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-SHS-OTW Case No. 1:23-cv-10211-SHS-OTW	**ECF CASE** 1:25-MD-03143-SHS/OTW

CLASS PLAINTIFFS' RULE 56.1 STATEMENT OF UNDISPUTED MATERIAL FACTS IN SUPPORT OF CLASS PLAINTIFFS' MOTION FOR PARTIAL SUMMARY JUDGMENT

TABLE OF CONTENTS

I. The Parties, Asserted Works, and Accused Models 1
A. Class Plaintiffs and Asserted Works 1
1. David Baldacci and Columbus Rose, Ltd. 2
2. Taylor Branch 2
3. Michael Connelly and Hieronymus Inc. 3
4. Sylvia Day and Sylvia Day LLC 3
5. Jonathan Franzen 4
6. Christopher Golden and Daring Greatly Corporation 4
7. Andrew Sean Greer 5
8. John Grisham and Belfry Holdings, Inc. 5
9. David Henry Hwang 6
10. George R.R. Martin & Wo & Shade, LLC 6
11. Jodi Picoult 7
12. Stacy Schiff and Gattopardo, LLC 7
13. James Shapiro 7
14. The Authors Guild (Estate of Mignon Eberhart) 8
B. The OpenAI Entities 8
C. OpenAI's Conversion from Nonprofit to Commercial Enterprise 10
D. Generative AI and Large Language Models 11
E. The Accused Models and ChatGPT 13
F. Microsoft 15
II. [redacted] 16
[redacted] 16
[redacted] 18
C. [redacted] 21
D. [redacted] 24
E. [redacted] 25
F. [redacted] 33

1. [redacted] 33
2. [redacted] 35
3. [redacted] 36
4. [redacted] 37
G. [redacted] 38
H. [redacted] 41
I. [redacted] 45
J. [redacted] 49
1. OpenAI's Contemporaneous Knowledge 49
2. Defendants' Admissions [redacted] 57
III. OpenAI Also Downloaded Books from [redacted] 61
A. [redacted] 61
B. [redacted] 66
C. OpenAI Also Acquired Books [redacted] 67
IV. [redacted] 71
A. [redacted] 71
B. [redacted] 79
C. [redacted] 83
D. [redacted] 85
E. [redacted] 85
F. OpenAI Created Multiple Copies of Books in Training 85
G. The Accused Models Reproduce Class Plaintiffs' Works [redacted] 87
V. [redacted] Distributed Books [redacted] 91
A. [redacted] 91
B. [redacted] 95
C. [redacted] 96
VI. Microsoft's Role in OpenAI's Infringement 97
A. [redacted]
B. Microsoft Formed a Partnership with OpenAI 101
C. [redacted] 104

D. [redacted] 108
E. [redacted] 110
F. [redacted] 113
G. Microsoft Benefitted Financially from OpenAI's Infringement 115
VII. The Licensing Market and Market Harm 116
A. [redacted] 116
B. Defendants' [redacted] Agreements 129
C. Other Licensing Activities 136
D. The Proliferation of AI-Generated Books 141
E. AI Poses Substantial Risks [redacted] 150
VIII. Copies of Plaintiffs' Works Are Present i[redacted] 153
[redacted]

Pursuant to Local Rule 56.1, Plaintiffs David Baldacci, Taylor Branch, Michael Connelly, Sylvia Day, Jonathan Franzen, Christopher Golden, Andrew Sean Greer, John Grisham, David Henry Hwang, Matthew Klam, George R.R. Martin, Jodi Picoult, Stacy Schiff, James Shapiro, The Authors Guild, Belfry Holdings, Inc., Columbus Rose, Ltd., Daring Greatly Corporation, Gattopardo, LLC, Hieronymus, Inc., Sylvia Day LLC, and Wo & Shade LLC (collectively, "Class Plaintiffs") respectfully submit this Statement of Undisputed Material Facts in Support of Class Plaintiffs' Motion for Partial Summary Judgment.

I. The Parties, Asserted Works, and Accused Models

A. Class Plaintiffs and Asserted Works

1. Class Plaintiffs are the bestselling and award-winning book authors identified below, the entities through which certain of those authors conduct their professional activities, and The Authors Guild, the oldest and largest professional organization for published writers in the United States.

2. The works at issue in this Statement are the copyrighted books listed in Appendix A to this Statement (the "Asserted Works"), which sets out the author, copyright holder, copyright registration number, first publication date, effective copyright registration date, and publishing royalty recipient for each work. While Plaintiffs' motion addresses the Asserted Works, Plaintiffs proudly represent the interests of copyright holders who were harmed by Defendants' conduct. Nelson Decl. Ex. 51 at 186:14-187:19; Nelson Decl. Ex. 5 at 220:7-221:14.

3. The copyright claimant registered each Asserted Work with the United States Copyright Office within five years of its first publication. Appx. A.

4. Each Asserted Work is currently available for sale. Appx. A.

1. David Baldacci and Columbus Rose, Ltd.

5. Plaintiff David Baldacci is an author, philanthropist, and lawyer who resides in Virginia. Nelson Decl. Ex. 104.

6. Baldacci has published more than fifty novels for adults, every one a national and international bestseller, and his books have sold more than 200 million copies in over eighty countries. *Id.*

7. Plaintiff Columbus Rose, Ltd. is a company through which David Baldacci conducts his professional activities, including the ownership and management of copyrights in his literary works. Nelson Decl. Ex. 5 at 104:6–19.

8. Columbus Rose, Ltd. owns the valid copyrights in forty published books authored by David Baldacci, including titles in the *Camel Club* series, the *King & Maxwell* series, the *Will Robie* series, and the *Amos Decker* series. Appx. A.

2. Taylor Branch

9. Plaintiff Taylor Branch is an author and historian who resides in Maryland. Nelson Decl. Ex. 69.

10. Branch's *Parting the Waters* won the Pulitzer Prize for History and the National Book Critics Circle Award, and it opens Branch's three-volume *America in the King Years* chronicle of Martin Luther King, Jr. and the civil rights movement. *Id.*

11. Branch has also received a MacArthur Fellowship and the National Humanities Medal. *Id.*

12. Taylor Branch owns the valid copyrights in two published books: *Parting the Waters: America in the King Years, 1954-63* and *Pillar of Fire: America in the King Years, 1963-65*. Appx. A.

3. Michael Connelly and Hieronymus Inc.

13. Plaintiff Michael Connelly is an author who resides in Florida. Nelson Decl. Ex. 110.

14. Connelly has written more than forty novels that have sold over ninety million copies worldwide, and his debut, *The Black Echo*, won the Mystery Writers of America Edgar Award for Best First Novel. *Id.*

15. He executive produces *Bosch*, *Bosch: Legacy*, and *Ballard* on Amazon Prime Video and *The Lincoln Lawyer* on Netflix, all of which are based on his books. *Id.*

16. Plaintiff Hieronymus Inc. is a corporation through which Michael Connelly conducts his professional activities, including the ownership and management of copyrights in his literary works. Nelson Decl. Ex. 15 at 33:8–13.

17. Hieronymus Inc. owns the valid copyrights in thirty-six published books authored by Michael Connelly, including the *Harry Bosch* series, the *Mickey Haller* (Lincoln Lawyer) series, and the *Renee Ballard* series. Appx. A.

4. Sylvia Day and Sylvia Day LLC

18. Plaintiff Sylvia Day is an author who resides in Nevada. Nelson Decl. Ex. 124.

19. Day has written more than twenty award-winning novels translated into forty-one languages and is a number one bestseller in twenty-nine countries. *Id.*

20. Plaintiff Sylvia Day LLC is a limited liability company through which Sylvia Day conducts her professional activities, including the ownership and management of copyrights in her literary works. Nelson Decl. Ex. 66 at 31:1–11; Nelson Decl. Ex. 137.

21. Sylvia Day LLC and Sylvia Day own the valid copyrights in twenty-nine published books authored by Sylvia Day, including the *Crossfire* series, which begins with *Bared to You*. Appx. A.

5. Jonathan Franzen

22. Plaintiff Jonathan Franzen is an author who resides in California. Nelson Decl. Ex. 100.

23. Franzen's *The Corrections* won the National Book Award and was ranked on the New York Times Book Review's 100 Best Books of the 21st Century. Nelson Decl. Ex. 101.

24. *Time* magazine featured Franzen on its cover as the "Great American Novelist," with an article marking the publication of his novel, Freedom. Nelson Decl. Ex. 100; Nelson Decl. Ex. 97.

25. Jonathan Franzen owns the valid copyrights in five published books, including *The Corrections*, *Freedom*, and *Crossroads*. Appx. A.

6. Christopher Golden and Daring Greatly Corporation

26. Plaintiff Christopher Golden is an author who resides in Massachusetts. Nelson Decl. Ex. 103.

27. Golden is a *New York Times* bestselling horror and thriller novelist who has been called the "the king of horror-thriller." *Id.*

28. His work has been published in dozens of languages, and he is the winner of the Bram Stoker Award. *Id.*

29. Plaintiff Daring Greatly Corporation is a Massachusetts-based corporation through which Christopher Golden conducts his professional activities, including the management of royalties in his literary works. Nelson Decl. Ex. 717 at 257:3–11.

