

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION**

ANDREA BARTZ, ANDREA BARTZ, INC.,
CHARLES GRAEBER, KIRK WALLACE
JOHNSON, and MJ + KJ, INC., individually and on
behalf of others similarly situated,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case No. 3:24-CV-05417-WHA

Action Filed: August 19, 2024

**DECLARATION OF STEVEN R.
PETERSON IN SUPPORT OF
DEFENDANT ANTHROPIC PBC'S
ADMINISTRATIVE MOTION FOR
SUMMARY JUDGMENT**

DECLARATION OF STEVEN R. PETERSON

I. QUALIFICATIONS AND ASSIGNMENT

A. Qualifications

1. My name is Steven R. Peterson. I am a Partner at Econic Partners. Econic Partners is an economic consulting firm that specializes in competition and regulation, among other areas. I received my A.B. in economics from the University of California, Davis, and my Ph.D. from Harvard University. While at Harvard, my areas of specialization were economic theory and industrial organization. Industrial organization is the study of competition between large firms that strategically influence the markets in which they operate. I have also served as an adjunct faculty member in the Department of Economics at Northeastern University where I taught courses on the economics of antitrust, regulation, and public policy.
2. I have consulted on the economics of antitrust and competition, mergers, estimation of damages, valuation, regulation, and public policy. I have extensive experience with the policy and economics of copyrights. I testified on behalf of the Radio Music License Committee (“RMLC”) on issues of market definition and market power in *Radio Music License Committee v. SESAC, Inc., SESAC, LLC, and SESAC Holdings, Inc.*, and I testified on public performance license rates for radio in the subsequent arbitration proceedings. I addressed similar issues in RMLC’s litigations against Global Music Rights (“GMR”) and Broadcast Music, Inc. I also testified before the United States Copyright Royalty Judges, Library of Congress, regarding appropriate rates for digital public performances of sound recordings in the *Web IV* (on behalf of the National Association of Broadcasters) and *Web V* (on behalf of Google) proceedings. And I submitted comments on behalf of the National Association of Broadcasters addressing the Department of Justice’s review of the American Society of Composers, Authors and Publishers (“ASCAP”) and Broadcast Music, Inc. (“BMI”) consent decrees. I have previously submitted declarations in *Concord Music Group, Inc., et al. v. Anthropic PBC* in both the Middle District of Tennessee and the Northern District of California addressing the use of copyrighted works to train artificial

intelligence (“AI”) models. A true and correct copy of my curriculum vitae is attached as Exhibit A.

B. Assignment and Materials Used

3. Counsel for Anthropic has asked me to evaluate the effect of Anthropic’s use of certain of Plaintiffs’ copyrighted works to train its large language model (“LLM”) on the value of those specific works. Counsel has also asked me to evaluate whether there is an existing or incipient viable or practicable market to license copyrighted materials to train LLMs. For purposes of my analysis, I assume that Plaintiffs’ books were used by Anthropic for training.
4. To perform my analyses, I reviewed Plaintiffs’ complaint, discovery materials, the deposition transcripts of Andrea Bartz (individually and as a 30(b)(6) witness for Andrea Bartz, Inc.), Charles Graeber, and Kirk Wallace Johnson (individually and as a 30(b)(6) witness for MJ + KJ, Inc.) (collectively “Named Plaintiffs”), and publicly available information on AI models, book publishing, and relevant matters. A list of the materials relied upon in the preparation of my report is attached as Exhibit B.

II. SUMMARY OF OPINIONS

5. Anthropic uses Plaintiffs’ books to train and create the LLM Claude.¹ Claude does not substitute for or compete with Plaintiffs’ books.
 - a. Anthropic uses Plaintiffs’ books to create a training corpus used to train an LLM. Training entails exposing computer software, a neural network, to text so it can learn the relationships between words and understand how language works.
 - b. Anthropic’s customers use Claude for a variety of purposes, examples of which include speeding drug development, improving the European Parliament’s access to documents with advanced search capabilities, and assisting with whale conservation efforts by analyzing acoustic data.² Claude can also write computer code, respond

¹ Anthropic has produced multiple versions of Claude. Throughout my Declaration, I refer to these versions collectively as Claude unless otherwise noted.

² Declaration of Jared Kaplan in Support of Anthropic’s Motion for Summary Judgement (“Kaplan Declaration”), ¶¶ 25-28.

to inquiries in a chat format, edit, analyze and summarize documents, analyze large volumes of data, draft original paragraphs or longer texts in response to prompts, and provide customer service.

c. Claude's *services* are not substitutes for Plaintiffs' books – that is, a small increase in the price of a Plaintiffs' book or all books would not lead readers to turn to Claude's services as a substitute. In fact, Claude is restricted from reproducing large portions of a book.³ Instead, in response to a price increase for Plaintiffs' books, readers would visit a library to borrow a book, or read magazines or the internet – all of which are much closer substitutes for new books than Claude's interactive services.

d. Claude's services are sold in different “product” markets than Plaintiffs' books. Even editing and copywriting services enabled by Claude are not market substitutes for Plaintiffs' books.

6. Writing books is a highly competitive industry with millions of titles published annually in the United States. Any effect of Claude's services in assisting authors' creation of new books on the value of existing books used to train Claude is speculative.

7. There is not an incipient market to license text to train LLMs. A competitive market for text to train LLMs is impracticable – that is, a competitive market cannot mediate the transactions necessary for an AI firm to license the volume of textual data necessary to train a highly capable model. This conclusion follows from the following subordinate conclusions.

a. AI firms would need millions of licenses just to cover the books owned by Plaintiffs' proposed class of copyright owners. Licensing the full training corpus necessary for an LLM with Claude's capabilities would also require licensing content on millions of internet domains and the number of licenses would rise still higher to the extent individual writers own the training rights to their blogs and other writing in subdomains.

³ Kaplan Declaration, ¶¶ 34, 66-68.

- b. The hypothetical market to license training data is characterized by incomplete information on copyright ownership and the fair market value of a license. Incomplete information raises the cost of identifying counterparties and verifying the rights that they claim to own. Administration of licenses and negotiations over license fees and terms raise transaction costs further.
- c. An AI firm would pay at most the increase in the value of its model resulting from incremental training data. This amount is small relative to the transaction costs of licensing additional data, in part because any individual book represents a tiny fraction of the overall data required to train an LLM. The low incremental value and relatively high transaction cost of licensing copyrighted works for use in training will lead to market failure because a market will not bring about transactions with a gross value that is lower than the cost of the transaction for the parties.
8. Agreements between AI firms and content owners publicized in the press do not support the conclusion that there is a viable market for training rights. Licensing some rights, if that has occurred, does not demonstrate the viability of licensing a suitable – and sufficiently large – training corpus in a market with incomplete information and other high transaction costs.
9. If licensing were required to use copyrighted text for LLM training, an important benefit of licensing would be avoiding copyright liability. However, in a fragmented licensing market of huge scale and incomplete information on the ownership of copyrights, an AI firm that licensed assiduously would still face substantial copyright liability. Licensing only a subset of the text required for a training corpus is costly but would still leave the LLM developer exposed to infringement risk from the copyright owners of any unlicensed works in the training corpus. Unless there is an existing or potential licensing market that can efficiently mediate the high number of license transactions required to assemble a sufficient corpus to train an LLM like Claude, the result will be market failure.

III. ANTHROPIC AND THE CLAUDE MODEL

10. Anthropic is an AI company that develops AI models known as LLMs.⁴ Their commercial LLM models are called “Claude.” Anthropic describes Claude as a “frontier” AI model, capable of providing end users with a variety of interactive services through “advanced reasoning,” “vision analysis,” “code generation,” and “multilingual processing,” among other features.⁵ Among other uses, customers use Claude to analyze acoustic information to accelerate whale conservation, improve the European Parliament’s access to government records through advanced search, speed clinical trials by reducing the time needed to produce reports, and to generate curriculum plans, quizzes, and behavior support to reduce administrative time and allow teachers to spend more time with students.⁶ According to Anthropic, “Claude tends to perform well at general, open-ended conversation; search, writing, editing, outlining, and summarizing text; coding; and providing helpful advice about a broad range of subjects.”⁷

11. Enterprise customers constitute a significant portion of Claude usage.⁸ Enterprise use of Claude involves coding and making business functions, such as accounting and data analysis, more efficient.⁹ Enterprise customers’ uses of Claude have little to do with Plaintiffs’ books.

⁴ See <https://www.anthropic.com/company> (accessed 3/18/2025) (“Anthropic is an AI safety and research company.”) and (“Anthropic is a collaborative team of researchers, engineers, policy experts, business leaders and operators, who bring our experience from many different domains to our work.”).

⁵ See <https://www.anthropic.com/claude> (accessed 3/18/2025).

⁶ Kaplan Declaration, ¶¶ 25-29.

⁷ “Public Comments of Anthropic PBC,” Before the United States Copyright Office, Notification of Inquiry Regarding Artificial Intelligence and Copyright, October 30, 2023 (“Anthropic Comments”), p. 2; See also Anthropic Comments, pp. 3-4, describing examples of productivity-enhancing uses of Claude since its launch.

⁸ Cristina Criddle, “Anthropic set to focus on business users in search for new revenues,” *Financial Times*, March 18, 2025, available at <https://www.ft.com/content/97e0ab06-8d83-4918-9079-3ed935bc1c63> (accessed 3/25/2025).

⁹ See, e.g., “Campfire accelerates accounting with Claude,” available at <https://www.anthropic.com/customers/campfire> (accessed 3/25/2025).

12. To better understand how individual, non-enterprise users interact with its AI models, Anthropic launched an analysis tool called Clio (Claude insights and observations). According to the results of a Clio study, published along with Clio’s launch announcement in December 2024, claude.ai users interact with Claude in a variety of ways, with the top use case, “Web & Mobile App Development,” accounting for just 10.4% of use cases.¹⁰ In addition, further analysis of how Claude is used reveals that users often rely on Claude to edit, rewrite, and brainstorm on original writing projects.¹¹ Notably, despite the variety of top use cases, each of the use cases that Anthropic describes is *generative* – that is, Claude is not designed to simply regurgitate text or facts, but to use its “knowledge” to help users more efficiently access, understand, create, and convey information.¹²
13. Like all LLMs, Claude processes large amounts of text to “understand” how words and portions of words work together and uses that “learning” to produce original output in response to a user-generated prompt. Claude is thus able to perform a range of assistive tasks.¹³
14. To perform the necessary textual processing to train an LLM, engineers first create a “neural network,” a type of software that analyzes large datasets and extracts statistical information from language to better understand how language works.¹⁴ LLMs “learn” how to recognize, interpret, and generate language through a “training” process that enables

¹⁰ See <https://www.anthropic.com/research/cliio> (accessed 3/18/2025). Other top use cases include “Content Creation & Communication” (9.2% of use cases), “Academic Research and Writing” (7.2% of use cases), “Education & Career Development” (7.1% of use cases), “Advanced AI/ML Applications” (6.0% of use cases), “Business Strategy & Operations” (5.7% of use cases), “Language Translation” (4.5% of use cases), “DevOps & Cloud Infrastructure” (3.9% of use cases), “Digital Marketing & SEO” (3.7% of use cases), and “Data Analysis & Visualization” (3.5% of use cases).

