

FACILITATIVE FAIR USE

MICHAEL MATTIOLI*

ABSTRACT

Recent generative artificial intelligence (AI) copyright lawsuits have ignited a crucial debate about the future of intellectual property. The technology involved is complex, but the essential conflict is simple: creators are trying to stop machines that are capable of imitating them from doing just that. These are familiar battle lines. From the loom to the first camera, inventions have long triggered anxieties about technology’s potential to displace artistry. Much of the commentary today focuses on AI’s potential to mimic creators, echoing these earlier concerns.

However, this framing misses a key insight: AI is a powerful tool for widely facilitating the production and dissemination of human creativity and artistry—two important goals of copyright law. Original case studies presented in this Article explore how AI is helping authors, book publishers, libraries, movie producers, and game developers fulfill these goals.

To balance the benefits of AI-assisted production and dissemination with creators’ rights, this Article introduces the concept of “Facilitative Fair Use.” This framework is both timely and urgent; without it, AI training will be limited to licensed works, potentially perpetuating bias in the copyright ecosystem and limiting the diversity of creative expression. Ultimately, this exploration reveals the debate’s true stakes: the kind of cultural landscape we want to foster.

TABLE OF CONTENTS

INTRODUCTION	2
I. THE TRAINING DATA DEBATE	6
<i>A. Is Training an AI on Copyrighted Works Fair Use?</i>	7
<i>B. Copyright’s Dissemination Imperative</i>	13
II. CASE STUDIES	16
<i>A. Book Publishing</i>	17
1. Assistive AI for Writers	17
2. Finding an Agent	18
3. Negotiating a Contract	19

* Professor of Law and Louis F. Niezer Faculty Fellow, Indiana University Maurer School of Law. My deep thanks to Kimberly Cogswell, Camilla Hrды, Marshall Leaffer, Mark Janis, Deborah Widiss, Annemarie Bridy, Timothy R. Holrbook, Justin Hughes, Michael W. Carroll, Amanda Levendowski, Andrew Hammond, Nicholas Almendares, and participants of the 2024 Intellectual Property Scholars Conference, hosted by Berkeley Law.

4. From the Author’s Desk to the Publisher’s Eye	19
5. AI-Assisted Editing.....	20
6. Printing, Marketing, and Release.....	21
7. Conclusion	22
<i>B. Movies</i>	22
1. Screenplay Adaptation.....	22
2. Studio Pitch and Script Analysis	22
3. Casting and Storyboarding.....	23
4. Accessibility.....	24
5. Conclusion	24
<i>C. Libraries</i>	25
1. AI-Powered Content Discovery.....	25
2. AI-Facilitated Research	26
3. Generating Reading Lists.....	26
4. Unlocking Visual Collections.....	27
<i>D. Video Games</i>	28
III. TOWARDS FACILITATIVE FAIR USE.....	29
<i>A. The Need for Facilitative Fair Use</i>	30
<i>B. Introducing Facilitative Fair Use</i>	33
<i>C. Facilitating Fairly</i>	36
CONCLUSION.....	39

INTRODUCTION

Chances are that you, the reader, are a robot.¹ Today, AI systems are the world’s most voracious readers, routinely consuming millions of books, academic papers, newspaper articles, and social media posts.² Unlike human readers, however, AI systems can imitate nearly everything they take in.³ This unique behavior has ignited a set of high-stakes copyright lawsuits and fueled fierce public debate about the future of

1. For examples of AI systems trained specifically on legal texts (such as law review articles like this one), see, for example, *Claudius. Where Everything Legal Starts.*, CLAUDIUS LEGAL INTEL., <https://www.claudius.ai/> (last visited Aug. 14, 2024) (a large language model trained on legal texts to assist with legal research); see *Harness the Power of Generative AI*, THOMSON REUTERS, <https://legal.thomsonreuters.com/en/products/westlaw-precision> (last visited Aug. 14, 2024) (an AI-driven legal research tool trained on legal sources and secondary legal materials, such as law review articles).

2. See, e.g., STEPHEN WOLFRAM, WHAT IS CHATGPT DOING . . . AND WHY DOES IT WORK? 9–10 (2023) (indicating that ChatGPT, a popular large language model, has been trained on “a few million books”); *What Are Large Language Models (LLMs)?*, IBM, <https://www.ibm.com/topics/large-language-models> (last visited Sep. 5, 2024) (“To ensure accuracy, this process involves training the LLM on a massive corpora of text (in the billions of pages), allowing it to learn grammar, semantics and conceptual relationships . . .”).

3. See, e.g., Brenden M. Lake & Marco Baroni, *Human-like Systematic Generalization Through a Meta-Learning Neural Network*, 623 NATURE 115, 119 (2023), <https://doi.org/10.1038/s41586-023-06668-3>. See generally Carlos Gómez-Rodríguez & Paul Williams, *A Confederacy of Models: A Comprehensive Evaluation of LLMs on Creative Writing*, ASS’N FOR COMPUTATIONAL LINGUISTICS 14504, 14506, 14512 (2023), <https://aclanthology.org/2023.findings-emnlp.966> (summarizing the evaluation of large language models on a creative writing task involving generating stories in the style of author John Kennedy Toole).

technology and creative expression.⁴ Given AI's immense power and potential, society has a significant stake in the outcome of these conflicts.

Many legal commentators examining this issue have focused on AI's unique capacity for mimicry.⁵ This is understandable. AI-generated imitations of well-known artists and authors have captured the public's attention and contributed to the recent slate of lawsuits.⁶ The potential for AI-generated works to compete with human authorship is an important and urgent problem. However, the urgency of this problem has largely overshadowed something helpful about AI: its ability to facilitate the production and distribution of creative works on a massive scale.

This Article illuminates a paradox: technologies that seem to threaten artists and authors also hold the potential to dramatically increase the reach of creative works.⁷ The case studies presented in Part II illustrate this potential through original examples from different creative sectors.⁸ For example, book publishers use AI to study market trends, forecast reader interests, and improve how easily readers can find books.⁹ Libraries employ AI-powered tools to manage vast collections, generate summaries and translations, and expand the reach of academic works.¹⁰ AI is streamlining script analysis and casting decisions in the film industry.¹¹ AI tools can help people with certain disabilities express themselves, expanding the

4. Class Action Complaint at 1, 3, *Andersen v. Stability AI Ltd.*, No. 3:23-cv-00201 (N.D. Cal. Jan. 13, 2023) (artists suing Stability AI for copyright infringement and unfair competition); Complaint at 1, *Getty Images (US), Inc. v. Stability AI, Inc.*, No. 1:23-cv-00135 (D. Del. Feb. 3, 2023) (Getty Images accusing Stability AI for copyright infringement for using its images to train AI model); Class Action Complaint at 1–2, *Authors Guild v. OpenAI Inc.*, No. 1:23-cv-08292 (S.D.N.Y. Sept. 19, 2023) (class action lawsuit involving authors accusing OpenAI of copyright infringement); Complaint at 3, *Concord Music Grp., Inc. v. Anthropic PBC*, No. 3:23-cv-01092 (M.D. Tenn. Oct. 18, 2023) (music publishers suing AI startup for using copyrighted lyrics to train AI model); Complaint at 1–2, *The New York Times Co. v. Microsoft Corp., OpenAI Inc.*, No. 1:23-cv-11195 (S.D.N.Y. Dec. 27, 2023) (The New York Times suing AI company for copying and using millions of new articles); First Amended Class Action Complaint at 2–3, *Alter v. OpenAI Inc.*, No. 1:23-cv-10211-SHS (S.D.N.Y. Dec. 19, 2023) (plaintiffs alleging OpenAI's use of copyrighted works to train AI models constitutes copyright infringement).

5. See, e.g., Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1311–12 (2019); Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. DAVIS L. REV. 893, 900 (2019); Haochen Sun, *Redesigning Copyright Protection in the Era of Artificial Intelligence*, 107 IOWA L. REV. 1213, 1239 (2022); Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 TEX. L. REV. 743, 766–67 (2021); Guy A. Rub, *Against Copyright Customization*, 107 IOWA L. REV. 677, 697–98 (2022).

6. See Will Oremus & Elahe Izadi, *AI's Future Could Hinge on One Thorny Legal Question*, WASH. POST (Jan. 4, 2024, 7:00 AM), <https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/> (discussing a lawsuit filed by The New York Times against OpenAI and Microsoft for allegedly using copyrighted articles to train AI models, raising questions about fair use in copyright law); J. Edward Moreno, *Boom in A.I. Prompts a Test of Copyright Law*, N.Y. TIMES (Dec. 30, 2023), <https://www.nytimes.com/2023/12/30/business/media/copyright-law-ai-media.html> (exploring the impact of AI tools like ChatGPT on copyright law, highlighting lawsuits by authors and a photo agency against AI developers for using their content to train systems, and discussing the potential consequences for the generative AI industry).

7. See *infra* Part II.

8. See *infra* Part II.

9. See *infra* Section II.A.

10. See *infra* Section II.C.

11. See *infra* Section II.B.

pool of authorial voices and perspectives.¹² These examples raise a fundamental question for courts: should AI's ability to spread creative expression influence copyright disputes over training data?

This Article argues that the answer is yes. It is widely recognized that copyright serves to encourage the production of expressive works. Dissemination, meanwhile, is another fundamental goal of copyright law, reflecting Enlightenment ideals about spreading knowledge for societal progress.¹³ Copyright's balance between granting creators exclusive rights and allowing fair use, which enriches the public domain, reflects these two principles.¹⁴

While no direct precedent for AI training exists, landmark cases involving the Sony Betamax¹⁵ (allowing home recording of TV) and Google Books¹⁶ (digitizing vast libraries) show how courts sometimes interpret the fair use doctrine to enable technologies that facilitate the copyright ecosystem.¹⁷ Fair use is a copyright doctrine that permits activities that would otherwise be infringement in limited cases.¹⁸ In both of these cases, courts decided that the doctrine of fair use protected technologies capable of expanding public access to copyrighted works.¹⁹ However, copyright's traditional four-factor test does not explicitly instruct judges to consider the impact of a defendant's use on the production or dissemination of copyrighted works generally. Instead, like sunlight behind a cloud, these copyright goals inform judges' reasoning diffusely.²⁰ As a result, it is unclear whether courts will hold that the use of copyrighted works to train AI systems constitutes fair use.²¹ Moreover, courts determine fair use on a

12. See *infra* Section II.A.6–7.

13. See, e.g., Annemarie Bridy, *Fearless Girl Meets Charging Bull: Copyright and the Regulation of Intertextuality*, 9 U.C. IRVINE L. REV. 293, 315, 324 (2019) (noting the importance of knowledge dissemination as a copyright goal and “the economic-rights orientation of American copyright, which favors vigorous commerce in cultural works as a means of promoting the dissemination of knowledge”).

14. Jeanne C. Fromer, *An Information Theory of Copyright Law*, 64 EMORY L.J. 71, 118 (2014) (“Encouraging broad dissemination is thus central to copyright law.”).

15. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 420 (1984).

16. Authors Guild v. Google, Inc. (*Google Books*), 804 F.3d 202, 207 (2d Cir. 2015).

17. See *infra* Part I (discussing these landmark cases in detail). I use the term “copyright ecosystem” in this Article to refer generally to the interconnected networks of people, institutions, technologies, laws, and market interactions that control the production and dissemination of copyrighted works.

18. 17 U.S.C. § 107.

19. *Sony*, 464 U.S. at 454 (“[T]o the extent time-shifting expands public access to freely broadcast television programs, it yields societal benefits.”); *Google Books*, 804 F.3d at 207 (“Google’s making of a digital copy to provide a search function is a transformative use, which augments 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 in the original works or derivatives of them.”).

20. One place it does so is under the “public benefit” consideration, discussed later in this Article. See, e.g., Amanda Levendowski, *Fairer Public Benefit in Copyright Law*, 47 CARDOZO L. REV. (forthcoming 2025) (manuscript at 10–11) (on file with author); Patrick K. Lin, *Fair’s Fair: How Public Benefit Considerations in the Fair Use Doctrine Can Patch Bias in Artificial Intelligence Systems*, 11 IND. J.L. & SOC. EQUAL. 229, 237–40 (2023).

21. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 3–4 (2023).

case-by-case basis, which casts uncertainty on how one court decision about training data affects other AI products and services.²²

To provide industry, the public, and courts with more certainty, this Article introduces the concept of “Facilitative Fair Use.” Under this proposed framework, courts would explicitly consider whether a defendant’s AI is used to enhance the production or dissemination of copyrighted works—aligning with copyright’s foundational goals.²³ While the proposal would help technology companies, it would also help creators in two ways: first, by barring AI uses that harm the traditional or reasonable markets for creators’ works, and second, by incentivizing the creation of AI systems that expand the market for copyrighted works.

Critics will correctly point out that AI differs from earlier technologies like the VCR and Google Books in some fundamental ways. For one thing, AI’s role in the production and dissemination process is indirect and less tangible than that of VCRs or search engines.²⁴ However, production and dissemination are processes—steps that connect authors to audiences.²⁵ Technological tools have always shaped how information flows from creators to audiences. As Part II shows, AI’s ability to intervene at earlier stages is an increasingly important part of the copyright ecosystem.²⁶ A second critique is that AI’s potential for dissemination does not directly benefit the specific works used for training. Instead, the machines rob Peter to pay Paul—harming some artists to aid others. This argument overlooks that copyright seeks to benefit the public broadly. The fact that it typically achieves this goal by benefiting copyright holders (e.g., through exclusive rights that can drive licensing and sales) is not central to the policy goal. This is why courts analyzing fair use sometimes look beyond the defendant’s impact on the plaintiff and instead consider how the defendant expanded public access to knowledge.²⁷

Yet there is another powerful reason for courts to adopt this Article’s proposal: it would avert a serious problem involving prejudices in AI training. Without the flexibility that fair use provides, copyright law would restrict AI developers to using only those works in the public domain and

22. *Google Books*, 804 F.3d at 213.