30. Christopher Golden owns the valid copyright in one published book: *Ararat*. Appx. A.

7. Andrew Sean Greer

31. Plaintiff Andrew Sean Greer is an author who resides in California. Nelson Decl. Ex. 70.

32. Greer's *Less* won the 2018 Pulitzer Prize for Fiction and was a national bestseller. *Id.*

33. Greer has also received a Guggenheim Fellowship, the California Book Award, and the New York Public Library Young Lions Award. *Id.*

34. Andrew Sean Greer owns the valid copyrights in four published books, including *Less* and its sequel *Less Is Lost*, as well as *The Confessions of Max Tivoli* and *The Story of a Marriage*. Appx. A.

8. John Grisham and Belfry Holdings, Inc.

35. Plaintiff John Grisham is an author who resides in Virginia. Nelson Decl. Ex. 85.

36. Grisham has published more than fifty consecutive number one bestsellers, and his books have been translated into nearly fifty languages. *Id.*

37. He has twice won the Harper Lee Prize for Legal Fiction and received the Library of Congress Creative Achievement Award for Fiction. *Id.*

38. Plaintiff Belfry Holdings, Inc. is a corporation through which John Grisham conducts his professional activities, including the ownership and management of copyrights in his literary works. Nelson Decl. Ex. 24 at 16:10–25.

39. Belfry Holdings, Inc. and John Grisham own the valid copyrights in twenty-three published books authored by John Grisham, including *A Time to Kill*, *The Firm*, *The Pelican Brief*, and *The Client*. Appx. A.

9. David Henry Hwang

40. Plaintiff David Henry Hwang is an author and playwright who resides in New York. Nelson Decl. Ex. 793.

41. He has been described by the New York Times as "a true original" and by *Time* magazine as "the first important dramatist of American public life since Arthur Miller." Nelson Decl. Ex. 122.

42. Hwang's *M. Butterfly* won the 1988 Tony Award for Best Play, making Hwang the first Asian American playwright to win that award, and was a Pulitzer Prize finalist. Nelson Decl. Ex. 94.

43. David Henry Hwang owns the valid copyrights in three published books, including *M. Butterfly*. Appx. A.

10. George R.R. Martin & Wo & Shade, LLC

44. Plaintiff George R. R. Martin is an author who resides in New Mexico. Nelson Decl. Ex. 83. Martin's *A Song of Ice and Fire* novels have sold more than 85 million copies worldwide and appear in 47 languages. Nelson Decl. Ex. 117.

45. He has won numerous Hugo and Nebula Awards, and the World Fantasy Convention honored him with its Life Achievement Award in 2012. Nelson Decl. Ex. 107.

46. HBO adapted *A Song of Ice and Fire* as *Game of Thrones*, which earned Martin three Emmy Awards as a producer. *Id.*

47. Plaintiff Wo & Shade, LLC is a limited liability company through which George R. R. Martin conducts his professional activities, including the ownership and management of copyrights in his literary works. Nelson Decl. Ex. 34 at 16:13–19.

48. Wo & Shade, LLC owns the valid copyrights in fifteen published books authored by George R.R. Martin, including novels in the *A Song of Ice and Fire* series. Appx. A.

11. Jodi Picoult

49. Plaintiff Jodi Picoult is an author who resides in New Hampshire. Nelson Decl. Ex. 86.

50. Picoult has written twenty-nine novels translated into forty languages, with an estimated 63.7 million copies in print. *Id.*

51. Her honors include the New England Bookseller Award for Fiction and the Sarah Josepha Hale Award. *Id.*

52. Jodi Picoult owns the valid copyrights in twenty-six published books, including *My Sister's Keeper*, *Nineteen Minutes*, and *Small Great Things*. Appx. A.

12. Stacy Schiff and Gattopardo, LLC

53. Plaintiff Stacy Schiff is an author who resides in New York. Nelson Decl. Ex. 95.

54. She is a Pulitzer Prize winner, and her later books, among them *Cleopatra: A Life* and *The Revolutionary: Samuel Adams*, have been national bestsellers. *Id.*

55. Schiff has also received fellowships from the Guggenheim Foundation and the National Endowment for the Humanities and was a Director's Fellow at the Cullman Center for Scholars and Writers at the New York Public Library. *Id.*

56. Plaintiff Gattopardo, LLC is a limited liability company through which Stacy Schiff conducts her professional activities, including the management of royalties in her literary works. Nelson Decl. Ex. 51 at 16:7–12.

57. Stacy Schiff owns the valid copyrights in three published books, including *The Witches* and *Cleopatra: A Life*. Appx. A.

13. James Shapiro

58. Plaintiff James Shapiro is an author who resides in New York. Nelson Decl. Ex. 82.

59. Shapiro is the Larry Miller Professor of English and Comparative Literature at Columbia University and the Shakespeare Scholar in Residence at the Public Theater in New York. Nelson Decl. Ex. 82; Nelson Decl. Ex. 109.

60. *A Year in the Life of William Shakespeare: 1599* won the Samuel Johnson Prize, and *The Year of Lear* won the James Tait Black Memorial Prize. Nelson Decl. Ex. 109.

61. James Shapiro owns the valid copyrights in three published books, including *A Year in the Life of William Shakespeare: 1599* and *The Year of Lear: Shakespeare in 1606*. Appx. A.

14. The Authors Guild (Estate of Mignon Eberhart)

62. Plaintiff the Authors Guild is a registered 501(c)(6) not-for-profit association that advocates for professional writers and provides services to more than 15,000 members. Nelson Decl. Ex. 46 at 21:13-22:9.

63. Founded in 1912, the Authors Guild is the oldest and largest professional organization for published writers in the United States. Nelson Decl. Ex. 73.

64. Mignon Eberhart was an author who published more than fifty mystery novels over a sixty-year career. Nelson Decl. Ex. 87.

65. She is a former president of the Mystery Writers of America and a recipient of the organization's Grand Master Award. *Id.*

66. Eberhart bequeathed the copyrights in her published books to the Authors Guild, which acquired those copyrights upon her death. Nelson Decl. Ex. 136.

67. The Authors Guild owns the valid copyrights in five published books authored by Mignon Eberhart. Appx. A.

B. The OpenAI Entities

68. Defendant OpenAI, Inc. (now known as OpenAI Foundation) is a Delaware corporation headquartered in California. OpenAI's Answer ¶ 44, ECF No. 724.

69. [redacted]

70. Defendant OpenAI GP, LLC is a Delaware limited liability company headquartered in California. OpenAI's Answer ¶ 45, ECF No. 724.

71. [redacted]

72. Defendant OpenAI OpCo LLC (formerly known as OpenAI LP) is a Delaware limited liability company headquartered in California. OpenAI's Answer ¶ 46, ECF No. 724.

73. [redacted]

74. Defendant OpenAI, LLC is a Delaware limited liability company headquartered in California. OpenAI's Answer ¶ 47, ECF No. 724.

75. [redacted]

76. Defendant OpenAI Global LLC is a Delaware limited liability company headquartered in California. OpenAI's Answer ¶ 48, ECF No. 724.

77. During at least some portion of the relevant time period, Microsoft Corporation held a minority stake in OpenAI Global LLC. *Id.*

78. Defendant OpenAI Holdings, LLC is a Delaware limited liability company. OpenAI's Answer ¶ 50, ECF No. 724.

79. [redacted]

80. On March 4, 2026, OpenAI moved to substitute OpenAI Group PBC for OpenAI Holdings, LLC because "on December 31, 2025, OpenAI Holdings, LLC merged into OpenAI Group PBC and no longer exists as a standalone entity." ECF No. 1386 at 2. The Court granted the motion on March 26, 2026. ECF No. 1464 at 1.

81. Defendant OpenAI Startup Fund I LP is a Delaware limited partnership headquartered in California. OpenAI's Answer ¶ 51, ECF No. 724.

82. Defendant OpenAI Startup Fund GP I LLC is a Delaware limited liability company headquartered in California. OpenAI's Answer ¶ 52, ECF No. 724.

83. Defendant OpenAI Startup Fund Management LLC is a Delaware limited liability company headquartered in San Francisco, California. OpenAI's Answer ¶ 53, ECF No. 724.

84. This Statement collectively refers to OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC and its successor by merger, OpenAI Group PBC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC as "OpenAI," and refers to OpenAI and Microsoft Corporation collectively as "Defendants."

C. OpenAI's Conversion from Nonprofit to Commercial Enterprise

85. [redacted]

87. [redacted]

93. The products OpenAI developed and commercialized through that for-profit structure are the GPT large language models and ChatGPT described below. *See infra* Section I.E.

