided that certain criteria are met. To determine whether a particular use is fair, courts weigh four nonexclusive factors:

- (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.<sup>33</sup>

These nonexclusive factors indicate in part that, when copyrighted works are used fairly under the statute, the new works should constitute an "invaluable contribution to the progress of science and cultivation of the arts."<sup>34</sup> Each factor proves key to understanding fair use.. To further explore the fair use status of ingesting entire copyrighted works into Generative AI tools, I first describe each factor with respect to relevant case precedent, then apply that precedent to support the conclusion that the ingestion of entire copyrighted works for the purpose of training Generative AI tools constitutes a transformative, fair use under U.S. Copyright law.

- A. Factor 1. "[T]he purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes."<sup>35</sup>

This first factor, classically referred to as the "soul of fair use," focuses on the purpose and character of the use.<sup>36</sup> The purpose and character of the use entails whether the new work is "transformative," or "add[ing] something new, with a further purpose or different character, altering the first with new expression, meaning, or message."<sup>37</sup> To be transformative, a new work cannot simply "supersede[] the objects of the original creation" or substitute the original; rather, that new work must "serve[] a new and different function."<sup>38</sup> In this respect, the first factor seeks an objective balance by comparing the purpose and character of the original work's use with the purpose and character of the allegedly infringing secondary use, including "whether th[at] use is commercial and, importantly, the reasons for copying."<sup>39</sup>

Justice Gorsuch's concurring opinion in *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith* explained why the "purpose and character of the use" should

---

<sup>33</sup> 17 U.S.C. § 107.

<sup>34</sup> *HathiTrust*, 755 F.3d at 90 (quoting *Acuff-Rose Music, Inc.*, 510 U.S. at 579) (citing Leval, 103 Harv. L. Rev. at 1111).

<sup>35</sup> 17 U.S.C. § 107.

<sup>36</sup> Leval, *supra* note 32, at 1116.

<sup>37</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 579 (quoting *Folsom v. Marsh*, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841)); *see also* Leval, *supra* note 32, at 1111 (noting a use that adds value to the original, "transformed in the creation of new information, new aesthetics, new insights and understandings" is "the very type of activity that the fair use doctrine intends to protect for the enrichment of society"); *HathiTrust*, 755 F.3d at 96 (holding the new work must do "something more than repackage or republish the original copyrighted work").

<sup>38</sup> *HathiTrust*, 755 F.3d at 96–97.

<sup>39</sup> *Andy Warhol Found.*, 598 U.S. at 543, 550 (stating the objective meaning of 2 Live Crew's song was relevant insofar as it concerned the reasons for copying, which was different from "any 'new expression, meaning, or message,' which is a subjective inquiry falling outside the fair use test).

turn on the “*purpose* the *creator* had in mind when producing his work” as opposed to any subjective intent or artistic purpose behind the original or infringing work.<sup>40</sup> Justice Gorsuch suggests that the words “purpose” and “use” were intentionally paired together in the preamble to the fair use factors, which asks courts “to assess whether the person asserting a fair-use defense seeks to ‘use’ a copyrighted work ‘for purposes such as criticism, comment, news reporting, teaching. . . , scholarship, or research.’”<sup>41</sup> In *Warhol*, the court found that (what was essentially) a recolored photograph of the pop star did not favor fair use because the use of both the orange recolored Prince and the original purple-colored Prince served the same purpose, i.e., publication in a “special edition magazine devoted to Prince[.]”<sup>42</sup> As the Court noted, “Goldsmith’s original works, like those of other photographers, are entitled to copyright protection, even against famous artists.”<sup>43</sup>

In his concurring opinion, Justice Gorsuch also affirmed that subjectivity has no place in the fair use analysis, pointing out that courts cannot simultaneously give copyright holders the exclusive right to create “derivative works” that “transfor[m]” or “adap[t]” while also giving others the right to do so on the grounds that those allegedly derivative works had different subjective meanings.<sup>44</sup> In the case of *Warhol*, any attempt to justify fair use of the recolored photograph simply because the new orange Prince subjectively spoke to the “impact of celebrity” while the original purple Prince supposedly spoke instead to Prince’s “unique human identity” would only “risk making a nonsense of the statutory scheme.”<sup>45</sup> With this clarification, the Court affirmed that neither the artists’ subjective intent nor the court’s subjective interpretation of a work’s meaning matters; rather, what matters is what the user of the new work *does* with the copyrighted work.<sup>46</sup> In *Warhol*, both photos were licensed for use in a for-profit magazine for the purpose of publishing an article concerning Prince.<sup>47</sup> With the purpose and character of the uses practically mirroring one another, factor one weighed against the Foundation’s use of Goldsmith’s original Prince photograph.<sup>48</sup>

Also relevant to the first factor analysis is that copying an entire work can be transformative when the result “serve[s] a different function from the original [copied work,]” even when that new work is of a commercial nature.<sup>49</sup> In *Campbell v. Acuff-Rose Music, Inc.*, the Court noted that the lower court erred in cutting short its fair use inquiry simply because the allegedly infringing song had a “blatantly commercial

---

<sup>40</sup> *Id.* at 553 (Gorsuch J., concurring) (emphasis added).

<sup>41</sup> *Id.* at 554.

<sup>42</sup> *Id.* at 513 (majority opinion).

<sup>43</sup> *Id.*

<sup>44</sup> *Id.* at 554-55 (Gorsuch, J., concurring).

<sup>45</sup> *Id.* at 566, 555.

<sup>46</sup> *Id.* at 545 (majority opinion).

<sup>47</sup> *Id.*

<sup>48</sup> *Id.*

<sup>49</sup> *Authors Guild v. Google, Inc.*, 804 F.3d 202, 217 (2d Cir. 2015) (quoting *HathiTrust*, 755 F.3d at 97).



purpose.”<sup>50</sup> Instead, the Court emphasized that all factors must be considered together, noting that “the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use.”<sup>51</sup>

B. Factor 2. “[T]he nature of the copyrighted work.”<sup>52</sup>

In examining the nature of the copyrighted work, courts look at “the level of creativity or functionality in the original work.”<sup>53</sup> This second factor suggests that creative works lie along a spectrum ranging from factual, informational works like directories and basic lists to highly expressive works like paintings, musical compositions, poems, and fictional stories.<sup>54</sup> The latter of these presumably lie closer to the core of copyright protection than the former, wherein fair use proves easier to establish.<sup>55</sup>