23. This proposal also reinforces an approach that courts occasionally follow. *See, e.g.*, Jane C. Ginsburg, *Copyright and Control Over New Technologies of Dissemination*, 101 COLUM. L. REV. 1613, 1617 (2001) (explaining that, historically, when copyright owners seek to eliminate a new kind of dissemination, and when courts do not deem that dissemination harmful to copyright owners, many courts have hedged on the side of declining to find infringement).

24. *See infra* Part II.

25. For an in-depth guide to the book publishing and distribution process, *see, for example*, JOHN B. THOMPSON, *MERCHANTS OF CULTURE: THE PUBLISHING BUSINESS IN THE TWENTY-FIRST CENTURY* 22, 25 (1st ed. 2010). For the film and television distribution process, *see, for example*, SCOTT KIRKPATRICK, *INTRODUCTION TO MEDIA DISTRIBUTION: FILM, TELEVISION, AND NEW MEDIA*, at xiv (1st ed. 2019). For an excellent source on the process of video game production and distribution, *see* STEVEN CONWAY & JENNIFER DEWINTER, *VIDEO GAME POLICY: PRODUCTION, DISTRIBUTION, AND CONSUMPTION* (2016).

26. *See infra* Part II.

27. *See supra* note 19 and accompanying text (discussing how public access to knowledge influenced two landmark fair use decisions).

whatever works they can license, severely limiting the diversity and breadth of training data.²⁸ This could warp the creative field by reinforcing existing prejudices and inequalities. A more expansive fair use regime would help AI developers build diverse datasets to consciously mitigate against historical biases. The robots are already here; the best we can do for ourselves is teach them well.

This Article proceeds in three parts. Part I delves into the current AI copyright debate, analyzing expert commentary on fair use and the current emphasis on AI's potential to mimic authors and artists. It then highlights an overlooked but fundamental goal of copyright law: dissemination. Part II examines AI's expanding role in content distribution, presenting case studies across publishing, film, libraries, and video games. Building on this analysis, Part III proposes a recalibration of fair use law—Facilitative Fair Use—to directly address AI's unique capabilities for knowledge dissemination. This proposal is essential in light of AI's potential to exacerbate biases and inequities in the creative landscape.

I. THE TRAINING DATA DEBATE

The debate over AI and copyright involves two opposing viewpoints. Authors argue that using their works without their permission to train AI systems exploits their creative efforts.²⁹ AI developers, on the other hand, contend that training an AI on copyrighted works is permissible under the copyright doctrine of fair use, a defense that permits the use of copyrighted works without an author's consent.³⁰ Section A of this discussion explains this debate and the relevant legal background. Section B identifies a gap in the debate that deserves deeper study: the potential impact of AI to facilitate the copyright ecosystem.

28. See *infra* Section III.C.

29. See training data lawsuits cited *supra* note 4.

30. These arguments have been expressed in corporate responses to requests for public comments on AI and Copyright. See Artificial Intelligence and Copyright Notice, 88 Fed. Reg. 59942 (Aug. 30, 2023); Adobe Inc., Comment Letter on the U.S. Copyright Office Notice of Inquiry and Request for Comments on Artificial Intelligence and Copyright (Oct. 30, 2023) [hereinafter Adobe Comments], <https://www.documentcloud.org/documents/24117931-adobe> (arguing that using copyrighted works as part of a dataset to train AI models should constitute fair use under current law because this intermediate copying serves the transformative purpose of distilling factual information and style concepts from the training data in order to fuel a new creative platform rather than directly copying protected aesthetic expression or artistic messages for the same purpose as the original works); Stability AI, Comment Letter on the U.S. Copyright Office Inquiry into Artificial Intelligence and Copyright (Oct. 2023) [hereinafter Stability Comments], <https://s3.documentcloud.org/documents/24117937/stabilityai-stable-diffusion-etc.pdf> (arguing that using copyrighted works as part of a dataset to train AI models constitutes fair use under current law and supports innovation because this intermediate copying serves the transformative purpose of enabling the AI models to learn unprotected ideas, facts, and structures within the training data rather than directly copying or collaging together protected aesthetic expression or full passages for the same purpose as the original works); see also Ben Brooks, Head of Public Policy, Stability AI, Statement to the U.S. Senate AI Insight Forum on Transparency, Explainability, and Copyright (Nov. 29, 2023), <https://stability.ai/news/copyright-us-senate-open-ai-transparency> (arguing training AI on copyrighted works is, or should be, protected as fair use); 17 U.S.C. § 107.

A. Is Training an AI on Copyrighted Works Fair Use?

Imagine this: after years of writing mystery novels known for witty dialogue and surprising twists, you discover your distinct style has helped train an AI. This AI now churns out stories imitating your unique style, flooding the market with cheaper imitations. Should society accept this as the cost of progress, or should the law protect authors from machines that replicate their creative work? This dilemma encapsulates authors' perceptions of the conflict between technology and art.³¹ Technology companies, conversely, argue that their AI systems analyze books not to compete with them but instead to learn from them.³² A challenging question lies at the heart of this debate: how should the law differentiate between imitation, inspiration, and misappropriation?

To navigate this debate, it is essential to understand some key terms. AI broadly refers to systems that simulate human intelligence.³³ Machine learning (ML) allows computers to learn from data without explicit programming.³⁴ For example, an ML system can learn to identify cats by analyzing large sets of cat photos without human guidance. ML systems trained on vast amounts of text are known as large language models (LLMs).³⁵ ML systems (including LLMs) are powerful because they can analyze various types of media—such as photos, books, videos, and code—and generate new content in those formats in ways that mimic human creation, a capability termed “generative AI.”³⁶ For simplicity, this Article will use “AI” to refer to all of these systems, occasionally specifying “LLMs” when discussing text-based AI systems.

It is important to note that most AI systems do not internally store the texts that they train on.³⁷ Instead of memorizing complete passages, AI neural networks reconfigure themselves to internalize the underlying linguistic patterns of the texts.³⁸ This process is more akin to organic learning than traditional data processing. For instance, a beautifully crafted phrase from a mystery novel might enhance an AI's ability to generate natural-sounding prose, even if the AI does not retain or reproduce that exact

31. See training data lawsuits cited *supra* note 4.

32. See OpenAI, *OpenAI and Journalism*, OPENAI (Jan. 8, 2024), <https://openai.com/blog/openai-and-journalism> (“Training is fair use . . .”); see also Adobe Comments, *supra* note 30; Stability Comments, *supra* note 30.

33. See *Artificial Intelligence Definitions*, STAN. UNIV. HUM.-CENTERED A.I. (Sept. 2020), <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>.

34. *Id.*

35. IBM, *supra* note 2.

36. *Id.*

37. WOLFRAM, *supra* note 2, at 56 (“After all, it’s certainly not that somehow ‘inside ChatGPT’ all that text from the web and books and so on is ‘directly stored.’ Because what’s actually inside ChatGPT are a bunch of numbers—with a bit less than 10 digits of precision—that are some kind of distributed encoding of the aggregate structure of all that text.”). This is why many LLMs “hallucinate” false information—their responses are based on probability, not traditional file storage. For a technical explanation, see DANIEL JURAFSKY & JAMES H. MARTIN, *SPEECH AND LANGUAGE PROCESSING* ch. 14, at 2, 14 (2024) (e-book), <https://web.stanford.edu/~jurafsky/slp3/14.pdf> (discussing what happens to training data after the training process).

38. WOLFRAM, *supra* note 2, at 56.

wording.³⁹ This is why feeding dictionary definitions to AI is ineffective; the systems thrive on the complexity of genuine human expression. In this way, LLM training parallels the way human writers learn to write. As Stephen King famously advised, “If you don’t have time to read, you don’t have the time (or the tools) to write.”⁴⁰

Author groups, however, see the situation more simply. They believe that training AIs on their works is copyright infringement.⁴¹ There are arguably three places where training an AI system like an LLM creates an infringing copy or derivative work: at the point of ingestion, during the process of training, and at the output. Ingestion refers to the step of providing training data—in this case, a copyrighted work—to the AI being trained.⁴² To do this, it is necessary to prepare a digital copy in a format the system can interpret. The second potentially infringing copy occurs at a more abstract level: the internal representation of the work imprinted on the neural network. Whether this representation can properly be called a “copy” for purposes of copyright is a hotly debated issue at the time of this writing.⁴³ The final potential infringement arises during the output-generation stage. Both LLMs and image generators can sometimes be coaxed (through user prompting) to produce outputs that are either infringing derivative works or partially verbatim copies of training data.⁴⁴ This phenomenon has also ignited considerable public debate.⁴⁵

There is no debate that at the point of ingestion, copyright-protected works are being used to train AI systems without permission. For example, an AI training dataset called “The Pile” is comprised of millions of unlicensed copyrighted texts scraped from the internet and includes a subset called “Books3” that contains around 190,000 published books.⁴⁶ Forensic analyses have concluded that AI systems like Meta’s Llama LLM and OpenAI’s ChatGPT are very likely trained on these sources.⁴⁷

39. *Id.*

40. STEPHEN KING, ON WRITING: A MEMOIR OF THE CRAFT 147 (2000).

41. Training data lawsuits cited *supra* note 4.

42. Sophie Jin, *The Importance of Data Ingestion and Integration for Enterprise AI*, IBM (Jan. 9, 2024), <https://www.ibm.com/think/topics/ai-for-data-integration>.

43. CHRISTOPHER T. ZIRPOLI, CONG. RSCH. SERV., LSB10922, GENERATIVE ARTIFICIAL INTELLIGENCE AND COPYRIGHT LAW 4–5 (2023).

44. Matthew Sag, *Copyright Safety for Generative AI*, 61 HOUS. L. REV. 295, 310–313 (2023).

45. Alex Reisner, *The Flaw That Could Ruin Generative AI*, THE ATL.: TECH. (Jan. 11, 2024), <https://www.theatlantic.com/technology/archive/2024/01/chatgpt-memorization-lawsuit/677099/>.

46. “Scraping involves extracting data from a website and copying it into a structured format, allowing for data manipulation or analysis.” *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180, 1186 (9th Cir. 2022); *see also* Alex Reisner, *These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech*, THE ATLANTIC (Sept. 25, 2023), <https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/> (“This summer, I acquired a data set of more than 191,000 books that were used without permission to train generative-AI systems by Meta, Bloomberg, and others.”).

47. Researchers examining this training data have found evidence that tools like ChatGPT are trained to alter their output to conceal the use of copyrighted material—a concerning practice.

In response, AI proponents invoke the hallowed doctrine of fair use.⁴⁸ This defense, which exempts activity that would otherwise constitute copyright infringement, hinges on four factors laid out in the 1976 Copyright Act: (1) the purpose and character of the use; (2) the nature of the underlying copyrighted work; (3) how much of the underlying work the defendant used; and (4) whether the defendant’s use threatens the economic potential of the original.⁴⁹ No single factor is determinative; courts weigh them together in a nuanced, case-specific fashion.⁵⁰ However, the first factor has often had the greatest influence on outcomes.⁵¹ Additionally, courts have sometimes favored fair use defendants who have shown their use provided a public benefit.

AI supporters see a compelling precedent in *Google Books*,⁵² a landmark 2015 copyright case in which the Second Circuit ruled that Google’s massive digitization of books for a searchable index was fair use.⁵³ The court found Google’s index transformative because the service augmented books through search capabilities rather than replacing them.⁵⁴ The court also emphasized the factual nature of many of the digitized works and Google’s display of short “snippets” instead of entire books—a purpose the court said was “transformative.”⁵⁵ This case is often cited as an

48. 17 U.S.C. § 107; see OpenAI, *supra* note 32; see also Adobe Comments, *supra* note 30, at 2–3; Stability Comments, *supra* note 30, at 2.

49. Section 107 states:

Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

- (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.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

17 U.S.C. § 107.

50. See Matthew Sag, *Predicting Fair Use*, 73 OHIO ST. L.J. 47, 52–53 (2012) (analyzing over 280 federal fair use cases decided between 1978–2011 and reporting on how much each factor related to the outcome); Gideon Parchomovsky & Philip J. Weiser, *Beyond Fair Use*, 96 CORNELL L. REV. 91, 100 (2010) (“[T]he fair use doctrine has decidedly remained an ‘equitable rule of reason,’ with all of its attendant uncertainty.”).

51. See, e.g., Niva Elkin-Koren & Orit Fischman-Afori, *Rulifying Fair Use*, 59 ARIZ. L. REV. 161, 182 (2017) (“The first fair-use factor is ‘the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes.’ This factor is often considered the heart of the four-factor inquiry.”).

52. 804 F.3d 202 (2d Cir. 2015).

53. *Id.* at 229.

54. *Id.*

55. *Id.*

example of a so-called technological use of a work because the copying is for functional, rather than expressive, ends.⁵⁶

At first glance, the parallels to AI seem strong. Like Google Books, AI models often need to access large amounts of copyrightable subject matter to be useful.⁵⁷ AI tools do not aim to replace original works, also like Google Books. The fact that AI systems primarily generate new insights rather than simply redistributing copyrighted works further strengthens a fair use defense.

The 1984 Supreme Court case *Sony Corp. of America v. Universal City Studios, Inc.*⁵⁸ also seems to superficially support AI fair use.⁵⁹ In that case, the Court held Sony had no copyright liability for selling Betamax video recorders, even though these machines could be used to duplicate copyrighted broadcasts.⁶⁰ The Court was convinced that “substantial noninfringing uses” of home video recording outweighed the risks to film and television studios.⁶¹ Courts today might similarly believe that AI training opens the door to countless noninfringing uses, such as translating books, making works more accessible to people with disabilities, or assisting with drafting emails and presentations.