D. Generative AI and Large Language Models

94. Large language models ("LLMs") are a kind of "generative artificial intelligence," a term that refers to systems that generate outputs, such as text and images, based on the inputs (i.e. data) on which the system was trained. Nelson Decl. Ex. 679.

95. The core of the LLM training process is called "pre-training." All of the in-suit models went through this "pre-training" process. Nelson Decl. Ex. 680.

96. During pre-training, LLMs are trained using a process by which the in-development model is exposed to incomplete text data, such as part of a sentence, drawn from a corpus of "training data." Nelson Decl. Ex. 681. The model then outputs a prediction of the next word—or fragment of a word, known as a "token"— in the sequence. *Id*. At the start of this pre-training process, outputs are effectively random. *Id*.

97. The model's predicted output is then checked against the training data. If the model's output does not match the training data, the model training code adjusts the model's mathematical components so that the incorrect guess is penalized and the correct answer is favored. Nelson Decl. Ex. 700.

98. These internal components are known as "weights" or "parameters." The measurement of the extent to which the model's prediction varies from the actual data in its training data is known as "loss." Nelson Decl. Ex. 700.

99. [redacted]

102. [redacted]

[redacted]

E. The Accused Models and ChatGPT

104. The models at issue are GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini (the "Accused Models"). Opinion & Order (Oct. 27, 2025), ECF 705 at 1.

105. [redacted]

[redacted]

109. [redacted]

116. [redacted]

F. Microsoft

121. Defendant Microsoft Corporation is a Washington-based corporation that develops and sells software, cloud computing services, and artificial intelligence products. [redacted]

122. [redacted]

II. [redacted]t

A. Library Genesis

126. Library Genesis ("LibGen") is a shadow library—a website that collects, hosts, and distributes digital copies of copyrighted books without charge and without authorization from the authors and publishers who hold the rights to those books. Seeley Decl. Ex. A at ¶¶ 15–16; Nelson Decl. Ex. 13 at 19:20–20:20; Nelson Decl. Ex. 64 at 22:14–23:12.

127. LibGen originated in Russia in 2008. Seeley Decl. Ex. A at ¶ 21. At its inception, it incorporated copyrighted books from earlier pirate libraries. *Id.*

128. [redacted]

129. The only person ever to submit a declaration in court on behalf of LibGen "refer[red] to the websites' activities as 'pirating.'" Nelson Decl. Ex. 726.

130. Courts have repeatedly ordered LibGen to shut down and have entered substantial judgments against it. In late October 2015, the District Court for the Southern District of New York ordered LibGen to shut down and to suspend use of the domain name libgen.org. *See id.* at *4–6. The site remained accessible through alternate domains. Seeley Decl. Ex. A at ¶ 21.

131. In 2017, Elsevier obtained a $15 million judgment and a permanent injunction against LibGen. *See* Nelson Decl. Ex. 808; Seeley Decl. Ex. A at ¶ 21.

132. In 2023, a group of United States educational publishers sued LibGen for infringement. Nelson Decl. Ex. 732 at 1. The publishers obtained a default judgment in 2024 for injunctive relief and $30 million. *See id.* at 5–6; Seeley Decl. Ex. A at ¶ 21.

133. The Office of the United States Trade Representative has repeatedly identified LibGen as a notorious market for counterfeiting and piracy. USTR first listed LibGen in its 2017 Out-of-Cycle Review of Notorious Markets. Seeley Decl. Ex. A at ¶ 24. USTR listed LibGen again in its 2019, 2020, 2021, 2023, 2024, and 2025 Reviews of Notorious Markets for Counterfeiting and Piracy. *Id.* at ¶ 44; Shan Decl. Ex. A. at ¶ 46; Nelson Decl. Ex. 728.

134. The 2023 Review reports that LibGen hosts 2.4 million non-fiction books and 2.2 million fiction books, "many of which are unauthorized copies of copyright protected content." Nelson Decl. Ex. 729.

135. United States law enforcement has also acted against LibGen and other shadow libraries. *See, e.g.*, Seeley Decl. Ex. A at ¶ 21; Nelson Decl. Ex. 731. In 2022, United States law enforcement seized hundreds of mirror site domains associated with LibGen. Seeley Decl. Ex. A at ¶ 21; Nelson Decl. Ex. 733. LibGen and its network of mirror sites remain active. Seeley Decl. Ex. A at ¶ 21.

136. LibGen divides its books collection into two categories, fiction and nonfiction. Shan Decl. Ex. A. at ¶ 47. LibGen assigns every uploaded file an Md5 hash, a unique code that identifies that file. *Id.* LibGen lists metadata for the files it collects, including title, author, ISBN, upload date, file format, and Md5 hash. *Id.*

137. As of October 2019, LibGen fiction contained 2.18 million books. *Id.* at ¶ 47. As of January 2020, LibGen nonfiction contained 2.46 million books. *Id.* During the period relevant to this action, LibGen hosted approximately 4.6 million books.

B. [redacted]

138. [redacted]

142. [redacted]

151. [redacted]

160. [redacted]

168. [redacted]

[redacted]

182. In February 2019, OpenAI published the paper Language Models are Unsupervised Multitask Learners, which described the training of GPT-2. Nelson Decl. Ex. 77.

183. [redacted]

188. [redacted]

E. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

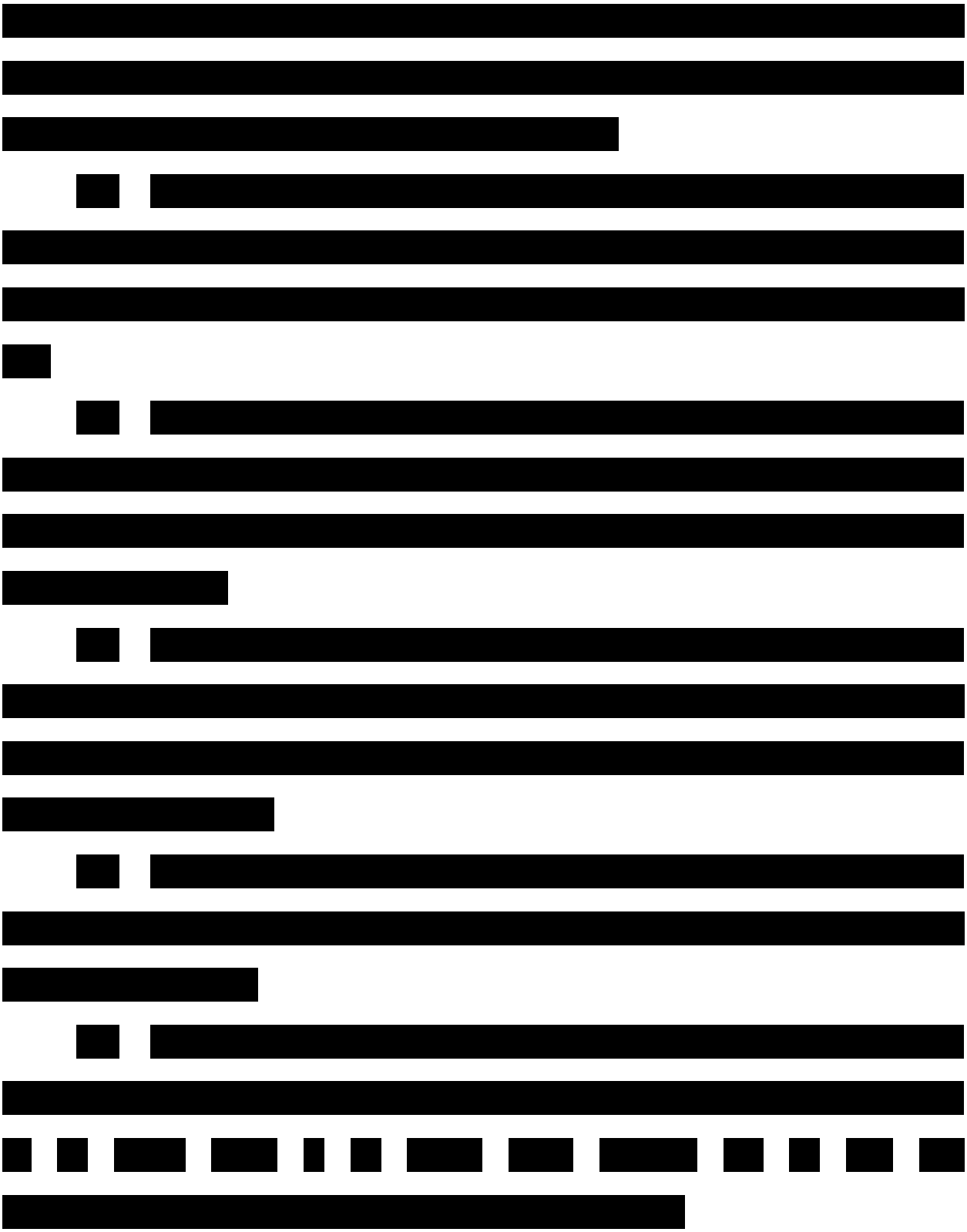
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

