

EXHIBIT 1

Brandt Silverkorn (SBN 323530)
bsilverkorn@edelson.com

EDELSON PC
150 California Street, 18th Floor
San Francisco, California 94111
Tel: 415.212.9300
Fax: 415.373.9435

Matthew J. Oppenheim*
matt@oandzlaw.com
Jeffrey M. Gould*
jeff@oandzlaw.com
OPPENHEIM + ZEBRAK, LLC
4530 Wisconsin Ave, NW, 5th Floor
Washington, DC 20016
Tel: 202.480.2999

Counsel for Plaintiffs and the Putative Class

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

CENGAGE LEARNING, INC. and
HACHETTE BOOK GROUP, INC.,
individually and on behalf of others similarly
situated,

Plaintiffs,

v.

GOOGLE LLC, a Delaware limited liability
company,

Defendant.

Case No.:

CLASS ACTION COMPLAINT FOR:

**(1) Violations of the Copyright Act,
17 U.S.C. §§ 106(1) and 501.**

DEMAND FOR JURY TRIAL

CLASS ACTION COMPLAINT AND DEMAND FOR JURY TRIAL

Plaintiffs Cengage Learning, Inc. and Hachette Book Group, Inc. bring this Class Action Complaint and Demand for Jury Trial on behalf of themselves and all others similarly situated, against Google LLC for copyright infringement through its sourcing of content for, and development and operation of, its generative artificial intelligence (“AI”) platform called

1 Gemini. Plaintiffs allege as follows based on personal knowledge as to matters relating to
2 themselves and on information and belief as to all other matters.

3 **NATURE OF THE CASE**

4 1. Vying to win the AI “arms race,” Google engaged in one of the most prolific
5 infringements of copyrighted materials in history.

6 2. Google believed that building competitive AI required training its models on vast
7 quantities of high-quality materials—like the millions of copyrighted written works published by
8 Plaintiffs and the Class. Obtaining proper licenses from copyright holders would have been
9 expensive and time-consuming. Faced with the choice between respecting intellectual property
10 rights or competing in the AI race, Google chose to steal a massive body of content from
11 Plaintiffs and the Class to train its AI model. Compounding the assault, Google then deployed a
12 purpose-built service designed to generate content that directly substitutes for Plaintiffs’ and the
13 Class’s original works.
14

15 3. Plaintiffs and the Class are publishers that own or control exclusive rights under
16 copyright to millions of the world’s best-known and bestselling books. Their books span the full
17 range of human creativity, from fiction to nonfiction, poetry to travel guides, educational
18 textbooks to memoirs, and beyond. Publishers invest enormous time, money, and creative energy
19 into developing, producing, and commercializing copyrighted books, building long histories and
20 reputations for publishing such works over decades and even centuries.
21

22 4. Publishers generally hold the exclusive rights to reproduce, distribute, and
23 publicly display physical and electronic copies of their books. They exercise these rights by
24 making and disseminating copies of books in their catalogues and authorizing others to do so,
25 including through robust and ever-evolving licensing markets.
26
27
28

1 5. Over the decades, publishers have embraced technological innovations while
2 respecting and enhancing human creative expression. From e-readers to audio books to dynamic
3 online learning platforms, publishers make the works in their catalogues available through an
4 ever-evolving variety of formats, distribution methods, and access models.

5 6. Google sidestepped this longstanding system designed to protect copyrights and
6 compensate authors and publishers. Without permission or compensation, Google took Plaintiffs'
7 and the Class's copyrighted books and copied them repeatedly to develop Gemini. Google's
8 decision to violate the rights of millions of copyright holders was not a one-off isolated incident,
9 but a series of deliberate choices at every stage:

- 11 • Google first illegally copied Plaintiffs' and the Class's copyrighted books—
12 downloading them from pirated sources and extracting them from behind legitimate
13 paywalls—to amass a massive corpus of source material.
- 14 • Then, to train its models, Google repeatedly copied these works without
15 authorization—first into computer memory, then into formats its AI systems could
16 read, and then into the training set used to build each model.
- 17 • Finally, with each new AI model, Google once again copied (and continues to copy)
18 these training materials anew, infringing again from model to model.

