UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: Case No. 1:23-cv-08292 Case No. 1:23-cv-10211	25-md-3143 (SHS) (OTW) Hon. Sidney H. Stein Hon. Ona T. Wang

DECLARATION OF GREGORY K. LEONARD IN SUPPORT OF OPENAI'S MOTION FOR SUMMARY JUDGMENT

I, Gregory K. Leonard, Ph.D, declare under penalty of perjury, pursuant to 28 U.S.C. § 1746, as follows:

1. I submit this Declaration in support of OpenAI Defendants' (collectively, "OpenAI") Motion for Summary Judgment.

I. QUALIFICATIONS

2. My name is Gregory K. Leonard. I am an economist and Vice President at Charles River Associates, 601 12th Street, Oakland, CA. Charles River Associates provides economic, financial, and business strategy consulting in the U.S. and other countries throughout the world. I received a Sc.B. in Applied Mathematics-Economics from Brown University in 1985 and a Ph.D. in Economics from the Massachusetts Institute of Technology in 1989. After receiving my Ph.D., I became an assistant professor at Columbia University. I subsequently moved into economic consulting and worked at several economic consulting firms prior to joining Charles River Associates.

3. My specialties within economics are applied microeconomics, the study of the behavior of consumers and firms, and econometrics, the application of statistical methods to economic data. I have published over sixty articles in scholarly and professional publications. Many of these articles address topics in industrial organization and the economics of intellectual property. I have also served as an expert witness in a number of litigation matters before U.S. District Courts, state courts, arbitration panels, and the (U.S.) International Trade Commission. In the area of economic analysis related to copyrights specifically, I served as an economics expert in *Oracle v. Google*, the *Phonorecords III*, *Phonorecords IV*, *Web V*, and *Web VI* proceedings before the Copyright Royalty Board, and arbitrations between the Radio Music Licensing Committee and, respectively, SESAC and ASCAP.

II. ASSIGNMENT

4. I have been asked by counsel for OpenAI to perform an economic analysis of the harm that Plaintiffs—purported owners of book copyrights—claim to have suffered as a result of OpenAI's alleged use of the Asserted Works. I understand that the Plaintiffs in this case are David Baldacci, Taylor Branch, Michael Connelly, Sylvia Day, Jonathan Franzen, Christopher Golden, Andrew Sean Greer, John Grisham, David Henry Hwang, George R.R. Martin, Jodi Picoult, Stacy Schiff, James Shapiro, the Authors Guild, and other companies that claim to have ownership stakes in some of the Asserted Works.

5. I previously prepared the Expert Report of Gregory K. Leonard, Ph.D., dated April 24, 2026 ("Leonard Opening Report"), the Expert Rebuttal Report of Gregory K. Leonard, Ph.D., dated May 29, 2026 ("Leonard Rebuttal Report"), and the Expert Reply Report of Gregory K. Leonard, Ph.D., dated June 26, 2026 ("Leonard Reply Report") in this matter.

6. I was deposed in this matter on July 22, 2026.

III. BACKGROUND

7. Plaintiffs allege that they have authored or have ownership stakes in a total of 212 Asserted Works.[1]

8. I understand that OpenAI made its GPT-3 large language model (LLM) available in its API in July 2020 and launched its LLM-powered chatbot known as ChatGPT to the public in November 2022.[2]

[1] Leonard Opening Report ¶ 34 & Ex. 1.

[2] Leonard Opening Report ¶ 38.

9. For purposes of this declaration, I collectively refer to OpenAI's LLMs at issue in this matter as the "In-Scope LLMs," which is defined to include OpenAI's GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini LLMs.[3]

IV. OPINIONS

10. I performed five types of empirical analyses to evaluate Plaintiffs' claims of harm based on OpenAI's alleged use of the Asserted Works: (1) an analysis of the sales of books in the U.S. over time in aggregate and for trade books from 2018 to 2024; (2) an analysis of the sales of top sellers after July 2020 and after November 2022; (3) an analysis of net sales trends of the Asserted Works after July 2020 and after November 2022; (4) an analysis using a synthetic control method; and (5) an analysis of book prices after July 2020 and after November 2022.[4] Analyses (1)–(4) support my conclusion that the introduction of the first In-Scope LLM (GPT-3) in July 2020 and the introduction of the chatbot ChatGPT to the public in November 2022 had no adverse impact on the sales of the Asserted Works.[5] Analysis (5) is consistent with my conclusion that any release of AI-generated works after July 2020 and after November 2022 had no effect on the prices of trade books like the Asserted Works.[6]

11. For purposes of my analyses, I mainly relied on third-party data, including annual reports from the Association of American Publishers ("AAP data"), individual book sales data in the United States from Circana ("Circana data"), and pricing data from Barnes & Noble ("Barnes & Noble data").[7] I understand that the Circana data produced in this matter is

[3] *Id.* ¶ 35.