2. **Who Could Access the Books**

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

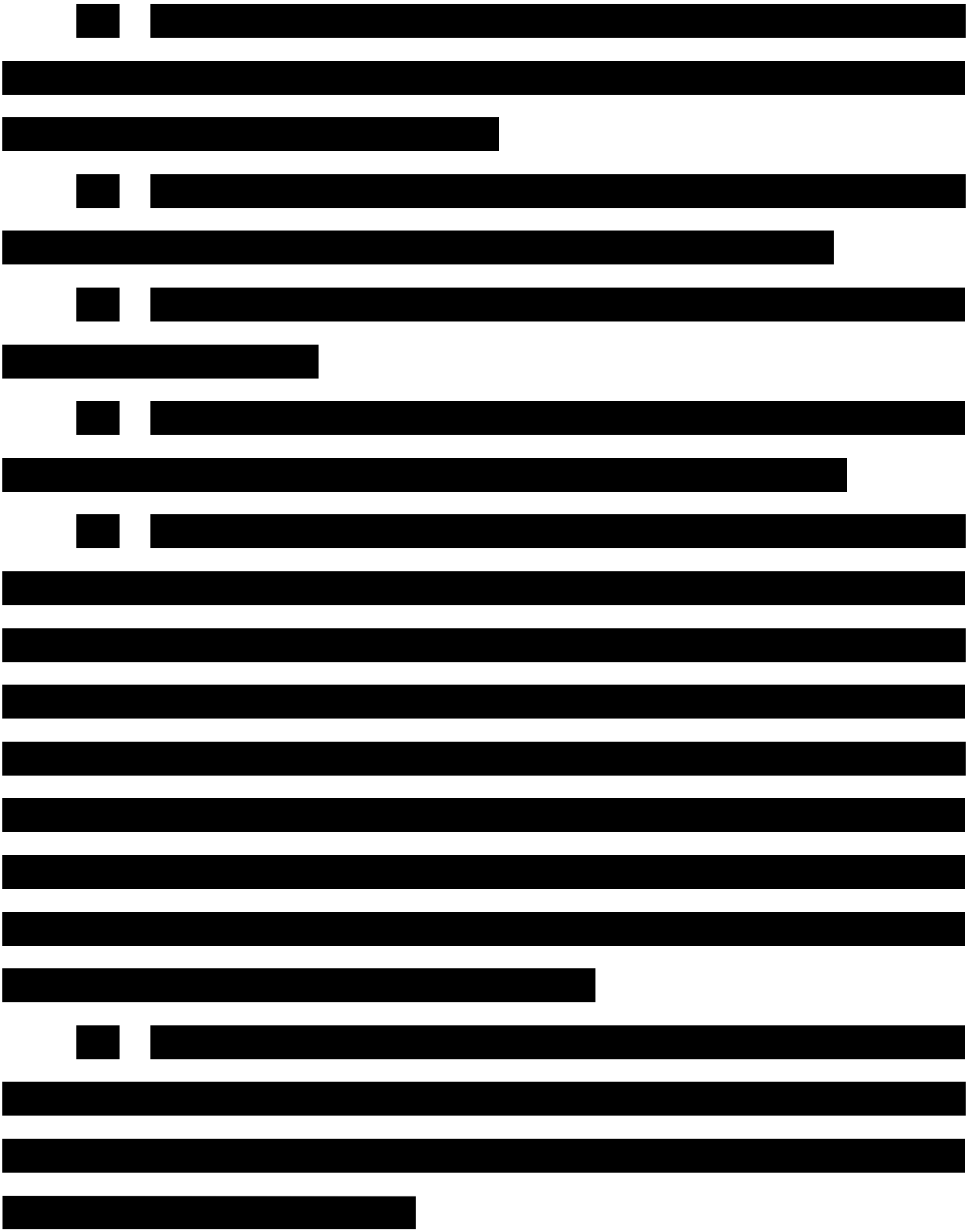
[redacted]

[redacted]

[redacted]



[redacted]



[redacted]

[redacted]

1. OpenAI's Contemporaneous Knowledge

330. [redacted]

336. [redacted]

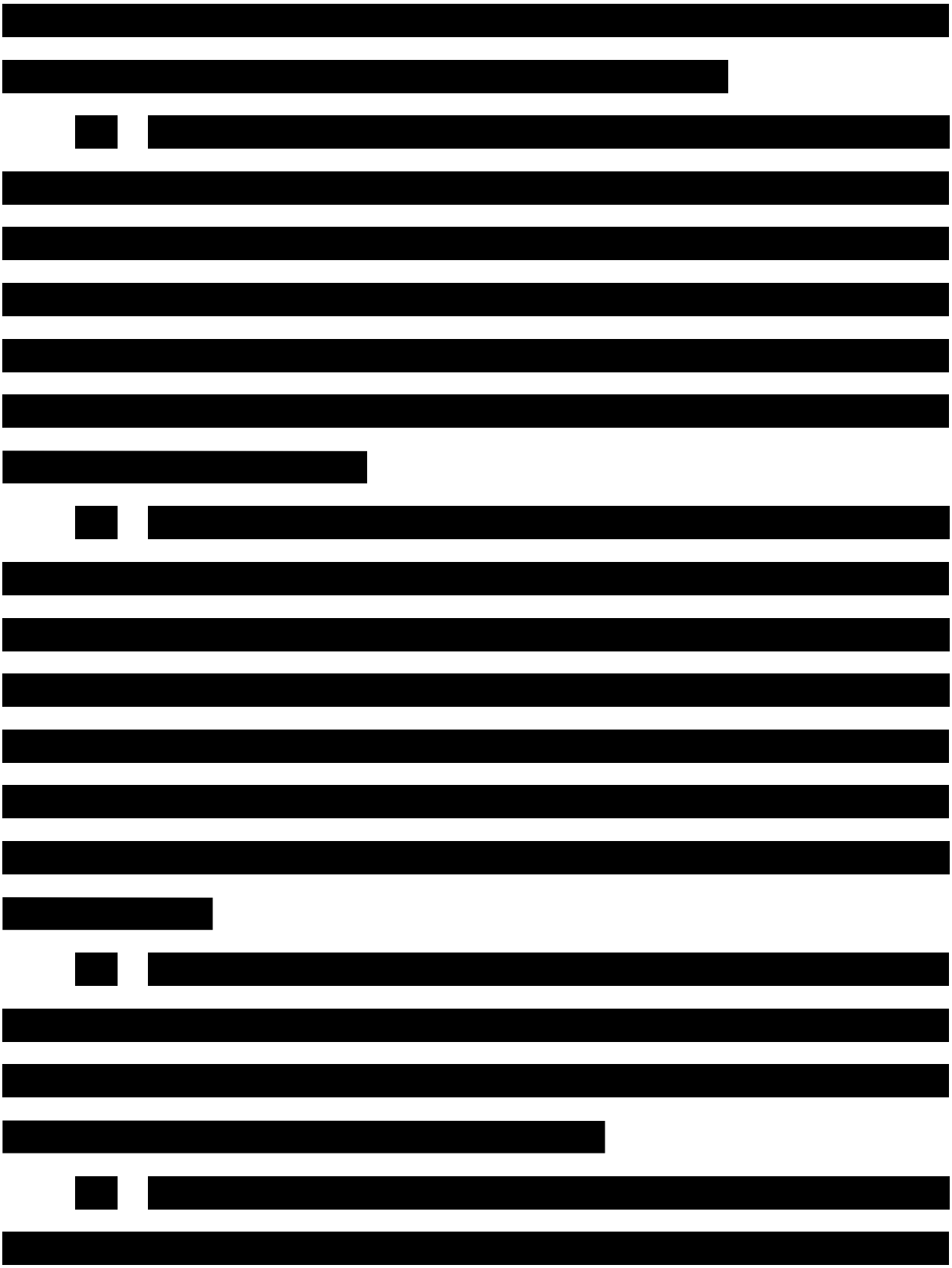
[redacted]

[redacted]



[redacted]





[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

A. OpenAI Knew Books Were High-Quality Training Data

448. In June 2018, OpenAI published a paper titled "Improving Language Understanding by Generative Pre-Training," which introduced OpenAI's first large language model, later called GPT-1. Nelson Decl. Ex. 78.

449. The paper states that OpenAI trained the model on a collection of books called "BooksCorpus," containing "over 7,000 unique" unpublished books from "a variety of genres including Adventure, Fantasy, and Romance." *Id.* at 4.

450. The paper further states: "Crucially, it contains long stretches of contiguous text, which allows the generative model to learn to condition on long-range information." *Id.* at 4. The paper contrasts BooksCorpus with an alternative dataset that was "shuffled at a sentence level – destroying long-range structure." *Id.* at 5.

[redacted]

[redacted]



460. [redacted]

466. [redacted]

[redacted]

478. [redacted]

[redacted]

486. The understanding that books were valuable training data was widely shared in the LLM industry, including among OpenAI's competitors Anthropic and Google, and among third-party researchers. Zhao Decl. Ex. A at ¶¶ 46–72.