¹¹ Kaplan Declaration, ¶ 23.

¹² Kaplan Declaration, ¶¶ 30-34.

¹³ See Kaplan Declaration, ¶¶ 20-29.

¹⁴ Kaplan Declaration, ¶ 35.

them to identify relationships and patterns between words.¹⁵ Claude is trained on a mix of datasets, most of which have been sourced from the Internet.¹⁶ Training datasets are extremely large because the goal of training is for the model to identify patterns and relationships between words across a very wide variety of contexts. Therefore, effective training depends on the volume and diversity of text.¹⁷ Testing shows that LLMs trained on larger volumes of data were more adept at generative tasks (such as creating new expressions, reasoning, and analysis).¹⁸

15. Anthropic estimates that the Claude models known as Claude 1, Claude 2, Claude 3, Claude 3.5, and Claude 3.7 were trained on datasets of billions of texts. Each successive version of Claude has been trained on a larger corpus. Claude 3.7 was trained on a corpus of [REDACTED] tokens.¹⁹ Tokens are basic semantic units that can correspond to words, parts of words, or characters.²⁰ Different tokenizers will tokenize the same sentence into different numbers of tokens, but using the most recent version of Anthropic's tokenizer, it takes approximately 1.3 tokens to represent a single English-language word.²¹

IV. THE ECONOMIC ANALYSIS OF FACTORS DETERMINING FAIR USE

16. I understand fair use is a legal doctrine that allows for the non-infringing use of copyrighted original works based on four factors set out in Section 107 of the Copyright Act and defined further by courts. The economic justification for the fair-use limitation on copyright monopolies is to allow use of copyrighted works in circumstances where overly broad

¹⁵ Kaplan Declaration, ¶ 36; Anthropic Comments, p. 6.

¹⁶ Anthropic Comments, p. 5.

¹⁷ Kaplan Declaration, ¶¶ 40-42.

¹⁸ “Tracing Model Outputs to the Training Data,” *Anthropic*, August 8, 2023 (“[P]atterns of generalization became more abstract with model scale.”) available at <https://www.anthropic.com/index/influence-functions> (accessed 3/19/2025).

¹⁹ Kaplan Declaration, ¶¶ 46, 53. The training corpus for Claude 3.7 contained roughly [REDACTED] base tokens that were sampled to train Claude.

²⁰ Kaplan Declaration, ¶ 39.

²¹ Kaplan Declaration, ¶ 39.

protections would not support the goals of “promot[ing] progress of science and the useful arts” that copyright protections are intended to advance.²²

17. The four factors set forth in the Copyright Act are:

- i. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
- ii. the nature of the copyrighted work;
- iii. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- iv. the effect of the use upon the potential market for or value of the copyrighted work.²³

18. My economic analysis primarily addresses the fourth factor concerning the effect of Anthropic’s asserted fair use on the “potential market for or value of the copyrighted work.”²⁴ There are two groups of markets in which Plaintiffs may assert that Anthropic’s use of their works to create Claude could reduce their value.

- a. First, consumers buy books (in various formats) to read them. Others may license the right to create derivative works that incorporate expressive elements of books. Plaintiffs may assert that Claude itself or the services that Claude provides to users could substitute for Plaintiffs’ books in the markets in which they are sold.
- b. Second, Plaintiffs may assert that the uncompensated use of Plaintiffs’ works to create Claude’s training data corpus could interfere with the development of an incipient market to license copyrighted works to train LLMs.

19. The impact of Claude on the value of Plaintiffs’ books in the first group of markets, sometimes described as “traditional” or “established” markets for the asserted works,

²² Pierre N. Leval, “Toward a Fair Use Standard,” *Harvard Law Review*, Vol 103, p. 1109 (“Notwithstanding the need for monopoly protection of intellectual creators to stimulate creativity and authorship, excessively broad protection would stifle, rather than advance, the objective.”).

²³ Copyright Act, Section 107.

²⁴ Copyright Act, Section 107.

depends on the degree to which Claude itself, and Claude's outputs in response to users' prompts, compete with sales or licenses of Plaintiffs' copyrighted books. Economically transformative uses are less likely to substitute for Plaintiffs' traditional or established markets for a copyrighted work and, therefore, are less likely to reduce their value.

20. The key economic issue that informs whether Claude interferes with the "potential" market to license works to train LLMs is whether there is plausibly a viable or practicable market to license books and other copyrighted works for the purpose of training LLMs. The lack of an established market for the use in question indicates that the use is new and likely transformative.

V. USING PLAINTIFFS' COPYRIGHTED WORKS TO TRAIN CLAUDE DOES NOT SUBSTITUTE FOR OR COMPETE WITH PLAINTIFFS' BOOKS

21. The fourth fair-use factor addresses whether creating Claude using Plaintiffs' copyrighted books diminishes their value in the markets in which they are sold. Anthropic's use of Plaintiffs' works would diminish their value if consumers who buy books view Claude as a substitute for Plaintiffs' works. Claude is not a substitute for books. Claude is designed to understand and generate human language, analyze documents, and assist with a wide range of tasks. And as noted, customers use Claude to write code, speed drug trials, make accounting more efficient, contribute to conservation efforts, as well as for writing and editing tasks. Books do not provide these services and are not a substitute for Claude, and Claude is not a substitute for books in any market, including the markets in which books are sold.

22. Put differently, training Claude using Plaintiffs' works has created a computer program that provides generative artificial intelligence services, not a book for consumers' reading pleasure, a screenplay adaptation, or any other existing use for books. The program and its services are different from books and do not substitute for or compete with books in any properly defined market for those works. In short, Claude is not a book. And Claude cannot be used as a book can.

23. Even if some authors may use Claude or other LLMs to help them write books, that fact does not make Claude a substitute for a book. If Claude is used to generate additional books, that increase in artistic output benefits consumers and is procompetitive. It is also unlikely to have more than a de minimis effect on any particular book's sales.

24. Below, I briefly describe the economics of competition and substitution before examining whether Claude or Claude's outputs compete with Plaintiffs' books or books generally.

A. The Economics of Competition and Substitution

25. In the "traditional" markets where Plaintiffs may contend Claude lowers the value of their works, the economic issue is substitution. Competition between two products occurs when consumers view them as substitutes. Two goods are substitutes when the increase in price of one leads to an increase in consumers' purchases of the other – that is, an increase in price leads consumers to substitute the good that did not experience a price increase for the now more costly good. The degree of consumer substitution determines the degree to which the goods compete.

26. The degree to which consumers are willing to substitute one product for another is the primary source of competition among sellers. A seller cannot profitably charge more than a seller of a similar good if many consumers are willing to shift their purchases to the cheaper good should a seller raise its price. It is consumers' willingness and ability to substitute one good for another that turns sellers into competitors.

27. Competition economics has methods of determining when products are sufficiently close substitutes to justify treating them as being in the same market. Products belong in the same market when a modest price increase – typically 5% to 10% – for one product would lead a substantial number of consumers to switch to a similar alternative product that does not experience a price increase.²⁵ If the substitution between products in the face of such a price increase is limited or nonexistent, the products are not in the same market.

²⁵ The threshold for establishing the boundary of the market is based on whether the price increase is profitable to the manufacturer of the product. If a large number of buyers substitute to other products, the price increase is not profitable and the market must be expanded. *See* U.S. Department of Justice and Federal Trade Commission, "Merger Guidelines," December 18, 2023,

28. For example, small Toyota sedans and small Honda sedans are in the same market with other small sedans because the products are similar, and an increase in the price of the Toyota would lead a reasonable number of customers to purchase a Honda or another manufacturer's small sedan. In short, a small Honda sedan is one of the good substitutes for a small Toyota sedan.

29. An increase in the price of all small sedans, however, would lead few consumers to buy a full-size pickup truck. A full-size pickup truck has many of the same uses as a small sedan. But a modest change in the price of small sedans would be more likely to lead small-sedan buyers to purchase a mid-size sedan or a small crossover vehicle than a much larger full-size pickup truck, which has features the typical driver of a small sedan does not value (like towing and storage capacity) while underperforming small sedans in ways that matter to sedan drivers (such as gas mileage). In this example, small sedans and full-size pickup trucks are different products that are sold in different markets because consumers do not consider them to be good substitutes.

30. I understand that in the case of copyright, a relevant issue is whether an asserted fair use of a copyrighted book, for example, is a substitute for the expressive content of that particular book.²⁶ As described below, Claude and its outputs do not substitute for or diminish the value of books generally or any book in particular.

Section 4.3, available at https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf (accessed 3/25/2025).

²⁶ See *Authors Guild v. Google, Inc.*, 804 F.3d 202, 224-25 (2d Cir. 2015) (holding that “the possibility, or even the probability or certainty, of some loss of sales does not suffice to make the copy an effectively competing substitute that would tilt the weighty fourth factor in favor of the rights holder in the original” and finding a lack of fourth-factor harm where “the type of loss of sale” was likely to occur “in relation to interests that are not protected by the copyright,” as where the defendant’s use “satisf[ied] the searcher’s need for the book” by providing a historical fact rather than authorial expression).

B. Claude and the Services It Provides Do Not Substitute for Books

31. Anthropic’s use of Plaintiffs’ books to train Claude does not substitute for or compete with Plaintiffs’ books. The training process does not expose the Plaintiffs’ books to consumers and cannot substitute for or compete with Plaintiffs’ books in consumer-facing markets.

32. Nor do Claude’s generative artificial intelligence services compete with Plaintiffs’ books or any book. No book can perform these interactive, generative services, and they are not a substitute for reading a book. If one of Plaintiffs’ books (or any book) were to increase modestly in price, a consumer shopping for a book would not choose to subscribe to Claude as a substitute. A much closer substitute to the now more expensive book is another book in the same or a similar genre, a book from the library or borrowed from a friend, or some other type of reading material, such as magazines or the Internet.²⁷ Similarly, a small increase in the price of Claude would not lead consumers using Claude’s generative AI services to buy and read a book as a substitute.

33. The differences between books and the services Claude provides lead to the conclusion that any book and Claude are different “products” – that is, an LLM like Claude is not a substitute for any book or collection of books. This gives a clear answer regarding the fourth factor: Claude’s services do not reduce the value of Plaintiffs’ works through substitution in their traditional markets.

34. This economic conclusion also informs the analysis of whether Anthropic’s use of Plaintiffs’ books (*i.e.*, the creation of Claude) is transformative. As a matter of economics, Claude’s services are sufficiently different in their characteristics from Plaintiffs’ books that those books have been sufficiently transformed to create new products that are used differently and for different purposes than books.