Case law further explains. In *Bill Graham Archives v. Dorling Kindersley Ltd.*, for instance, which concerned the reproduction of the Grateful Dead musical group’s copyrighted concert posters and tickets, the second factor favored the copyright holder because the images in question were “creative artworks, which traditionally serve as the core of intended copyright protection.”<sup>56</sup> However, the court acknowledged this factor was of limited use “where the creative work of art is being used for a transformative purpose.”<sup>57</sup> The transformative purpose in this case was to create a musical biography about the Grateful Dead.<sup>58</sup> Even when copied in their entirety, the publisher’s use of the poster and ticket images in the musical group’s biography was completely different than the original use of those posters and tickets to advertise and facilitate concerts.<sup>59</sup>

While generally receiving little airtime compared to the other three factors, the second factor was the primary focus in *Google, LLC v. Oracle America, Inc.*<sup>60</sup> In this case, parties argued whether and where to position the computer programming element “declaring code” along the second factor’s copyright protection spectrum.<sup>61</sup> In

---

<sup>50</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 569. See also H.R. REP. NO. 94-1476 at 66 (1976) (clarifying that the commercial or non-profit character of an activity “is not intended to be interpreted as any sort of not-for-profit limitation on educational uses of copyrighted works” but rather “an express recognition that, as under the present law, the commercial or non-profit character of an activity, while not conclusive with respect to fair use, can and should be weighed along with other factors in fair use decisions.”).

<sup>51</sup> *Id.* at 579. See also H.R. REP. NO. 94-1476 at 65 (noting “there is ample case law recognizing the existence of the doctrine and applying it.”).

<sup>52</sup> 17 U.S.C. § 107.

<sup>53</sup> *Google LLC*, 141 S. Ct. at 1215.

<sup>54</sup> 17 U.S.C. § 107; see *Campbell*, 510 U.S. at 586; see also *Google LLC*, 141 S. Ct. at 1215.

<sup>55</sup> *Campbell*, 510 U.S. at 586.

<sup>56</sup> *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 612 (2d Cir. 2006).

<sup>57</sup> *Id.*

<sup>58</sup> *Id.* at 606.

<sup>59</sup> *Id.* at 608–09.

<sup>60</sup> *Google LLC*, 141 S. Ct. 1183 (2021).

<sup>61</sup> *Id.* at 1192.

this 6-2 ruling, the Court held that Google's replication of 11,500 lines of code from Oracle's desktop/laptop platform constituted a fair use because the code was used to create the new Android Operating System (OS), a platform for the mobile/smartphone environment, which was distinctly different from the desktop/laptop environment from which Oracle's code was copied.<sup>62</sup>

While acknowledging that Oracle's Java SE code was protected under copyright, the Court determined the subset of code that Google copied was nonetheless transformative due to its functional nature in being "inherently bound together with uncopyrightable ideas (general task division and organization) and new creative expression (Android's own implementing code)."<sup>63</sup> By copying declaring code shortcuts from 37 Java packages and writing their own code for most implementing programs, Google helped Java programmers program, debug, modify, and build upon existing code for the new Android smartphone interface.<sup>64</sup> This transformative use, especially with questions as to whether the code was even copyrightable at all, tipped the balance of this second factor in favor of fair use.<sup>65</sup>

C. Factor 3. "[T]he amount and substantiality of the portion used in relation to the copyrighted work as a whole."<sup>66</sup>

The third fair use factor considers the amount and substantiality of material copied from the allegedly infringed work, specifically whether more material than necessary was copied and the qualitative significance of that copied portion.<sup>67</sup> Even the copying of a small amount of material can fall outside the scope of fair use when that material constitutes the "heart" of the work, as seen in *Harper & Row, Publishers v. Nation Enterprises*.<sup>68</sup> In this case, the Court held that Nation magazine's publication of 300 words of direct quotations from Gerald Ford's yet-unpublished memoir constituted infringement, as that small portion was of "central importance[.]" essentially constituting the "heart of the book."<sup>69</sup>

While copying even a small amount of material can constitute infringement, copying a large amount—even an entire work—can still constitute a fair use, particularly when that use is transformative.<sup>70</sup> In *Authors Guild, Inc. v. HathiTrust*, for example, authors and author associations ("Authors") sued a collection of higher education research libraries and other nonprofit institutions ("Hathitrust") for digitizing more than ten million copyrighted works for the purpose of creating a full-text

---

<sup>62</sup> *Id.* at 1202 (calling smartphones a "very different context"). Also implicated was whether the copied code constituted an uncopyrightable "idea, procedure, process, system, method of operation, concept, principle, or discovery" under 17 U.S.C. § 102(b).

<sup>63</sup> *Id.* at 1202.

<sup>64</sup> *Id.* at 1193.

<sup>65</sup> *Id.* at 1197 (electing to "assume, but purely for argument's sake, that the entire Sun Java API falls within the definition of that which can be copyrighted.").

<sup>66</sup> 17 U.S.C. § 107.

<sup>67</sup> *Id.*

<sup>68</sup> *Harper & Row, Publs., Inc. v. Nation Enters.*, 557 F. Supp. 1067, 1072 (S.D.N.Y. 1983).

<sup>69</sup> *Id.*

<sup>70</sup> *Google LLC*, 141 S. Ct. at 1189.

searchable database.<sup>71</sup> The resultant HathiTrust Digital Library (“HDL”) allowed the general public to search for terms across all the repository’s digital copies, with search results producing only the page numbers and number of instances in which those terms appeared (unless the copyright holder authorized wider access).<sup>72</sup> The HDL also allowed persons with print disabilities to access electronic formats that converted text into spoken words.<sup>73</sup> As a third benefit, the HDL enabled a process to create replacement copies as allowable under 17 U.S.C. § 109(a), which allows libraries and other lawful owners of copyrighted works to sell, lend, or dispose of that specific copy without seeking permission from the copyright owner.<sup>74</sup>

The District Court awarded summary judgment to HathiTrust on account of the HDL’s “transformative” nature in using the copyrighted works to create something new, with particular credit to HDL’s “invaluable” contribution to advancing knowledge and realizing the ideals of the Americans with Disabilities Act of 1990 in making printed materials more accessible to individuals with print disabilities.<sup>75</sup> On appeal, the Second Circuit affirmed that the creation of digital copies of entire copyrighted works for the purpose of creating a full-text search of books was a “quintessentially transformative” fair use, as HDL did not “add into circulation any new, human-readable copies of any books[,]” but rather simply enabled a “word search.”<sup>76</sup> Even though entire works were copied, the digitization of those entire works was necessary to create the searchable database.<sup>77</sup> Even the results of the word search, in displaying only page numbers and number of occurrences, were “different in purpose, character, expression, meaning, and message from the page (and the book) from which it [was] drawn.”<sup>78</sup>

Another case, *Authors Guild v. Google, Inc.*, involved Google’s alleged infringement in creating Google Books, a search tool that allowed users to see “snippets” of text within books that contained specific search terms.<sup>79</sup> In creating this free, publicly available tool, Google made digital copies of tens of millions of books submitted by major libraries.<sup>80</sup> In exchange, and under agreement that those digital copies would not be used in a way that would violate copyright, Google provided the libraries with a digital copy of the books they submitted.<sup>81</sup> To build the Google Books tool, Google effectively created two digital copies of copyrighted works – one for use in their search engine, and the one provided to libraries.<sup>82</sup>

---

<sup>71</sup> *HathiTrust*, 755 F.3d at 90.