Despite these potential similarities, *Sony*, *Google Books*, and similar precedents offer limited clarity on AI training disputes. Consider the first factor of the fair use test, which looks at the purpose and character of the defendant’s use. Under this factor, transformative uses are more likely to qualify as fair use, while uses that reproduce original works are not. AI systems typically transform original works because they use them to learn how to communicate. However, recent research has shown that LLMs can sometimes inadvertently reproduce portions of copyrighted works that they were trained on.⁶² This capability weighs against fair use. Moreover,

56. David Newhoff, *Google v. Oracle IX: A Poor Fair Use Defense Is an Opportunity for Guidance*, THE ILLUSION OF MORE (Sept. 21, 2020), <https://illusionofmore.com/google-v-oracle-ix-a-poor-fair-use-defense-is-an-opportunity-for-guidance/>.

57. OpenAI has said in comments before the House of Lords that AI development would be “impossible” without access to copyrighted works. OPENAI, HOUSE OF LORDS COMMUNICATIONS AND DIGITAL SELECT COMMITTEE INQUIRY: LARGE LANGUAGE MODELS 4 (2023), <https://committees.parliament.uk/writtenevidence/126981/pdf/> (“Because copyright today covers virtually every sort of human expression—including blog posts, photographs, forum posts, scraps of software code, and government documents—it would be impossible to train today’s leading AI models without using copyrighted materials. Limiting training data to public domain books and drawings created more than a century ago might yield an interesting experiment, but would not provide AI systems that meet the needs of today’s citizens.”).

58. 464 U.S. 417 (1984).

59. *Id.* at 454–55.

60. *Id.* at 454–56 (“When these factors are all weighed in the ‘equitable rule of reason’ balance, we must conclude that this record amply supports the District Court’s conclusion that home time-shifting is fair use.”).

61. *Id.* at 442.

62. Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Úlfar Erlingsson, Alina Oprea, & Colin Raffel, *Extracting Training Data from Large Language Models*, ARXIV, June 2021, at 1, 5, <https://arxiv.org/pdf/2012.07805> (explaining and reporting results from experiments on this form of AI “memorization”).

if an author argues that they created a work in part for the purpose of training AI systems, then this would suggest that the purpose of the unlicensed uses of that work is the same as the purpose of the original work.⁶³

The second factor, which asks whether the defendant has used factual work or creative expression, is also murky. Courts could easily view ingested novels and artworks as quintessential protected expression.⁶⁴ But AI extracts factual linguistic patterns rather than the creative essence of what it takes in, which arguably weighs in favor of fair use.⁶⁵ Consuming entire works, as AI systems do, runs afoul of the third factor, which examines the extent of use. However, courts may overlook extensive copying if they believe that AI's purpose is likely transformative, as the court did in *Google Books*.

On the fourth factor, tech companies contend that AI-generated works do not displace author sales.⁶⁶ Typically, courts consider whether a defendant's use will usurp "traditional, reasonable, or likely to be developed" markets.⁶⁷ Again, it is difficult to know how a court will analyze this factor. On the one hand, training an AI system is not a traditional market for copyrighted works. However, there has been a recent flurry of licensing agreements between copyright holders and AI companies.⁶⁸ These agreements could contribute to a "likely to be developed" market in the view of a court.⁶⁹ However, what if subtly influential training data empowers AI to compete with human creativity in the future?⁷⁰ Courts cannot guarantee against this market risk. In summary, it is anyone's guess how courts will apply the fair use doctrine to AI.⁷¹

63. Given the recent flurry of licenses between AI companies and copyright holders, such an argument might be plausible.

64. *Sag*, *supra* note 50, at 61 ("Over the past thirty years, courts have typically distilled the amorphous 'nature' of the work into two more tractable considerations: whether the plaintiff's work is creative as opposed to informational; and whether the work is unpublished, as opposed to published.").

65. WOLFRAM, *supra* note 2, at 56 (explaining that all that is stored are statistical relationships).

66. See Adobe Comments, *supra* note 30, at 8; Stability Comments, *supra* note 30, at 5–6, 13.

67. *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614 (2d Cir. 2006) (quoting *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930 (2d Cir. 1994)).

68. See, e.g., Aruni Soni, *AI Models Force Media Firms to Pick Licensing or Litigation*, BLOOMBERG L. (Aug. 5, 2024, 3:11 AM), <https://news.bloomberglaw.com/ip-law/generative-ai-forces-media-firms-to-pick-licensing-or-litigation> (enumerating and interviewing the author about AI training data licenses).

69. It is important to note that such a "market" might be developing primarily as a risk-mitigation strategy in the shadow of looming litigation, rather than as an organic response to demand for property rights. In a forthcoming paper, I argue that courts should consider these markets differently when possible.

70. Moreover, a majority of authors may believe that the use of their works for training LLMs is properly part of the "market" for their books. *Survey Reveals 90 Percent of Writers Believe Authors Should Be Compensated for the Use of Their Books in Training Generative AI*, THE AUTHORS GUILD (May 15, 2023), <https://authorsguild.org/news/ai-survey-90-percent-of-writers-believe-authors-should-be-compensated-for-ai-training-use/> (reporting that 90% of writers believe they should be compensated for the use of their works in training generative AI, and 65% support a collective licensing system for this purpose).

71. Mark Lemley aptly sums up the state of the AI fair use debate: "We aren't saying courts, or the court of public opinion, will definitely reject the fair use defense as it is currently understood, only that there is a risk they will do so." Lemley & Casey, *supra* note 5, at 769.

This uncertainty has spurred a growing cohort of scholars to search for ways that AI can thrive without disrupting copyright incentives. For example, Mark Lemley and Bryan Casey present a compelling argument that transformative fair use principles should protect most AI data uses.⁷² The authors contend that ML systems are disinterested in the copyrightable expression within works, and instead seek to extract unprotected ideas and facts.⁷³ Similarly, Amanda Levendowski argues fair use facilitates “fairer AI” because copyright enforcement could entrench AI biases by limiting datasets.⁷⁴ In a more recent example, Matthew Sag argues that training generative AI on copyrighted works likely falls under fair use, but he also points out edge cases where fair use is less likely to apply (such as when models memorize and produce images of copyrightable characters).⁷⁵ Sag helpfully proposes a set of best practices to mitigate the risk of infringement.⁷⁶

Pamela Samuelson’s work serves as a foundation for all of these discussions.⁷⁷ Samuelson has observed that fair use cases tend to cluster around policy-relevant themes such as learning, access to information, and critical commentary.⁷⁸ Focusing on the learning theme, she explains how courts have balanced the need to promote access to knowledge with the rights of copyright holders.⁷⁹ This insight is relevant because AI is making new forms of learning possible beyond the traditional classroom setting. For instance, AI can promote learning by conveying knowledge to users in ways that are highly personalized and tailored to a user’s learning

72. *Id.* at 770–72.

73. *Id.* at 772 (“ML systems generally copy works, not to get access to their creative expression (the part of the work the law protects), but to get access to the uncopyrightable parts of the work—the ideas, facts, and linguistic structure of the works. A self-driving car, for instance, doesn’t care about the composition or lighting of your photograph, or indeed about what you were likely actually intending to depict in your photo. It cares about the fact that there’s a stop sign in it.”).

74. Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem*, 93 WASH. L. REV. 579, 586–87, 590 (2018).

75. Matthew Sag, *Fairness and Fair Use in Generative AI*, 92 FORDHAM L. REV. 1887, 1914 (2024) (“Generative AI models that do not, in their ordinary and routine operation, copy (or produce copies of) the original expression in their training data are an example of non-expressive use.”). Sag refers to the phenomenon in which AI systems learn the “idea” of a character and generate new images of that character when prompted with a mere description of the character’s traits as the “Snoopy Problem.” *Id.* at 1911–12.

76. Sag argues that users of copyrighted works for training purposes should not bypass an existing market for access to those works. He also suggests that technology companies would do well to provide copyright owners with opportunities to opt out of having their works used. Additional best practices included: filtering prompts and outputs of models, taking steps to avoid memorization, avoiding inducement of downstream users, implementing security measures, and working to ensure that the outputs of a model do not substitute in the markets that training sources operate in. *Id.* at 1917–18, 1920–21.

77. Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537, 2546–47 (2009) [hereinafter Samuelson, *Unbundling*]; Pamela Samuelson, *Fair Use Defenses in Disruptive Technology Cases*, UCLA L. REV. (forthcoming 2024) (manuscript at 3–4) [hereinafter Samuelson, *Fair Use Defenses*], <https://ssrn.com/abstract=4631726>.

78. Samuelson, *Unbundling*, *supra* note 77, at 2537, 2546.

79. *Id.* at 2546–47.

style.⁸⁰ Also, learning is deeply intertwined with dissemination. AI can, as the case studies in this Article show, promote learning by supercharging the traditional machinery of copyright distribution, making knowledge accessible to diverse and potentially underserved communities.

Samuelson's *Fair Use Defenses in Disruptive Technology Cases* also provides important insights. Samuelson examines how courts have grappled with contemporary uses of copyrighted works enabled by advances like photocopying, home taping, and internet search engines.⁸¹ She emphasizes the importance of the market harm analysis in these cases and cautions against presumptions of harm that could stifle innovation.⁸² These insights provide a crucial lens for understanding how courts might think about the fourth factor of the fair use test in generative AI disputes.

The training data lawsuits place courts in a quandary: how do we balance authors' rights without chilling innovation? Fair use, the central legal battleground, is fraught with uncertainty. Although past cases like *Sony* and *Google Books* offer some guidance, their relevance to the current disputes is unclear. This legal uncertainty leaves both creators and tech companies in limbo. However, commentators have underplayed a critical issue in this debate: AI's ability to facilitate the production and dissemination of copyrighted works. The following Section is a short primer that provides historical perspective on this issue.

B. Copyright's Dissemination Imperative

The notion that copyright law aims to spread knowledge—and not solely to protect creators—may surprise some casual observers of today's debates over AI and fair use. However, this principle has a rich historical lineage and has shaped our copyright system from its earliest days. To understand its enduring power and its relevance to the training data debate, let us embark on a brief (but sweeping) journey through history.

Our journey starts in a modest workshop in fifteenth-century Germany. Johannes Gutenberg, a man of faith and ingenuity, toils on his

80. This Article is squarely focused on dissemination, but a focus on how interaction with AI chatbots can facilitate learning is a critical next step for scholarship to examine. AI systems, particularly those leveraging machine learning and adaptive algorithms, analyze vast amounts of data to understand a user's learning preferences, strengths, and weaknesses. Ilie Gligorea, Marius Cioca, Romana Oancea, Andra-Teodora Gorski, Hortensia Gorski, & Paul Tudorache, *Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review*, EDUC. SCIS., Dec. 2023, at 3–4. These systems can dynamically adjust content delivery, pacing, and difficulty levels to optimize the learning process. For instance, AI-driven tutoring systems can provide real-time feedback and customized exercises based on a student's performance, while personalized curriculum platforms can recommend learning paths that align with a user's goals and interests. *Id.* at 3–4, 21.

81. Samuelson, *Fair Use Defenses*, *supra* note 77, at 10.

82. *Id.* at 82 (“Before *Sony*, courts and commentators often assumed harm from the existence of unauthorized copies and one school of thought posited that harm was irrelevant to copyright liability and entitlement to relief. *Sony* represented a sea change in judicial understanding of the market effects factor in fair use cases, later refined in *Campbell* and *Google v. Oracle*. Since *Sony*, courts have consistently looked for evidence of actual or likely harm before ruling against fair use defenses.” (footnote omitted)).

mechanical printing press, dreaming of making the scriptures accessible to all.⁸³ Little did he know that his invention would spark change far beyond the religious realm.

Gutenberg's printing press sparked an explosion of the printed word.⁸⁴ Books, once reserved for the wealthy elite, grew cheaper and more abundant. Library holdings swelled.⁸⁵ Literacy rates rose.⁸⁶ These developments laid the foundation for the Age of Enlightenment, an intellectual revolution fueled by the sharing of ideas.⁸⁷

Enlightenment thinkers like John Locke understood that progress depended on the free flow of information; a society where elites hoarded knowledge was a society destined to stagnate.⁸⁸ These principles found their way across the Atlantic and into the minds of America's founders. The architects of the new American nation—most notably James Madison—recognized copyright as a means to spread knowledge.⁸⁹ Reflecting this belief, the first Copyright Act was titled “An Act for the Encouragement of Learning.”⁹⁰ Although the Act embraced a narrow set of subject matter—just maps, charts, and books—it covered all books, regardless of content or genre. In doing so, the Act embraced a broad definition of knowledge, encompassing not only informational or instructive books that could help citizens to participate in civic life but also fictional works. This was insightful. Fiction provides not only pleasure, but also fosters emotional intelligence, empathy, and critical thinking—all of which are crucial tools for a healthy and thriving citizenry. Over time, copyright has embraced this broad view of “knowledge.” Today, the Copyright Act includes a wide array of expressive works, including films, music, software, choreography, and even architecture.⁹¹

It is easy to see how copyright incentivizes authors to create new works—i.e., by providing limited exclusive rights to authors—but how does copyright ensure these works are disseminated? Here is where

83. See generally RICHARD ABEL, *THE GUTENBERG REVOLUTION: A HISTORY OF PRINT CULTURE* 23–24, 26, 29–30, 40–42, 75–77, 79 (2011); SAMUEL WILLARD CROMPTON, *THE PRINTING PRESS: TRANSFORMING POWER OF TECHNOLOGY* 20–21 (2004).