²⁷ Non-infringing competition between works in the same genre, style, or format, is consistent with the economic goal of incentivizing the creation of new works for the public’s benefit through a limited copyright monopoly. Moreover, such competition occurs regularly in the markets for books of various types and genres.

C. Claude Users' Output Does Not Compete Directly with Plaintiffs' Copyrights

35. One of the services that Claude provides is writing assistance in response to user prompts.²⁸

Plaintiffs assert that books created by Claude users compete with their copyrighted books used to train Claude.²⁹

36. Critically, Plaintiffs do not claim that Claude output infringes the copyrights covering Plaintiffs' books.³⁰ The outputs prompted by Claude users therefore do not directly usurp Plaintiffs' monopolies over the original writing and expression in their books, including Plaintiffs' rights to create derivative works.

37. Authors looking to write a profitable book are always free to write a non-infringing book that incorporates similar ideas to an existing book in an effort to, for example, create a superior work or to profit from a trend in the popularity of a particular type or style of book. Authors cannot copyright ideas.³¹ Thus, the writer of a spy novel, for example, is always subject to potential competition from another author's non-infringing spy novel.

38. In fact, authors' efforts to reach readers have led to roughly 3 million books being published in the United States each year with the majority of those being self-published. In 2023, over 2.6 million of the 3 million books published in the United States self-published, the fifth

²⁸ Kaplan Declaration ¶ 23.

²⁹ First Amended Class Action Complaint, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, December 4, 2024 ("Complaint"), ¶ 53 ("As LLMs have become more advanced—and enabled to train on more and more copyrighted material—they are able to generate more content and more sophisticated content. The result is that it is easier than ever to generate rip-offs of copyrighted books that compete with the original, or at a minimum dilute the market for the original copyrighted work.").

³⁰ Complaint, ¶ 63.

³¹ 17 U.S.C. § 102(b) (2018) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.").

consecutive year with more than 2 million self-published books in the United States.³² In 2005, roughly 280,000 books were published in the United States, indicating that the number of books published has grown by a factor of 10 since 2005.³³ This rapid growth in book publishing occurred prior to the public availability of LLMs in November 2022.³⁴ Of course, the three million books published each year have to compete for attention with the stock of books published in prior years, including prior years' best sellers and classic titles.

39. Even prior to LLMs, more books were published each year than could capture the attention of any appreciable number of readers. Most books are not stocked in bookstores and are sold as e-books or are printed on demand. Circana estimates that 767 million physical books were sold in 2023, which indicates the average physical sales of each title is quite low.³⁵

40. The distribution of book sales is highly skewed by title. A 2022 New York Times article explained that "In 2021, fewer than one percent of the 3.2 million titles that BookScan tracked sold more than 5,000 copies," and that "Penguin Random House executives said that just 35 percent of books the company publishes are profitable. Among the titles that make money, a very small sliver – just 4 percent – account for 60 percent of those

³² Andrew Albanese and Jim Milliot, "Self-Publishing's Output and Influence Continue to Grow," *Publishers Weekly*, November 8, 2024, available at <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96468-self-publishing-s-output-and-influence-continue-to-grow.html> (accessed 3/17/2025); Jim Milliot, "Self-Publishing Is Thriving, According to Bowker Report," *Publishers Weekly*, February 17, 2023, available at <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/91574-self-publishing-is-thriving-according-to-bowker-report.html> (accessed 3/17/2025).

³³ "Title Output Up," *Publishers Weekly*, June 18, 2007, available at <https://www.publishersweekly.com/pw/print/20070618/5223-title-output-up.html> (accessed 3/17/2025).

³⁴ Bernard Marr, "A short History of ChatGPT: How We Got to Where We Are Today," *Forbes*, May 19, 2023, available at <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/> (accessed 3/19/2025).

³⁵ "Adult Fiction Outperforms the U.S. Book Market in 2023, Circana Reports," *Circana*, February 2, 2024, available at <https://www.circana.com/intelligence/press-releases/2024/adult-fiction-outperforms-the-u-s-book-market-in-2023-circana-reports/> (accessed 3/19/2025).

profits.”³⁶ While many self-published books have low sales, self-published books in the aggregate have billions of dollars in sales.³⁷ Thus, the book market is characterized by top writers whose books have high sales and a long tail of book titles with almost no sales. The implication is that competition among authors vigorously exploits the available profitable opportunities.

41. The possibility that Claude will help authors inject further competition into a market that is already highly competitive is a socially beneficial increase in artistic output. That said, with about 3 million book titles being published each year in the United States before LLMs such as Claude were available, I see no reason to conclude that the introduction of Claude or other LLMs that can assist authors will have anything other than a trivial impact on the value of any of Plaintiffs’ copyrights.

D. Anthropic’s Use of Plaintiffs’ Works to Create Claude Has Produced a New Product and Is Economically Transformative

42. I understand that courts have interpreted the first factor of the fair-use analysis to depend on whether the use in question is “transformative.”³⁸ A use is “transformative” if it alters the original work’s purpose or character.³⁹ Courts have further reasoned that if a use seeks to achieve the same purpose as the original, the resulting work or product will tend to

³⁶ Elizabeth Harris, Alexandra Alter, and Adam Bednar, “A Trial Put Publishing’s Inner Workings on Display. What Did We Learn?,” *The New York Times*, August 19, 2022, available at <https://www.nytimes.com/2022/08/19/books/prh-penguin-random-house-trial.html> (accessed 3/18/2025).

³⁷ Andrew Albanese and Jim Milliot, “Self-Publishing’s Output and Influence Continue to Grow,” *Publishers Weekly*, November 8, 2024, available at <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96468-self-publishing-s-output-and-influence-continue-to-grow.html> (accessed 3/17/2025) (“...[I]ndustry observers estimate that annual sales of self-published is in the billions of dollars. Amazon, for example, maintains that in the 10 years Kindle Unlimited has been active, [Kindle Direct Publishing] authors have earned more than \$3.5 billion in royalties[.]”).

³⁸ *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 595 (2023).

³⁹ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).

substitute for or compete with the original.⁴⁰ Conversely, a use that serves a different purpose than the original work is less likely to compete with the original.

43. Thus, whether a use is transformative depends in part on the degree of substitution between the original copyrighted work and the product resulting from an asserted fair use.⁴¹ With this definition, the courts have added an economic element to the first factor and linked the first and fourth factors from an economic perspective. Thus, my analysis of the fourth fair-use factor also informs the question of whether using Plaintiffs works to create Claude is transformative.

44. In short, the economic interpretation of transformativeness is that, if an asserted fair use of an original work leads to secondary works, products, or services that are not substitutes for the original, the use has transformed the original into new “products” that are sold in different markets than the original. As products sold in separate markets, transformative new products have little to no effect on the value of existing original works in the markets for those pre-existing works. Claude is sold in separate markets than the books used to train the model, and Claude has little to no effect on the value of books in the traditional markets for those works. These factors demonstrate that Claude is economically transformative.

VI. THERE IS NOT AN INCIPIENT MARKET TO LICENSE COPYRIGHTED TEXT TO TRAIN LLMS

45. Plaintiffs allege that Anthropic’s asserted fair use of Plaintiffs’ copyrights interferes with the development of a market to license text to train LLMs. If the asserted fair use made it more difficult for Plaintiffs to license their works to other LLM developers in an incipient and economically coherent market for such rights, their work’s value could be diminished

⁴⁰ *Warhol*, 598 U.S. at 528 (“In that way, the first factor relates to the problem of substitution—copyright’s *bête noire*. The use of an original work to achieve a purpose that is the same as, or highly similar to, that of the original work is more likely to substitute for, or supplan[t], the work.” (internal quotation marks and citation omitted)).

⁴¹ *See, e.g., Warhol*, 598 U.S. at 528-29; *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994) at 591.

from forgone licensing revenue. However, as described below, AI firms could not contact and license from all of the licensors necessary to build a useful training corpus.

46. The question of whether there is an incipient market to license a copyrighted work for a particular use depends economically on whether a market to transfer the required rights is practicable. For a market to develop and for most or all advantageous transactions in it to be realized, parties need good and freely available information on who owns the rights to be licensed and the value of those rights. In addition, transaction costs need to be low relative to the value of the rights being licensed. If the market is too costly and inefficient to allow for both viable buyers and viable sellers, it will not exist.

47. A viable licensing market for AI training data needs to be economically coherent.

Coherence implies that the licensing market needs to sustain both the licensors and the licensees – that is, an AI firm licensing diligently and in good faith in the hypothetical competitive licensing market where licenses for AI training are required should be legally and financially viable. If AI firms cannot license sufficient data to operate profitably without significant potential copyright liability, they would not be present to license training rights in the first place. For this reason, claims that an incipient market for licensing training data exists because it would be possible to license some data, such as large collections of books, song lyrics, or news articles, are reductive – that is, such claims ignore the negative implications of the remaining unlicensed training data for the licensee.

48. Put another way, the relevant question is not whether AI firms could license some books or other texts in a licensing market, but whether they could license all the copyrighted works necessary to assemble data sufficient to build current and future models.

49. A competitive market would not be capable of mediating the necessary transactions if AI firms were legally required to license copyrighted training data. To mediate the tens of millions of transactions required to license the diversity of text data needed to train current and future cutting-edge LLMs, AI firms would need readily available, reliable information on who owns training rights to particular texts. Moreover, other transaction costs, such as negotiation and administration of rights and payments, would need to be low because of the

huge scale of the licensing effort. In this case, the hypothetical market to license AI training data has none of the characteristics needed for it to function.⁴²

50. Currently, no practicable market exists to license rights for training LLMs. The evidence of transactions between LLM developers and copyright owners related to training data is limited, with no indication that any substantial amount of text relative to the size of a training corpus has been licensed. Anthropic approached publishers in 2024 to investigate licensing books for training and found that at most small portions of publishers' catalogs were available to license,⁴³ which is consistent with Named Plaintiffs' testimony that

44

No AI firm could expect to license a substantial number of books for its training corpus in a market with highly dispersed rights and high transaction costs.

51. Rights owners naturally want to be paid for both transformative and non-transformative uses of their works, which means they will argue that there could, in some scenario, be a licensing market for an economically transformative use like AI training. This argument should be resisted when no such market is observed to avoid rights holders expanding their rights even when no licensing market is discouraged by fair use.