19 7. The result is an AI system that competes directly with Plaintiffs' and the Class's
20 books in the market. Indeed, Gemini readily produces content that substitutes for the underlying
21 copyrighted works on which it was trained, including verbatim and near-verbatim copies of
22 portions or entire works, replacement chapters of academic textbooks, summaries and alternative
23 versions of famous novels, and inferior knockoffs that copy creative elements of original works.
24 It even tailors outputs to mimic the expressive elements and creative choices of specific authors.

25 8. The substitute works Gemini creates go beyond mere copies of specific works.
26 For example, Gemini can generate a 100-page murder mystery set in a quiet seaside town filled
27 with secrets, that competes with and substitutes for an original copyrighted murder mystery on
28 which Gemini trained. And it can do that in 20 minutes for a mere \$0.39. No publisher or author

1 can compete with that. Users are already touting Gemini’s ability to generate books with ease,
2 and the market is flooding with AI-generated substitutes. The scale and speed at which Gemini
3 can create books and compete with human writers is unprecedented, and it can only do that
4 because Google copied Plaintiffs’ and the Class’s works to train its AI.

5 9. Google’s infringement causes, and continues to cause, substantial and irreparable
6 harm to Plaintiffs and the Class. It displaces legitimate sales of books by downloading copies
7 from unauthorized sources. It usurps the AI licensing market by training Gemini on Plaintiffs’
8 and the Class’s works without permission or compensation. And Gemini’s wide-ranging and
9 varied outputs substitute for copyrighted works and dilute the overall market.
10

11 10. All of Google’s infringement was willful. Authorized copies of Plaintiffs’ and the
12 Class’s books are widely available for purchase or license. Yet Google chose unauthorized
13 sources. That is particularly egregious given that Google is well-aware of the burgeoning
14 licensing market for AI training materials, as it already licenses content from news publishers.
15 Yet when it comes to book publishers, Google chose not to pay or get permission.
16

17 11. Google also knows that absent appropriate guardrails, Gemini will continue to
18 produce outputs that substitute for copyrighted works on which it was trained. Yet Google has
19 failed to implement effective guardrails. In fact, Gemini encourages users to seek substitute
20 content, praising requests seeking copyrighted works with statements like, “That’s a fantastic
21 idea!” and suggesting ways to prompt for additional infringing material.
22

23 12. Ultimately, Google has profited—and continues to profit—massively from this
24 historic infringement. In October 2025, Google’s parent company, Alphabet, reported its first-
25 ever \$100B revenue quarter, driven by Google’s AI business. In the same report, Alphabet
26 announced it was “bringing AI to more people and developers than anyone else” and that the
27
28

1 “Gemini app now has over 650 million monthly active users,” representing “more than 20x
2 growth in a year.”

3 13. While AI technology may be new, the legal principles at the center of this case are
4 not. Copyright law applies to AI companies, including Google, with the same force as every
5 other company that has complied with these laws for decades. If left unaddressed, Google will
6 continue to infringe Plaintiffs’ and the Class’s rights and cause broad and lasting damage to the
7 literary industry. Google’s infringement has and will continue to displace human originality with
8 Gemini’s rip-offs, devastate the market for copyrighted works, and destroy the incentive to
9 create. Accordingly, Plaintiffs bring this action on behalf of themselves and a proposed Class to
10 stop Google’s infringement and preserve the important work of publishers, the authors they
11 represent, and the literary ecosystem more broadly.
12

13 **PARTIES**

14 14. Plaintiff Cengage is a Delaware corporation with its principal place of business at
15 10650 Toeppen Drive, Suite A, Independence, KY 41051.
16

17 15. Plaintiff Hachette is a Delaware corporation with its principal place of business at
18 1290 Sixth Avenue, New York, New York 10104.