<sup>72</sup> *Id.* at 91.

<sup>73</sup> 17 U.S.C. § 121.

<sup>74</sup> *HathiTrust*, 755 F.3d at 92; 17 U.S.C. § 109(a).

<sup>75</sup> *Id.* at 93 (quoting *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d 445, 464 (S.D.N.Y. 2012)).

<sup>76</sup> *Id.* at 97.

<sup>77</sup> *Id.* at 98.

<sup>78</sup> *Id.* at 97.

<sup>79</sup> *Authors Guild*, 804 F.3d at 207.

<sup>80</sup> *Id.*

<sup>81</sup> *Id.*

<sup>82</sup> *Id.*

As the Google Books case went to litigation, the court of first instance awarded Google summary judgment based on Google's fair use defense argument, a judgment which the Second Circuit later affirmed.<sup>83</sup> Amongst other arguments on appeal, the authors and similarly situated plaintiffs (which included both mainstream and educational textbook publishers) argued that Google's digital copying of entire books was not transformative within the meaning of *Campbell* because the digital copies effectively substituted the plaintiffs' works.<sup>84</sup> Plaintiffs also argued that Google's "ultimate commercial profit motivation" weighed against fair use because the Books project essentially contributed to Google's commercial bottom line.<sup>85</sup> Just as in *HathiTrust*, however, the court determined that Google's use was transformative insofar as the digitization of entire works was necessary to create the new system that helped users locate words and phrases across millions of books.<sup>86</sup>

Similarly, in *Bill Graham Archives*, the Second Circuit held the copying of entire concert posters and tickets qualified as a fair use because the 480-page coffee table biography served the "transformative purpose of enhancing the biographical information . . . , a purpose separate and distinct from the original artistic and promotional purpose for which the images were created."<sup>87</sup> As the court related, "copying the entirety of a work is sometimes necessary to make a fair use of the image."<sup>88</sup> Therefore, no bright line exists when considering the amount copied under the third factor, with "the extent of permissible copying var[ying] with the purpose and character of the use."<sup>89</sup>

D. Factor 4. "[T]he effect of the use upon the potential market for or value of the copyrighted work."<sup>90</sup>

The fourth fair use factor examines market effects, or how the use of the allegedly infringing work impacts the market for the copyrighted work.<sup>91</sup> As the Second Circuit related, "copyright is a commercial right, intended to protect the ability of authors to profit from the exclusive right to merchandise their own work."<sup>92</sup> In other words, a new work that reduces the market demand for the original work and deprives the copyright holder of profits would likely not constitute a fair use because that new work would effectively circumvent the very reasons why copyright law exists in the first place.

---

<sup>83</sup> *Id.* at 208.

<sup>84</sup> *Id.* at 207.

<sup>85</sup> *Id.*

<sup>86</sup> *Id.* at 221.

<sup>87</sup> *Bill Graham Archives*, 448 F.3d at 610.

<sup>88</sup> *Id.* at 613 ("See *Kelly [v. Arriba Soft Corp.]*, 336 F.3d [811,] 821 ([ ] images used for a search engine database are necessarily copied in their entirety for the purpose of recognition); *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 24 (1st Cir. 2000) ([ ] to [have] cop[ied] any less than [the entire image] would have made the picture useless to the story)").

<sup>89</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 586-87 (1994).

<sup>90</sup> 17 U.S.C. § 107.

<sup>91</sup> *Id.*

<sup>92</sup> *Authors Guild*, 804 F.3d at 214.

This fourth factor requires courts to consider “not only the extent of market harm caused by the particular actions of the alleged infringer, but also ‘whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market’ for the original.”<sup>93</sup> Markets play a key role in this determination, as the substitution must occur in the *same* market as that of the original.<sup>94</sup> In *Sony v. Universal City Studios*, for instance, which concerned the consumer use of Sony’s Betamax videocassette recorder to tape copyrighted television programs, Universal failed to prove that Sony reduced the “value of the Studios’ copyrighted works in *present* markets.”<sup>95</sup> Instead, the Court recognized that the Sony Betamax recorder had created a new market for “persons who desire to view television programs at times other than when they are broadcast.”<sup>96</sup>

In addition to market effects and substitution, the fourth factor also considers whether the benefits to the public outweigh any harms incurred by the copyright holder.<sup>97</sup> In *Sony*, for instance, even if harm in the market of television programming had been proven, that harm was outweighed by the public’s interest in the “free flow of ideas, information, and commerce” that comes with the access and use of new technology.<sup>98</sup> This weighing of public interests against private interests underscores the case-by-case nature of fair use, as well as how the incentive-based premise of copyright law is designed to ultimately benefit the public.<sup>99</sup> In essence, while all four factors are involved in making the determination, the “ultimate test of fair use . . . is whether the copyright law’s goal of promoting the Progress of Science and useful Arts would be better served by allowing the use than by preventing it.”<sup>100</sup>

In summary, use is key throughout the fair use analysis, and as Justice Gorsuch stated in *Warhol*, the four factors “suppl[y] courts with a sequential chain of questions about the particular challenged use—starting with its *purpose* and *character* (in the first factor) and ending with its *effect* (in the fourth).”<sup>101</sup> No factor should “be treated

---

<sup>93</sup> *Campbell*, 510 U.S. at 590 (quoting *Nimmer* § 13.05[A][4], p. 13–102.61 (footnote omitted)); *accord*, *Harper & Row, Publs.*, 471 U.S. at 569 (1985); *Folsom v. Marsh*, 9 F. Cas. 342, 349 (1841).

<sup>94</sup> 4 NIMMER ON COPYRIGHT § 13F.08 (2024).

<sup>95</sup> *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984). Noteworthy to Sony was also the recognition of the “substantial non-infringing use” doctrine, where a product or technology that is capable of infringing uses may nonetheless escape infringement claims when that technology displays significant non-infringing uses.

<sup>96</sup> *Id.*

<sup>97</sup> *MCA, Inc. v. Wilson*, 677 F.2d 180, 183 (2d Cir. 1981) (noting the need to balance benefits to the public against losses to the copyright owner).

<sup>98</sup> *Sony Corp.*, 464 U.S. at 429.