487. When describing the value of receiving an additional 1T tokens, an Anthropic researcher stated that "books represent very high quality pretraining data." Nelson Decl. Ex. 740.

488. Anthropic researchers found that long-context (including books) are crucial to an LLM's ability to produce high-quality text "rather than low-quality internet speak." Nelson Decl. Ex. 741, at -434.

489. This is what Anthropic described as "The Good Writing Hypothesis," which amounts to the idea that in order to be effective "[m]odels just need good writing" in their training data. *Id.* at -433.

490. Anthropic's financial forecast described books as part of how it would compete with OpenAI. Anthropic stated that it could compete by "continu[ing] to release the same products as OpenAI . . . but offer them to customers at a lower price." Nelson Decl. Ex. 742.

491. When describing how it would effectuate this strategy, Anthropic stated that data quality would be a "key contributor." *Id.*

492. As Anthropic described it "[a]dding higher quality data (particularly books that aren't publicly available on the internet) allows smaller and cheaper models to outperform larger models." *Id.*

493. Books were listed first on Anthropic's list of sources of additional "High-Quality Data." Nelson Decl. Ex. 741.

494. [redacted]

B. [redacted]

[redacted]

499. [redacted]

505. [redacted]

[redacted]

C. [redacted]

[redacted]

[redacted]

[redacted]

G. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

B. [redacted]

[redacted]

[redacted]

[redacted]

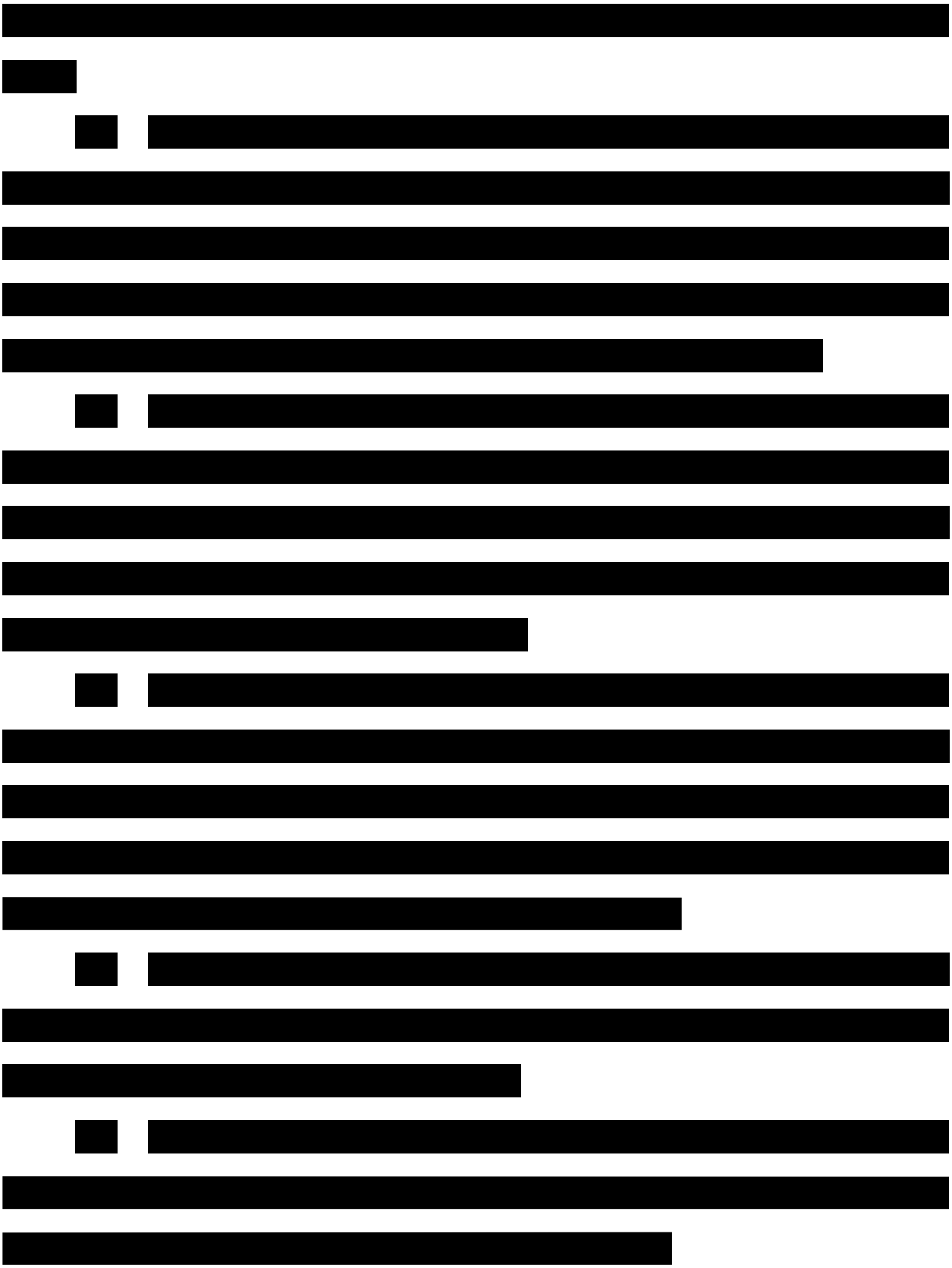
VI. Microsoft's Role in OpenAI's Infringement

A. [redacted]

[redacted]

[redacted]

[redacted]



[redacted]

B. Microsoft Formed a Partnership with OpenAI

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

D. [redacted]

[redacted]

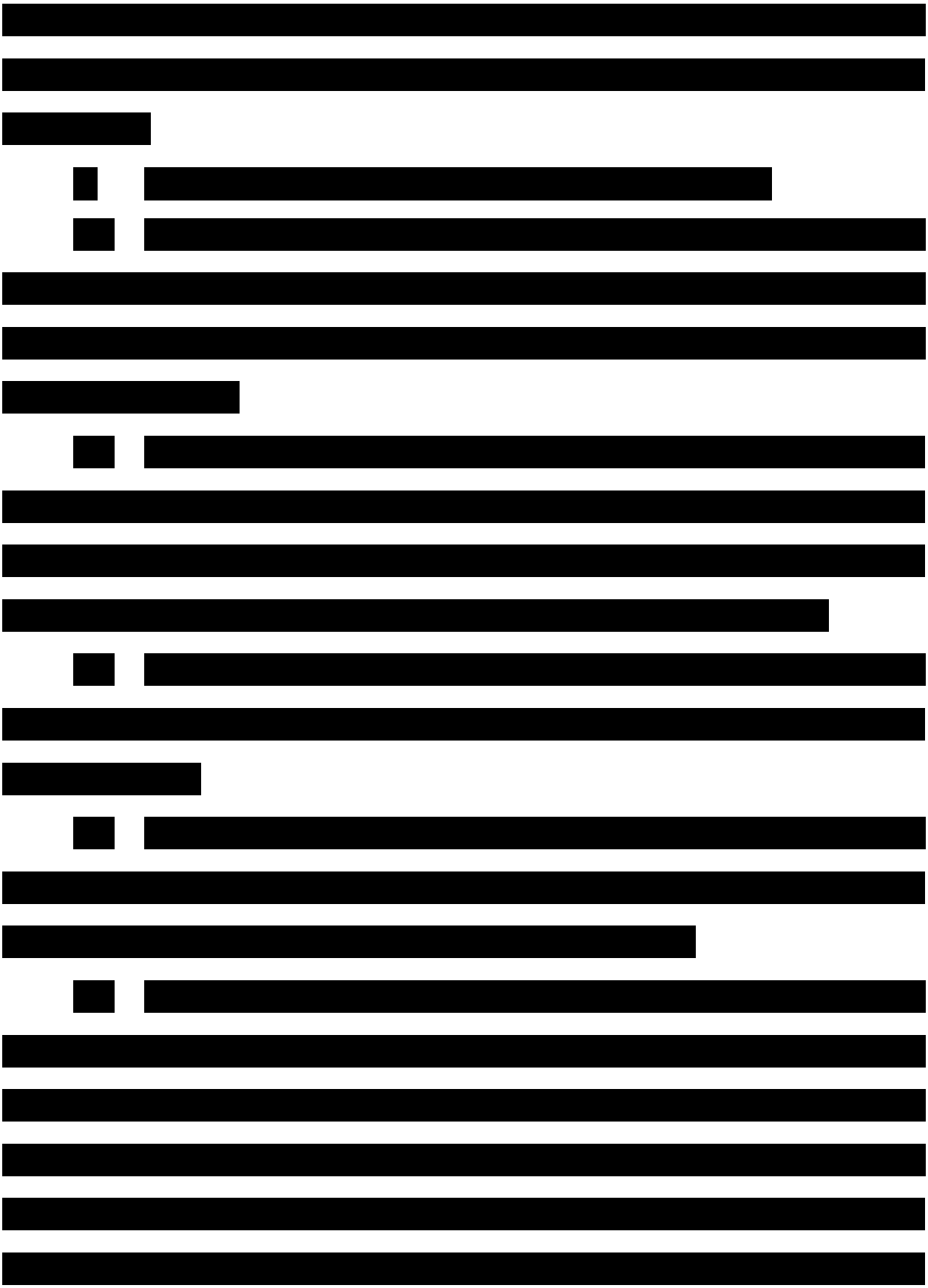
[redacted]