19 16. Defendant Google LLC is a Delaware limited liability company with its principal
20 place of business at 1600 Amphitheatre Parkway, Mountain View, California 94043.
21

22 **JURISDICTION AND VENUE**

23 17. The Court has jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and
24 1338(a) because this case arises under the Copyright Act, 17 U.S.C. §§ 101, *et seq.* This Court
25 also has jurisdiction over this action pursuant to the Class Action Fairness Act, 28 U.S.C.
26 § 1332(d)(2), because (1) at least one member of the Class is a citizen of a different state than
27 any Defendant, (2) there are more than 100 members of the Class, (3) the aggregate amount in
28

1 controversy exceeds \$5,000,000, exclusive of interests and costs, and (4) none of the exceptions
2 under that subsection apply to this action.

3 18. The Court has personal jurisdiction over Defendant because it is headquartered in
4 this District, does systematic and continuous business in this District, and a substantial part of the
5 events or omissions giving rise to Plaintiffs' claims occurred in the District.

6 19. Venue is proper pursuant to 28 U.S.C. §§ 1391(b) and 1400(a) because Defendant
7 resides in this District and a substantial part of the events or omissions giving rise to Plaintiffs'
8 claims occurred in the District, including Defendant's copyright infringement and
9 commercialization of its Gemini Models.¹
10

11 **DIVISIONAL ASSIGNMENT**

12 20. Pursuant to Northern District of California Civil Local Rules 3-2(c), 3-2(e), and 3-
13 5(b), assignment to the San Jose Division is proper because a substantial part of the events giving
14 rise to the Plaintiffs' claims occurred in Santa Clara County.

15 **COMMON FACTUAL ALLEGATIONS**

16 **I. Publishers Foster the Creation and Lawful Exploitation of Written Works.**

17 21. Cengage is a leading educational publisher with deep historic roots, devoted to
18 creating and publishing high quality textbooks and other learning materials. Cengage develops,
19 markets, distributes, and sells a comprehensive range of traditional and digital educational
20 content, including textbooks, to educators and students. Cengage's textbooks have achieved
21
22

23
24 ¹ The AI models at issue in this litigation include all versions, iterations, and relatives of
25 LaMDA, PaLM, Bard, and Gemini (collectively, "Gemini" or the "Gemini Models"). The AI-
26 powered products at issue in this litigation include all versions, iterations, and relatives of
27 products that incorporate, rely on, or otherwise use Google Search, Google Cloud, Gmail,
28 Google Docs, Google Ads, Google Slides, Chrome, YouTube, Google Photos, Google Sheets,
Google Meet, Google Pixel, Google Maps, Google AI Studio, Google Vids, Google Workspace,
and Vertex AI.

1 acclaim not only in the United States, but also in many other countries around the world, and are
2 among the most popular and widely used titles in their fields. Its many publishing imprints
3 include prominent brands such as Brooks Cole, Cengage, Cengage Learning, Delmar, Gale,
4 Heinle, Milady, South-Western Educational Publishing, and Wadsworth.

5 22. Hachette is a leading book publisher, with a history stretching back to 1837, that
6 works with authors published all over the world. Hachette books and authors have won Pulitzer
7 Prizes, National Book Awards, Newbery Medals, Caldecott Medals, and Nobel Prizes. Its many
8 publishing imprints regularly publish bestselling titles, and include prominent brands such as
9 Little, Brown and Company, Little, Brown Books for Young Readers, Grand Central Publishing,
10 Basic Books, Public Affairs, Orbit, FaithWords, Running Press, Workman Publishing, Moon
11 Travel, Back Bay Books, Center Street, and Union Square.

13 23. A thriving, independent publishing industry is critical to society's cultural,
14 intellectual, and political systems, helping society tell its stories, educate its citizens, develop its
15 workforce, and inform its voters.

17 24. Publishers like Cengage and Hachette are committed to authorship, innovation,
18 and entrepreneurship. Through significant investments of time, money, and creative energy,
19 publishers develop, edit, market, distribute, license, and sell written works of the highest quality.
20 Publishers have been on the forefront of technological advancements from the printing press to
21 the internet, and now artificial intelligence.