<sup>99</sup> *See* Leval, *supra* note 32, at 1135–36 (“The stimulation of creative thought and authorship for the benefit of society depends assuredly on the protection of the author’s monopoly. But it depends equally on the recognition that the monopoly must have limits. Those limits include the public dedication of facts (notwithstanding the author’s efforts in uncovering them); the public dedication of ideas (notwithstanding the author’s creation); and the public dedication of the right to make fair use of material covered by the copyright.”).

<sup>100</sup> *Castle Rock Entm’t, Inc. v. Carol Publ’g Group*, 150 F.3d 132, 141 (2d Cir. 1998) (quoting U.S. CONST. art. I, § 8, cl. 8).

<sup>101</sup> *Andy Warhol Found.*, 598 U.S. at 555 (Gorsuch, J., concurring).

in isolation, one from another[.]”<sup>102</sup> nor should courts “speculate about artistic ambitions or aesthetics.”<sup>103</sup> The use of the new work remains central, with the fourth factor “assess[ing] whether the challenged use (as revealed by its purpose, character, amount of source material used, and effect) serves as a complement to or a substitute for a copyrighted work.”<sup>104</sup>

So, while Generative Artificial Intelligence tools may be new to the general public, a long line of established case law provides guidance on one of the most pressing, constantly recurring issues facing new AI technologies: does the ingestion of full copyrighted works in training Generative AI tools constitute a transformative, fair use? In this next section, I apply the insight from the aforementioned cases to conclude that, yes, the ingestion of copyrighted works for the purpose of training Generative AI tools does in fact constitute fair use. After applying case law, I discuss implications for both creators and copyright law as humanity moves forward in an increasingly artificially intelligent world.

### III. Application of Fair Use Case Law to Generative Artificial Intelligence (AI) Tools

In this section, I apply case law insight to support the conclusion that the ingestion of full copyrighted works for the purpose of training Generative AI tools constitutes a transformative, fair use under U.S. Copyright law. See factor-by-factor analysis with respect to Generative AI below:

- A. Factor one weighs in favor of fair use, because the purpose and character of the use of works ingested into Generative AI fundamentally differ from the use of the Generative AI tools themselves.

Factor one examines “the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes.”<sup>105</sup> This analysis requires one to separate (1) the use of the creative works ingested into Generative AI tools from (2) the use of Generative AI tools themselves.<sup>106</sup> This “comparatively modest inquiry focuse[s] on how and for what reason a person is using a copyrighted work in the world, not on the moods of any artist or the aesthetic quality of any creation.”<sup>107</sup> In making this distinction, the two uses differ.

The creative works ingested into Generative AI tools serve numerous purposes – self-expression; social commentary; entertainment; criticism; catharsis; cultural preservation; documentation; education; innovation; aesthetic enjoyment; spiritual

---

<sup>102</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 578.

<sup>103</sup> *Andy Warhol Found.*, 598 U.S. at 555 (Gorsuch, J., concurring).

<sup>104</sup> *Id.*

<sup>105</sup> 17 U.S.C. § 107. Confusion seems to arise when the purpose and character of the use of Generative AI tools are conflated with the purpose and character of the output generated by the users of those tools, a separate question touched upon in the implications section but less relevant to the current analysis.

<sup>106</sup> *See Andy Warhol Found.*, 598 U.S. at 556 (Gorsuch, J., concurring).

<sup>107</sup> *Id.* at 554.

exploration; activism; or even the simple attempt to earn a living. As the cases outlined in the previous section suggest, most creators would likely agree they do not create works to feed Generative AI tools or to amass so many works for the purpose of remixing them in response to user prompts. In this way, the purpose and character of the two uses differ.

This first factor inquiry also considers whether “such use is of a commercial nature or is for nonprofit educational purposes[,]” with a company’s “blatantly commercial purpose” not necessarily being dispositive.<sup>108</sup> This is one area where the uses of creative works and Generative AI tools overlap: both usually have a shared purpose to generate commercial or economic value. Just as many “[creators’] livelihoods are dependent on sharing their content on the internet[,]” companies engaged with Generative AI also look to generate profit.<sup>109</sup> Conversely, while most Generative AI tools make versions available to the public for free (or, more precisely, in exchange for the user’s input, data which serves to further refine the model), the latest and most sophisticated versions are typically subscription-based. So, even though Generative AI tools may have various commercial applications, the widespread, free availability of those tools to the public—even if not the most recent versions—contributes to the tools’ transformative character. This transformative character, as discussed in the following section, is possibly the most compelling reason that the first factor weighs in favor of fair use.

Turning from the purpose to the character of the use, the first factor also considers whether the challenged use of Generative AI tools is transformative, an inquiry which asks advocates and defenders of creators to stare, perhaps uncomfortably, into the very soul of fair use. Factor one is called the soul of fair use for speaking to “whether the particular quotation is of the transformative type that advances knowledge and the progress of the arts or whether it merely repackages, free riding on another’s creations.”<sup>110</sup> Generative AI tools that facilitate conversations, structure communications, generate ideas, and assist with mundane tasks simply serve a different purpose from the creative works upon which they were trained.<sup>111</sup> By adding value and creating “new information, new aesthetics, new insights and understandings,” the use of full copyrighted works to train Generative AI models arguably constitutes “the very type of activity that the fair use doctrine intends to protect for the enrichment of society.”<sup>112</sup>

---

<sup>108</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 574–79; see also H.R. REP. NO. 94-1476 at 66 (1976) (clarifying the commercial character of an activity “is not intended to be interpreted as any sort of not-for-profit limitation on educational uses of copyrighted works” but rather “an express recognition that, as under the present law, the commercial or non-profit character of an activity, while not conclusive with respect to fair use, can and should be weighed along with other factors in fair use decisions.”).

<sup>109</sup> Second Amended Complaint at 18, *Leovy v. Google LLC*, No. 3:23-cv-3440-AMO (N.D. Cal. 2024).

<sup>110</sup> Leval, *supra* note 32, at 1116.

<sup>111</sup> *See Is Generative AI a Gamechanger?* J.P. MORGAN (Feb. 14, 2024), <https://www.jpmorgan.com/insights/global-research/artificial-intelligence/generative-ai>.

<sup>112</sup> Leval, *supra* note 32, at 1111.

The first factor analysis for Generative AI is not as clear as that in *Warhol*, where the use of both Goldsmith's purple Prince photo and Warhol's infringing orange Prince photo exhibited the exact same purpose and character in being licensed to the same for-profit magazine for the same express purpose of publishing an article concerning Prince.<sup>113</sup> With Generative AI, in contrast, a multitude of corpora are ingested and used to generate humanlike output geared toward a multitude of unique, user-prompted purposes. The use is simply more varied and more complicated than the straightforward use in *Warhol*.