84. CROMPTON, *supra* note 83, at 64 (discussing the printing press's instrumental influence on the Enlightenment).

85. See Elizabeth L. Eisenstein, *In the Wake of the Printing Press*, 35 Q.J. LIBR. CONG. 183, 187 (1978) (noting that the press benefited libraries).

86. See Jeremiah E. Dittmar, *Information Technology and Economic Change: The Impact of the Printing Press*, 126 Q.J. ECON. 1133, 1139 (2011) (“The availability of inexpensive texts was a key prerequisite for the spread of literacy . . .”).

87. See WOLFGANG LUTZ & REINER KLINGHOLZ, *EDUCATION FIRST!: FROM MARTIN LUTHER TO SUSTAINABLE DEVELOPMENT* 33–58 (2017) (explaining how the printing press laid the foundation for the Enlightenment).

88. See PETER B. KAUFMAN, *THE NEW ENLIGHTENMENT AND THE FIGHT TO FREE KNOWLEDGE* 17–18, 46–47, 81–83 (2021).

89. In this spirit, James Madison eloquently wrote, “Knowledge will forever govern ignorance . . .” Letter from James Madison to W.T. Barry (Aug. 4, 1822), in 9 *THE WRITINGS OF JAMES MADISON* 71 (Gaillard Hunt ed., 1910).

90. Copyright Act of 1790, ch. 15, 1 Stat. 124 (1790) (current version at 17 U.S.C. § 101).

91. 17 U.S.C. § 102(a) (listing copyrightable subject matter).

copyright gets clever: within the protected space it grants to creators, the law also carves out pathways for sharing knowledge. These deliberate exceptions allow for expanded access without a copyright holder’s permission. Provisions for libraries and schools⁹² allow people in educational settings to share knowledge more freely than they otherwise would. Compulsory licensing schemes, particularly in music, open new creative possibilities by mandating access to existing work.⁹³ Accommodations for people with visual impairments, blindness, or reading disabilities⁹⁴ promote greater accessibility of knowledge. These are just a few examples of how copyright law actively promotes the spread of ideas despite granting exclusive rights to creators.⁹⁵

Fair use, discussed in Section I.A, is an especially important way that copyright promotes the dissemination of knowledge. While dissemination is not explicitly listed in the four-factor fair use test, the *Sony* and *Google Books* decisions demonstrate how courts have factored it into the equation. In *Google Books*, the Second Circuit stated that copyright’s ultimate goal is “to expand public knowledge and understanding,” and fair use supports that purpose.⁹⁶ The court deemed Google’s use of book snippets to be means of “augment[ing] public knowledge.”⁹⁷ Similarly, the *Sony* Court valued increased public access, finding that “time-shifting” broadened access to free television broadcasts.⁹⁸ Both courts strongly linked the issue of public access to the fourth factor—the market factor.⁹⁹ This makes sense: the market for a copyright holder’s work might sometimes be affected by technologies that expand public access to copyrighted works in general. But both courts explained how copyright law navigates this apparent tension. In *Google Books*, the Second Circuit viewed Google’s use of snippets from books as an example of expanding public access to knowledge without limiting the copyright owner’s likely or reasonable market.¹⁰⁰ Likewise, the Supreme Court in *Sony* saw VCR time-shifting as

92. 17 U.S.C. § 108.

93. 17 U.S.C. § 115(a)(1)(A) (providing a compulsory licensing scheme for musical recordings).

94. 17 U.S.C. § 121.

95. For more examples and a deep analysis, see Jacob Noti-Victor, *Copyright’s Law of Dissemination*, 44 CARDOZO L. REV. 1769, 1771–72 (2023).

96. *Google Books*, 804 F.3d 202, 212 (2d Cir. 2015).

97. *Id.* at 207.

98. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 454 (1984) (“[T]o the extent time-shifting expands public access to freely broadcast television programs, it yields societal benefits.”). Time-shifting refers to using a VCR to record a television broadcast for later viewing. *Id.* at 423.

99. *Id.* at 454; *Google Books*, 804 F.3d at 212–13, 223, 225.

100. *Google Books*, 804 F.3d at 224–25 (“Even if the snippet reveals some authorial expression, because of the brevity of a single snippet and the cumbersome, disjointed, and incomplete nature of the aggregation of snippets made available through snippet view, we think it would be a rare case in which the searcher’s interest in the protected aspect of the author’s work would be satisfied by what is available from snippet view, and rarer still—because of the cumbersome, disjointed, and incomplete nature of the aggregation of snippets made available through snippet view—that snippet view could provide a significant substitute for the purchase of the author’s book.”); *Sony*, 464 U.S. at 454 (“The District Court’s conclusions are buttressed by the fact that to the extent time-shifting expands public access to freely broadcast television programs, it yields societal benefits.”).

a technology that substantially expanded public knowledge without limiting copyright incentives.¹⁰¹ If there is one lesson in these precedents, it is that when technologies expand access to knowledge, the fourth factor can favor fair use, potentially outweighing commercial concerns.

This highlights a significant gap in the training data debate. So far, the debate has focused largely on whether training AI on copyrighted works constitutes a transformative fair use under the first fair use factor. However, AI's potential as a tool that can facilitate the production and dissemination of copyrighted works has received scant attention. Consider scenarios where AI assists with tasks like contract drafting, content tagging, and even preliminary vetting of book submissions. These tasks are vital links in the knowledge distribution chain. With its potential to summarize, curate, and make complex data searchable, AI could dramatically enhance access to knowledge embedded within protected material. The next Part examines this possibility in greater detail.

II. CASE STUDIES

This Part explores the evolving role of AI within the copyright ecosystem by examining its impact on book publishing, filmmaking, libraries, and video games. Current documented practices, potential uses of AI, and ongoing discussions within industry publications inform the hypothetical case studies presented here. These case studies demonstrate the pervasive role AI tools already play in indirectly disseminating copyrighted works, revealing the tension between existing law and this new reality.

Technology discussions can easily veer into the abstract. To keep these case studies concrete and relatable, this Article uses hypotheticals.¹⁰² Section A and B will shadow the journey of a book from initial idea to final published work to film adaptation, frequently stopping along the way to learn about the nuanced realities of AI in these creative sectors.¹⁰³ Section C will pop into a library using AI to better connect patrons with library holdings.¹⁰⁴ Section D will peek behind the curtain at a software company using AI to expand a video game's distribution.¹⁰⁵ This narrative approach has an added benefit in this context: it paints a clear picture of the

101. *Sony*, 464 U.S. at 443, 446.

102. For similar examples of fictional, or hypothetical case studies, see, for example, Peter K. Yu, *From Pirates to Partners (Episode II): Protecting Intellectual Property in Post-WTO China*, 55 AM. U. L. REV. 901, 906 (2006) (using hypothetical case studies to illustrate methods of protecting intellectual property without legal enforcement); Robert Weber, *A Theory for Deliberation-Oriented Stress Testing Regulation*, 98 MINN. L. REV. 2236, 2276 (2014) (presenting four hypothetical case studies to illustrate how financial "stress tests" work); Johnny Rex Buckles, *Curbing (or Not) Foreign Influence on U.S. Politics and Policies Through the Federal Taxation of Charities*, 79 MD. L. REV. 590, 590 (2020) (presenting hypothetical case studies examining foreign influence on elections vis-à-vis charitable donations); Robert D. Stone, *The Cloudy Crystal Ball: Genetics, Child Abuse, and the Perils of Predicting Behavior*, 56 VAND. L. REV. 1557, 1560 (2003) (presenting a "hypothetical child abuse investigation as a vehicle to identify legal and ethical issues for physicians and state child welfare departments").

103. See *infra* Sections II.A–B.

104. See *infra* Section II.C.

105. See *infra* Section II.D.

dissemination process—a process that, up to this point, this Article has discussed only generally.

This analysis draws from various sources, including industry journals, trade press, and news outlets. This method provides an accessible and useful snapshot of current industry products and practices involving AI rather than an exhaustive technical deep dive. The case studies focus on AI technologies that are trained on copyrighted works. The very services at issue in the ongoing copyright lawsuits described in Part I are some examples.¹⁰⁶ Other examples that are not the subject of litigation are still relevant to the underlying question of how AI systems trained on copyrightable material steer distribution.

A. Book Publishing

AI is redefining the book industry. From the first flicker of an author’s imagination to the final pages turned by a reader, algorithms influence the journey books undertake. To learn more, this Section follows the hypothetical story of a writer with dyslexia embarking on the publication of her debut novel.

1. Assistive AI for Writers

Amelia Wright, a young fantasy writer, embodies a paradox that is common in creative fields. She has dyslexia, which is often a barrier to clear written communication, yet she is also a gifted storyteller who crafts fantastical worlds. Recent research suggests that people with dyslexia process information differently.¹⁰⁷ They often excel at pattern recognition and out-of-the-box thinking—ideal traits for a fantasy writer.¹⁰⁸ This is where assistive AI writing tools come in. One of the most popular and widely known is Grammarly, an online service with an AI writing assistant.¹⁰⁹ Grammarly’s AI is trained on a massive text dataset that includes news articles, academic papers, and copyrighted works like books and creative

106. See training data lawsuits cited *supra* note 4.

107. Jillian Petrova, *The Many Strengths of Dyslexics*, DYSLEXIA HELP: SUCCESS STARTS HERE, <https://dyslexiahelp.umich.edu/dyslexics/learn-about-dyslexia/what-is-dyslexia/the-many-strengths-of-dyslexics> (last visited Oct. 14, 2024) (emphasizing the strengths of individuals with dyslexia, including a strong memory for stories, excellent puzzle-solving skills, brilliant spatial reasoning, great conversational skills, tremendous empathy, imaginative and abstract thinking, the ability to think outside the box, critical thinking, and the ability to analyze stories told or read to them).

108. *Id.*

109. Grammarly keeps its specific sources of training data confidential, but a chat with the AI was still informative. When asked for more detail, Grammarly’s AI responded: “I was trained on a diverse range of sources, including books, articles, websites, and other textual resources. The training data was collected from various sources, and it includes a wide range of topics, styles, and formats. This diversity helps me to better understand and generate natural language responses to user inputs.” When asked if these sources included Books3, Grammarly’s AI replied, “As an AI-powered assistant, I am not aware of the specific datasets that were used to train me. However, it is possible that the dataset known as Books3, or the Book Corpus, was included in the diverse range of internet data that I was trained on.” Grammarly AI, in conversation with the author (Sept. 18, 2024).

writing.¹¹⁰ This training data allows Grammarly to offer writers nuanced suggestions, proposing clearer phrasing and more appropriate vocabulary choices while respecting an author’s intended style. A journalist and long-time Grammarly user recently reported that the service has helped her “cheat” dyslexia.¹¹¹

In addition to Grammarly, Amelia might use an LLM like ChatGPT or Sudowrite to receive early feedback on her story’s structure.¹¹² Both tools use the GPT-3.5 and GPT-4 language models (subject to ongoing litigation), which are trained on a large corpus of successful works.¹¹³ This training data allows them to scrutinize story elements like pacing, readability, and narrative flow.¹¹⁴ Sudowrite can pinpoint overused words, awkward sentence structures, and sections where a novel’s pace could be improved.¹¹⁵

In addition to the hypothetical Amelia, actual writers with dyslexia have reported that these tools help them in many ways.¹¹⁶ “Because I’m dyslexic, it takes me a really long time to get an article down on paper,” explained a journalist in a recent survey. “So the hack I’ve come up with is, I’ll dictate my entire article. Then I’ll have ChatGPT basically correct my spelling and grammar.”¹¹⁷

These examples show that AI is helping more authorial voices to be heard. By allowing diverse minds—minds society has historically marginalized—to contribute to the copyright ecosystem, AI expands the boundaries of knowledge dissemination.

2. Finding an Agent

Amelia is ready to seek a publisher for her completed novel, *The Unwritten Spell*. Traditionally, finding the right agent has meant navigating a

110. Rahul Roy-Chowdhury, *Ushering in a New Era of Communication Assistance with Generative AI*, GRAMMARLY BLOG (Mar. 8, 2023), <https://www.grammarly.com/blog/grammarlygo-augmented-intelligence/> (announcing the release of Grammarly’s generative AI-powered communication assistant, designed to enhance productivity and communication for individuals and businesses by providing contextually aware suggestions, enabling on-demand writing assistance, and supporting the entire communication lifecycle).

111. Lisa Wood Shapiro, *How Technology Helped Me Cheat Dyslexia*, WIRED (June 18, 2018, 7:00 AM), <https://www.wired.com/story/end-of-dyslexia/>.

112. *Try the AI Writing Tool with Unparalleled Story Smarts*, SUDOWRITE, <https://www.sudowrite.com/> (last visited Sept. 12, 2024) (“Sudowrite is the non-judgmental AI writing partner you always wanted. . . . Sudowrite never sleeps. . . [or] runs out of ideas.”); *ChatGPT*, OPENAI, <https://openai.com/chatgpt> (last visited Sept. 12, 2024).

113. *Frequently Asked Questions*, SUDOWRITE, <https://www.sudowrite.com/faq> (last visited Sept. 12, 2024) (“We use over two dozen AI models, including Claude 2 by Anthropic, multiple open models, and several variants of GPT-3.5, and GPT-4, transformer models created by OpenAI.”).

114. SUDOWRITE, *supra* note 112 (explaining these features).

115. *Id.*

116. Drew Harwell, Nitasha Tiku, & Will Oremus, *Stumbling with Their Words, Some People Let AI Do the Talking*, WASH. POST (Dec. 10, 2022, 7:00 AM), <https://www.washingtonpost.com/technology/2022/12/10/chatgpt-ai-helps-written-communication/>.