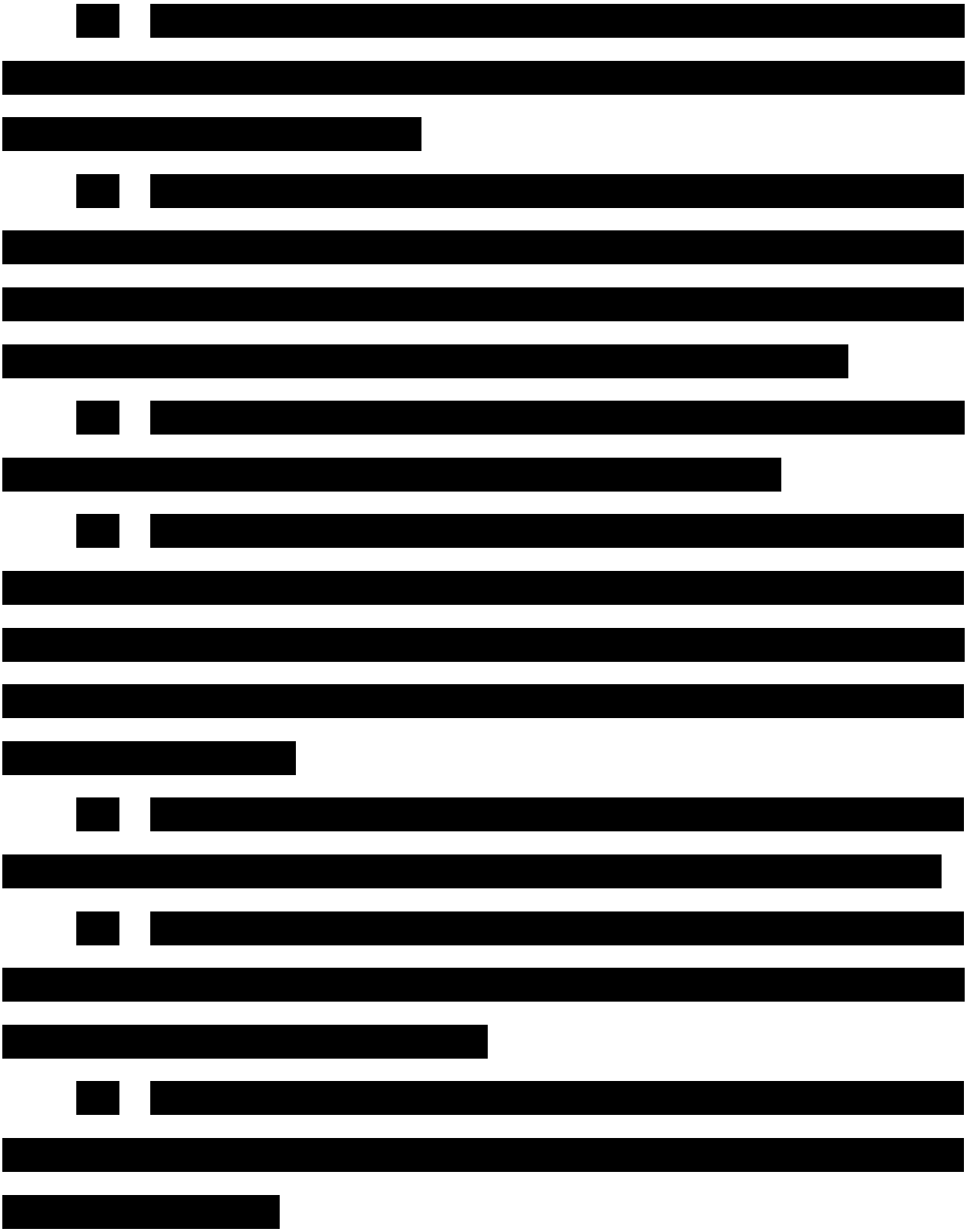
[redacted]

[redacted]

[redacted]

[redacted]





[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

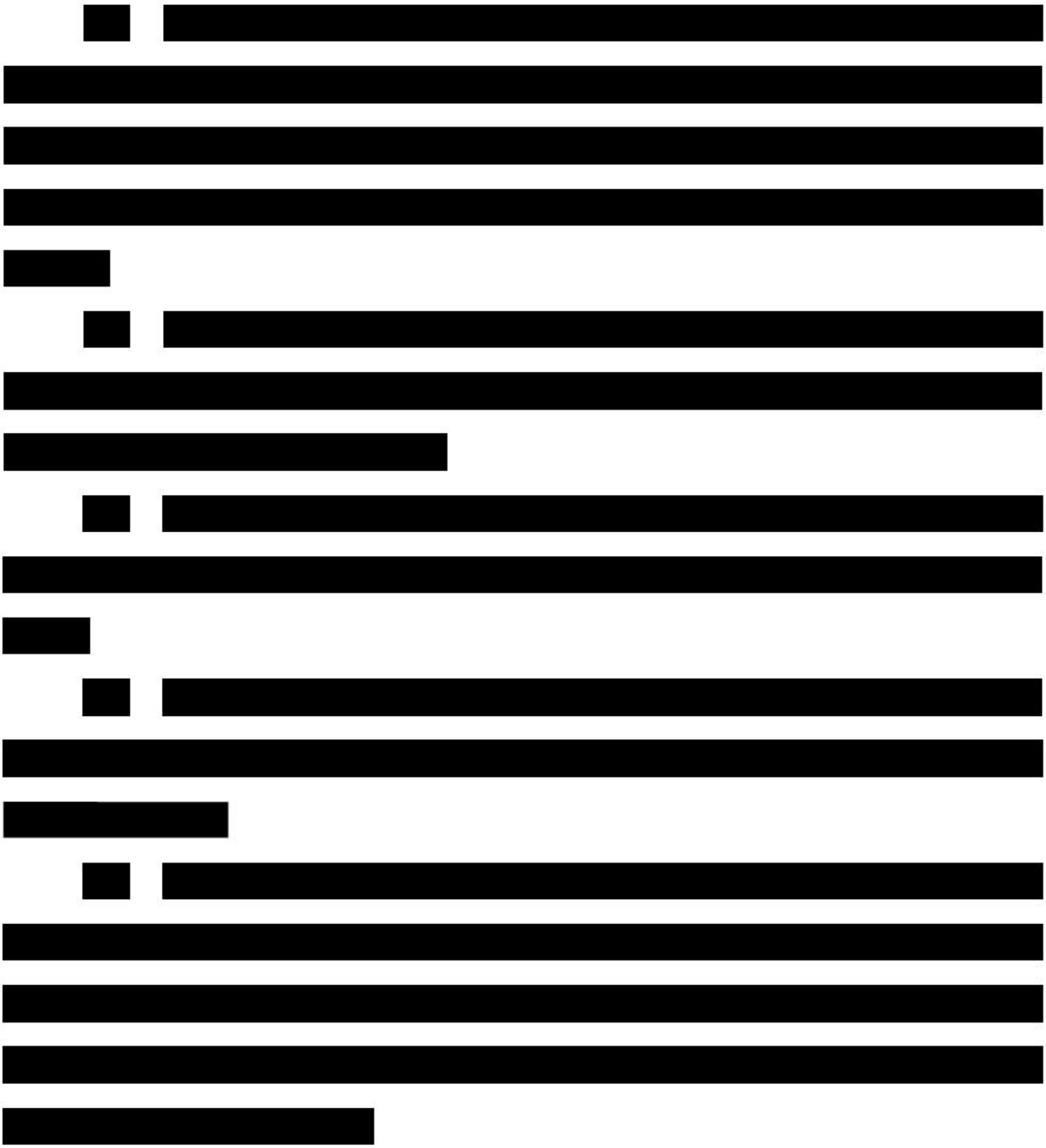
C. Other Licensing Activities

815. The market for licensing content for AI-related uses, including LLM training and output generation, is established and growing. Lasinski Decl. Ex. B at ¶ 139.

816. The practice of licensing data and content for training computer systems predates large language models. Lasinski Decl. Ex. B at ¶ 49.

817. Early examples include the Reuters RCV1 corpus in 2000, later distributed by NIST in 2004 under license agreements, and the Linguistic Data Consortium's "English Gigaword" dataset in 2003, containing approximately 1.8 billion words. *Id.*

818. That practice has continued and expanded with the rise of generative AI, with AI developers seeking licensed content across multiple verticals including news and media publishers, educational platforms, academic publishers, and books. *Id.* at ¶¶ 49–54.