In summary, with exception to the sometimes shared commercial purpose of ingested works and Generative AI tools, the purpose and character of the respective uses differ. As scholar Matthew Sag notes, applying copyright's fundamental principles in the context of new technologies necessarily implies that copying expressive works for non-expressive purposes should not be counted as infringement and must be recognized as fair use.<sup>114</sup> Generative AI tools are non-expressive per se (as separated from their output, which has the potential to be expressive if used as such, an inquiry that again falls outside the scope of this analysis), and the purpose and character of the two uses simply differ. Therefore, with the two uses displaying such different purposes and characters, factor one argues in favor of Generative AI tools as a fair, transformative use.

- B. Factor two also weighs in favor of fair use, even when expressive works are ingested, because of the transformative nature of Generative AI tools.

The second factor presents a challenge because "the nature of [a] copyrighted work" hinges on the nature of the allegedly infringing work in question, whereas Generative AI tools are trained on countless works spanning the entire copyright spectrum. AI tools are trained on non-expressive works such as databases, statutes, recipes, technical documentation, historical records, catalogs, and survey data, each of which lie far from "the core of intended copyright protection."<sup>115</sup> Just as the court in *Google, LLC v. Oracle America, Inc.* determined the second factor weighed in favor of fair use because the copied code may not even have been copyrightable, the second factor clearly weighs in favor of fair use with respect to Generative AI tools that ingest more "informational or functional" works in training their models.<sup>116</sup>

On the other end of the copyright spectrum are expressive works, those creative works which lie closer to "the core of intended copyright protection."<sup>117</sup> AI tools are trained using myriad of expressive works, for instance the works of fiction currently serving as the subject of complaint in *Authors Guild v. OpenAI*. So, while non-

---

<sup>113</sup> See *Andy Warhol Found.*, 598 U.S. at 511.

<sup>114</sup> See Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC'Y OF THE U.S.A. 291, 294 (2019).

<sup>115</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 586.

<sup>116</sup> See *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 50 (2021) (quoting 4 M. NIMMER & D. NIMMER, Copyright §13.05(A)(2)(a) (2019)).

<sup>117</sup> *Bill Graham Archives*, 448 F.3d at 612.



expressive works generally present no issue, where expressive works are involved, the second factor weighs less in favor of fair use.

To untangle this issue, courts have held that a work's level of expressiveness proves of limited usefulness when those works are used for a transformative purpose.<sup>118</sup> In *Bill Graham Archives*, for example, the New York Southern District Court held that the Grateful Dead concert posters and ticket images were “transformatively different” when used to illustrate a timeline of the band for the purpose of commemorating historic events in a musical biography.<sup>119</sup> If the use of Grateful Dead concert posters and tickets was transformative when used to talk about the exact same Grateful Dead band, then so too should be the use of countless unrelated expressive “classical love and romance” and “middle school angst” works be transformative when used to generate a response to “ten ideas for the next Lifetime Afterschool Special.” Extending the analogy, so too should a language model trained on a corpus of legal documents to generate dialogues for fictional courtroom drama be transformative, as would be a model comprising a dataset of historical political speeches to generate chatbot responses for inquiries on local civic events and county ordinances. Simply put, when used for a transformative purpose, a work's level of expressiveness simply proves less relevant.<sup>120</sup> On balance, this second “nature of the copyrighted work” factor also weighs in favor of fair use, regardless of whether the ingested works are informational or expressive, due to the generally transformative nature of Generative AI tools.

- C. Factor three weighs in favor of fair use, because copying entire works is likely necessary to achieve Generative AI tools' transformative purpose(s).

The third factor inquiry into the “amount and substantiality of the portion used” considers whether more material than necessary was copied and the qualitative significance of that copied portion “in relation to the copyrighted work as a whole.”<sup>121</sup> *Harper & Row* provides little guidance here, as most Generative AI tools ingest entire works, thus making moot whether anything of central “importance” or the “heart of the [work]” was ingested.<sup>122</sup> Instead, *HathiTrust* proves more instructive, as the entirety of over ten million works were copied to create the HDL full-text searchable database.<sup>123</sup> That said, *HathiTrust* also differs in that the purpose of the HDL was to direct users to the very copyrighted works upon which the system was built.<sup>124</sup> Generative AI tools, on the other hand, use entire copyrighted works to generate

---

<sup>118</sup> See *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (concluding that images used for a search engine database are necessarily copied in their entirety for the purpose of recognition). See also *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 24 (1st Cir. 2000) (concluding that to copy any less than entire images would have rendered the images useless to the story).

<sup>119</sup> *Bill Graham Archives*, 448 F.3d at 609.

<sup>120</sup> See *Kelly*, 336 F.3d at 8221 (copying entire images for a search engine database was necessary for the purpose of recognition). See also *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d at 24 (copying anything less than the entire image would have made the picture useless to the story).

<sup>121</sup> *Id.*

<sup>122</sup> *Harper & Row, Publ'r.*, 471 U.S. at 549, 565.

<sup>123</sup> *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 90 (2nd Cir. 2014).

<sup>124</sup> *Id.* at 91–92.

humanlike output, with generally zero reference back to the works that contributed to generating that expression.

Although courts consider each factor separately, transformation remains the common thread woven through each factor. In *Authors Guild v. Google, Inc.*, the Second Circuit found that Google’s copying of tens of millions of books was essential to create the decidedly transformative, publicly available Google Books search tool.<sup>125</sup> In order to return “snippets” of all the books reflecting the user’s search terms, Google needed to have copied entire works.<sup>126</sup> The court held this copying was a fair use because the Google Books search tool presented a “highly transformative purpose” that “augment[ed] public knowledge by making available information *about* Plaintiffs’ books without providing the public with a substantial substitute for matter protected by the Plaintiffs’ copyright interests[.]”<sup>127</sup> Equally effective was this argument in *HathiTrust*, as the Second Circuit again found factor three weighing in favor of fair use because the entire amounts copied were necessary to create the HDL full-text search system.<sup>128</sup> Just as Google “added something new and important” by helping Java programmers write for mobile platforms, the copying of entire works was likely also necessary for AI tools to generate human-like output and further the “creative ‘progress’ that is the basic constitutional objective of copyright itself.”<sup>129</sup> Generative AI tools require massive amounts of data to analyze syntax and learn patterns that help the tool “understand” context and generate outputs that respond to an endless variety of contexts.<sup>130</sup> In summary, because the output produced by a Generative AI tool often looks nothing like the works used to train the model, and because the copying of entire works is likely necessary to achieve Generative AI’s transformative use of those works, on balance, the third factor also weighs in favor of fair use.

- D. Factor four also weighs in favor of fair use because a market that preferences the personalized output of Generative AI would likely create a separate market from that of readers who prefer works from individual creators.