117. Journalist’s quote taken from Francesca Paris & Larry Buchanan, *35 Ways Real People Are Using A.I. Right Now*, N.Y. TIMES (Apr. 14, 2023), <https://www.nytimes.com/interactive/2023/04/14/upshot/up-ai-uses.html>.

labyrinth of guesswork and rejection,¹¹⁸ but AI is changing that. Agent-matching platforms like QueryTracker and Publishers Marketplace analyze authors' work by scrutinizing genre, style, and deeper elements of plot and theme.¹¹⁹ These tools draw their knowledge from a vast pool of textual training data, including published novels and past agent deals.¹²⁰ They use this training to perform a kind of literary matchmaking, comparing users' drafts against successful submissions, agent profiles, and the ever-shifting winds of the book market.¹²¹

These platforms claim they can recommend agents who appreciate authors' unique voices, and they also help authors tailor their query letters, highlighting the specific elements that might resonate with a given agent's tastes. This process pushes Amelia to identify and amplify what makes her work compelling. Ultimately, these tools increase Amelia's chances of finding the right representative to bring her story to a wider audience.

3. Negotiating a Contract

Amelia finds a promising agent, but she wants to be careful about handing over the reins of her creative work. She turns to a service called Spellbook.¹²² Powered by LLMs like GPT-4 and trained on vast datasets of copyrightable subject matter, including legal documents, Spellbook helps users understand and modify contracts.¹²³ Traditionally, only writers who could afford to hire a lawyer had access to these insights and abilities. But with Spellbook, Amelia uploads the agent's proposed contract, and the software reviews the document and suggests changes.¹²⁴

4. From the Author's Desk to the Publisher's Eye

Let us explore how AI is shaping the book market. For a publisher like Evermore Press, a small press specializing in fantasy-themed stories, AI-based market research tools trained on massive datasets of editorial reviews, book metadata, and social media sentiment analysis offer useful insights.¹²⁵ These tools help Evermore understand bestseller trends, niche

118. See, e.g., MICHAEL LARSEN, HOW TO GET A LITERARY AGENT 36 (2006) (discussing editor and agent rejections).

119. See *What is a Literary Agent?*, QUERYTRACKER, https://querytracker.net/help/agents/what_is_an_agent/#process (last visited Sept. 6, 2024); see also *Publishers Marketplace Writer's Guide*, PUBLISHERS MARKETPLACE, <https://www.publishersmarketplace.com/help/WritersGuide.cgi> (last visited Sept. 6, 2024).

120. *Agent Reports*, QUERYTRACKER, <https://querytracker.net/help/reports/> (last visited Sept. 1, 2024); see *Publishers Marketplace*, *supra* note 119.

121. See sources cited *supra* note 119.

122. *Draft and Review Contracts 10x Faster with AI*, SPELLBOOK, <https://www.spellbook.legal/> (last visited Oct. 14, 2024) ("Spellbook uses AI to review and suggest terms for your contracts, right in Microsoft Word.")

123. *Id.*

124. *Id.*

125. TONY SAIKALY, THE TRANSFORMATIVE ROLE OF ARTIFICIAL INTELLIGENCE IN THE PUBLISHING INDUSTRY 5–6 (2023), https://fadel.com/wp-content/uploads/The-Transformative-Role-of-AI-in-the-Publishing-Industry_4.pdf ("AI-driven market analysis is particularly adept at

audiences, and evolving reader preferences. Suppose Evermore discovers from these tools that fantasy novels with strong female protagonists are growing in popularity. The hunt is on to find a promising manuscript that fits this criteria.

Amelia's agent submits *The Unwritten Spell* through Evermore Press's submission portal. During this initial screening phase, Evermore uses an LLM as a gatekeeper to sift through the submission "slush pile."¹²⁶ Algorithms trained on a vast corpus of successful manuscripts allow the software to quickly assess literary merit, commercial potential, and alignment with a publisher's interests.¹²⁷ These tools flag promising works, but a publisher's decision ultimately moves a manuscript forward. Even then, AI tools analyze the submission alongside current bestsellers and broader market trends.¹²⁸ This provides publishers and editors with detailed and helpful comparisons.¹²⁹ While the human touch remains essential, these AI-powered analyses are helping publishers discover and publish works that meet the market's evolving demands, ensuring stories find their way to eager readers.

Amelia's hard work and the publisher's AI-driven analysis pay off: Evermore Press extends a publication offer. AI continues to play a role, even in the contract negotiation phase. Again, LLMs in services like Spellbook help authors like Amelia and their agents assess contractual terms, potentially empowering them to advocate for favorable rights.¹³⁰ In doing so, this type of tool increases the likelihood of a fair and transparent agreement that paves the way for a book's successful launch.

5. AI-Assisted Editing

The editorial process at Evermore Press is a fruitful collaboration between human experts and AI assistants.¹³¹ After Amelia's editor provides

identifying emerging trends and shifts in audience preferences. By examining patterns in reader behavior, AI can pinpoint topics, genres, or themes that are gaining traction. This knowledge empowers publishers to act swiftly and strategically to meet evolving market demands. . . . AI has emerged as a game-changer, enabling publishers to gather and interpret vast amounts of data to make data-driven decisions that shape their acquisitions, content strategies, and distribution channels.”)

126. A “slush pile” refers to a collection of unsolicited manuscripts submitted to a publisher, typically from writers without agents or established connections. See *Slush Pile*, CAMBRIDGE DICTIONARY, <https://dictionary.cambridge.org/us/dictionary/english/slush-pile> (last visited Jan. 1, 2025); Thad McIlroy, *AI Is About to Turn Book Publishing Upside-Down*, PUBLISHERS WKLY. (June 2, 2023), <https://www.publishersweekly.com/pw/by-topic/digital/content-and-e-books/article/92471-ai-is-about-to-turn-book-publishing-upside-down.html> (anticipating that ChatGPT will soon be “patron saint” of the slush pile).

127. *Id.*

128. SAIKALY, *supra* note 125, at 5–6, 8.

129. *Id.* at 8.

130. SPELLBOOK, *supra* note 122.

131. Elizabeth A. Harris & Alexandra Alter, *A.I.'s Inroads in Publishing Touch Off Fear, and Creativity*, N.Y. TIMES (Aug. 2, 2023), <https://www.nytimes.com/2023/08/02/books/ais-inroads-in-publishing-touch-off-fear-and-creativity.html> (“At the same time, there’s already been a boom in publishing start-ups that are tapping artificial intelligence to create, package, edit and market books, said Thad McIlroy, an industry analyst who has studied the impact of A.I., and tracked nearly 50 such companies.”).

detailed high-level guidance on plot structure and character arcs, AI tools offer support in several ways. As in the earlier stages of drafting, AI-powered word processing software like Sudowrite analyzes the manuscript for narrative pacing, flags potential plot inconsistencies, and suggests deeper character motivations. Additionally, a grammar and style checker, such as Grammarly, provides assistance through multiple rounds of editing.

6. Printing, Marketing, and Release

Evermore Press carries out an AI-powered marketing strategy for *The Unwritten Spell*.¹³² First, an AI analyzes book sales data and an LLM reads and interprets online reviews to pinpoint potential readers, mainly young adults and fantasy enthusiasts. A generative AI tool (e.g., ChatGPT) develops targeted promotional strategies, crafting sample social media posts like “Dive into a world of magic and mystery with *The Unwritten Spell!*” and drafting ad copy that echoes the book’s tone. A generative AI also suggests eye-catching promotional materials, such as a vibrant book cover featuring a mystical forest and a captivating tagline. This approach helps Amelia’s story reach readers most likely to enjoy and appreciate it, maximizing its impact. As a result, the book sells well.

The domestic success of *The Unwritten Spell* leads Evermore Press to explore translation options to reach international readers. They use LLMs like ChatGPT and Claude.ai to draft the novel in Spanish, French, and Japanese.¹³³ These tools, trained on extensive multilingual texts, provide accurate translations that human translators later refine to preserve the whimsy and charm of Amelia’s characters. This blend of AI and human expertise speeds up the translation process, making it more cost-effective than traditional translation methods.¹³⁴

Amelia and Evermore Press are dedicated to making *The Unwritten Spell* accessible to a wide range of readers. They use an LLM to produce summaries of varying complexity, helping readers with cognitive or learning disabilities understand the book’s plot and themes. For people who are visually impaired, an LLM converts the text into an audiobook or into electronic braille format.¹³⁵ These initiatives ensure that individuals with

132. *Id.* (“A.I.-powered programs are already being used to try to solve that problem, helping readers find a greater array of books. Open Road Integrated Media, which provides marketing services to publishers, announced recently that it will use the technology to continually optimize the metadata behind individual titles, so they appear more prominently in search engines and on retail websites.”).

133. *Meet Claude*, CLAUDE.AI, <https://www.anthropic.com/claude> (last visited Sept. 6, 2024) (Claude.ai is an AI assistant powered by an LLM that engages in human-like conversations to help users with a wide array of tasks); CHATGPT, <https://chatgpt.com> (last visited Sept. 10, 2024).

134. Chris Arrant, *AI Is Coming to Manga, and Looking to Remove Human Translators from the Equation*, POPVERSE (Sept. 20, 2023), <https://www.thepopverse.com/manga-ai-translation-localization-relettering-orange>.

135. Although there are no services specifically designed to generate braille from text, general purpose LLMs such as Claude.ai can be used effectively for this task. The author tested this simply by asking Claude.ai to translate a portion of Alice’s Adventures in Wonderland formatted in braille into

disabilities can immerse themselves in Amelia's world, expanding the book's audience and enhancing its distribution.

7. Conclusion

Amelia is fictional, but her general experience is not. AI is transforming the literary landscape by expanding the pool of writers and altering the relationship between authors and their audiences. From assistive tools that aid writers with dyslexia to algorithms that match authors with compatible agents, AI offers support at various stages of the publication process. It assists in contract negotiations, helps publishers identify promising manuscripts, and analyzes market trends for effective book promotion. Finally, AI enhances accessibility by facilitating translations and making books available in formats suitable for people with disabilities. As AI continues to evolve, we can expect further changes in how books are disseminated, raising intriguing questions about the future of authorship and readership.

B. Movies

1. Screenplay Adaptation

The success of *The Unwritten Spell* attracts interest from Hollywood, raising the possibility of a film adaptation. While adaptations often involve external screenwriters, Amelia feels deeply connected to her story and wants to maintain creative control. Accordingly, she decides to try crafting the screenplay herself. To aid in this endeavor, she uses an LLM like ChatGPT or Sudowrite for scriptwriting assistance.¹³⁶ As Amelia analyzes successful screenplays within her genre, Sudowrite helps refine her prose into compelling cinematic exchanges. Sudowrite also can offer insights into visual storytelling, character arcs, and principles of the three-act structure.¹³⁷

2. Studio Pitch and Script Analysis

With her screenplay complete, Amelia and her agent pitch the project to various studios. Aurora Studios, known for its imaginative and visually striking films, expresses interest. However, Amelia's screenplay faces stiff competition within the studio's extensive pool of submissions.

To aid its evaluation process, Aurora Studios uses AI-powered script analysis tools. Two of these tools, ScriptBook and Cinelytic, specialize in analyzing vast datasets of successful scripts to provide studios with

English. The tool did so without any errors. Claude.ai conversation with the author (Sept. 18, 2024) (on file with author).

136. See *ChatGPT*, *supra* note 112 (While OpenAI, ChatGPT's creator, has not disclosed whether the tool has been explicitly trained on screenplays, when prompted with this question, the tool itself reports that it may have been. A screenshot of this exchange is on file with the author); *SUDOWRITE*, *supra* note 112.

137. *SUDOWRITE*, *supra* note 112.

in-depth insights.¹³⁸ ScriptBook, trained on over 30,000 real scripts, focuses on script analysis, offering predictions on a film’s financial forecast, character likability scores, potential audience profiles, and even critic ratings.¹³⁹ Cinelytic, on the other hand, is designed to forecast box-office performance.¹⁴⁰ This tool analyzes data from thousands of movies and entertainment industry professionals to make its predictions.¹⁴¹ These tools identify Amelia’s screenplay to Aurora Studios as a strong contender for selection and production.

After some discussions, Aurora Studios and Amelia decide to work together to turn *The Unwritten Spell* into a movie. As described in the previous case study, Spellbook is again helpful at this stage for reviewing and suggesting changes to their agreement.

3. Casting and Storyboarding

Preproduction kicks into gear, and the search for actors to embody Amelia’s beloved characters begins. Traditionally, casting has been an art—a realm where intuition, experience, and (let’s face it) industry connections reigned supreme. But Aurora Studios embraces technology. They use an AI-powered casting tool called Largo.ai.¹⁴² This service dissects an actor’s every nuance—facial expressions, vocal inflections, and even overall cinematic presence.¹⁴³ By analyzing actors’ emotional range and aligning their past roles with those in Amelia’s world, these tools promise they can reveal hidden gems overlooked by traditional casting methods.¹⁴⁴

As Aurora Studios makes casting decisions for the film, the director and production team collaborate on visually translating the screenplay into a cinematic blueprint. Traditional hand-drawn storyboards can be time-consuming and slow to produce. Aurora Studios, always seeking to be innovative, turns to Storyboarder.ai—a generative AI tool designed to speed up the storyboarding process.¹⁴⁵ The service can generate storyboards with remarkable speed, allowing filmmakers to experiment with

138. James Vincent, *Warner Bros. Signs AI Startup That Claims to Predict Film Success*, THE VERGE (Jan. 9, 2020, 3:38 AM), <https://www.theverge.com/2020/1/9/21058094/ai-film-decision-making-warner-bros-signs-cinelytic> (discussing Warner Bros.’ partnership with Cinelytic to use AI in film decision-making); James Vincent, *Hollywood Is Quietly Using AI to Help Decide Which Movies to Make*, THE VERGE (May 28, 2019, 9:10 AM), <https://www.theverge.com/2019/5/28/18637135/hollywood-ai-film-decision-script-analysis-data-machine-learning>.