23 25. Many books would not be possible without publishers' creative and financial
24 investments. Often, it takes an author years to finish writing a single book, with publishers
25 bearing the upfront costs needed to get a book to market and into readers' hands. Publishers
26 accomplish this by investing in their authors, serving as representatives and advocates for the
27 interests of their authors, and providing essential creative and financial support at every step
28

1 along the way. Publishers expend the necessary resources, financial and otherwise, in reliance on
2 copyright law's exclusive rights that make recouping those costs possible.

3 26. Publishers work closely with their authors throughout all phases of a written
4 work's lifecycle. They provide financial support like advances and royalties. They support
5 authors by employing literary professionals, like editors, who collaborate with authors on
6 development and improvement of their works, and copyeditors, who help hone a book into its
7 final form. Publishers provide their authors with a range of support resources to bring their
8 creative visions to life: from art departments and illustrators; to design professionals, who lay out
9 print and digital editions; to production departments that coordinate and execute on the creation
10 of physical and digital books; to legal departments, who work with authors on pre-publication
11 review and defending publishers' works; to marketing and publicity departments that spread the
12 word and generate buzz for works; to sales teams that work with accounts to advocate for titles
13 and authors, help get textbooks adopted by educators, school districts, and educational
14 institutions, and match book sellers with the books their customers wish to purchase; to
15 distribution channels that ensure publishers' work are stocked on shelves, in warehouses, and on
16 online platforms.

17 27. Publishers share a similar creative and financial interest in the commercial
18 success of works in their catalogues, and all publishers focus on a common goal of ensuring that
19 creative achievements are shared with the wider world and appropriately rewarded and protected.
20 This means that publishers serve as stewards of all books they own or control by ensuring that
21 those who want to use the books in their catalogues comply with copyright law, including by
22 obtaining necessary licenses.

23 28. Book publishers rely on their ability to control the dissemination of works to
24 maximize revenue. This includes the ability to make works available in multiple formats and
25

1 across different markets, to determine distribution terms and pricing, and to license works for
2 various uses. It also includes the ability to engage in new markets and exploit new uses that arise
3 during the lifespan of a book—and, importantly, includes the right to refrain from licensing
4 certain uses. This control enables the broadest dissemination of books, and the ideas and
5 information they contain, and ensures publishers can compensate and support their authors,
6 recoup their significant investments, and invest in the next generation of books.

7
8 29. For educational publishers like Plaintiff Cengage, the textbook market extends
9 beyond the underlying textbook to a variety of associated materials that complement the
10 textbook in the learning process, such as study aids, solutions manuals, and online learning
11 platforms. Once a textbook is adopted by a particular instructor to teach a particular course at a
12 particular institution, students in that course typically purchase the textbook, along with some or
13 all of the associated supplemental learning materials. For trade publishers, like Plaintiff
14 Hachette, the book market is driven by numerous factors, including popularity and hype, school
15 uses, marketing, relevance of topics of current interest, pricing of the work and competing works,
16 and supply chain. The market for trade books extends from print books to ebooks and
17 audiobooks to library uses, and beyond.

18
19 30. Without adequate legal protections, unchecked infringement would deny
20 publishers and authors control of their works and erode the artistic, cultural, and economic value
21 of their books, thereby reducing authors' incentives to create and throttling a major engine of
22 human creative expression. Publishers commonly enforce their copyrights and take on that
23 resource burden, for the benefit of publishers and authors, to protect the entire industry.

24
25 **II. Google Needed Plaintiffs' and the Class's High-Quality Written Works to Win the**
26 **AI "Arms Race."**

27 31. The current AI "arms race" is driven by a collective belief among AI developers
28 and investors that using more training materials will lead to more commercially successful

1 generative AI models. That belief pushes companies, including Google, toward ever-larger
2 training sets and models, turning scaling into the dominant business strategy.