The fourth factor examines “the effect of the use upon the potential market for or value of the copyrighted work.”<sup>131</sup> Generative AI tools differ from search tools like Google Books and the HDL, which are cases where “[the] theory of market harm does not work . . . , because the full-text search function does not serve as a substitute for the books that are being searched.”<sup>132</sup> As such, while still overall weighing in

---

<sup>125</sup> *Authors Guild*, 804 F.3d at 229.

<sup>126</sup> *Id.* at 217.

<sup>127</sup> *Id.* at 207.

<sup>128</sup> *Authors Guild*, 755 F.3d at 98.

<sup>129</sup> *Google LLC*, 141 S. Ct. at 1203 (citing *Feist Publications Inc., v. Rural Telephone Service Co.*, 499 U. S. at 349–50 (noting “[t]he primary objective of copyright is not to reward the labor of authors, but ‘[t]o promote the Progress of Science and useful Arts.’”) (quoting U. S. Const., Art. I, §8, cl. 8)).

<sup>130</sup> *Id.* at 1203. *See also Feist*, 499 U. S. at 349–50 (noting “[t]he primary objective of copyright is not to reward the labor of authors, but ‘[t]o promote the Progress of Science and useful Arts.’”).

<sup>131</sup> 17 U.S.C. § 107.

<sup>132</sup> *HathiTrust*, 755 F.3d at 100.

favor of fair use for the reasons discussed below, this fourth factor produces the least favorable analysis with respect to Generative AI tools.

With Generative AI, market harms are more plausible because the tools can generate new works on demand that could potentially substitute or dilute the value of copyrighted works. The text-to-image model in Stable Diffusion 2.0, for instance, currently generates images with resolutions of 2048x2048 pixels or better, which provides adequate detail for products like digital banners, website graphics, printed postcards, flyers, and other marketing materials.<sup>133</sup> In absence of this tool, consumers would turn to image services like Shutterstock or Getty Images.<sup>134</sup> Similarly, if AI-generated works were to flood and dilute the market, creators' works would likely be more difficult to find, inevitably impacting the value of those works. In fact, Amazon's recent policy, limiting self-published Kindle eBooks to three a day per account, foreshadows and attempts to safeguard against this potential for market dilution.<sup>135</sup>

That said, Generative AI tools may have indeed created a separate market all their own. Just as the Sony Betamax recorder created a new market for "persons who desire to view television programs at times other than when they are broadcast," so too have Generative AI tools potentially created a myriad of new markets, including a market for individuals who prefer to read hyper-personalized works.<sup>136</sup> Case in point—nowhere is the bookstore section for a fancy vegan dinner murder mystery that can be read in fewer than two hours starring guests to include myself, my sister Judy, Colonel Mustard, Calamity Jane, and the Wizard of Oz, who all gather at Kilkenny Castle on Christmas Eve 1885. Given the potential for AI to establish multiple "new markets[.]" courts may elect to examine the aspect of market effects on a case-by-case basis.

The potential for AI tools to "hallucinate" and generate an infringing use is also less an issue, as *Sony's* substantial non-infringing use doctrine instructs that, just because a Generative AI tool is capable of infringing uses, the product or technology should not be found to be infringing in light of its significant non-infringing uses.<sup>137</sup> It would be different if Generative AI tools like ChatGPT produced full copyrighted texts at the user's prompt, but just as Google Books only showed "snippets" to control the amount of text visible to searchers, so too have companies like OpenAI pointedly removed the ability to return verbatim excerpts from copyrighted texts.<sup>138</sup>

---

<sup>133</sup> *Stable Diffusion 2.0 Release*, STABILITY AI (Nov. 24, 2023), <https://stability.ai/news/stable-diffusion-v2-release>.

<sup>134</sup> Ironically, many of these services have also introduced their own AI-related components. See Carl Franzen, *Shutterstock Debuts an AI Image Editor for its 750-Million Picture Library*, VENTUREBEAT (Oct. 26, 2023), <https://venturebeat.com/ai/shutterstock-debuts-an-ai-image-editor-for-its-750-million-picture-library/>.

<sup>135</sup> Sascha Brodsky, *The AI-Generated Books Trend is Getting Worse*, INFORMA (Jan. 24, 2024), <https://aibusiness.com/responsible-ai/the-ai-generated-books-phenomenon-is-getting-worse>.

<sup>136</sup> Sony Corp., 464 U.S. at 498; see generally *Id.*

<sup>137</sup> *Id.* at 442.

<sup>138</sup> See Class Action Complaint at 12, *Authors Guild v. OpenAI*, No. 1:23-cv-08292 (S.D.N.Y. Sept. 19, 2023).

The effect of the use on the potential market is also diminished when the output of the challenged use significantly differs from the creative works used in creating the transformative system. Such was the case in *Hathitrust*, where the results produced by the HDL search looked nothing like the original works used to create the tool (see image below).<sup>139</sup> So, too, does most of the output of Generative AI tools look nothing like the works used in each tool’s respective training. For example, Generative AI models may draw from numerous corpora to generate several haiku at the user’s prompting, while none of the works in those corpora may actually themselves be haiku. Similarly, a Generative AI model may draw from millions of recipes to generate an adventure story about a capricious carrot, militant mushroom, and opinionated onion that ends with all the vegetables gleefully grasping tendrils and leaping together into a pot of boiling water.



Figure 1. J.A. 681 ¶ 80 (Wilkin Decl.), at 91. The HDL’s output had no similarities to the creative works copied in the process of its building.

Another nuance distinguishing both *Hathitrust* and the fair use of copyrighted works in training Generative AI tools is their potential to serve the public interest. Service to the public interest carries great weight, and the more the use of a copyrighted work promotes the “Progress of Science and the Useful Arts,” such as fostering creativity, innovation, education, and advancing visual and other arts, the more likely that new work will constitute a fair use.<sup>140</sup> The digital library offered by the Hathitrust non-profit allowed the general public to search and access digital copies of public domain titles.<sup>141</sup> The HDL also served a public interest function by allowing

<sup>139</sup> J.A. 681 ¶ 80 (Wilkin Decl.) (copied from *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91 (2d Cir. 2014)).

<sup>140</sup> U.S. Const. art. I, § 8, cl. 8.

<sup>141</sup> *How to Search & Access*, HATHITRUST, <https://www.hathitrust.org/the-collection/search-access/#hathitrust-access-chart> (last visited Mar. 17, 2024).

print-disabled patrons of member libraries and institutions to access electronic formats of otherwise inaccessible texts.<sup>142</sup>

Even if the recreational reader market turned to reading AI-generated works over those written by individual human creators, such a shift arguably creates a separate market, one transformative and allowable under fair use. By making versions freely available to the general public, Generative AI tools also serve the public interest just as significantly as the HDL in *Hathitrust* and Google Books project.<sup>143</sup> For these reasons, reasons which only further support and reinforce the transformative nature of using entire copyrighted works to train Generative AI tools, factor four also weighs in favor of fair use.