[4] Leonard Reply Report ¶ 86.

[5] *Id.*

[6] *Id.*

[7] Leonard Opening Report ¶¶ 163–165, 217.

an output from Circana's broader BookScan database, which is widely considered the most comprehensive and accurate source for book sales and is relied upon by industry participants such as publishers, authors, and agents to make business decisions.[8]

A. Analysis 1: Sales of Books in the U.S. Over Time

12. Figure 1 below plots the sales of books in the U.S. over time in aggregate and for trade books specifically using the AAP data.[9] The first green vertical line indicates the year in which GPT-3 was first made available on OpenAI's API (July 2020).[10] The subsequent green vertical line represents the year in which the chatbot ChatGPT was available to the public (November 2022).[11] The graph shows that book sales have been generally trending upward since 2020.[12]

[8] *Id.* ¶ 164. The sales and royalty reports produced by Plaintiffs and their publishers had several inconsistencies and limitations that made them unsuitable for analyzing sales trends. *See id.* ¶ 161.

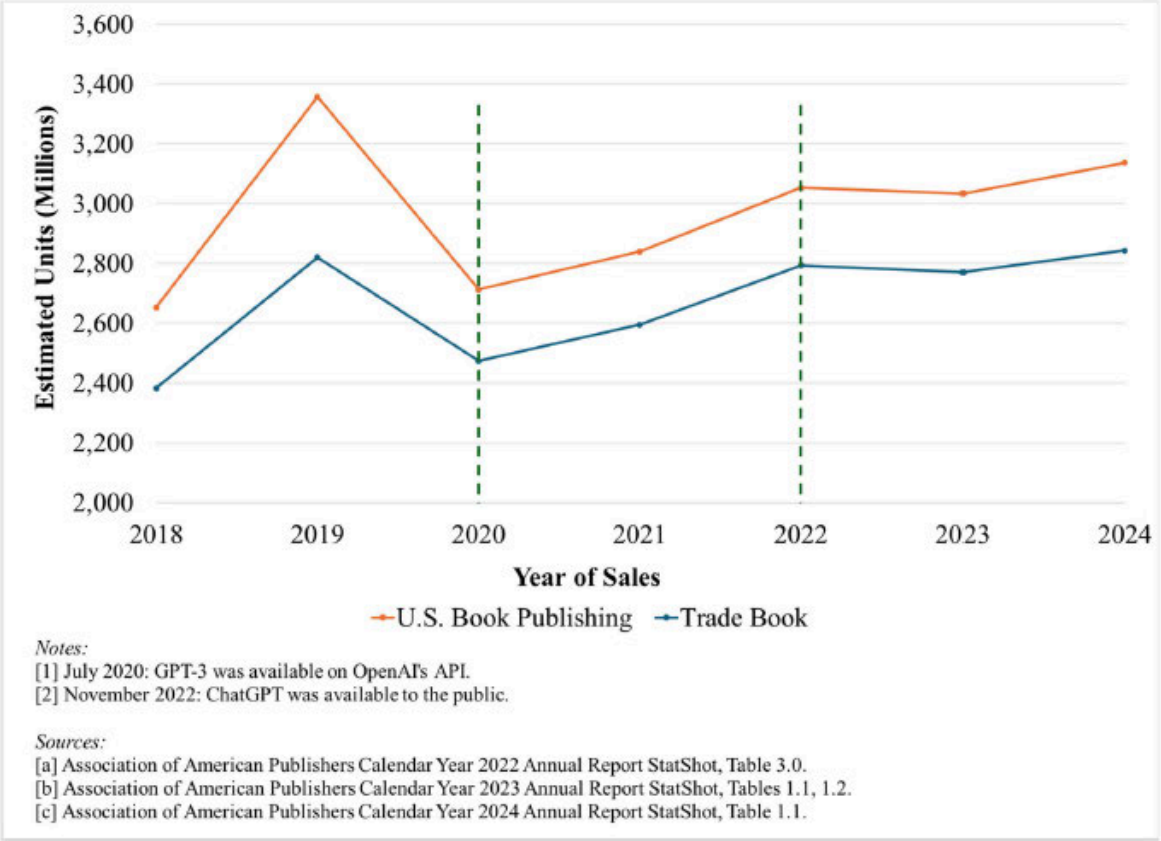
[9] *Id.* ¶ 169.