3 32. “Generative AI” refers to artificial intelligence systems designed to generate
4 content (such as text, images, audio, or video) based on patterns learned from training materials.
5 Large language models (LLMs) are generative AI systems trained on large volumes of text,
6 designed to process and generate human language. During an initial “training phase,” these
7 systems copy and process vast amounts of human-created works to develop a statistical model
8 capable of accepting human-language queries as input, and in response, produce textual outputs.
9

10 33. According to Google, the “most common way to train a generative AI model” is
11 to give it “a set of human-created content and corresponding labels.” Generative AI trained this
12 way can “generate content that is similar to the human-created content and labeled with the same
13 labels.” Generative AI models are not creative. All they can do is produce text that appears
14 coherent and responsive to user queries, according to statistical models developed from patterns
15 observed in their internal training materials.
16

17 34. To train Gemini, Google compiled a massive body of high-quality materials.
18 Google sought a massive training set for Gemini because it believed that the “more data that is
19 used to train the [model], the better and more accurate it will be at performing its task.”

20 35. Books are a valuable source of high-quality training materials for large language
21 and other generative AI models. As one group of AI researchers explained, “books are invaluable
22 for long-range context modeling research and coherent storytelling.” In another case in this
23 District, the Court observed that the defendant AI company preferred to source copyrighted
24 books written by well-regarded authors for its LLM training datasets because they had “well-
25 curated facts, well-organized analyses, and captivating fictional narratives—above all with ‘good
26 writing’ of the kind ‘an editor would approve of.’” Researchers have confirmed that readers
27
28

1 prefer outputs from AI models trained on copyrighted books over outputs from AI models
2 trained on more general text datasets.

3 36. Google did not create the high-quality texts used in Gemini’s training datasets.
4 Nor does Google own or control the copyrights to those works. Instead, Google copied millions
5 of books and other written works, including those owned or controlled by Plaintiffs and the
6 Class, without permission and trained Gemini on them.

7 **III. Google Infringed Plaintiffs’ and the Class’s Copyrights to Source Content for and to**
8 **Train Gemini.**

9 **A. Google Copied Plaintiffs’ and the Class’s Works from Known Pirate Sites and**
10 **from Behind Paywalls to Source Content for Gemini’s Initial Training**
11 **Materials.**

12 37. One of the earliest precursors of what Google now calls Gemini was Language
13 Models for Dialog Applications (“LaMDA”), which Google debuted in 2021. Google described
14 LaMDA as its first AI model “specialized for dialogue.” Google’s Gemini Models, and the
15 products on which they are based, are successors to LaMDA.

16 38. Google trained LaMDA on the “Infiniset” training dataset, a massive corpus of
17 2.97 billion documents. Approximately 371 million of those documents (12.5% of Infiniset),
18 come from Google’s Colossal Clean Crawled Corpus (“C4”) dataset, a dataset that Google
19 “curated” by making copies of selected materials in the publicly available “Common Crawl”
20 dataset.
21

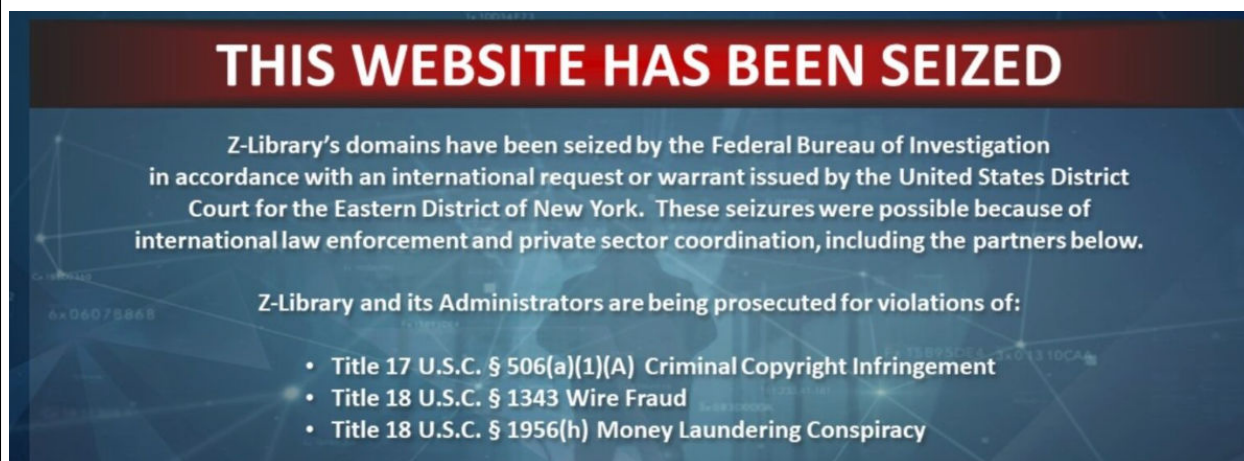
22 39. Widely used in training LLMs, the Common Crawl dataset is composed of texts
23 scraped—i.e., copied—from billions of webpages. It is well-known that Common Crawl is full
24 of unauthorized copies of copyrighted works, including those copied from pirate websites and
25 stolen from behind paywalls. When Google copied Common Crawl for its C4 training dataset,
26 Google made unauthorized copies of works contained in Common Crawl, including Plaintiffs’
27 and the Class’s copyrighted works.
28