52. Conferring copyright protection in a case where a transformative use and other factors indicate it does not interfere with traditional markets for copyrights harms consumer welfare. In such a case, copyright owners are unaffected in their traditional markets by an

⁴² I have not found evidence of a frontier LLM, *i.e.*, an LLM comparable to Claude that was trained exclusively on licensed or non-copyrighted texts. Pending lawsuits against other LLM developers like Google, Meta, OpenAI, NVIDIA, Databricks, and Microsoft allege that each of them has engaged in infringement based on the decision to train on unlicensed copyrighted works. *See, e.g., In re Google AI Generative Copyright Litig.*, Case No. 23-cv-3440 (N.D. Cal.); *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417 (N.D. Cal.); *In re OpenAI ChatGPT Litig.*, Case No. 23-cv-03223 (N.D. Cal.); *Authors Guild v. OpenAI, Inc.*, Case No. 23-cv-08292 (S.D.N.Y.); *Dubus v. NVIDIA Corp.*, Case No. 24-cv-02655 (N.D. Cal.); *In re Mosaic LLM Litig.*, Case No. 24-cv-01451 (N.D. Cal.); *New York Times v. Microsoft Corp.*, Case No. 23-cv-11195 (S.D.N.Y.).

⁴³ Declaration of Tom Turvey in Support of Anthropic's Motion for Summary Judgement ("Turvey Declaration") ¶ 31.

⁴⁴ *See e.g., Ex. 2 (Bartz Dep. Tr.)* at 168:21-172:4, 176:18-182:14; *Ex. 1 (Graeber Dep. Tr.)* at 136:3-25; *Ex. 3 (Johnson Dep. Tr.)* at 142:19-146:3.

innovator's use of their works, but the development of useful products that consumers value is discouraged. In short, the unwarranted expansion of the copyright monopoly harms the public by forcing an innovator to pay unnecessarily for access to information.⁴⁵

53. Below, I address in more detail how information on copyright ownership, transaction costs, and other factors indicate that there is no incipient market capable of mediating the transactions necessary for AI firms to license the content they need to train current and future cutting-edge LLMs.

54. Suggestions that collective licensing could arise and resolve the deficiencies of the licensing market are effectively admissions that the licensing market AI firms face would not allow them to license the texts they need under existing conditions. Of course, a licensing collective would be a cartel of authors. Such a cartel would expand the monopoly grants of the authors and could raise licensing rates above competitive levels.

A. Markets Are Costly to Set Up and to Use

55. Markets are costly to set up and to use. Centralized markets, such as the stock market or a commodity exchange require traders, market makers, record keepers, technology, real estate to house employees, and so forth.⁴⁶ These costs preclude the establishment of many

⁴⁵ William M. Landes and Richard A. Posner, "An Economic Analysis of Copyright Law," *The Journal of Legal Studies*, Vol. 18, No. 2. June, 1989, p. 335.

⁴⁶ Dennis Carlton, "The Theory of Allocation and Its Implications for Marketing and Industrial Structure: Why Rationing is Efficient," *The Journal of Law & Economics*, October 1991, 34(2):231-262, p. 233 ("One can learn a great deal about the cost of a price system by studying institutions whose business is to create such markets. For example, if one studies the Chicago Board of Trade, one observes the following: (1) a large building in an expensive part of town; (2) many people involved in each transaction (for example, brokers, floor traders, employees of a clearinghouse, consultants); (3) office buildings housing the people involved in each transaction; and (4) elaborate record keeping. Undoubtedly the greatest cost is the time cost of all the people involved. A significant fraction of the economy of the city of Chicago is devoted to the making of markets that clear by price. If a magic spell could be cast to make transactions costless, the Chicago economy would be devastated, at least in the short run. This emphasizes how far from costless the making of markets really is.").

potential markets.⁴⁷ For markets that are not explicitly organized, market participants may expend substantial resources identifying counterparties, and when the details of transactions are not standardized, may expend time and resources negotiating prices and terms. If the value of a transaction is less than the cost of negotiating and consummating it, parties will not have an interest in seeking each other out, and transactions will not occur.⁴⁸

B. The Hypothetical Market to License LLM Training Data

1. Training Requires Huge Amounts of Text, Making Licensing Infeasible

56. Cutting-edge LLMs require huge amounts of text for training. Depending on the dispersion of rights to license copyrighted text for training, obtaining a training corpus would require millions and likely tens of millions of license transactions. The scale of the needed licensing effort results from the size of training datasets. For example, Claude 3.7 Sonnet was trained on approximately [REDACTED] tokens, from a base corpus of [REDACTED] tokens representing roughly [REDACTED] words.⁴⁹ AI firms continue to search for more data to train the next generations of LLMs.

57. Few people have experience with very large numbers, so it is difficult to understand how large the number one trillion is. In terms of books, Google estimated in 2010 that there

⁴⁷ Dennis Carlton, “Contracts, Price Rigidity, and Market Equilibrium,” *Journal of Political Economy*, October 1979, 87(5):1034-1062, p. 1035. Bernard Salanié, *Microeconomics of Market Failures*, The MIT Press, 2000, Chapter 13.

⁴⁸ Dennis Carlton, “The Theory of Allocation and Its Implications for Marketing and Industrial Structure: Why Rationing is Efficient,” *The Journal of Law & Economics*, October 1991, 34(2):231-262, p. 233 (“Aside from the fact that there are costs to setting up such markets, it is also true that many markets, once created, fail. Roughly one-half of all successful futures markets fail within ten years of their introduction. Since these markets clearly produce at least some benefits to their creators, the high failure rate must indicate the presence of significant costs of operation that outweigh these benefits. Furthermore, only a small fraction of products have ever had futures markets. Again, this emphasizes that there must be significant costs to the creation of these markets, otherwise the benefits flowing from them would suggest a proliferation of such markets.”).

⁴⁹ Kaplan Declaration, ¶ 39.

1 were approximately 130 million books in the world.⁵⁰ If each book had on average 100,000
2 words, which is roughly the number of words in a modern novel,⁵¹ all the books in the
3 world contained 13 trillion words in 2010, [REDACTED] the corpus used to train the latest
4 release of Claude.

5 58. To further illustrate the volume of licenses that would be required for training data used in
6 AI training, it's worth considering the Common Crawl Archive. This is just one (among
7 many⁵²) of the datasets commonly used "for training natural language processing (NLP)
8 models" including Claude.⁵³ The most recent Common Crawl Archive "contains 2.6 billion
9 web pages (or 402 TiB of uncompressed content). Page captures are from 47.6 million
10 hosts or 38.5 million registered domains and include 1 billion new URLs."⁵⁴ In total, the

11
12 ⁵⁰ Leonid Taycher, "Books of the world, stand up and be counted! All 129,864,880 of you,"
13 *Google Book Search*, August 5, 2010, available at
14 <https://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>
15 (accessed 19/2025) (Discussing the difficulty of identifying unique books, "One definition of a
16 book we find helpful inside Google when handling book metadata is a 'tome,' an idealized bound
17 volume. A tome can have millions of copies (e.g., a particular edition of 'Angels and Demons' by
18 Dan Brown) or can exist in just one or two copies (such as an obscure master's thesis languishing
19 in a university library). This is a convenient definition to work with, but it has drawbacks. For
20 example, we count hardcover and paperback books produced from the same text twice, but treat
21 several pamphlets bound together by a library as a single book.").

22 ⁵¹ "How Many Words in a Novel (Updated for 2025)," *Reedsy Studio*, September 19, 2024,
23 available at <https://reedsy.com/studio/resources/how-many-words-in-a-novel> (accessed
24 3/26/2025). [REDACTED] (Deposition of Andrea Bartz,
25 *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others*
26 *similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California,
27 San Francisco Division, 3:24-cv-05417, March 7, 2025 ("Ex. 2 (Bartz Dep. Tr.)") at 49:1-8 ("[REDACTED]
28 [REDACTED]").

29 ⁵² See e.g. Kaplan Declaration, ¶ 53.

30 ⁵³ See Harsh Singhal, "Every Data Professional Should Know About the Common Crawl
31 Project," *DataScienceFM*, December 22, 2022, available at [https://datascience.fm/every-data-
32 professional-should-know-about-the-common-crawl-project/](https://datascience.fm/every-data-professional-should-know-about-the-common-crawl-project/) (accessed 3/18/2025); Kaplan
33 Declaration, ¶ 53.

34 ⁵⁴ Thom Vaughan, "February 2025 Crawl Archive Now Available," *Common Crawl*,
35 February 23, 2025, available at [https://commoncrawl.org/blog/february-2025-crawl-archive-now-
36 available](https://commoncrawl.org/blog/february-2025-crawl-archive-now-available) (accessed 3/18/2025).

full Common Crawl dataset contains “over 250 billion webpages” and is “essentially a periodic copy of the entire Internet” spanning the past “18 years.”⁵⁵

59. As these numbers readily illustrate, even when copyright owners can be readily identified and contacted, the volume of licenses required makes mass licensing of copyrights impracticable. The U.S. Copyright Office agrees that licensing for mass digitization “is essentially impossible, not necessarily because of the lack of identifying information or the inability to contact the copyright owner, but because of the sheer number of individual permissions required.”⁵⁶

2. AI Firms Would Not Be Able to Identify Many Rights Holders

60. Experience shows that the scale of the licensing effort needed to license a current training corpus makes the effort “essentially impossible.”⁵⁷ Beyond the scale of the licensing effort, many rights holders are unknown or cannot be found with reasonable effort. Named

Plaintiffs testified that

⁵⁸ Each author would need to be

identified. Moreover,

⁵⁹

⁵⁵ Kaplan Declaration, ¶ 48.

⁵⁶ “Orphan Works and Mass Digitization,” *United States Copyright Office*, June 2015, p. 1, available at <https://www.copyright.gov/orphan/reports/orphan-works2015.pdf> (accessed 3/19/2025).

⁵⁷ “Orphan Works and Mass Digitization,” *United States Copyright Office*, June 2015, p. 1, available at <https://www.copyright.gov/orphan/reports/orphan-works2015.pdf> (accessed 3/19/2025).

⁵⁸ See e.g., Ex. 2 (Bartz Dep. Tr.) at 168:21-172:4, 176:18-182:14; Ex. 1 (Graeber Dep. Tr.) at 136:3-25; Ex. 3 (Johnson Dep. Tr.) at 142:19-146:3.

⁵⁹ See e.g., Ex. 2 (Bartz Dep. Tr.) at 168:21-172:4, 176:18-182:14; Ex. 1 (Graeber Dep. Tr.) at 136:3-25; Ex. 3 (Johnson Dep. Tr.) at 142:19-146:3.

61. Information on who owns copyrights is often unavailable or poor. Many existing training datasets come from sources on the Internet.⁶⁰ Studies show that the copyrights in existing training datasets containing text from the Internet are not well documented, which makes licensing the data in existing datasets impracticable.⁶¹ Beyond the problem of poor documentation, it may be impossible to identify and contact the copyright owners of many works because a work may be “orphaned,” meaning its copyright owner cannot reliably be identified even with diligent effort.⁶² The implication is that using text available on the Internet and seeking licenses for those texts is not a feasible approach to satisfying the licensing obligations that Plaintiffs seek to impose on AI firms. This implies that AI firms would need to first identify potential licensors, and then directly obtain and license text from enough of them to avoid infringing copyrights if training were not fair use.