[redacted] the AI content licensing market is also reflected in a substantial volume of publicly reported deals.

828. As shown in Figure 2 of Mr. Lasinski's Rebuttal Report (Attachment 8.0), the major AI developers—including OpenAI, Microsoft, Google, Meta, Amazon, Perplexity, and others—have publicly announced content licensing agreements with rights holders across news publishers, academic publishers, and other content providers, covering both training and output uses, with rights holders receiving compensation in exchange for granting access to their content.

Figure 2: Publicly Disclosed Licensors and Licensees of Access and Use Agreements to Data and Content for LLM Training and Other AI Uses

Licensees/Grantees			
Alphabet/Google	Dow Jones	Midjourney	ProRata.ai
Amazon	LexisNexis	Mistral	Snowflake
Anthropic	Meta	OpenAI	Udio
Bria	Microsoft	Perplexity	
Rights Holders			
270 Media	Folha de S. Paulo	Mexico News Daily	Süddeutsche Zeitung
Adweek	Forbes	Minkabu the Infinoid	The Associated Press
Agence-France-Presse	Foreign Policy	Mumsnet	The Atlantic
Agencia EFE	Fortune	New York Times	The Daily Caller
Arena Group	Fox News	News Corp	The Economist
Atlas Obscura	Future	News Corp Australia	The Globe and Mail
Automattic	Gannet	News UK	The Guardian
AWP Finanznachrichten AG	Gear Patrol	News/Media Alliance	The Independent
Axel Springer	GEDI	NewsPicks	The Los Angeles Times
Axios	Graham Holdings Company	Newsquest Media Group	The National
Barron's	Guardian Media Group	Newstex	The News/Media Alliance
Beijing Review	HarperCollins	NTV	The Texas Tribune
Blavity	Healthline	Packt	The Times of India
Business Insider	Hearst	Politico	The Wall Street Journal
Buzzfeed	Hearst Magazines	Prisa Media	The Washington Examiner
CB Insights	Hello!	Private Equity News	The Washington Post
CNN	Hong Kong Economic Times	Pro Football Network	TheStreet
Condé Nast	Humanoid	Prospect	TIME
Der Spiegel	IBT Media	Reach PLC	Tipranks
DotDash Meredith/People	Industry Dive	Recurrent Ventures	Universal Music
Dow Jones Newswires	Infobae	Reddit	USA Today
DPCMO	Investor's Business Daily	Reuters	Various Authors
DPReview	Kompass	RTL Deutschland	Vox
El Pais	Le Figaro	Schibsted Media Group	Warner Records
Entrepeneur	Le Monde	Shutterstock	Washington Examiner
Exame	Lee Enterprises	Skift	Washingtonian
Exponential View	Map Happy	Sky News	Wordsmith
Fast Company, Inc.	MarketWatch	Snopes	World History Encyclopedia
Financial News	McClatchy Media	Stack Overflow	Worth
FiscalNote	MediaHuis	Stern	
Financial Times	MediaLab	StyleBlueprint	

Lasinski Decl. Ex. B at ¶ 62 & Attachment 8.0.

829. Publicly reported deals include OpenAI's agreements with the Associated Press (July 2023), Axel Springer (December 2023), Le Monde (March 2024), the Financial Times (April 2024), Reddit (May 2024), News Corp (May 2024), TIME (June 2024), Condé Nast (August 2024), GEDI (September 2024), and Hearst (October 2024), among others. *Id.* at Attachment 8.0.

830. The emergence of AI content aggregators and intermediary platforms further demonstrates the development of market infrastructure to facilitate licensing between content owners and AI developers. Lasinski Decl. Ex. B at ¶¶ 63–69 & Attachment 8.0.

831. Multiple third-party platforms have emerged to aggregate content for AI licensing, including the Copyright Clearance Center; ProRata.ai, which provides attribution technology to identify when publisher content is used in generative AI answers and to compensate content owners; LexisNexis, which has entered agreements with news outlets for AI-related uses; and the Perplexity Publishers' Program, which provides revenue sharing when publisher content is referenced in AI-generated answers. Lasinski Decl. Ex. B at ¶¶ 63–69 & Attachment 8.0; Nelson Decl. Ex. 677; Nelson Decl. Ex. 674; Nelson Decl. Ex. 675; Nelson Decl. Ex. 676.

832. These intermediary platforms reduce transaction costs and facilitate licensing arrangements that would be difficult for individual authors and small publishers to negotiate directly with AI developers. Lasinski Decl. Ex. B at ¶¶ 63–69 & Attachment 8.0.

833. Created by Humans launched an AI rights licensing platform for authors in January 2025. Nelson Decl. Ex. 121.

834. The platform connects rights holders with AI developers and permits rights holders to grant both training rights and reference rights. Lasinski Decl. Ex. B at ¶ 65.

[redacted]

837. In September 2025, Anthropic PBC entered into a class action settlement agreement with authors in *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal.), resolving claims arising

from Anthropic's downloading of books from Library Genesis and the Pirate Library Mirror. Nelson Decl. Ex. 758.

838. The settlement established a $1.5 billion settlement fund and a Work List of 482,460 works, which corresponds to approximately $3,100 per work, and it released Anthropic for past use of the works without granting a license for future use. *Id.*

839. By April 16, 2026, rights holders had filed claims for 440,490 works, which corresponds to approximately $3,400 per claimed work. Nelson Decl. Ex. 759.

840. As of May 2026, rights holders had claimed 92.77 percent of the works eligible under the settlement. Nelson Decl. Ex. 760, at 7.

D. The Proliferation of AI-Generated Books

841. Amazon is the largest online bookstore in the world. Lasinski Decl. Ex. B ¶ 380. As of 2022, Amazon accounted for an estimated 40 percent of print book sales in the United States and approximately 67 percent of worldwide e-book sales, and generated approximately $28 billion in revenue from worldwide book sales. *Id.*

842. In February 2023, Reuters reported that ChatGPT "appears ready to upend the staid book industry as would-be novelists and self-help gurus looking to make a quick buck are turning to" ChatGPT to generate entire books for sale on Amazon, and described an author who produced a children's book in a matter of hours. Nelson Decl. Ex. 123.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

849. [redacted]

[redacted]

[redacted]

[redacted]

855. Reimers and Waldfogel found that the number of new titles appearing each month on Amazon nearly tripled between 2022 and late 2025, rose by a factor of ten in some categories, and that the increase coincided with the diffusion of large language models. Nelson Decl. Ex. 762, at 2-3.

856. Using AI detection across a random sample of titles, they found detected AI use to be roughly zero through 2022, rising to 30 percent in 2023, 45 percent in 2024, and surpassing 60 percent during 2025. *Id.*

857. They further found that the average quality of books released after the introduction of large language models in 2022 is lower than that of books released before 2022, that AI-containing books garner substantially less usage per title and receive lower star ratings, and that direct evidence shows AI-containing books have lower quality, "and their rising share . . . drives the overall decline" in average book quality. *Id.* at 2-3, 21-23 & Abstract; Lasinski Decl. Ex. B at ¶¶ 376–77.

858. A 2026 study titled "Generative AI Floods and Dilutes the Market for Books" analyzed 14,419 self-published genre fiction titles released between January 2023 and March 2026 and found that books with substantial AI text, defined as more than 25 percent of the full text detected as AI, spread into every genre tracked and accounted for 20.0 percent of titles in the sample. Nelson Decl. Ex. 763, at 2–3, 11–12; Lasinski Decl. Ex. D at ¶ 6.

859. Those titles accounted for 12.1 percent of sales and 11.3 percent of revenue, and their share of observed sales rose from near zero in early 2023 to roughly 20 percent by the second quarter of 2026. *Id.* at 2–3.

860. Books with light AI text, defined as 25 percent or less of the full text detected as AI, accounted for a further 17.1 percent of the titles in the sample, 16.2 percent of sales, and 16.2 percent of revenue. *Id.* at 3.

861. The study observed that "a 12% share of sales is crucial, especially when it comes from 20% of the catalog that competes for a revenue pool which grew far more slowly than the catalog itself." *Id.*

862. Between January 2023 and March 2026, the cumulative catalog of released titles on the Amazon self-published genre-fiction marketplace grew 38.3-fold, and the number of books

selling in a quarter grew 19.2-fold, while quarterly sales grew only 7.3-fold and revenue only 8.9-fold. Nelson Decl. Ex. 763, at 3.

863. The study found that the number of books released "expanded faster than the sales and revenue available to absorb them," that revenue per book declined among books with no AI text as well, and that the declines were concentrated in the genres where AI exposure was greatest. *Id.* at 3-4, 8; Lasinski Decl. Ex. D at ¶ 6.

864. Books with substantial AI text "did not stay at the margins" and reached "the scarce positions at the top of the market." Nelson Decl. Ex. 763.

865. The share of new top-25 entrants in each genre, ranked by median Amazon sales rank, with substantial AI text rose from near zero to 31 percent, and books with substantial AI text accounted for 10 percent of the top 5 percent of the market. *Id.*

866. The study concluded that "the market now holds many books with substantial AI text that sell little, alongside a smaller set of them that reaches the very top." *Id.* at 3.