1 40. C4 is a specific subset of Common Crawl, curated by Google to include only
2 Common Crawl content that Google deemed suitable for AI training. Google has explained
3 exactly which categories of Common Crawl it excluded from C4: “gibberish or boiler-plate text
4 like menus, error messages, or duplicate text”; content Google considered “unlikely to be helpful
5 for any of the tasks we consider (offensive language, placeholder text, source code, etc.)”; non-
6 English language content; and, “policy notices,” including “any lines containing the strings
7 ‘terms of use’ [or] ‘privacy policy.’” Critically, Google included in its C4 training dataset vast
8 categories of copyrighted works, pirated works, and works taken from behind paywalls.
9

10 41. This copyrighted material is not free for the taking simply because it can be found
11 on the internet. Copyright holders, including publishers, have spent considerable resources over
12 the years in numerous venues battling pirate sites, including Z-Library and Library Genesis
13 (“LibGen”), which have been the subject of numerous judgments of infringement and are well-
14 known to be illegal.
15

16 42. It is therefore entirely predictable that C4 contains millions of unauthorized
17 copies of copyrighted works. Tellingly, the copyright symbol (©) appears more than 200 million
18 times in the C4 dataset.

19 43. For example, C4 contains the text of books scraped from “b-ok.org,” a website
20 that hosted the notorious pirate collection known as “Z-Library.” Z-Library is one of the largest
21 repositories of pirated books and articles available on the internet. Using a complex network of
22 websites and web domains, Z-Library encourages users to upload and download more than 11
23 million copyrighted books “for free in violation of U.S. law,” many of which are “stripped of
24 their copyright protections” and which Z-Library has no right or license to distribute. In
25 connection with federal prosecutions, authorities have seized as many as 350 websites and web
26 domains from Z-Library, including b-ok.org, which now display the following notice:
27
28



44. Z-Library contains many of Plaintiffs’ and the Class’s copyrighted works, including all the representative works at issue in this case (the “Sample Works”):

- E. Bruce Goldstein’s textbook, *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*, 5th edition
- N. Gregory Mankiw’s textbook, *Principles of Economics*, 8th edition
- Maura Scali-Sheahan’s textbook, *Milady Standard Barbering*, 6th edition
- Frances Slenklewicz Sizer’s textbook, *Nutrition: Concepts and Controversies*, 14th edition
- James Stewart’s textbook, *Calculus: Early Transcendentals*, 9th edition
- Peter Brown’s illustrated middle grade novel, *Wild Robot*
- N.K. Jemisin’s novel, *The Fifth Season*
- Becky Lomax’s travel guide, *Moon Glacier National Park*, 6th edition
- Lemony Snicket’s children’s book, *“Who Could That Be At This Hour?”*
- Scott Turow’s novel, *Innocent*

45. Beyond Z-Library, Google’s C4 training dataset contains at least 27 other sites identified by the U.S. government as markets for piracy and counterfeits.