[10] *Id.*

[11] *Id.*

[12] *Id.*

Figure 1: U.S. Book Publishing and Trade Book Unit Sales



B. Analysis 2: Sales of Top Sellers

13. I analyzed how the sales of top sellers changed after GPT-3 was first made available on OpenAI's API (July 2020) and after the chatbot ChatGPT was available to the public (November 2022).[13] For each month, I identified the top 100 selling ISBNs in that month and added up their sales in the Circana data.[14] For each of the July 2020 and November 2022 introduction dates, I summed the top 100 sales for four different time periods: (1) the second year before the introduction, (2) the year before the introduction, (3) the year following the introduction, and (4) the second year following the introduction.[15] These

[13] *Id.* ¶ 170.

[14] *Id.*

[15] *Id.* & Ex. 2.

results were summarized in Exhibit 2 to my Opening Report, which is reproduced below. For both introduction dates, the top 100 ISBNs [redacted][16]

Exhibit 2
Total Net Unit Sales of 100 Top-Selling ISBNs
Monthly Top Sellers Over Four 12-Month Time Periods

Date	Top-selling ISBN	24-13 Months Before	12-1 Months Before	1-12 Months After	13-24 Months After
July 2020	Top 100	[redacted]			
November 2022	Top 100				

14. In general, the identity of the top 100 ISBNs [redacted], but focusing on top sellers, we can see that the [redacted].[17] This is consistent with the conclusion that the introduction of the first In-Scope LLM (GPT-3) in July 2020 and the introduction of the chatbot ChatGPT to the public in November 2022 had no adverse impact on the sales of the Asserted Works.

C. Analysis 3: Net Sales Trends of the Asserted Works

15. I analyzed the extent to which the net sales trends of the Asserted Works changed after GPT-3 was first made available on OpenAI's API (July 2020) and after the chatbot ChatGPT was available to the public (November 2022).[18] I used econometric methods to compare the trend exhibited by an Asserted Work's net sales in the year prior to the

[16] Leonard Opening Report ¶ 170.

[17] *See id.* ¶ 170.

[18] *Id.* ¶ 171.

GPT-3 API / chatbot ChatGPT introduction to the trend in the year following the two introduction dates.[19]

16. I used the Circana sales data for this analysis.[20] For each month, I calculated the net unit sales per day for each Asserted Work.[21] I used the 12 months prior to and the 12 months following the GPT-3 API / chatbot ChatGPT introductions to identify the preexisting trend for each Asserted Work and the change in that trend.[22] I also used the sales and royalty reports produced by Plaintiffs if those data met specific criteria that made them usable for this analysis.[23]

17. The 2020 and 2022 analyses account for a total of [redacted] of 212 Asserted Works.[24] For most Asserted Works, a statistical test does not reject the hypothesis that the sales trend did not change after the two introduction dates of July 2020 and November 2022.[25] For the November 2022 introduction date, only [redacted] of the Asserted Works were found to have a statistically significant change in trend.[26] Of these, [redacted] had an increase in trend and [redacted] had a decrease in trend.[27] For the July 2020 introduction date, only [redacted] of the Asserted

[19] *Id.*

[20] *Id.* ¶ 172.

[21] *Id.*

[22] *Id.*

[23] *Id.* ¶ 173.

[24] *Id.* I excluded Asserted Works from this analysis that were not present in the Circana data and did not have sufficient data in the royalty reports produced by Plaintiffs to evaluate trends because there was insufficient data to evaluate trends in sales. *Id.* ¶ 175. I also excluded Asserted Works that sold fewer than 10,000 units over the 2017-2025 time period because these sales figures are too small to decouple from statistical noise. *Id.*

[25] Errata to Leonard Opening Report ¶ 2.

[26] *Id.*

[27] *Id.*

Works were found to have a statistically significant change in trend.[28] Of these, [redacted] had an increase in trend and [redacted] had a decrease in trend.[29] These results were summarized in Amended Exhibit 3a in the Errata to my Opening Report, which is reproduced below.