62. The lack of information regarding copyright owners coupled with the scale of the required licensing effort makes it essentially impossible to license training data for a cutting-edge LLM through a competitive market with negotiations between copyright owners and AI

⁶⁰ Jesse Dodge, et al., “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus,” *Paul G. Allen School of Computer Science & Engineering, University of Washington*, September 2021 (arXiv:2104.08758v2).

⁶¹ Shayne Longpre, et al., “The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI,” *Data Provenance Initiative*, November 2023, (arXiv:2310.16787v3), p. 2; Jesse Dodge, et al., “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus,” *Paul G. Allen School of Computer Science & Engineering, University of Washington*, September 2021 (arXiv:2104.08758v2).

⁶² Report of Gloriana St. Clair, *The Authors Guild, Inc. v. Google Inc.*, No. 1:05-cv08136 (S.D.N.Y.), Dkt. 1042-1 ¶ 29 (experiment showed that “only half of the publishers... responded to [license] request” despite efforts to send initial request letters and follow-up letters); see also “Orphan Works and Mass Digitization,” United States Copyright Office, June 2015, p. 1, available at <https://www.copyright.gov/orphan/reports/orphan-works2015.pdf> (accessed 3/19/2025) (“a user’s ability to seek permission or to negotiate licensing terms is compromised by the fact that despite his or her diligent efforts, the user cannot identify or locate the copyright owner. Second, in the case of mass digitization – which involves making reproductions of many works, as well as possible efforts to make the works publicly accessible – obtaining permission is essentially impossible, not necessarily because of the lack of identifying information or the inability to contact the copyright owner, but because of the sheer number of individual permissions required (emphasis added).”); p. 38 (“Studies of library collections of printed, published books and similar works estimate that between 17% and 25% of published works and as much as 70% of specialized collections are orphan works.”).

firms. This is particularly the case because of the low value of the licenses that result from the small contribution that even large publishers would make to the value of the LLM.

3. In the Hypothetical Market to License Training Text, License Rates Would Be Low Relative to Transaction Costs, Making the Market Impracticable

63. Plaintiffs assert that copyright covers the use of text to train LLMs, and the Named Plaintiffs testified that [REDACTED]⁶³ Each author's potential contribution to the capability of the licensee's LLM is small because an individual author has only a small amount of text to license. However, the cost of identifying and negotiating even straightforward license agreements would generally exceed the training value of a rights holder's text.⁶⁴

64. A licensor negotiating with an LLM developer is too small to influence the AI firm's decision as to whether to develop the model, because the impact of adding the licensor's text to the training corpus on the LLM's capabilities is negligible.⁶⁵ Thus, the licensor would be unable to detect any difference in performance of the LLM from licensing its text, which means the LLM would have the same effect on the licensor whether it licenses or not. To the extent the licensor is unaffected by the decision to license its works for training, the licensor's cost of licensing its works to train an LLM, other than transaction costs, is small.

65. Anthropic's data acquisition strategies indicate incremental text has low value in the context of the trillions of tokens required to train an LLM like Claude. [REDACTED]

⁶³ Complaint, ¶¶ 73-75. *See also*, Ex. 2 (Bartz Dep. Tr.) at 168:21-172:4, 176:18-182:14; Deposition of Charles Andrew Graeber, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, March 5, 2025 ("Ex. 1 (Graeber Dep. Tr.)") at 136:3-25; Deposition of Kirk Wallace Johnson, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, March 6, 2025 ("Johnson Dep. Tr.") at 142:19-146:3.

⁶⁴ Transaction costs are all costs and opportunity costs borne by either party to a transaction to reach an agreement.

⁶⁵ Kaplan Declaration, ¶ 40.

1
2
3
4
5
6
7
8 66.
9
10
11
12
13
14
15

16 67. Based on the evidence I have seen, the transaction costs associated with licensing would
17 exceed the value of Plaintiffs' and other licensors' tokens. Consider the transaction costs
18 involved on both sides of a transaction to license authors' books for training. Named
19 Plaintiffs each have several books in various genres.⁶⁹ If a typical book contains tokens
20 worth [REDACTED] to Anthropic before its administrative costs, an author with three
21 books would have tokens that Anthropic would pay [REDACTED] to add to its training

22 ⁶⁶ Turvey Declaration ¶ 27.

23 ⁶⁷ Turvey Declaration ¶ 27.

24 ⁶⁸ Ex. 45 (WTP Valuation Model).

25 ⁶⁹ Ex. 2 (Bartz Dep. Tr.) at 128:23-130:5, explaining that [REDACTED]
26 [REDACTED]; Ex. 1 (Graeber Dep. Tr.) at 106:18-107:25, explaining that

27 [REDACTED]; Ex. 3 (Johnson Dep. Tr.) at 110:3-111:3 and 158-21:159:3, explaining that [REDACTED]
28 [REDACTED]

database. An author with 10 books would have tokens worth [REDACTED] before transaction costs. Thus, a typical author with two to ten books would control tokens that Anthropic would be willing to pay [REDACTED] to acquire before its administrative costs.⁷⁰

68. The parties' transaction costs would, in most cases, exceed the value of an author's tokens. Anthropic would have to hire staff to identify rights holders, contact them, verify their rights, and make an offer. Licensors' lack of information regarding the value of their works for training could frequently lead to disagreement and costly negotiations. These costs need to be deducted from Anthropic's willingness to pay the copyright owner.

69. Plaintiffs testified at deposition that [REDACTED]⁷¹ which suggests that reaching a deal with them would require possible offers and counteroffers – a cost to both parties. Plaintiff Graeber indicated [REDACTED]⁷² The value of Mr. Graeber's time, his agents' time, and Anthropic's staff's time almost certainly exceed Anthropic's gross willingness to pay for the two books Mr. Graeber could license.⁷³ Similarly, Plaintiff Bartz indicated that [REDACTED]

⁷⁰ Plaintiff Bartz has authored four books, Plaintiff Graeber has authored two books, and Plaintiff Johnson has authored three books (Complaint, ¶¶ 12-14).

⁷¹ See, e.g., Ex. 2 (Bartz Dep. Tr.) at 43:9-17; Ex. 1 (Graeber Dep. Tr.) at 32:16-33:14; Ex. 3 (Johnson Dep. Tr.) at 51:11-52:1.

⁷² See, e.g., Ex. 1 (Graeber Dep. Tr.) at 32:16-33:14 ("[REDACTED]").

⁷³ Charles Graeber, "Bio," available at <https://www.charlesgraeber.com/bio> (accessed 3/26/2025).

[REDACTED]

[REDACTED]⁷⁴ With an average hourly rate for lawyers in New York, where Ms. Bartz lives,⁷⁵ exceeding \$300 per hour,⁷⁶ Ms. Bartz’s legal costs alone could readily exceed Anthropic’s gross willingness to pay for the text in the four books she could license.⁷⁷ Similarly, Plaintiff Johnson indicated that [REDACTED]
[REDACTED]⁷⁸ With an average hourly rate for lawyers in California, where Plaintiff Johnson lives,⁷⁹ exceeding \$300 per hour,⁸⁰ Mr. Johnson’s legal costs could readily exceed Anthropic’s willingness to pay for the text in the three books that he could license.⁸¹

70. Notably the costs enumerated here do not account for the cost of delay from negotiating licenses. If Anthropic needed just one million licenses to build an LLM training corpus, it would take more than 27 years to negotiate them if Anthropic could negotiate 100 licenses per day, seven days per week. Even if Anthropic could somehow execute 1,000 licenses per

⁷⁴ Ex. 2 (Bartz Dep. Tr.) at 43:9-17 (“[REDACTED]
[REDACTED]”).

⁷⁵ Complaint, ¶ 12.

⁷⁶ Catherine Brock, “How Much is a Lawyer? Hourly Rates by State and More,” *LawPay*, November 12, 2024, available at <https://www.lawpay.com/about/blog/lawyer-hourly-rate-by-state/> (accessed 3/26/2025).

⁷⁷ Andrea Bartz, “Books,” available at <https://www.andreabartz.com/books/> (accessed 3/26/2025).

⁷⁸ Ex. 3 (Johnson Dep. Tr.) at 49:10-18 (“[REDACTED]
[REDACTED]”).

⁷⁹ Complaint, ¶ 14.

⁸⁰ Catherine Brock, “How Much is a Lawyer? Hourly Rates by State and More,” *LawPay*, November 12, 2024, available at <https://www.lawpay.com/about/blog/lawyer-hourly-rate-by-state/> (accessed 3/26/2025).

⁸¹ Kirk Wallace Johnson, “Books,” available at <http://kirkwallacejohnson.com/books/> (accessed 3/26/2025).

day, negotiations would still take roughly 2.7 years. And Anthropic will certainly require more than one million licenses.

⁸² The capabilities of LLMs are growing quickly, so the delay in licensing training data would be a material cost to AI firms.

C. Collective Licensing Is Not a Viable Response to High Transaction Costs

71. Plaintiffs suggest that without Anthropic's alleged infringement, blanket licenses to use copyrighted text for training would be available through clearinghouses.⁸³ The suggestion that collective licensing could arise and resolve the deficiencies of the licensing market is an admission that the licensing market AI firms currently face does not realistically enable them to license the texts they need. However, Plaintiffs are wrong that collective licensing covering the huge number of texts LLMs require for training is incipient.

72. The establishment of collective licensing of training rights is speculative for several reasons. First, Plaintiffs indicate that one such clearinghouse exists, called the Copyright Clearance Center ("CCC").⁸⁴ However, CCC does not suggest that it could license sufficient materials to train a cutting-edge LLM.⁸⁵ Second, Named Plaintiffs testified that [REDACTED] which indicates that training rights are highly dispersed.⁸⁶ Third, publishers, the companies that could most readily use existing commercial relationships with authors to gather training rights, have not done so and do not have training rights to many of the books in their catalogs.⁸⁷ The reason that training rights

⁸² Turvey Declaration, ¶ 22.

⁸³ Complaint, ¶ 48.

⁸⁴ Complaint, ¶ 48.

⁸⁵ See, e.g., <https://www.copyright.com/resource-library/communities/ai-copyright-licensing/> (accessed 3/25/2025).

⁸⁶ See Ex. 2 (Bartz Dep. Tr.) at 168:21-172:4, 176:18-182:14; Ex. 1 (Graeber Dep. Tr.) at 136:3-25; Ex. 3 (Johnson Dep. Tr.) at 142:19-146:3.

⁸⁷ Turvey Declaration, ¶¶ 31-35.

1 remain dispersed is that the effort to gather the vast number of copyrights required into
2 collectives would face comparable transaction costs to those the LLM developers face when
3 licensing directly.