On the one hand, Generative AI companies with their “ultimate commercial profit motivation[s]” may appear less inclined to serve the public interest in keeping the most recent versions of their tools behind a paywall.<sup>144</sup> Despite that, these companies still serve the public interest by making some version of their tools freely available, and the widespread usefulness of those tools arguably outweighs any negatives attributable to commercial motivation. Even the earliest versions of tools like ChatGPT still give anyone with internet access the opportunity to receive personal-assistant-like help in automating routine tasks that free time and energy for more complex, creative work.<sup>145</sup> Generative AI tools answer questions, draft emails, create schedules, and summarize documents. These tools quickly and efficiently sift through mass amounts of data to find relevant insight, generate literature reviews, and extract key points, all at a speed unrivaled by any human counterpart. Furthermore, unlike human labor, Generative AI tools are available around-the-clock, assisting humans whenever they need, in whatever manner they need, with no frustrations as to how many iterations are required to arrive at what the user deems to be an acceptable end product.

One must also acknowledge Generative AI tools’ potential to support functions vital to the public interest, specifically those that ease the workload of teachers and other public servants. Most teachers work long hours outside the classroom to plan lessons, activities, and other educational experiences.<sup>146</sup> Generative AI tools can help alleviate these burdens by drafting and honing lesson plans, generating ideas for activities that engage students, and helping craft themes, create shopping lists, and facilitate logistics for the myriad of educational and social experiences teachers end up

---

<sup>142</sup> *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91 (2d Cir. 2014). Here, the public interest function was limited in only serving print-disabled patrons of those member institutions, rather than the public at large. In contrast, anyone with internet access can take advantage of freely available Generative AI tools.

<sup>143</sup> *Authors Guild*, 804 F.3d at 207.

<sup>144</sup> *Id.*

<sup>145</sup> *Id.* at 207-09. When building its own Java API, Sun also repurposed existing interfaces.

<sup>146</sup> Tim Walker, *Survey: Teachers Work More Hours Per Week Than Other Working Adults*, NEA NEWS (Sept. 18, 2023), <https://www.nea.org/nea-today/all-news-articles/survey-teachers-work-more-hours-week-other-working-adults>.

responsible for planning.<sup>147</sup> For students, Generative AI tools can provide personalized feedback, generate practice questions, and present interactive scenarios that help learners grasp concepts.<sup>148</sup>

The public benefits extend far beyond the individual. Generative AI tools are helping address society's most complex problems, from fighting global warming to predicting extreme weather events and detecting fires before they race out of control.<sup>149</sup> AI tools are detecting, predicting, and treating a variety of medical disorders.<sup>150</sup> Startups are also using Generative AI to help reduce greenhouse gas emissions, optimize energy use, and facilitate the shift to a zero-carbon economy.<sup>151</sup> Both for the individual and at scale, Generative AI tools are serving the public interest, and without the ability to ingest massive amounts of copyrighted works, these tools would likely not serve that interest as effectively. In this respect, the public interest functions that some tools provide outweighs any possible detriment stemming from an "ultimate commercial profit motivation."<sup>152</sup>

#### IV. Implications & Future Directions

While the ingestion of full copyrighted works into Generative AI tools may be acceptable under the fair use exception to U.S. copyright law, copyright law itself remains solid, and creators still have ample recourse against individuals who infringe upon their exclusive rights. For example, creators still have a valid copyright infringement claim against individuals claiming authorship over AI-generated output that exhibits a substantial similarity to the creator's copyrighted work.<sup>153</sup> Courts are poised

---

<sup>147</sup> See David Wiley, *Teaching Assistants that Actually Assist Instructors with Teaching*, IMPROVING LEARNING BLOG (Jul. 13, 2023), <https://opencontent.org/blog/archives/7271>. See also Beth McMurtrie & Beckie Supiano, *ChatGPT has Changed Teaching. Our Readers Tell Us How*, THE CHRON. OF HIGHER EDUC. (Dec. 11, 2023), <https://www.chronicle.com/article/chatgpt-has-changed-teaching-our-readers-told-us-how>.

<sup>148</sup> See Will Gendron, *AI is Going to Offer Every Student a Personalized Tutor, Founder of Khan Academy Says*, BUS. INSIDER (May 4, 2023), [https://www.businessinsider.com/ai-will-give-every-student-personalized-tutor-sal-khan-academy-2023-5; Is ChatGPT Better Than a Human Tutor? Survey Says. . . Yes, GOV'T TECH. \(May 25, 2023\), https://www.govtech.com/education/k-12/is-chatgpt-better-than-a-human-tutor-survey-says-yes](https://www.businessinsider.com/ai-will-give-every-student-personalized-tutor-sal-khan-academy-2023-5; Is ChatGPT Better Than a Human Tutor? Survey Says. . . Yes, GOV'T TECH. (May 25, 2023), https://www.govtech.com/education/k-12/is-chatgpt-better-than-a-human-tutor-survey-says-yes) (summarizing survey results from 801 students who reported studying with ChatGPT was more effective than a human tutor).

<sup>149</sup> Miles O'Brien, *Scientists Harness Power of Artificial Intelligence to Battle Wildfires*, PBS NEWSHOUR (Mar. 27, 2024), <https://www.pbs.org/newshour/show/scientists-harness-power-of-artificial-intelligence-to-battle-wildfires>.

<sup>150</sup> Niam Yaraghi, *Generative AI in Health Care: Opportunities, Challenges, and Policy*, BROOKINGS (Jan. 8, 2024), <https://www.brookings.edu/articles/generative-ai-in-health-care-opportunities-challenges-and-policy/> (discussing the ever-increasing reliance on AI-assisted decision-making in the health care industry).

<sup>151</sup> Lisbeth Kaufman and Benoit de Chateauevieux, *How Climate Tech Startups Use Generative AI to Address the Climate Crisis*, AWS STARTUPS BLOG (Feb. 19, 2024), <https://aws.amazon.com/blogs/startups/how-climate-tech-startups-use-generative-ai-to-address-the-climate-crisis/>.

<sup>152</sup> *Authors Guild*, 804 F.3d at 207.

<sup>153</sup> Under 17 U.S.C. Section 501(a), "anyone who violates any of the exclusive rights of the copyright owner. . . is an infringer of the copyright or right of the author comment as the case may be." See generally 4 Modern Federal Jury Instructions-Civil P 86B.02 (2024): To establish copyright

to continue supporting these claims, as seen in the Northern District of California's recent dismissal of *Kadrey v. Meta Platforms, Inc.*<sup>154</sup> This suit, which concerned the copying of authors' books in training Meta's Large Language Model Meta AI (LLaMA) language model, was dismissed because "[w]ithout any plausible allegation of an infringing output, there can be no vicarious infringement," (emphasis omitted).<sup>155</sup> This wording implies that an infringing output would have produced a valid claim, reinforcing that creators still have recourse under copyright law against works that infringe their exclusive rights.