Amended Exhibit 3a
Change in Net Sales Trends Following July 2020 and November 2022
12 Month Pre-Periods and 12 Month Post-Periods

Date	# of Asserted Works in Sample	# Positive Coefficients		# Negative Coefficients	
		Total	(q-value < 0.05)	Total	(q-value < 0.05)
July 2020	[redacted]				
November 2022					

18. Overall, the results do not suggest a systematic or widespread decrease in sales trends for the Asserted Works after the release of GPT-3 in OpenAI's API or after the release of GPT-3.5 powering the first publicly available version of the chatbot ChatGPT.[30]

D. Analysis 4: Synthetic Control Method

19. I compared the sales of the Asserted Works after each of the GPT-3 API / chatbot ChatGPT introduction dates to the sales of "control" books that were not included in OpenAI's training datasets for the In-Scope LLMs.[31] The effect of an Asserted Work's inclusion in the training datasets can be measured by the difference between (1) the sales of an Asserted Work in the period following the In-Scope LLMs' introduction and (2) what would be predicted for sales of the Asserted Work based on the sales of control

[28] *Id.*

[29] *Id.* & Amended Ex. 3a.

[30] Errata to Leonard Opening Report ¶ 2.

[31] Leonard Opening Report ¶ 176.

books.[32] The control books are used to account for the effects on book sales of factors other than the Asserted Works' inclusion in the training datasets.[33]

20. I used the "synthetic control" (SC) method for performing the analysis.[34] In 2017, prominent economists Susan Athey and Guido Imbens described this method as "arguably the most important innovation in the policy evaluation literature in the last 15 years."[35] The SC method involves predicting the post-treatment outcome for a treated unit by taking a weighted average of the outcome of the control units (during the post-treatment time period), where greater weight is given to those control units that are a closer match to the treated unit.[36]

21. The assessment of closeness of the match is based on pre-treatment outcomes and, potentially, other characteristics of the units.[37] For example, a control unit with a pattern of outcomes in the pre-treatment period that more closely matches the pattern of outcomes of the treated unit is deemed to be a closer match, and is assigned a greater weight, than a control unit with a pattern of outcomes in the pre-treatment period that less closely matches those of the treated unit.[38] In the present case, the outcome of interest is net sales.[39] The pre-treatment outcomes used in the weight determination are the sales in the weeks prior to July 2020 (when GPT-3 was made available via an API) and

---

[32] *Id.*

[33] *Id.*

[34] *Id.* ¶ 178.

[35] *Id.*

[36] *Id.*

[37] *Id.*

[38] *Id.*

[39] *Id.*

November 2022 (the release of the chatbot ChatGPT to the public).[40] Other characteristics used in the match assessment include the book's Book Industry Standards and Communications ("BISAC") code, number of pages, an indicator of whether the book is part of a series, and the publication date.[41]

22. BISAC codes classify works by topic and are the standard in the U.S. used by companies throughout the supply chain.[42] BISAC codes are organized hierarchically, with broad Level 1 categories that are further subdivided into more specific subject areas (e.g., BISAC Level 1 (FICTION) is subdivided into Level 2 categories for Romance, Fantasy, Mystery & Detective, Science Fiction, Thrillers, etc.).[43]

23. The "treatment" of interest in this instance was the inclusion of a book in the training datasets of the In-Scope LLMs.[44] I focused on Plaintiffs' Asserted Works, which I assumed, for purposes of my analysis, were included in the training datasets.[45] The "treatment" group included [redacted] of the 212 Asserted Works that met five criteria: (1) the work was in the Circana data; (2) the work was sold separately; (3) the work had net sales of more than 10,000 units combined from January 1, 2017, to January 31, 2026; (4) the work had at least 12 months of positive net sales prior to June 2020; and (5) there were more than a specified number of control books with the same BISAC Level 2 code.[46]

---

[40] *Id.*

[41] *Id.*

[42] *Id.* ¶ 167.

[43] *Id.*

[44] *Id.* ¶ 179.

[45] *Id.*

[46] *Id.* ¶¶ 179–180.