4 73. A licensing collective would bring together Plaintiffs' and others' copyrights to license
5 them to LLM developers. Such a collective would coordinate the pricing of many rights
6 holders, effectively acting as a cartel of rights holders that could raise prices above the
7 competitive level.⁸⁸ The market power of the collective would be an expansion of the
8 monopolies granted by the collective's members because LLM developers would no longer
9 face competition among licensors.⁸⁹ In short, when a licensing collective sets an above-
10 competitive price for all of its copyrights, LLM developers would not be able to respond by
11 substituting one license for another. The exercise of market power to elevate prices harms
12 consumer welfare.

13 **D. Evidence of Limited Transactions To Acquire LLM Training Data Does Not**
14 **Establish a Competitive Market that Could Mediate the License Transactions**
15 **Needed to Train a Cutting-Edge LLM and Highlights the Infeasibility of Such a**
16 **Market**

17 74. Plaintiffs argue that other AI firms have entered licenses for training.⁹⁰ The agreements that
18 Plaintiffs highlight are not in the record, but regardless of their structure they would not
19 establish the incipency of a licensing market for LLM training data. A review of the trade
20 press discussing these agreements also indicates that they may give access to data that the
21 licensee would not be able to obtain without the cooperation of the licensor or that they
22 permit the public display of copyrighted text, which is a separate use that could require a

23 ⁸⁸ Lynne Pepall, *et al.*, *Contemporary Industrial Organization*, John Wiley & Sons, Inc.,
24 2011, 234-238.

25 ⁸⁹ Stephen Breyer, "The Uneasy Case for Copyright, A Study of Copyright in Books,
26 Photocopies, and Computer Programs, *Harvard Law Review*, Vol. 84, No. 2, December, 1970, p.
27 318; *See also* B. Douglas Bernheim and Michael D. Whinston, "Common Marketing Agency as a
Device for Facilitating Collusion," *The RAND Journal of Economics*, Vol. 16, No. 2, Summer,
1985.

28 ⁹⁰ Complaint, ¶ 48.

license.⁹¹ Transactions driven by these types of considerations do not speak to the existence of a market for training licenses.

75. The economically relevant feature of these transactions is that they exemplify the futility of attempting to license training text for an LLM. For example, Plaintiffs assert that the OpenAI agreements they cite license training data. If true, however, OpenAI's licensing efforts, which have been costly, have not spared it significant copyright litigation and potential related liability. OpenAI appears to be embroiled in perhaps nine copyright suits related to its LLMs.⁹² Thus, OpenAI may have forestalled litigation with some of the parties with which it has deals, but it has not avoided costly and disruptive lawsuits asserting millions of infringed works and up to \$150,000 per work infringed in potential statutory damages.⁹³

⁹¹ Plaintiffs allege "OpenAI, Google, and Meta" "paid... to obtain licenses to reproduce copyrighted material for LLM training" alleging such deals with "Axel Springer, News Corporation, the Associated Press, and others. (Complaint, ¶ 48). The agreement between OpenAI and Axel Springer allows OpenAI "to show parts of articles in ChatGPT responses" and gives OpenAI "access to the news service's archives." (Will Oremus and Elahe Izadi, "AI's future could hinge on one thorny legal question," *The Washington Post*, January 4, 2024, available at <https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/> (accessed 3/25/2025)). The agreement between OpenAI and NewsCorp gave OpenAI "permission to display content from News Corp mastheads in response to user questions and to enhance its products." ("News Corp and OpenAI Sign Landmark Multi-Year Global Partnership," *News Corp*, May 22, 2024, available at <https://newscorp.com/2024/05/22/news-corp-and-openai-sign-landmark-multi-year-global-partnership/> (accessed 3/25/2025)). The agreement between OpenAI and the Associated press "sees OpenAI licensing part of AP's text archive, while AP will leverage OpenAI's technology and product expertise[.]" (Matt O'Brien, "ChatGPT-maker OpenAI signs deal with AP to license news stories," *AP*, July 13, 2023, available at <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a> (accessed 3/25/2025)). The public information on these three agreements do not suggest that they are for licensing content for training purposes rather than accessing material otherwise unavailable to OpenAI.

⁹² "Master Chart of copyright, DMCA, and other claims in suits vs. AI," *ChatGPT is Eating the World*, March 23, 2025 available at <https://chatgptiseatingtheworld.com/2025/03/23/master-chart-of-copyright-dmca-and-other-claims-in-suits-v-ai/> (accessed 3/25/2025), showing seven cases (before consolidation) with direct copyright claims, and two additional cases with DMCA claims.

⁹³ Statutory damages of \$150,000 per work on 1 million books results in damages of \$150 billion, which is roughly the value of OpenAI, for example. See Cade Metz, "Open AI Completes

76. The agreements between firms building LLMs and content providers that Plaintiffs raise do not indicate that the market to license text to train LLMs is practicable. Those agreements cover small amounts of training data and do little to reduce the potential copyright liability of the licensees, who continue to face multiple lawsuits alleging that LLM training infringes copyrights. Publicized agreements do not change the conclusion that there is not a viable market to license data sufficient to train an LLM like Claude.

77. For the foregoing reasons, it is my opinion that, as a matter of economics, Anthropic's use of Plaintiffs' books to create Claude does not harm any existing or likely-to-be-established market for Plaintiffs' books.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct. Executed on March 27, 2025, in Arlington, Massachusetts.



STEVEN R. PETERSON

Deal That Values Company at \$157 Billion,” *The New York Times*, available at <https://www.nytimes.com/2024/10/02/technology/openai-valuation-150-billion.html>.

CERTIFICATE OF SERVICE

I, Douglas A. Winthrop, am the ECF user whose identification and password are being used to file the foregoing **DECLARATION OF STEVEN R. PETERSON IN SUPPORT OF DEFENDANT ANTHROPIC PBC's MOTION FOR SUMMARY JUDGMENT.**

Dated: March 27, 2025

/s/ Douglas A. Winthrop

EXHIBIT A

CURRICULUM VITAE

Steven R. Peterson, Ph.D.

PROFESSIONAL EXPERIENCE

Econic Partners, LLC
Boston, MA
Partner, February 2025 - Present

Compass Lexecon
Boston, MA
Executive Vice President, April 2013 – February 2025
Managing Director, Boston Office, January 2007 – June 2010
Senior Vice President, January 2006 – March 2013
Managing Director, August 1999 – December 2005

The Economics Resource Group, Inc.
Senior Economist, 1992 – July 1999
Economist, 1990 – 1992

Northeastern University, Boston, MA
Adjunct Faculty, 2011 – 2017

EDUCATION

Harvard University, Cambridge, MA
Ph.D. in Economics

University of California, Davis
A.B., Economics, Highest Honors

TESTIMONY AND CONSULTING EXPERIENCE

Anthropic PBC

Concord Music Group, Inc., et al., v. Anthropic PBC, in the United States District Court for the Northern District of California San Jose Division, Declaration of Steven R. Peterson in Support of Defendant's Opposition to Plaintiff's Renewed Motion for Preliminary Injunction (August 22, 2024). Declaration of Steven R. Peterson, Ph.D. in Support of Defendant's [Proposed] Surreponse to Plaintiffs' Motion for Preliminary Injunction. (October 7, 2024).

ExxonMobil Canada Properties, ExxonMobil Canada Resources, Chevron Canada Resources, Petro-Canada Hibernia Partnership, Murphy Oil Company Ltd. and Equinor Canada Ltd.

In the Matter of an Arbitration Under the Hibernia Development Project Incidental Net Profits Interest Agreement Dated November 10, 1990, and the Commercial Arbitration Act, R.S.C., 1985, C. 17 (2nd Supp.) Between: *ExxonMobil Canada Properties and Resources, Chevron Canada Resources, Petro-Canada Hibernia Partnership, Murphy Oil Company Ltd. and Equinor Canada Ltd., Claimants, and His Majesty the King in Right of Canada, as represented by Natural Resources Canada and the Canada Development Investment Corporation, Respondent*, Expert Report of Steven R. Peterson, Ph.D. (March 22, 2023), Reply Expert Report of Steven R. Peterson, Ph.D. (September 13, 2024).

Anthropic PBC

Concord Music Group, Inc., et al., v. Anthropic PBC, in the United States District Court for the Middle District of Tennessee Nashville Division, Declaration of Steven R. Peterson, Ph.D. in Support of Defendant's Opposition to Plaintiffs' Motion for Preliminary Injunction (January 16, 2024).

Converge Midstream, LLC

Converge Midstream, LLC v. Magellan Crude Oil Pipeline Company, L.P. and Magellan Midstream Partners, L.P., in the United States District Court, Harris County, Texas, 113th Judicial District, Expert Report of Steven R. Peterson, Ph.D. (October 27, 2023), Deposition (January 10, 2024).

YouTube, LLC

Maria Schneider, Uniglobe Entertainment, LLC, and AST Publishing Ltd., individually and on behalf of all others similarly situated v. YouTube, LLC and Google LLC (Case No.: 3:20-cv-04423-JD), in the United States District Court, Northern District of California, San Francisco Division. Preliminary Expert Rebuttal Report of Steven R. Peterson (September 22, 2022). Expert Rebuttal Report of Steven R. Peterson (December 29, 2022). Deposition (January 20, 2023).

Canadian National Railway

Before the Surface Transportation Board, In re Application of the National Railroad Passenger Corporation under 49 U.S.C. § 24308(a), Docket No. FD 35743, Verified Statement of Steven R. Peterson (May 27, 2022). Reply Verified Statement of Steven R. Peterson (July 26, 2022). Rebuttal Verified Statement of Steven R. Peterson (August 25, 2022).

MGM Studios

Consultant to MGM on acquisition by Amazon Inc. (2022).

Equal Justice Under Law

Nathan Wright et al., v. Family Support Division of the Missouri Department of Social Services et al., in the United States District Court for the Eastern District of Missouri Eastern Division. Affidavit (November 15, 2019). Deposition (December 4, 2019). Rebuttal Affidavit (January 31, 2020).

Google, LLC

In re: *Determination of Rates and Terms for the Digital Performance of Sound Recording and Making of Ephemeral Copies to Facilitate Those Performances*, before the United States Copyright Royalty Judges, The Library of Congress. Written Direct Testimony (September 23, 2019, Amended December 10, 2019). Written Rebuttal Testimony (January 10, 2020, Amended February 7, 2020). Trial Testimony (August 25-26, 2020).

Radio Music License Committee

Radio Music License Committee v. Broadcast Music, Inc., in the United States District Court for the Southern District of New York. Expert Report (July 22, 2019). Expert Rebuttal Report (August 19, 2019). Deposition (September 11, 2019).

Legal Aid Justice Center

Damian Stinnie et al., v. Richard Holcomb in his official capacity as the Commissioner of the Virginia Department of Motor Vehicles, in the United States District Court for the Western District of Virginia, Charlottesville Division. Hearing Testimony (November 15, 2018). Expert Report (May 1, 2019). Deposition (May 24, 2019).