139. *Democratizing Storytelling Through the Art of AI*, SCRIPTBOOK, <https://www.scriptbook.io> (last visited Sept. 11, 2024) (reporting over 30,000 scripts were used to train the system).

140. *Future-Proofing the Entertainment Industry*, CINELYTIC, <https://www.cinelytic.com> (last visited Sept. 11, 2024).

141. *Id.*

142. LARGO, <https://home.largo.ai> (last visited Sept. 11, 2024) (“Screenplay analysis, project package optimisation, character analysis & casting proposition, and financial forecasts.”).

143. *See id.*

144. *See id.*

145. *From Idea to Storyboard in Minutes*, STORYBOARDER.AI, <https://storyboarder.ai> (last visited Sept. 11, 2024) (describing an AI-powered tool that generates storyboards for filmmakers); *see Online Storyboard Creator*, KROCK.IO, <https://krock.io/online-storyboards-creator/> (last visited Sept. 11, 2024) (providing a product with similar capabilities).

camera angles, lighting, and scene composition with unprecedented ease.¹⁴⁶ This technological leap saves time and resources, ultimately bringing Amelia's story to moviegoers more quickly.

4. Accessibility

Amelia and Aurora Studios are committed to ensuring everyone can experience the magic of *The Unwritten Spell* on the big screen. Powered by an AI speech recognition tool, the film has flawless closed captioning.¹⁴⁷ In the future, these tools might be able to carefully consider a movie's sound design and pacing to ensure the captions convey each scene's spoken words and emotional atmosphere.¹⁴⁸ The simpler automated captioning provided by today's technology still benefits viewers with hearing impairments and those viewing the film in noisy environments.

Aurora Studios also invests in AI-powered audio descriptions to expand the film's reach further.¹⁴⁹ Using ML technologies capable of viewing and understanding scenes in the film, they create a detailed narrative track that vividly describes key visual elements, character actions, and scene transitions.¹⁵⁰ Visually impaired viewers can now experience aspects of the film that they otherwise would have missed, such as the actors' subtle facial expressions. Finally, LLMs help translate Amelia's dialogue into multiple languages, broadening accessibility on a global scale.

5. Conclusion

From the initial adaptation pitch to the film's release, AI played a pivotal role in disseminating *The Unwritten Spell*. It helped Amelia, a writer with dyslexia, maintain creative control during the screenplay adaptation process. AI tools analyzed successful scripts, providing insights that informed her work and helped her hone a compelling pitch for studios. During negotiations, AI tools scanned contracts and brought Amelia and the studio to a deal more quickly, and perhaps even more carefully, than would have been possible in the past.¹⁵¹ At the casting stage, AI's analysis of actors' past performances augmented traditional casting methods,

146. STORYBOARDER.AI, *supra* note 145.

147. See, e.g., *Speech to Text in Premiere Pro*, ADOBE, <https://helpx.adobe.com/premiere-pro/using/speech-to-text.html> (last visited Sept. 12, 2024) (describing Adobe Premiere Pro, a standard in the film industry, which has an AI-powered feature to automatically generate transcripts and captions for videos); see also *Descript*, DESCRIPT, <https://www.descript.com> (last visited Sept. 12, 2024) (providing an overview of Descript, a software tool that provides simple and automatic captioning with AI).

148. ADOBE, *supra* note 147.

149. See Cliff Weitzman, *Can AI Video Editors Add Subtitles, Open Captions, or Closed Captions to a Video?*, SPEECHIFY: VIDEO STUDIO (Oct. 11, 2023), <https://speechify.com/blog/how-use-ai-add-subtitles/>.

150. See Press Release, Perkins School for the Blind, *Audible Sight: A New Method for Creating Audio Descriptions* (Jan. 25, 2024), <https://www.perkins.org/resource/audible-sight-a-new-method-for-creating-audio-descriptions> (describing Audible Sight, an AI-powered tool that creates audio descriptions for those with visual impairments).

151. See SPELLBOOK, *supra* note 122 (claiming Spellbook helps draft contracts "10x faster.").

potentially leading to the discovery of overlooked talent. AI-generated visuals streamlined the storyboarding process, saving time and resources. Finally, AI made the film accessible through closed captions that considered pacing and sound design, detailed audio descriptions for visually impaired viewers, and seamless translation for global audiences. This case study demonstrates how AI can expand the reach of stories, lower costs,¹⁵² and diversify the voices¹⁵³ that we see on screen.

C. Libraries

1. AI-Powered Content Discovery

AI is transforming how libraries store, organize, and disseminate knowledge. While historical advancements like the Dewey Decimal System excelled in managing human-generated content,¹⁵⁴ AI offers something new: the ability to analyze, understand, and even generate text. This ability significantly expands the effectiveness of library services and redefines the role of the modern librarian. To see this in action, we will follow a day in the life of Alex Lee, the fictional head librarian of the Raintree County Public Library.

After arriving at work one morning, Alex encounters a patron who is eager to find more fantasy novels like those of a particular author (the patron had already exhausted the library's collection of books by that author). Recognizing an opportunity to utilize Talpa, a new AI-powered library search tool, Alex walks the patron through the process.¹⁵⁵ After explaining this tool's capabilities, Alex assists the patron in phrasing the request as a natural language query: "Can you recommend fantasy novels with strong female protagonists set in quaint villages?"¹⁵⁶ In response, Talpa provides dozens of books that fit the description. This ability to process natural language queries, combined with an in-depth understanding of the library's collection, goes far beyond what a traditional library search

152. See Jessica Toonkel & Sarah Krouse, *Who Owns SpongeBob? AI Shakes Hollywood's Creative Foundation; Artificial-Intelligence Tools Raise Questions About Intellectual-Property Rights That Are Under Scrutiny by Courts and Entertainment Executives*, WALL ST. J. (Apr. 4, 2023), <https://www.wsj.com/articles/ai-chatgpt-hollywood-intellectual-property-spongebob-81fd5d15>

("The conversations we are having aren't about how generative AI is going to create the story for you, but it is how it can make things faster and cheaper around the edges," said Andre James, the global head of media and entertainment practice at Bain & Co., which has partnered with OpenAI. "And those edges are hundreds of millions of dollars." Many companies are already using AI tools for dubbing and subtitles, he said, which represent a \$2.6 billion market.")

153. See Interview by Anthony Cawood with Nadira Azermai, CEO of ScriptBook, at SCREENWRITING OPPORTUNITIES (Apr. 21, 2020), <https://screenwritingopportunities.wordpress.com/2020/04/21/interview-with-nadira-azermai-ceo-of-scriptbook> (In this interview, Nadira Azermai, the CEO of ScriptBook, discusses how AI can benefit the film industry by providing unbiased, objective analysis of scripts. She believes that AI can democratize storytelling by removing subjective biases and offering a more inclusive platform for diverse voices.)

154. MELVIL DEWEY, *DEWEY DECIMAL CLASSIFICATION AND RELATIVE INDEX*, at xxv–xvii (Benjamin A. Custer ed., Forest Press 1979) (1876).

155. *What is Talpa Search?*, TALPA AI, <https://www.talpasearch.com/about> (last visited Sept. 12, 2024).

156. *Id.* (demonstrating this kind of functionality).

system can accomplish. Within the list of suggestions, Alex notices a new title: *The Unwritten Spell*. After reading the book's description, the patron decides to check it out.

2. AI-Facilitated Research

That afternoon, a high school student approaches Alex's desk, seeking help with a research paper on the Dust Bowl. Together, Alex and the student visit an AI-based search engine called Perplexity.ai.¹⁵⁷ Their dialogue with the service goes as follows:

User: Can you recommend a book about the Dust Bowl for a high school student?

Perplexity: For a high school student interested in learning about the Dust Bowl, I would highly recommend *Children of the Dust Bowl: The True Story of the School at Weedpatch Camp* by Jerry Stanley. This book offers several advantages that make it particularly suitable for young readers. . . . The book brings history to life by following specific families and individuals, making the historical events more relatable and memorable for students. . . . The book is relatively short and written in an engaging, easy-to-understand style that is appropriate for high school students.¹⁵⁸

Alex and the student locate the book in the library's holdings. Alex looks the book over and decides that it is a good match for what the student is looking for.

Within some libraries today, search engines like Perplexity and chatbots like ChatGPT address the challenge of limited staff availability.¹⁵⁹ They can assist at any hour and can handle common inquiries, allowing librarians to focus on more complex requests.¹⁶⁰ However, it is essential to remember that the quality of ChatGPT's output depends on the clarity of the question asked. Librarians play a vital role in guiding patrons on how to effectively interact with these tools and stressing the importance of critically evaluating AI-generated information.¹⁶¹

3. Generating Reading Lists

Late in the day, Alex receives an email request from a local book club seeking reading recommendations. The club is interested in historical

157. PERPLEXITY AI, <https://www.perplexity.ai> (last visited Sept. 12, 2024).

158. This dialogue is a verbatim record of the author's dialog with this service at the time of writing. Conversation on file with the author.

159. See Adebowale Jeremy Adetayo, *Artificial Intelligence Chatbots in Academic Libraries: The Rise of ChatGPT*, 40 LIBR. HI TECH NEWS, Mar. 2023, at 18, 18, <https://doi.org/10.1108/LHTN-01-2023-0007> (reporting that AI chatbots can make library services more accessible to users with time or mobility constraints by assisting library patrons in accessing materials, placing holds, and completing other tasks at any time of day).

160. See Subaveerapandiyan A, *Application of Artificial Intelligence (AI) in Libraries and Its Impact on Library Operations Review*, LIBR. PHIL. & PRAC., June 2023, at 7, <https://digitalcommons.unl.edu/libphilprac/7828> (describing the use of AI Chatbots).

161. *Id.* at 5.

fiction novels set in Tudor England. Alex has always enjoyed crafting personalized reading lists and believes that the library's new content generation tool¹⁶² can make her better at this task. Experimenting with the tool, Alex provides the book club's request as a prompt. Within moments, the AI generates a diverse list of relevant titles, including well-known classics and intriguing new releases. Alex reviews and edits the list using her own expertise and sends it to the book club.

AI's ability to produce book summaries, reading lists, articles, and even blog posts offers a valuable time-saving resource for librarians. These tools can personalize content based on user preferences, enriching the library experience. While librarians should always carefully review AI-generated content for accuracy and relevance, AI technology can augment librarian expertise and make the creation of library resources more efficient.

4. Unlocking Visual Collections

After a busy day, Alex takes a moment to explore the library's local history collection. A worn, leather-bound photo album on a high shelf catches her eye. Its pages are filled with images detailing the county's rich past, but the album is crumbling and difficult to navigate. Recognizing an opportunity to apply the library's image-tagging tool,¹⁶³ Alex selects a few sample photos and uploads them to the platform. Powered by computer vision models like Google Cloud Vision or Clarifai,¹⁶⁴ the tool analyzes the images and generates preliminary descriptive tags such as "people," "buildings," and "vehicles." Encouraged, Alex knows that by carefully tagging a small selection of photos herself, she can train the AI to achieve even greater accuracy and enhance the discoverability of these unique historical records.

AI-assisted image tagging and metadata generation have the potential to unlock the rich visual collections held within libraries.¹⁶⁵ Computer vision technologies, even those not specifically designed for historical images, can provide a starting point for categorizing vast backlogs of photographs, maps, and other visual documents. With the help of librarians, who can apply detailed labels to a subset of images, these tools can learn to identify increasingly complex patterns and themes.¹⁶⁶ This enhances

162. A general-purpose LLM such as ChatGPT or Claude.ai, for instance.

163. One of these tools is Microsoft Syntex. See *Overview of Microsoft Syntex*, MICROSOFT LEARN, <https://learn.microsoft.com/en-us/microsoft-365/syntex/syntex-overview> (last visited Sept. 12, 2024).

164. *Vision AI: Extract Insights from Images, Documents, and Videos*, GOOGLE CLOUD, <https://cloud.google.com/vision> (last visited Sept. 10, 2024); *Use AI Auto Tagging to Improve Asset Naming Consistency*, CLARIFAI, <https://www.clarifai.com/customers/online-travel-agency> (last visited Sept. 10, 2024).

165. See Catherine Nicole Coleman, Claudia Engel, & Hilary Thorsen, *Subjectivity and Discoverability: An Exploration with Images*, in *THE RISE OF AI: IMPLICATIONS AND APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN ACADEMIC LIBRARIES* 83 (Sandy Hervieux & Amanda Wheatley eds., 2022).

166. *Id.* at 88 (describing this use).

discoverability for researchers and patrons and creates connections between different visual materials, uncovering new insights within the library's collection.

The integration of AI into library workflows significantly enhances the dissemination of knowledge to the public. AI-powered search tools like Talpa provide more precise and personalized content discovery experiences. Chatbots powered by ChatGPT assist with basic inquiries, empowering patrons to find answers whenever needed. Content generation tools facilitate the creation of tailored reading lists and summaries, enriching the library experience. Finally, AI-assisted image tagging unlocks the potential of visual collections, making them more accessible for research and discovery.

“AI should not be seen as a replacement for human interaction but rather as a tool for augmenting human impact,” stresses David Leonard, President of the Boston Public Library.¹⁶⁷ As AI capabilities evolve, libraries and librarians will play a pivotal role in guiding its responsible implementation.¹⁶⁸ A recent poll by the Association of Research Libraries reported an optimistic outlook, with over half of respondents stating they anticipate significant improvements in library services by mid-2024.¹⁶⁹

D. Video Games

Video game developers often face a difficult choice: focus resources on a single platform or adapt their work to reach a wider audience. To understand how AI helps disseminate video games, this Section looks at a hypothetical game developer.