These [redacted] Asserted Works make up [redacted] of the total net unit sales associated with all of the Asserted Works present in the Circana data.[47]

24. I grouped the [redacted] Asserted Works into 15 cells based on their BISAC Level 2 code.[48] I then selected up to 25 potential control books from each cell.[49] Cells with fewer than 10 potential control books were dropped from the analysis.[50]

25. The "control books" are books that were not in the training datasets of the In-Scope LLMs.[51] The control books included in this analysis met the following criteria: (1) the work sold more than 10,000 net units combined in the time period of January 1, 2017, to January 31, 2026; (2) the work had at least 12 months of positive net sales prior to June 2020; (3) the work was not present in any of the training datasets of the In-Scope LLMs made available in this litigation; (4) the work had the same BISAC Level 2 code as the cell of Asserted Works; (5) the work was available in Kindle format; (6) the work was an actual book; and (7) the works in the [redacted] cell needed to be more than 100 pages so they were more comparable to the Asserted Works in this cell.[52]

26. I ran the SC method for each of the 106 Asserted Works.[53] A summary of the results is presented in Exhibit 4a to my Opening Report, which is reproduced below.[54]

---

[47] *Id.* ¶ 180.

[48] *Id.* ¶ 183.

[49] *Id.* ¶ 184.

[50] *Id.*

[51] *Id.* ¶ 181.

[52] *Id.* ¶ 189.

[53] *Id.* ¶ 191.

[54] *Id.* ¶ 192. A p-value is a statistical calculation defined as "The smallest significance level at which the null hypothesis can be rejected. Equivalently, the largest significance level at which the null hypothesis cannot be rejected." *See* Jeffrey Wooldridge, *Introductory Econometrics*, 2012, p. 854.

Exhibit 4a

Sales Gaps Following July 2020 and November 2022

Post-Periods Starting July 2020 and December 2022

Date	# of Asserted Works in Sample	# with Net Sales > Synthetic Control			# with Net Sales < Synthetic Control		
		Total	(p-value <0.05)	(q-value <0.05)	Total	(p-value <0.05)	(q-value <0.05)
July 2020	[redacted]						
November 2022							

27. As shown above, for the November 2022 introduction date, [redacted] Asserted Work was found to have sales that exceeded the sales of its SC by a statistically significant amount (as determined by the p-value) and [redacted] Asserted Works were found to have sales less than their respective SCs by a statistically significant amount.[55] The remaining [redacted] Asserted Works were found to have sales consistent with their respective SCs.[56] For the July 2020 introduction date, [redacted] Asserted Works were found to have net sales greater than their respective SCs by a statistically significant amount (as determined by the p-value) and [redacted] Asserted Works were found to have net sales less than their respective SCs by a statistically significant amount.[57] The remaining [redacted] Asserted Works were found to have sales consistent with their respective SCs.[58]

28. Because of the multiple hypotheses being tested, I adjusted the standard p-values (which apply when only a single hypothesis is being tested) using generally accepted methods and assessed the statistical significance using the adjusted p-values (known as the "sharpened q-values").[59] Based on this analysis, none of the Asserted Works are found to have a statistically significant difference from their respective SCs for the November

[55] Leonard Opening Report ¶ 193.

[56] *Id.*

[57] *Id.*

[58] *Id.*

[59] *Id.*, n. 427.

2022 introduction date.[60] If a work does not have a statistically significant difference from its respective SCs, then we can conclude that the treatment effect (i.e., use of the work to train the In-Scope LLMs) had no effect on post-treatment outcomes for the treated unit (i.e., sales).[61]

29. The results of the SC method support the conclusion that the introduction of the first accused In-Scope LLM in July 2020 and the introduction of the chatbot ChatGPT to the public in November 2022 had no adverse impact on the sales of the Asserted Works.[62]

E. Analysis 5: Book Prices

30. I used the Barnes & Noble data to analyze whether book prices declined after July 2020 (when GPT-3 was made available via an API) and November 2022 (when the chatbot ChatGPT was made available to the public) and any resulting release of AI-generated books.[63] The price of a book that is part of a series (such as the Black Book series by James Patterson) tends to be similar to the price of the preceding book in the series.[64] I analyzed whether this pattern was disrupted for pairs of books in a series where the earlier book in the pair was released before GPT-3 API / chatbot ChatGPT introduction and the later book in the pair was released afterwards.[65] If the release of AI-generated works drove down the prices of the Asserted Works and other human-authored trade books, we would expect to see that a series book released after GPT-3 API / chatbot

[60] *Id.* ¶ 193. For the July 2020 introduction date, [redacted] of the Asserted Works were found to have a statistically significant difference from their respective SCs. *Id.*, n. 438.

[61] *Id.* ¶ 190.

[62] *Id.* ¶ 194.

[63] *Id.* ¶ 217.