Furthermore, while works freely available on the internet (regardless of copyright status) may be fair game for ingestion, companies that employ unlawful circumvention tools to access copyrighted content can face legal repercussions under the Digital Millennium Copyright Act (DMCA).<sup>156</sup> While the DMCA serves multiple purposes, one of its main goals is to prohibit the circumvention of technological measures that control access to copyrighted works, like decrypting an encrypted work or bypassing a password without the copyright owner's authority.<sup>157</sup> While the DMCA certainly targets the circumvention of digital protections guarding copyrighted material, the statute does not concern itself with the subsequent post-circumvention use of those materials, underscoring the possibility that even copyrighted works accessed unlawfully can still be lawfully used under copyright's fair use exception.<sup>158</sup>

In certain states, creators may also have valid state-level claims under "identity misappropriation" or the right of publicity, which speaks to "the inherent right of every human being to control the commercial use of his or her identity."<sup>159</sup> California, for example, prohibits any person from knowingly using another's name, voice, signature, photograph, or likeness for commercial purposes without prior consent.<sup>160</sup> Indiana provides similar protections for a personality's name, voice, signature, photograph, image, likeness, distinctive appearance, gestures, or mannerisms.<sup>161</sup>

---

infringement, the plaintiff must prove that (1) copying occurred of a copyrighted work, (2) the infringer had access to that work, and (3) substantial similarity between the copyrighted work and the alleged infringement.

<sup>154</sup> *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417-VC, 2023 U.S. Dist. LEXIS 207683, at \*2 (N.D. Cal. Nov. 20, 2023).

<sup>155</sup> *Id.*

<sup>156</sup> See 17 U.S.C. § 1201(c)(3) (noting that the DMCA "does not . . . exempt from liability circumvention tools otherwise deemed unlawful," quoting *Realnetworks, Inc. v. DVD Copy Control Ass'n*, 641 F. Supp. 2d 913, 941 (N.D. Cal.)).

<sup>157</sup> See 17 U.S.C. § 1201(a)(1)(A) ("No person shall circumvent a technological measure that effectively controls access to a [copyrighted work.]")

<sup>158</sup> *Id.*

<sup>159</sup> JT McCarthy, *McCarthy on Trademarks and Unfair Competition* (1992), § 28.01[1].

<sup>160</sup> Cal. Civ. Code § 3344 (Deering, Lexis Advance through the 2024 Regular Session Ch 1).

<sup>161</sup> Ind. Code Ann. § 32-36-1-1 (Burns, Lexis Advance through P.L. 4-2024 of the Second Regular Session of the 123rd General Assembly). (Noting that the Indiana statute does not apply to the use of a personality's attributes in literary works, theatrical works, musical compositions, film, radio, or television programs, or in connection with the broadcast or reporting of an event or a topic of general or public interest).

Consequently, recourse potentially exists even outside of copyright law, and these right of publicity claims are not necessarily preempted by federal copyright law because a person's image, identity, and personality are not in themselves copyrightable subject matter.<sup>162</sup>

Finally, the uncertainties surrounding artificial intelligence beg the question as to whether the copyright law should be amended to account for the introduction of these new technologies. The answer is no. While Generative AI tools certainly add new dimension to copyright law, they ultimately prove no more novel than those presented by other once-new technologies like the Betamax recorder, peer-to-peer file sharing, or even 3D printing technologies. Even with the advent of Generative AI, copyright law remains capable of balancing and safeguarding the rights of creators against the ultimate goal to promote the "Progress of Science and the Useful Arts" for the greater benefit of society.<sup>163</sup> While there are opportunities for courts to lay precedent solidifying the ingestion of copyrighted works as a fair use, any challenges presented by Generative AI seem to be completely workable within the current boundaries of the statute.

## V. Conclusion

Considering all four factors together in light of copyright's "ultimate test" in whether a challenged use promotes the "Progress of Science and the Useful Arts[.]" the public is better served by allowing the ingestion of copyrighted works rather than by preventing that use.<sup>164</sup> Generative AI tools are novel, innovative technologies that push the boundaries of natural language processing and human-computer interaction. The alternative to declaring the ingestion of entire copyrighted works would be to require companies to license each individual work used in training their models, which would amount to less robust systems and certainly "stifle the very creativity which [copyright] law is designed to foster."<sup>165</sup> Such a request also falls outside the boundaries of fair use, which mentions nothing about fairly using only those works which one commercially acquired. Rather, the statute refers to the fair use of *any* copyrighted work.<sup>166</sup>

In analyzing the fair use status of copyrighted works ingested for use in Generative AI tools, the public interest function remains key. The results of this analysis would differ if companies placed their tools behind a paywall. For now, though, by making versions available to the public, Generative AI tools serve copyright's fundamental goal in furthering human creativity, stimulating productive thought, and

---

<sup>162</sup> 1 The Law of Electronic Commercial Transactions § 3.11 (noting that claims that allege violations of the publicity right based on mere copying of photographic or digital images may well be copyright claims, rather than independent state law tort claims).

<sup>163</sup> U.S. Const. art. I, § 8, cl. 8.

<sup>164</sup> *Acuff-Rose Music, Inc.*, 510 U.S. at 578 ("All are to be explored, and the results weighed together, in light of the purposes of copyright.").

<sup>165</sup> *Stewart v. Abend*, 495 U.S. 207, 236 (1990).

<sup>166</sup> *HathiTrust*, 755 F.3d at 95.



providing pathways to help individuals express themselves across a variety of applications.

Beyond this, Generative AI tools also provide an essential social justice function, with many tools providing resources previously only available to those who could afford the human labor needed to generate ideas, assist with writing, handle logistical tasks, etc.<sup>167</sup> So while some may argue such functions dilute and “excessively diminish[] the incentives for creativity,” the reality is just the opposite.<sup>168</sup> Generative AI tools open the doors to creativity for those who can afford it and those who cannot. In weighing exclusive creator rights against the public interest potential for Generative AI to further the “Progress of Science and useful Arts[,]”<sup>169</sup> a clear winner emerges. The winner—and the ever-increasingly obvious, true soul of fair use—is the public interest.

---

<sup>167</sup> Betsy Rosenblatt, *Considering the Role of Fairness in Copyright Fair Use*, 61 HOUS. L. REV. 261, 293 (2023).

<sup>168</sup> Leval, *supra* note 32, at 1110 (noting a fair use “must be of a character that serves the copyright objective of stimulating productive thought and public instruction without excessively diminishing the incentives for creativity.”).

<sup>169</sup> U.S. Const. art. I, § 8, cl. 8.