Imagine an independent studio that wants to bring its critically acclaimed puzzle game to more players. Their original game was designed for PCs (laptop and desktop computers), but they have now set their sights on consoles and mobile devices. This ambitious move comes with a challenge: making their game run smoothly across different hardware and operating systems. To tackle this, the studio uses AI tools trained on massive datasets that may include code. One of these tools is the popular coding assistant GitHub Copilot (powered by GPT-4).¹⁷⁰

One of the studio's first challenges is making their game look as good on a console or phone as it does on a computer. Each device has its own

167. *Leadership Brief: Explorations of Generative AI for Library Systems*, URB. LIBRS. COUNCIL 1 (2023), https://www.urbanlibraries.org/files/AI_Leadership-Brief_October2023.pdf.

168. See Emily Udell, *The World of AI: How Libraries Are Integrating and Navigating This Powerful Technology*, AM. LIBRS. (Mar. 1, 2024), <https://americanlibrariesmagazine.org/2024/03/01/the-world-of-ai/> (emphasizing that while AI can automate certain tasks, librarians will remain vital, particularly in managing ethical concerns, data privacy, and ensuring equitable access to information).

169. LEO S. LO & CYNTHIA HUDSON VITALE, *EVOLVING AI STRATEGIES IN LIBRARIES: INSIGHTS FROM TWO POLLS OF ARL MEMBER REPRESENTATIVES OVER NINE MONTHS 5–6* (2024), <https://doi.org/10.29242/report.aipolls2023>.

170. *The World's Most Widely Adopted AI Developer Tool*, GITHUB, <https://github.com/features/copilot> (last visited Sept. 10, 2024).

way of handling graphics, like a unique language. GitHub Copilot has studied different systems' graphics "languages," and game developers can supply it with a massive library of existing game code.¹⁷¹ Like a translator offering various ways to say the same thing, this tool suggests alternative coding techniques to make the visuals shine on consoles and phones.

The studio does not want its game to feel slow or clunky on any device. To tackle this, they use GitHub Copilot again to analyze the game's code and predict how it will run on different consoles or phones. The tool suggests ways to streamline how the game loads levels or renders complex effects. The studio's coding team always double-checks these suggestions to ensure their changes have the desired impact.

Porting a video game to a new platform can be like moving to a new house—unexpected and costly problems pop up at every turn.¹⁷² To stay ahead of these issues, the studio creates their own in-house LLM and trains it on the code of games they have ported in the past. This tool has learned to recognize patterns in code that often cause bugs when moved to different platforms. Like an inspector looking for potential issues with a new home, this AI highlights trouble that the code might encounter on a new platform. This early warning system helps software developers fix problems before they break the game and delay its release.

For a project like game porting, AI assistance means fewer bugs and a faster workflow, which is crucial for a small studio. By helping game creators port games to new devices so they will run smoothly, AI can expand a game's reach.

III. TOWARDS FACILITATIVE FAIR USE

This Part introduces Facilitative Fair Use, a legal framework that applies to technologies that help facilitate the copyright ecosystem. This framework establishes a presumption of fairness, rebuttable by evidence of market substitution.¹⁷³ The discussion explores the rationale for and mechanics of Facilitative Fair Use and addresses potential counterarguments and shortcomings. Finally, this Part examines how Facilitative Fair Use can foster wider dissemination while addressing biases within creative industries.

171. *Id.*

172. The term "porting" refers to the practice of adapting a video game to run on a computer or game system it was not originally designed to run on. *See, e.g.,* Alena Porokh, *How to Port Games: Business and Technical Aspects Unveiled*, KEVURU GAMES (Aug. 8, 2023), <https://kevuru-games.com/blog/how-to-port-games-business-and-technical-aspects/>.

173. The term "market substitution" in this context refers to competition—i.e., the potential for a secondary work to compete in the market with an original work. *See, e.g.,* Michael G. Anderson, Paul F. Brown, & Andrew P. Cores, *Market Substitution and Copyrights: Predicting Fair Use Case Law*, 10 U. MIA. ENT. & SPORTS L. REV. 33, 39 (1993).

A. The Need for Facilitative Fair Use

This Section applies the principles and goals of fair use to the examples presented in Part II. It concludes that these examples support copyright law and policy by facilitating the production and dissemination of knowledge in two ways: first, by expanding the number and variety of authorial voices, and second, by increasing efficiencies in the production and dissemination of works. The discussion then highlights the uncertainty surrounding how courts will analyze fair use arguments in the ongoing training data lawsuits. This uncertainty underscores the need for fair use to adapt in the era of AI.

The *Sony* decision articulates why the dissemination of knowledge is central to copyright law. Recall from Part I that the *Sony* Court concluded that private, time-shifted viewing via VCRs increased the audience for copyrighted broadcasts.¹⁷⁴ Writing for the majority (and quoting an earlier decision), Justice Stephens noted that copyright serves to benefit the public first and authors second: “The sole interest of the United States and the primary object in conferring the monopoly lie in the general benefits derived by the public from the labors of authors.”¹⁷⁵

This principle persisted into the internet era. In *Google Books*, the Second Circuit determined that Google’s project, despite involving the copying of millions of books, was fair use.¹⁷⁶ The court emphasized, “The ultimate goal of copyright is to expand public knowledge and understanding.”¹⁷⁷ It recognized that while authors are important beneficiaries of copyright, the public—whose access to knowledge copyright aims to advance through authorship rewards—is the ultimate and primary beneficiary.¹⁷⁸

The case studies in Part II illustrate how AI enhances knowledge dissemination across various sectors.¹⁷⁹ For example, in the publishing industry, AI tackles the challenge of sorting through vast volumes of submissions.¹⁸⁰ It does this by analyzing market trends and reader preferences. These functions require an understanding of language, which in turn requires AI systems trained on large amounts of text. As a result, publishers can spot promising works more quickly than they could in the past, thereby expanding the range of creative content accessible to the public. Furthermore, AI plays a role in contract drafting, editorial analysis, targeted

174. See *supra* Section I.B.

175. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).

176. *Google Books*, 804 F.3d 202, 229 (2d Cir. 2015) (“In sum, we conclude that: (1) Google’s unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses.”).

177. *Id.* at 212.

178. *Id.*; *Sony*, 464 U.S. at 432 (“But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good. ‘The sole interest of the United States and the primary object in conferring the monopoly,’ this Court has said, ‘lie in the general benefits derived by the public from the labors of authors.’”).

179. See *supra* Part II.

180. See *supra* Section II.A.

marketing, and recommendation systems. Collectively, these tools help disseminate the knowledge contained in literary works.

Similarly, AI powers the dissemination of diverse stories and creative visions in the film industry.¹⁸¹ Platforms like ScriptBook claim their AI can identify screenplays that would have otherwise gone overlooked.¹⁸² Filmmakers can bring a wider variety of stories to life by lowering production costs through tools like AI storyboard generators. Additionally, AI enables studios to understand their audiences better, ensuring that films are disseminated more broadly.

This drive to expand knowledge access extends to libraries.¹⁸³ ChatGPT provides users with natural language search, personalized results, and concise summaries, which make complex information more accessible. Platforms like Talpa go further, tailoring curated content to individual needs. These AI-powered tools empower libraries in their core mission of knowledge dissemination.

AI-assisted dissemination's potential bridges technical divides in video games, enabling games to reach wider audiences.¹⁸⁴ By translating graphics code for different platforms, optimizing performance, and even identifying potential bugs, AI streamlines the process of porting games to new devices. This expands a game's potential audience.

Some might disagree with this analysis. They might argue that comparisons between AI and past technologies like VCRs or Google Books are flawed due to AI's ability to create new content from copyrighted material. This ability raises the specter of market competition in ways that seem more dire than home video recorders or book search engines. While it is valuable to consider this difference, a complete fair use analysis would account for any harm to authors. The fact that market harm might well exist in the AI context does not negate the potential helpful impact AI has on dissemination. Technologies have the ability to both help and harm us in different ways.

Critics also might correctly point out that, unlike the VCR and Google Books, AI's role in dissemination is more complex and often indirect. Nonetheless, this does not weaken this Article's analysis. Dissemination is a process—a series of steps that connect authors to audiences, as the case studies vividly illustrate. The fact that AI intervenes at earlier stages of this process does not make its role in the dissemination process any less critical than the role of consumer-facing technologies.

Another possible critique is that AI's dissemination potential does not directly benefit the specific works used for training. But Google Books arguably benefited the very authors whose books it copied by expanding

181. *See supra* Section II.B.

182. *See supra* Section II.B.2.

183. *See supra* Section II.C.

184. *See supra* Section II.D.

the reach of those volumes. Firstly, the case studies in Part II show that AI can still be helpful to books that have been published. For instance, it can expand a work's market by generating translations, adaptations, or new marketing strategies. Secondly, this concern overlooks the fact that, in the view of copyright law, dissemination is all about benefiting the public—not just individual authors. This is why courts analyzing fair use have looked beyond the impact on the plaintiff to also consider how technology expands public access to knowledge.

So what is the problem? If AI-powered dissemination aligns so clearly with copyright's goals, will courts see it that way too? Can we not trust courts to analyze AI's training data use on a case-by-case basis and reach fair outcomes? Unfortunately, the law is not so predictable.¹⁸⁵

Consider TVEyes,¹⁸⁶ a technology company that offered its subscribers an easily searchable video database of television news broadcasts.¹⁸⁷ While providing some public benefit, its service captured near-complete copies of content and allowed playback of lengthy clips.¹⁸⁸ On the surface, TVEyes and Google Books share some similarities: both created large, searchable databases of copyrighted works. However, in a 2018 decision, the Second Circuit—the same court that decided *Google Books*—held that TVEyes's system was not fair use.¹⁸⁹ The court's analysis hinged on a few differences. Google's snippets were short excerpts from book pages, while TVEyes's clips were perhaps long enough to be more interesting or helpful to users, undermining the primary market for copyrighted broadcasts.¹⁹⁰ This case illustrates the precarious balance of the fair use doctrine: seemingly minor differences can significantly influence a court's conclusion as to whether a technology permissibly disseminates knowledge or impermissibly violates copyright.

The uncertainty surrounding the application of fair use to AI presents a threat to innovation. Consider a startup developing an AI tool for summarizing scientific articles. The tool, designed to enhance public understanding of complex research, could face legal challenges due to unclear fair use guidelines. This uncertainty might lead the startup to abandon the project, depriving the public of a valuable resource for disseminating knowledge. This risk of a chilling effect illustrates the need for courts to provide clearer fair use standards that recognize the role of AI in advancing copyright law's goal of knowledge dissemination.

185. See Parchomovsky & Weiser, *supra* note 50, at 100 (“[T]he fair use doctrine has decidedly remained an ‘equitable rule of reason,’ with all of its attendant uncertainty.”).

186. Fox News Network, LLC v. TVEyes, Inc., 883 F.3d 169, 173–74 (2d Cir. 2018).

187. *Id.*

188. *Id.* at 175.

189. *Id.* at 180.

190. *Id.* at 177, 179–80.

B. Introducing Facilitative Fair Use

The fair use doctrine must better address AI's role in spreading creative works. Specifically, the fourth factor of the fair use test, which focuses on market impact, may lead courts to overlook or downplay AI's potential to help produce and disseminate expressive works.

This Article provides courts with a new framework designed to better align copyright law with AI's role in the copyright ecosystem: Facilitative Fair Use. In simple terms, this framework would lead courts to favor AI systems that help copyright holders disseminate knowledge to the public. The rule instructs courts to consider facilitative effects in the existing four-factor fair use analysis. Importantly, this leaves open the possibility for evidence of market substitution to prevent a fair use decision.

Similarly tailored considerations already exist in other areas of fair use analysis. For instance, under the first factor of fair use analysis, a defendant's use of a copyrighted work for parody may weigh that factor in favor of fair use.¹⁹¹ Similarly, transformative uses that add new meaning are more likely to be fair uses.¹⁹² However, these considerations in favor of fair use are not absolute: plaintiffs can overcome them with evidence of countervailing factors, such as commercial substitution.¹⁹³

To successfully argue in favor of Facilitative Fair Use, a defendant must demonstrate that a purpose and character of the use is to disseminate knowledge. Relevant evidence would include documentary records showing that the AI was trained to help authors transfer their ideas into a tangible medium of fixation, to increase the discoverability of copyrighted works, to expand audience reach via translation, or to enhance access through tools like summarization or indexing. For instance, a publishing house training an LLM on copyrighted books could satisfy this standard if its AI tool was intended to bring overlooked works to the attention of new readers. Similarly, training an LLM on copyrighted books could qualify as a Facilitative Fair Use if a purpose behind that use was to create a tool designed to help people with dyslexia or ADHD transfer their thoughts to the page more effectively.

In reality, the analysis would often be difficult because LLMs are trained with more than one purpose in mind; they are general-purpose tools. Thus, the same AI that could expand the reach of one author could produce works that compete with another. While this is a challenge, it is not a new problem under fair use. In *American Geophysical Union v. Texaco Inc.*,¹⁹⁴ the Second Circuit grappled with Texaco's unlicensed copying and use of journal articles for multiple reasons, some of which favored fair use (e.g., archiving) and some of which did not (e.g., systematic

191. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).

192. *Google Books*, 804 F.3d 202, 214–17, 219 (2d Cir. 2015).

193. *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 532–33 (2023).

194. 60 F.3d 913 (2d Cir. 1994).

photocopying to avoid paying for additional subscriptions or licenses).¹⁹⁵ Ultimately, the court considered all potential purposes and found that the dominant purpose was to avoid paying for copies, which weighed against fair use.¹⁹⁶ Courts have followed a similar primary approach in other cases, and the approach would no doubt be a necessary piece of a Facilitative Fair Use analysis.