867. The share of top-ranked positions held by books with no AI text fell from approximately 88 percent in the lowest-exposure genre-months to approximately 63 percent in the highest. *Id.* at 8.

868. Among top-selling titles, books with substantial AI text incorporate the distinctive rare expressions of existing books more heavily than books with no AI text do, reaching a mean of 45.0 percent against 37.7 percent at the top-50 tier, and for AI-containing books that overlap rises with revenue while for books with no AI text it does not. *Id.* at 7–8.

869. The 2026 market-flooding survey documented the "depressive effect of the massive entry of AI content titles on the sales of human-authored books," and tied "the diminution of

human-authored books' earnings to the volume of market entry of AI titles, without regard to their comparative quality." Nelson Decl. Ex. 763.

870. Books with "no AI text," as a category, earned less than human-written books had earned before AI-content books began to compete with them, and the decline in revenue per book could not be explained as a "mix-shift" "where the average drops only because the catalog fills with many lower-revenue AI books." *Id.* at 3–4, 10.

871. Because "the effect is one of scale," the study stated that "one may surmise that 'as more and more AI-generated books are written,' the crowding effect will increase." *Id.* at 11.

872. A separate study found that readers prefer the outputs of AI trained on copyrighted books over the work of expert human writers. Nelson Decl. Ex. 764; Lasinski Decl. Ex. B at ¶ 366 & n.1243.

873. Contemporaneous press reporting documents the same market effects. Within months of ChatGPT's release, the volume of AI-generated submissions led Amazon to limit self-publishing to three titles per day per author. Nelson Decl. Ex. 127; Nelson Decl. Ex. 71.

874. By March 2024, NPR reported that AI books were crowding the marketplace on Amazon, quoting Authors Guild Chief Executive Officer Mary Rasenberger that "every new book seems to have some kind of companion book, some book that's trying to steal sales," and that the publishers posting those books "already made some money" by the time Amazon identified them. Nelson Decl. Ex. 116.

875. Rasenberger testified that the proliferation of AI-generated works makes it "much harder for the author-written books . . . to be discovered," because "it's just really hard for books . . . to come to the attention of the consumer." Nelson Decl. Ex. 46 at 21:20–22:16.

876. She further explained that Amazon will not remove books solely because they are AI-generated; the books must be infringing or below Amazon's quality standards. *Id*. at 84:10–85:18.

877. The Authors Guild and Class Plaintiffs have experienced AI-generated books being released at the same time as a human-authored work in order to take sales from that work.

878. Rasenberger testified that "[a]gents tell [the Authors Guild] that whenever there is a new book, there are multiple AI-generated books that appear right around publication that are clearly trying to take income, trying to take some sales from the real author's book." *Id*. at 86:21–88:24.

879. The Authors Guild was notified by Stephanie Land of an AI-generated mimic of her memoir. *Id*. at 77:2–15.

880. Jane Friedman complained of AI-generated books published under her name. Nelson Decl. Ex. 138.

881. AI knockoffs of Seth Harp's work appeared on Amazon, including on its Prime Deals savings page. Nelson Decl. Ex. 140.

882. AI-generated works have also piggybacked on recent book releases, Nelson Decl. Ex. 139, and AI-generated books have been published under an author's name one month before that author's memoir was released, Nelson Decl. Ex. 141.

883. Trade press reported that the same dynamic was accelerating and devaluing book publishing, and that new AI-driven imprints such as Spines aimed to publish 8,000 titles in 2025, approximately four times the number of titles published annually by Simon & Schuster. Nelson Decl. Ex. 106; Nelson Decl. Ex. 126.

884. Reporting on novelists' own experience followed. Half of the United Kingdom novelists surveyed said AI is likely to replace their work entirely. Nelson Decl. Ex. 102. Commentators asked what follows if readers like AI-generated fiction. Nelson Decl. Ex. 115. And The New York Times published a quiz inviting readers to distinguish AI writing from human writing. Nelson Decl. Ex. 120.

885. In March 2026, The New York Times reported that AI writing was "not only appearing in cheap self-published e-books that are flooding Amazon but is seeping into even traditionally published fiction," that more than 3.5 million books were self-published in the prior year, up from 2.5 million in 2024, and that Hachette had withdrawn the novel *Shy Girl* after suspected AI use. Nelson Decl. Ex. 89; Nelson Decl. Ex. 128.

886. Two months later, the Times reported that a story that had just won a literary prize was suspected of being AI-generated. Nelson Decl. Ex. 118.

887. In June 2026, best-selling author Tim Ferriss reported that his agent, "who has decades of statements to compare against, put it bluntly: 2025 was the first big drop, 2026 looks more severe, and the only thing that's really changed in that timeframe is the acceleration of AI." Nelson Decl. Ex. 96.

888. In July 2026, The Atlantic reported that *Daggermouth*, which had spent months on USA Today's bestseller list and ranked number one in Amazon's science-fiction romance category, was identified as containing substantial AI text. Nelson Decl. Ex. 125.

889. The New York Times reported that author Jess Michaels said "[t]he glut of the A.I. work makes discoverability very difficult," that competitors were "directly pirating my work, putting it into an engine so that you can produce a 'Jess Michaels-type book,' and then glutting the

market with it," and that "[i]t essentially erases actual authors"; her sales had been "so dismal that she might stop publishing." Nelson Decl. Ex. 119.

890. And The Wall Street Journal reported that the manuscript *Call Me, I'll Hide the Body* ignited a "frenzied auction" for United States rights that produced a two-book deal worth $2 million, after which the author's agency notified publishers that it could no longer support the book because of concerns about the author's possible AI use. Nelson Decl. Ex. 129.

891. Class Plaintiffs face the same harm. Taylor Branch testified that Defendants' models pose "a risk to [his] livelihood today" because "AI can generate automatically without involvement of any human being an artificial product that draws on my copyrighted material without a license and offer it to anybody to slap their names on it and sell it and thereby displace the market for me." Nelson Decl. Ex. 7 at 126:6–19.

892. James Shapiro testified that AI harms and will harm his ability to write and sell books, and that he believed the substitution did him "significant harm." Nelson Decl. Ex. 54 at 67:11–20, 90:1–91:5.

893. David Baldacci testified that AI-generated books were released contemporaneously with his novel *Nash Falls* and that those books had an Amazon rating evidencing that they had been purchased by others. Nelson Decl. Ex. 5 at 227:3–228:6.

894. Michael Connelly testified that he has "seen outlines produce for sequels to my books, specific books, that didn't come from me." Nelson Decl. Ex. 15 at 75:13–76:2.

895. John Grisham testified that he believes artificial intelligence poses a "substantial risk" to his livelihood because authors "could be replaced by AI-generated books . . . which are very cheap to produce." Nelson Decl. Ex. 24 at 174:23–175:17.

896. Stacy Schiff testified that she believes ChatGPT is an existing threat to her livelihood because it can replace her work, and that she believed her future books were less likely to succeed. *Id*. at 43:19–22.

897. Jonathan Franzen testified that OpenAI's conduct "would have a negative impact on [his] livelihood" due to its "theft of [his] copyrighted work [and] the use of it for commercial purposes without licensing[.]" Nelson Decl. Ex. 17 at 107:19–108:4.

E. AI Poses Substantial Risks That Defendants Recognized

898. As Plaintiffs' expert Dr. Ben Zhao documents, AI, and specifically LLMs, pose substantial risks to labor markets, mental health, education, cybersecurity, and public safety, [redacted] Zhao Decl. Ex. B ¶¶ 112–13.

899. AI is upending domestic and international labor markets, and AI systems acting with "greater autonomy across domains" "would likely accelerate labour market disruption." Nelson Decl. Ex. 765.

900. [redacted]

[redacted]

906. Pending complaints allege that OpenAI rushed GPT-4o to release without sufficient safety precautions and that ChatGPT encouraged users to commit suicide. Nelson Decl. Ex. 769; Nelson Decl. Ex. 770.

907. LLMs generate entire essays in seconds with minimal prompting, and AI-assisted cheating is now widespread in K-12 and higher education, with many college students reporting that LLMs have replaced the learning process in numerous classes. Nelson Decl. Ex. 771.

908. Detection of AI-based plagiarism is unreliable, and schools must balance academic integrity against the risk of disciplining innocent students on the basis of false positives. Nelson Decl. Ex. 773.

909. LLMs are easily weaponized to autonomously plan and execute complex, multi-step network attacks without human instruction. Nelson Decl. Ex. 773.

[redacted]

[redacted]

VIII. [redacted]

[redacted]

[redacted]



Shan Decl. Ex. M.



[redacted]

Date: September 4, 2026

Respectfully submitted,

/s/ Justin A. Nelson

SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Telephone: 713.651.9366
jnelson@susmangodfrey.com
Interim Lead Counsel for Class Plaintiffs