Keweenaw Bay Indian Community

Keweenaw Bay Indian Community v. Nick A. Khouri et al., in the United States District Court for the Western District of Michigan. Expert Report (May 25, 2018). Supplemental Report (August 3, 2018). Deposition (September 6, 2018).

United States Soccer Federation, Inc.

North American Soccer League, LLC v. United States Soccer Federation, Inc. in the United States District Court for the Eastern District of New York. Declaration (October 16, 2017). Reply Declaration (November 1, 2017).

Radio Music License Committee

In an Arbitration before the Hon. Vaughn Walker, Kenneth R. Feinberg, Esq., and Lee A. Freeman, Jr., Esq. between: *SESAC, Inc., SESAC, LLC, and SESAC Holdings, Inc., Claimants v. Radio Music License Committee, Respondent*, Expert Report (December 23, 2016). Reply Expert Report (January 23, 2017). Hearing Testimony (February 23 and March 10, 2017).

Radio Music License Committee

Radio Music License Committee v. Global Music Rights, LLC. in the United States District Court for the Eastern District of Pennsylvania. Declaration (November 18, 2016). Declaration (July 21, 2017).

TransCanada Corporation

Consultant to TransCanada Corporation on its acquisition of Columbia Pipeline Group (March – June 2016).

Sanum Investments Limited

In an Arbitration Under the Rules of the Singapore International Arbitration Centre between: *Sanum Investment Limited, Claimant, versus ST Group Co., Ltd., Sithat Xaysoulivong, S.T. Vegas Co., Ltd., S.T. Vegas Enterprise Ltd., Xaya Construction Co., Ltd. and Xaysana Xaysoulivong, Respondents*, Witness Statement, with Joseph P. Kalt and Eric Henson (April 20, 2016).

National Association of Broadcasters and Pandora Media, Inc.

In re: *Determination of Royalty Rates and terms for Ephemeral Recording and Digital Performance of Sound Recordings*, before the United States Copyright Royalty Judges, Library of Congress. Expert Report (February 23, 2015), Deposition Testimony (March 24, 2015), Hearing Testimony (May 14, 2015).

National Association of Broadcasters

In re: *Antitrust Consent Decree Review: American Society of Composers, Authors and Publishers/Broadcast Music, Inc.* Comments on Behalf of the National Association of Broadcasters (August 6, 2014).

Leidos, Inc.

United States of America v. Leidos, Inc., in the United States District Court for the District of Columbia. Expert Report (March 28, 2014).

Radio Music License Committee

Radio Music License Committee v. SESAC, Inc., SESAC, LLC, and SESAC Holdings, Inc. in the United States District Court for the Eastern District of Pennsylvania. Declaration (November 14, 2013). Deposition (December 4, 2013). Trial Testimony (December 10, 2013).

BJ's Wholesale Club, Inc.

Irene Cappalli v. BJ's Wholesale Club Inc. in the United States District Court, District of Rhode Island. Expert Report (March 15, 2012). Deposition (April 27, 2012). Declaration (February 12, 2013).

National Marine Fisheries Service

Consultant to National Marine Fisheries Service on market power and excessive-share limits in the Surf Clam and Ocean Quahog fisheries (2010 – 2011).

Energy Intensive Manufacturers Working Group

Coalition for Responsible Regulation, Inc. et al. v. United States Environmental Protection Agency, in the United States Court of Appeals for the District of Columbia Circuit. Declaration (September 14, 2010).

Amex Construction Company, Inc.

ExxonMobil Oil Corporation v. Amex Construction Company, Inc., in the United States District Court, Northern District of Illinois, Eastern Division. Expert Report (February 15, 2010). Deposition (March 2, 2010).

Delta Air Lines, Inc.

Consultant to Delta Air Lines on LaGuardia/Reagan National Airport slot swap with U.S. Airways (2009 – 2010).

Imperial Credit Industries, Inc.

In re: Imperial Credit Industries, Inc., in the United States Bankruptcy Court, Central District of California, Santa Ana Division. Rebuttal Report (April 27, 2007). Trial Testimony (May 22, 2008).

Delta Air Lines, Inc.

Consultant to Delta Air Lines on Delta-Northwest merger (2007 – October 2008).

Greater Lakeside Corporation

The Higbee Company v. Greater Lakeside Corporation, Causeway LLC of Delaware, Broadway Management Corporation, and Jeffrey Feil, in the United States District Court for the Eastern District of Louisiana. Expert Report (September 18, 2007). Supplemental Expert Report (September 25, 2007). Deposition (October 19, 2007).

TransCanada Corporation

Consultant to TransCanada Corporation on its acquisition of ANR Group (2007).

Exxon Mobil Corporation

JAAM, Inc., d/b/a Tigerland Exxon v. Exxon Mobil Corporation and Mon Valley Petroleum, Inc., in the United States District Court for the Western District of Pennsylvania. Expert Report (January 12, 2007).

Finova Capital

In Re: Finova Capital Corporation and Finova Mezzanine Capital, Inc., in the United States Bankruptcy Court for the District of Delaware. Expert Report (May 19, 2006). Deposition (August 2, 2006).

Volvo Cars of North America, Inc.

Bay Ridge Volvo American, Inc. et al. v. Volvo Cars of North America, Inc., in the United States District Court Southern District of New York. Expert Report (June 1, 2005). Deposition (August 17, 2005). Supplemental Expert Report (November 11, 2005).

Israel Electric Corporation, Ltd.

Israel Electric Corporation Ltd. vs. the Public Utilities Authority, the Minister of National Infrastructures, the Minister of Finance, the Israel Securities Authority and the Government Corporations Authority (Request for Injunction): In the Israeli Supreme Court, No. /04, August 2004. Statement (August 30, 2004), with Joseph P. Kalt and Paul B. Vasington.

Flying J, Inc.

Flying J, Inc. v. Comdata Network, Inc., in the United States District Court of Utah (Northern Division). Declaration (June 22, 2004). Damages Report (June 22, 2004). Deposition (October 6, 2004). Hearing Testimony (November 19, 2004).

Musicmatch, Inc.

Gracenote, Inc. v. Musicmatch Inc., In the United States District Court Northern District of California (Oakland Division). Expert Report (February 17, 2004). Declaration (February 24, 2004). Deposition (March 2004).

Monica Pappas, Bill DeVitt, and Monica Pappas Associates

The Healthcare Financial Group, Inc., v. Monica Pappas DeVitt et al., in the District Court, Arapahoe County, Colorado. Filed written expert testimony on lost-profits damages (February 2003).

Ticketmaster Corporation

Evaluated damages from asserted anti-competitive conduct (2003).

Amoco Production Company, Amerada Hess Corporation, and Shell Western E&P, Inc.

Assessed fair market value of CO₂ for payment of royalties. Analyzed issues of market structure of CO₂ industry and marketability of CO₂ at the well (2002).

American Airlines

Conducted analysis of market structure, capacity additions, and pricing in an antitrust suit asserting predatory conduct (2001).

For a Mutual Insurance Company

Conducted market research and performed benchmarking analyses to establish pricing approach and prices for new internet services (2000).

Bass Enterprises Production Company

Assessed fair market rental value of oil-bearing property temporarily taken by the federal government (2000).

Boeing Company

Filed declaration of behalf of Boeing Company (Delta Launch Services, Inc.) for a NASA administrative proceeding regarding release of contract information under the Freedom of Information Act (2000).

Honeywell, Inc.

Conducted study of damages arising from monopolization in the market for ring laser gyroscope inertial navigation systems. Conducted analysis of damages arising from patent infringement (1998).

British Airways, Plc.

Conducted study of the competitive effects of British Airways' proposed alliance with American Airlines. Advised on and assisted with presentations before the European Commission (1998).

HarperCollins Publishers

Brother Records, Inc., et al., v. HarperCollins Pub. Inc., et. al. Filed written expert testimony on damages in libel litigation (December 1997).

Northeast Utilities

Before the Federal Energy Regulatory Commission, OA97-237-000, ER 97-1079-000, and EC97-35-000. Conducted analysis of competition in the New England generation market. Filed affidavit in support of NU's Answer to Requests to Reject or Condition Approval of Market-Based Rates (with Frank A. Felder) (July 1997).

McDonnell Douglas Corporation

McDonnell Douglas Corporation v. National Aeronautics and Space Administration, in the U.S. District Court for the District of Columbia. Filed affidavit describing how the public release of cost and price information affects negotiations and competition in markets for launch services (November 1996).

Pennzoil

Before the Federal Energy Regulatory Commission, Docket No, IS95-35-000. Provided written direct testimony (October 1996) and oral testimony (January 1997) on the cost of capital of oil pipeline facilities.

Pennzoil

Before the Federal Energy Regulatory Commission, Docket No. IS94-37-000 and Docket No. IS 94-23-000. Provided written direct testimony (April 1995) and oral testimony (November 1995).

BP Exploration (Alaska) Inc.

Modeled the costs and benefits associated with increased enhanced oil recovery activities within the Prudhoe Bay Unit (1995).

Burlington Northern Industries-Santa Fe Pacific Corporation

Performed cost-benefit analysis of the proposed Burlington Northern/Santa Fe merger. Analyzed the benefits accruing to shippers from expanded single-line service (1994 – 1995).

PUBLICATIONS AND RESEARCH

“Using Economics to Identify Common Impact in Antitrust Class Certification,” American Bar Association, Section of Antitrust Law, Economics Committee Newsletter, Vol. 11, No. 1, Spring 2011 (with Andrew Lemon).

“Rigorous Analysis to Bridge the Inference Gap in Class Certification” (with Andrew Lemon), *Journal of Competition Law and Economics*, March 2011.

“Oil Price Volatility and Speculation” (with Kenneth Grant), *The Energy Daily*, August 25, 2009.

“Understanding Today’s Crude Oil and Product Markets” (with Kenneth Grant and David Ownby), American Petroleum Institute, 2006.

“Understanding Natural Gas Markets” (with Charles Augustine and Bob Broxson), American Petroleum Institute, 2006.

“Regulatory Failure in the California Electricity Crisis” (with Charles Augustine), *The Electricity Journal*, August/September 2003.

“Market Power Analysis in a Dynamic Electric Power Industry” (with F. Felder), *The Electricity Journal*, April 1997.

“Testing the Merits of Providing Customized Risk Management” (with Frank A. Felder and Sarah E. Tobiason), 17th Annual North American Conference of the United States Association for Energy Economics, International Association for Energy Economics, October 1996.

“Competition Between Regulators and Venue Shopping by Natural Gas Pipelines in California” (with Joseph P. Kalt), 14th Annual Conference of the Advanced Workshop in Regulation and Public Utility Economics, May 1995.

“Environmental Regulation and International Competitiveness: What Does the Evidence Tell Us?” (with Adam B. Jaffe, Paul R. Portney, and Robert N. Stavins), *Journal of Economic Literature*, Vol. 33, March 1995.