[64] *Id.*

[65] *Id.*

ChatGPT introduction would have a lower price than would be expected based on (1) the price of its predecessor in the same series and (2) the pattern of series book pricing that existed before (or after) the introduction of ChatGPT.[66]

31. I used the fields in the Circana data for series name and series number (i.e., the order of titles in a given series) to identify series that met certain criteria: (1) the books in the series are in mass market paperback format and have BISAC codes of "[redacted]" and "[redacted]"; (2) the series includes [redacted] and no more than [redacted] (3) the series has books with publication dates before July 2020 and books with publication dates after November 2022.[67] This process identified [redacted] series and [redacted] Asserted Works as books within these series.[68]

32. To perform the analysis, I started with the first and second books from each of these series.[69] For each book, I calculated the average Barnes & Noble sales price over the first three months of its sales after release.[70] I then categorized the series into three groups: (1) both the first and second books were released prior to GPT-3 API / chatbot ChatGPT introduction, (2) both the first and second books were released after GPT-3 API / chatbot ChatGPT introduction, or (3) the first book was released prior to GPT-3 API / chatbot ChatGPT introduction and the second book was released after ChatGPT.[71] If the introduction of GPT-3 in an API or chatbot ChatGPT to the public led to "flooding" that in turn led to price decreases, the price of the second book relative to the first book would

[66] *Id.*

[67] *Id.* ¶ 218.

[68] *Id.*

[69] *Id.* ¶ 219.

[70] *Id.*

[71] *Id.*

be lower for series in group 3 (where the second book, but not the first book, was potentially affected by any "flooding") than for series in group 1 (where neither book was potentially affected by any "flooding") or group 2 (where both books were potentially affected by any "flooding").[72]

33. I ran a regression of the logarithm of the second book's price on the logarithm of the first book's price; indicator variables for which group the series was in; and the ratio of the Consumer Price Index ("CPI") value when the second book was published to the CPI value when the first book was published (this is to account for changes in general inflation between when the two books were published).[73] The results are in Exhibit 5a to my Opening Report and show that group 3 [redacted] than groups 1 or 2.[74] That is, the introduction of ChatGPT and any resulting "flooding" did not appear to [redacted].[75] I repeated the analysis using the second and third books in each series and found [redacted], which are shown in Exhibit 5b to my Opening Report.[76] Exhibits 5a and 5b are reproduced below.

[72] *Id.*

[73] *Id.* A CPI value reflects how prices for a type of good (here, books) have changed relative to a base period.

[74] *Id.* ¶ 220.

[75] *Id.*

[76] *Id.*

Exhibit 5a
Pricing Analysis
Book 1 v. Book 2 within Series

Date	Statistic	log(Book 1 Price)	Group 3 Dummy	log(CPI Ratio)
July 2020		[redacted]		
	Coefficient			
	SE			
	p-value			
November 2022				
	Coefficient			
	SE			
	p-value			

Exhibit 5b
Pricing Analysis
Book 2 v. Book 3 within Series

Date	Statistic	log(Book 2 Price)	Group 2 Dummy	Group 3 Dummy	log(CPI Ratio)
July 2020		[redacted]			
	Coefficient				
	SE				
	p-value				
November 2022					
	Coefficient				
	SE				
	p-value				

34. In sum, this empirical analysis is consistent with the proposition that any release of AI-generated works after the introduction of GPT-3 in an API / chatbot ChatGPT to the public had no effect on the prices of trade books like the Asserted Works.[77]

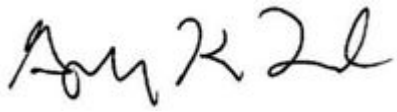
* * *

35. I have also reviewed the Opening Expert Report of Mr. Michael Lasinski served on April 24, 2026; the Rebuttal Expert Report of Mr. Lasinski served on May 29, 2026; the Reply

[77] *Id.* ¶ 221.

Expert Report of Mr. Lasinski served on June 26, 2026; and the Supplemental Expert Report of Mr. Lasinski served on August 2, 2026. In these reports, Mr. Lasinski did not perform any empirical analysis of his own to demonstrate that Plaintiffs lost sales of the Asserted Works due to OpenAI's alleged use of the Asserted Works. For instance, he did not conduct any control-group study to determine the purported impact that the release of GPT-3 in OpenAI's API in July 2020 or the release of the chatbot ChatGPT in November 2022 had on sales of the Asserted Works or pricing of the Asserted Works.

I declare under penalty of perjury that the foregoing is true and correct, and that I have personal knowledge of these facts and am competent to testify thereto if called upon to do so. Executed this 3rd day of September, 2026, in Hillsborough, California.



GREGORY K. LEONARD