In addition to the impact on the first factor, Facilitative Fair Use would also impact the fourth factor of the fair use test: the market impact. As discussed earlier, generative AI systems' potential to produce substitutive works should weigh against fair use. However, as this Article has shown, these systems can also create new markets and expand existing markets. Through the lens of Facilitative Fair Use, courts would be encouraged to weigh these competing potentials for good and for harm. Although this analysis would not be simple, courts should be guided by a copyright foundation: fair use should be carefully tailored so as not to diminish the incentive that copyright provides to encourage the creation of new works.

To better understand how Facilitative Fair Use would work in practice, consider two hypothetical scenarios involving the training of a large language model like GPT-4. First, imagine that a developer trains GPT-4 on a diverse dataset of novels, aiming to improve its ability to identify literary themes, writing styles, and potential plot developments. The goal is to create an AI model capable of generating insightful summaries, crafting audience-targeted promotional materials, and offering creative suggestions for adaptations that expand the reach of the original works. In this case, using copyrighted works for training directly aligns with the goal of knowledge dissemination. Under the Facilitative Fair Use framework, this type of training would likely lead a court to weigh the first factor of the fair use analysis in a defendant's favor.

Now imagine a developer training GPT-4 on works by well-known authors. The developer aims to create an AI model that can closely mimic the authors' styles and generate new novels that are commercially substitutive from the authors' existing works. Here, the training process focuses on direct replication rather than on enhancing the dissemination of copyrighted material. These factors and the potential for significant market substitution would likely push the first and second factors of the fair use analysis against fair use.

Note that within the Facilitative Fair Use framework, a developer's intent in using copyrighted works for training plays a significant role. To succeed, a developer must demonstrate a clear purpose for enhancing knowledge dissemination, such as improving searchability, audience reach, or any of the (many) examples of aiding dissemination in Part II. If

195. *Id.* at 915–31.

196. *Id.* at 920.

the developer cannot directly demonstrate such intent, it could still meet the presumption by providing extensive evidence that its AI, despite not being designed for dissemination, is widely used to facilitate the spread of knowledge. This empirical evidence would need to demonstrate measurable benefits in terms of expanded discoverability, accessibility, or audience size and composition for copyrighted works.

Facilitative Fair Use is designed to benefit creators. AI-powered tools that expand audience reach, improve discoverability, and offer new creative possibilities can translate into greater sales. By making works more accessible across languages, formats, and audiences, AI can generate new revenue streams and boost the long-term visibility of an author's work. This, in turn, aligns with copyright's core goal of promoting creativity by ensuring that innovation increases public access and provides greater public benefits.

At the same time, Facilitative Fair Use offers valuable incentives for technology companies. By providing greater clarity within the fair use doctrine, courts can encourage developers to create AI tools designed explicitly for knowledge dissemination. This helps the public and gives technology companies (and their investors) increased confidence by reducing legal risks.

Some might contend that Facilitative Fair Use would unfairly favor technology companies, shielding AI applications that significantly harm the market from original copyrighted works. While this is a valid concern, courts would retain the responsibility to analyze potential market impacts thoroughly. If an AI creates substitutes for original works or if its claimed facilitative benefits are negligible, then fair use would not apply. In this way, this new consideration functions like the existing considerations in the fair use framework that were mentioned at the start of this Section (a tool to help courts think about a special class of uses), not an end-run around the existing four-factor framework. This protects authors' economic interests.

Others might argue that this proposal could diminish authors' control over how their creations are used. This concern overestimates the freedom existing law gives authors. As Jane Ginsburg has observed, "Historically, when copyright owners seek to eliminate a new kind of dissemination, and when courts do not deem that dissemination harmful to copyright owners, courts decline to find infringement."¹⁹⁷ This perspective should guide the consideration of whether the potential long-term benefits of greater dissemination outweigh concerns about control. By expanding the reach and visibility of their work, AI can generate new audiences for creators' work, unlock hidden revenue streams, and ultimately bolster the incentives for creation.

197. Ginsburg, *supra* note 23, at 1613.

Additional critics might point out that quantifying the facilitative impact of certain AI technologies can be challenging. The concept of “facilitation” itself might seem abstract for legal purposes. To address this concern, this proposal suggests that evidence could demonstrate facilitation.¹⁹⁸ This evidence could include metrics showing increased discoverability, expanded audience reach, and greater accessibility of the original work. However, courts may need to evolve and refine their evidentiary standards to assess the impact of AI on the broader copyright ecosystem. The Facilitative Fair Use framework provides a starting point.

It is important to note that the relationship between AI and copyright law is rapidly evolving, and any new framework carries the potential for unforeseen consequences. Facilitative Fair Use is designed with this in mind. Rather than providing a rigid, static solution, it introduces a fundamental bias towards facilitation within the flexible frame of fair use. The framework expects courts to refine and adapt this doctrinal tool as new AI applications and their impacts on the creative landscape emerge.

Facilitative Fair Use offers a crucial update to copyright law in the era of AI. By recognizing the value of AI tools that enhance knowledge production and dissemination, it promotes copyright’s core goal of promoting knowledge sharing. It also leaves courts empowered to protect authors from AI tools that produce competing works.

C. Facilitating Fairly

Legal experts have long been concerned with the possibility that AI will amplify existing societal biases, a fear made more urgent as AI’s influence spreads across creative industries.¹⁹⁹ The training data used to develop AI models lies at the heart of this issue. Limited or biased datasets—for example, datasets that reflect the historical underrepresentation of certain groups in publishing—threaten to recreate societal biases in the AI itself. This could lead to a range of harmful outcomes, from overlooking diverse authors in manuscript selection to recommendation systems that fail to promote works outside the mainstream. The result is not simply the perpetuation of old injustices but a feedback loop where AI becomes increasingly narrow in its view of creativity.

Intellectual property scholars have expressed concerns about how copyright law may inadvertently exacerbate issues of algorithmic bias in

198. See *supra* Sections III.A–B.

199. See, e.g., Alice Xiang, *Reconciling Legal and Technical Approaches to Algorithmic Bias*, 88 TENN. L. REV. 649, 651–57 (2021); Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1673–83 (2020); Peter K. Yu, *Artificial Intelligence, the Law-Machine Interface, and Fair Use Automation*, 72 ALA. L. REV. 187, 188–91 (2020); Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L.J. 2218, 2221–27 (2019); Anya E. R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1259–67 (2020); Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1391–94 (2019).

AI systems.²⁰⁰ In a 2018 article, Amanda Levendowski explored how copyright restrictions on data access can inadvertently encourage reliance on biased sources to train AI systems.²⁰¹ This groundbreaking work sounded an early alarm about the risk of perpetuating historical biases in creative industries through AI tools dependent on nondiverse training data. As Levendowski observed, robust fair use rights are essential for empowering developers to create more representative datasets and inclusive models.²⁰² Patrick K. Lin echoed this concern in 2023, writing that problematic training data can introduce biases into AI systems.²⁰³ However, developers face legal risks and incentives to utilize biased data sets due to copyright restrictions and the threat of copyright infringement litigation.²⁰⁴

Think of AI as a student learning about the world. If that student reads books written by just a small group of similar authors, the student's understanding of the world will be limited. Now, imagine that the student grows up to become a publisher or librarian. They will likely choose to publish or stock books that reflect the narrow worldview they were exposed to when they were young. This means even less diverse literature gets made and shared, influencing the next generation of students, publishers, and librarians. This creates a vicious cycle where bias limits the range of voices we hear and reinforces the idea that only certain types of stories are worth telling.

For a real-world example, consider how this could play out with AI in book publishing. Systemic biases have long plagued the book industry, influencing which authors receive attention and support.²⁰⁵ This is particularly evident in the lack of racial diversity, with white authors disproportionately represented in the works published by major houses.²⁰⁶ The publishing industry's practice of basing advances on historical sales data—which often reflects systemic bias—creates disproportionate barriers for authors of color.²⁰⁷ While recent activism has pushed for diversity and inclusion initiatives, true equity will require structural changes throughout the industry.

200. Levendowski, *supra* note 74, at 589, 597, 606, 610–11, 613–14, 616, 619.

201. *Id.* at 589, 597, 606, 610, 613–14, 616.

202. *Id.* at 621–22, 630.

203. Lin, *supra* note 20, at 229–30.

204. *Id.* at 230, 234–35.

205. A 2020 social media campaign called #PublishingPaidMe revealed systemic biases and underpayment of Black authors in the publishing industry. The campaign showed that while Black authors dominate current bestseller lists, they receive significantly lower advances than white authors. The campaign attributed these disparities to institutional biases. *See, e.g.*, Constance Grady, *Black Authors Are on All the Bestseller Lists Right Now. But Publishing Doesn't Pay Them Enough*, VOX (June 17, 2020, 10:50 AM), <https://www.vox.com/culture/2020/6/17/21285316/publishing-paid-merdiversity-black-authors-systemic-bias>.

206. *Id.*

207. *Id.*

Similar biases exist in screenplays, where a history of male dominance has created significant gender disparities.²⁰⁸ The same problems have long affected the video game industry.²⁰⁹ Efforts to challenge these biases and promote diversity are ongoing, but to achieve lasting change, the industry needs to address the structural barriers that limit opportunities for women screenwriters.

These historical biases in publishing, film, and video game development provide the backdrop for understanding the dangers AI poses in these creative industries. An AI-powered manuscript selection tool trained mainly on works by white male authors, for instance, may systematically undervalue submissions from a more diverse set of authors. This perpetuates historical biases and creates a feedback loop in which the AI becomes less and less likely to recognize and value work that does not fit its narrow worldview. The case studies in this Article demonstrate how deeply LLMs and generative AI are becoming integrated into publishing, film, libraries, and video game distribution. Without careful intervention, there is a substantial risk that AI tools will distort the creative landscape and further exclude marginalized voices.

Technology companies have a responsibility to combat biases in AI by consciously focusing on training their AI models with diverse and representative datasets. This means companies must actively curate their training data to expose AI systems to a wide range of perspectives. Only by “seeing” the full spectrum of human creativity can AI tools begin to recognize, appreciate, and value works that fall outside the traditional mold.

This is where the doctrine of fair use becomes essential. Without it, companies developing AI systems will have access to only works that they licensed or works already in the public domain. As discussed earlier, training an LLM may entail using hundreds of millions of works. Licensing this number of works would likely be logistically and financially infeasible: a prospective licensee seeking to train an AI would need to identify and then contact every copyright holder and negotiate licensing terms. Moreover, there is no reason to expect that every copyright holder approached in this way would be willing to license. A technology company with the wherewithal to embark on such a quest—likely only the largest of companies—would still have access to far fewer works than they would in a fair use regime. The fewer the works, the more challenging it could

208. See, e.g., Jennifer Langston, *New Tool Quantifies Power Imbalance Between Female and Male Characters in Hollywood Movie Scripts*, UNIV. OF WASH. NEWS (Nov. 13, 2017), <https://www.washington.edu/news/2017/11/13/new-tool-quantifies-power-imbalance-between-female-and-male-characters-in-hollywood-movie-scripts/> (discussing a machine learning tool developed by University of Washington researchers to analyze gender bias in movie scripts that consistently found subtle biases giving male characters more power and agency).

209. See, e.g., Megan Farokhmanesh, *2024 Was the Year the Bottom Fell Out of the Games Industry*, WIRED (Dec. 20, 2024, 3:10 PM), <https://www.wired.com/story/2024-was-the-year-the-bottom-fell-out-of-the-games-industry/> (discussing gender disparities and harassment in the video game industry).

be to prevent bias.²¹⁰ Fair use empowers technologists to curate representative datasets, combat bias, and address historical imbalances in creative industries.

AI offers significant potential to enhance creative expression, but only if we address the dangers of bias. Facilitative Fair Use provides the flexibility needed to include copyrighted works in training datasets that are specifically designed to combat bias and promote diverse perspectives. This Article's proposal is not just helpful; it is urgently needed to help ensure that AI-driven feedback loops will not warp the creative ecosystem.

CONCLUSION

This Article has aimed to address a critical gap in the ongoing debate over copyright law and AI. By foregrounding the issue of knowledge dissemination, this Article offers a useful perspective on how AI influences the creative ecosystem. The empirically grounded case studies reveal that AI is already deeply embedded in the machinery of content distribution. This finding is significant for two reasons. Firstly, it highlights an opportunity for AI to enhance the spread of knowledge, showing that the debate over AI and copyright extends beyond mere mimicry. Secondly, it sounds an alarm that reveals the potential for AI to reinforce historical biases in creative industries.

This Article presents Facilitative Fair Use to address both the opportunity and the risk. This legal framework seeks to expand the fair use doctrine to accommodate AI's role in knowledge dissemination. It would require courts to presume fair use when copyrighted works train AI systems that enhance knowledge dissemination. Importantly, this presumption is subject to rebuttal in cases of market harm. By embracing Facilitative Fair Use, courts can foster a more inclusive and diverse creative landscape while mitigating the risk that AI tools will perpetuate historical biases.

Gutenberg's invention of the printing press revolutionized the dissemination of knowledge, laying the foundation for the Age of Enlightenment. Today, we are witnessing the rise of another transformational technology with the potential to reshape the way we share and acquire knowledge: AI. However, the law must give technology companies enough freedom to steer AI responsibly. Sometimes, the best way to foster creativity is to loosen copyright's grip.

210. See Levendowski, *supra* note 74, at 589 ("Copyright law causes friction that limits access to training data and restricts who can use certain data. This friction is a significant contributor to biased AI. The friction . . . encourages AI creators to use biased, low-friction data (BLFD) for training AI systems, like the word2vec toolkit, despite those demonstrable biases. As Google's decision not to freely release the Google News corpus [that is used to train the word2vec toolkit] reveals, copyright law can also curtail the implementation of bias mitigation techniques Copyright law can even preclude potential competitors from converting the customers of dominant AI players.").