“Implementation of the Core of a Two Person Exchange Economy without Integer Games or Refinements of Nash Equilibrium” (with Simon Grant, Stephen King, and Ben Polak), *Economics Letters*, 1992.

OTHER REPORTS AND PRESENTATIONS

“The Economic Impact of Delta Air Lines Seattle Expansion,” (with Bryan Keating). August 14, 2015.

“Antitrust Analysis of Aftermarkets,” American Bar Association, Section of Antitrust, 2010 Spring Meeting (with Edward Schwartz and Paula Render).

“Do Environmental Regulations Impair Competitiveness? A Critical Review of Economic Studies” (with Barry Galef and Kenneth Grant). Prepared by ICF Consulting Group and The Economics Resource Group, Inc., for the Office of Policy Analysis and Review, Office of Air and Radiation, U.S. Environmental Protection Agency, September 1995.

“Indexing Natural Gas Pipeline Rates” (with Amy B. Candell, Joseph P. Kalt, Sheila M. Lyons, and Stephen Makowka). Explored indexing as a form of Incentive regulation for natural gas pipelines and created the Pipeline Producer Price Index that could be used to implement indexing proposals. The Economics Resource Group, Inc., April 1995.

“Environmental Regulations and the Competitiveness of U.S. Industry” (with A. Jaffe, P. Portney and R. Stavins), U.S. Department of Commerce, Economics and Statistics Administration, Washington, DC, NTIS No. PB-93-193514, July 1993.

HONORS AND AWARDS

Jacob K. Javits Fellow, Harvard University, 1987 – 1991

Phi Beta Kappa, University of California, Davis, 1987

EXHIBIT B

Materials Relied Upon

Academic

- B. Douglas Bernheim and Michael D. Whinston, “Common Marketing Agency as a Device for Facilitating Collusion,” *The RAND Journal of Economics*, Vol. 16, No. 2, Summer, 1985.
- Bernard Salanié, *Microeconomics of Market Failures*, The MIT Press, 2000.
- Dennis Carlton, “Contracts, Price Rigidity, and Market Equilibrium,” *Journal of Political Economy*, October 1979, 87(5):1034-1062.
- Dennis Carlton, “The Theory of Allocation and Its Implications for Marketing and Industrial Structure: Why Rationing is Efficient,” *The Journal of Law & Economics*, October 1991, 34(2):231-262.
- Jesse Dodge, et al., “Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus,” *Paul G. Allen School of Computer Science & Engineering, University of Washington*, September 2021 (arXiv:2104.08758v2).
- Lynne Pepall, et al., *Contemporary Industrial Organization*, John Wiley & Sons, Inc., 2011.
- Pierre N. Leval, “Toward a Fair Use Standard,” *Harvard Law Review*, Vol 103.
- Shayne Longpre, et al., “The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI,” *Data Provenance Initiative*, November 2023, (arXiv:2310.16787v3).
- Stephen Breyer, “The Uneasy Case for Copyright, A Study of Copyright in Books, Photocopies, and Computer Programs,” *Harvard Law Review*, Vol. 84, No. 2, December, 1970.
- William M. Landes and Richard A. Posner, “An Economic Analysis of Copyright Law,” *The Journal of Legal Studies*, Vol. 18, No. 2. June, 1989.

Bates Documents

- WTP Valuation Model, ANT_BARTZ_000279410.

Declarations

- Declaration of Jared Kaplan in Support of Anthropic’s Motion for Summary Judgement.
- Declaration of Tom Turvey in Support of Anthropic’s Motion for Summary Judgement.

Deposition Transcripts

- Deposition of Andrea Bartz, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, March 7, 2025.

- Deposition of Charles Andrew Graeber, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, March 5, 2025.
- Deposition of Kirk Wallace Johnson, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, March 6, 2025.

Legal - Law & Other Proceedings

- Copyright Act, Section 107.
- 17 U.S.C. § 102(b) (2018)
- *Authors Guild v. Google, Inc.*, 804 F.3d 202, 224-25 (2d Cir. 2015).
- *Authors Guild v. Open AI*, Case No. 23-cv-08292 (S.D.N.Y.).
- *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994).
- *Dubus v. NVIDIA Corp.*, Case No. 24-cv-02655 (N.D. Cal.).
- *In re Google AI Generative Copyright Litig.*, Case No. 23-cv-3440 (N.D. Cal.).
- *In re Open AI ChatGPT Litig.*, Case No. 3:23-cv-03223 (N.D. Cal.).
- *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417 (N.D. Cal.).
- *Makkai v. Databricks, Inc.*, Case No. 24-cv-02653 (N.D. Cal.).
- *New York Times v. Microsoft*, Case No. 23-cv-11195 (S.D.N.Y.).
- Report of Gloriana St. Clair, *The Authors Guild, Inc. v. Google Inc.*, No. 1:05-cv08136 (S.D.N.Y.), Dkt. 1042-1.
- *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S.
- *Zhang v. Google LLC*, Case No. 24-cv-02531 (N.D. Cal.).

Legal Filings - This Proceeding

- First Amended Class Action Complaint, *Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson, individually and on behalf of others similarly situated, v. Anthropic PBC*, United States District Court, Northern District of California, San Francisco Division, 3:24-cv-05417, December 4, 2024.

Publicly Available

- “Adult Fiction Outperforms the U.S. Book Market in 2023, Circana Reports,” *Circana*, February 2, 2024, available at <https://www.circana.com/intelligence/press-releases/2024/adult-fiction-outperforms-the-u-s-book-market-in-2023-circana-reports/> (accessed 3/19/2025).
- “Campfire accelerates accounting with Claude,” available at <https://www.anthropic.com/customers/campfire> (accessed 3/25/2025).

- “How Many Words in a Novel (Updated for 2025),” *Reedsy Studio*, September 19, 2024, available at <https://reedsy.com/studio/resources/how-many-words-in-a-novel> (accessed 3/26/2025).
- “Master Chart of copyright, DMCA, and other claims in suits vs. AI,” *ChatGPT is Eating the World*, March 23, 2025 available at <https://chatgptiseatingtheworld.com/2025/03/23/master-chart-of-copyright-dmca-and-other-claims-in-suits-v-ai/> (accessed 3/25/2025).
- “News Corp and OpenAI Sign Landmark Multi-Year Global Partnership,” *News Corp*, May 22, 2024, available at <https://newscorp.com/2024/05/22/news-corp-and-openai-sign-landmark-multi-year-global-partnership/> (accessed 3/25/2025).
- “Orphan Works and Mass Digitization,” *United States Copyright Office*, June 2015, available at <https://www.copyright.gov/orphan/reports/orphan-works2015.pdf> (accessed 3/19/2024).
- “Public Comments of Anthropic PBC,” Before the United States Copyright Office, Notification of Inquiry Regarding Artificial Intelligence and Copyright, October 30, 2023.
- “Title Output Up,” *Publishers Weekly*, June 18, 2007, available at <https://www.publishersweekly.com/pw/print/20070618/5223-title-output-up.html> (accessed 3/17/2025).
- “Tracing Model Outputs to the Training Data,” *Anthropic*, August 8, 2023.
- Andrea Bartz, “Books,” available at <https://www.andreabartz.com/books/> (accessed 3/26/2025).
- Andrew Albanese and Jim Milliot, “Self-Publishing’s Output and Influence Continue to Grow,” *Publishers Weekly*, November 8, 2024, available at <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96468-self-publishing-s-output-and-influence-continue-to-grow.html> (accessed 3/17/2025).
- Bernard Marr, “A short History of ChatGPT: How We Got to Where We Are Today,” *Forbes*, May 19, 2023, available at <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/> (accessed 3/19/2025).
- Cade Metz, “Open AI Completes Deal That Values Company at \$157 Billion,” *The New York Times*, available at <https://www.nytimes.com/2024/10/02/technology/openai-valuation-150-billion.html> (accessed 3/19/2025).
- Catherine Brock, “How Much is a Lawyer? Hourly Rates by State and More,” *LawPay*, November 12, 2024, available at <https://www.lawpay.com/about/blog/lawyer-hourly-rate-by-state/> (accessed 3/26/2025).
- Charles Graeber, “Bio,” available at <https://www.charlesgraeber.com/bio> (accessed 3/26/2025).

- Cristina Criddle, “Anthropic set to focus on business users in search for new revenues,” *Financial Times*, March 18, 2025, available at <https://www.ft.com/content/97e0ab06-8d83-4918-9079-3ed935bc1c63> (accessed 3/25/2025).
- Elizabeth Harris, Alexandra Alter, and Adam Bednar, “A Trial Put Publishing’s Inner Workings on Display. What Did We Learn?,” *The New York Times*, August 19, 2022, available at <https://www.nytimes.com/2022/08/19/books/prh-penguin-random-house-trial.html> (accessed 3/18/2025).
- Harsh Singhal, “Every Data Professional Should Know About the Common Crawl Project,” *DataScienceFM*, December 22, 2022, available at <https://datascience.fm/every-data-professional-should-know-about-the-common-crawl-project/> (accessed 3/18/2025).
- <https://www.anthropic.com/claude> (accessed 3/18/2025).
- <https://www.anthropic.com/company> (accessed 3/18/2025).
- <https://www.anthropic.com/research/clio> (accessed 3/18/2025).
- <https://www.copyright.com/resource-library/communities/ai-copyright-licensing/> (accessed 3/25/2025).
- Jim Milliot, “Self-Publishing Is Thriving, According to Bowker Report,” *Publishers Weekly*, February 17, 2023, available at <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/91574-self-publishing-is-thriving-according-to-bowker-report.html> (accessed 3/17/2025).
- Kirk Wallace Johnson, “Books,” available at <http://kirkwjohnson.com/books/> (accessed 3/26/2025).
- Leonid Taycher, “Books of the world, stand up and be counted! All 129,864,880 of you,” *Google Book Search*, August 5, 2010, available at <https://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html> (accessed 3/19/2025).
- Matt O’Brien, “ChatGPT-maker OpenAI signs deal with AP to license news stories,” *AP*, July 13, 2023, available at <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a> (accessed 3/25/2025).
- Thom Vaughan, “February 2025 Crawl Archive Now Available,” *Common Crawl*, February 23, 2025, available at <https://commoncrawl.org/blog/february-2025-crawl-archive-now-available> (accessed 3/18/2025).
- U.S. Department of Justice and Federal Trade Commission, “Merger Guidelines,” December 18, 2023, Section 4.3, available at https://www.ftc.gov/system/files/ftc_gov/pdf/2023_merger_guidelines_final_12.18.2023.pdf (accessed 3/25/2025). Will Oremus and Elahe Izadi, “AI’s future could hinge on one thorny legal question,” *The Washington Post*, January 4, 2024, available at <https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/> (accessed 3/25/2025).