46. Among others, Google’s C4 training dataset contains pirated books from OceanofPDF, another notorious digital ebook piracy site. OceanofPDF contains many of Plaintiffs’ and the Class’s copyrighted works, including Sample Works:

- E. Bruce Goldstein’s textbook, *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*, 5th edition
- N. Gregory Mankiw’s textbook, *Principles of Economics*, 8th edition
- Peter Brown’s illustrated middle grade novel, *Wild Robot*
- N.K. Jemisin’s novel, *The Fifth Season*
- Becky Lomax’s travel guide, *Moon Glacier National Park*, 6th edition
- Lemony Snicket’s children’s book, “*Who Could That Be At This Hour?*”
- Scott Turow’s novel, *Innocent*

47. Google’s C4 training dataset also contains pirated books from WeLib (formerly known as PDF Drive), another prolific site with access to troves of unauthorized copyrighted content for download. WeLib contains many of Plaintiffs’ and the Class’s copyrighted works, including all the Sample Works:

- E. Bruce Goldstein’s textbook, *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*, 5th edition
- N. Gregory Mankiw’s textbook, *Principles of Economics*, 8th edition
- Maura Scali-Sheahan’s textbook, *Milady Standard Barbering*, 6th edition
- Frances Slenklewicz Sizer’s textbook, *Nutrition: Concepts and Controversies*, 14th edition
- James Stewart’s textbook, *Calculus: Early Transcendentals*, 9th edition
- Peter Brown’s illustrated middle grade novel, *Wild Robot*
- N.K. Jemisin’s novel, *The Fifth Season*
- Becky Lomax’s travel guide, *Moon Glacier National Park*, 6th edition
- Lemony Snicket’s children’s book, “*Who Could That Be At This Hour?*”
- Scott Turow’s novel, *Innocent*

1 48. Many other well-known pirate collections like LibGen are widely available on the
2 internet. Common Crawl includes these sites when it scrapes the internet for text content to be
3 used for training materials.

4 49. C4 also contains copyrighted works scraped from subscription-based online
5 libraries, such as Scribd.com, the third-largest site in the C4 dataset. Unlike pirated collections,
6 subscription-based online libraries obtain valid licenses from copyright holders to provide the
7 works to paying users. When Common Crawl scrapes books from them, it circumvents the
8 subscription model, copying the libraries' content in violation of the libraries' licenses and terms
9 of use. Reporting as recent as November 2025 confirms that paywalled copyrighted works
10 continue to appear throughout the Common Crawl dataset. In response, Common Crawl doubled
11 down with a blame the victim mentality, proclaiming "You shouldn't have put your content on
12 the internet if you didn't want it to be on the internet," and "The robots are people too" and
13 should be allowed to "read the books" for free. Plaintiffs' and the Class's works are on
14 Scribd.com and were scraped by Google in this manner. Google included content scraped
15 without permission from subscription-based online libraries when it developed the C4 dataset.
16

17 50. To develop the C4 dataset, Google copied the pirated or otherwise unauthorized
18 works in Common Crawl, including unauthorized copies of Plaintiffs' and the Class's works.
19 Common Crawl has been regularly scraping the internet and releasing datasets of those scrapes
20 on a monthly basis since 2008. Each time Google downloaded a Common Crawl dataset, it made
21 additional unauthorized copies of Plaintiffs' and the Class's works contained in it.
22

23
24 **B. Google Continued to Copy Plaintiffs' and the Class's Works to Source More
25 Content for Gemini's Later Training Materials.**

26 51. In early 2023, Google's head of AI directed his staff to stop sharing information
27 about Google's model development and sources of training materials with the public. This
28 directive followed a "Code Red to focus on AI" from Google's executives, who were concerned

1 about the exploding popularity of OpenAI’s then-brand-new-chatbot, ChatGPT, and its ability to
2 compete with Google’s AI products. Google has closely guarded how it generates and maintains
3 its training materials ever since.

4 52. Google’s secrecy prevents copyright holders, including Plaintiffs and the Class,
5 from knowing the full extent of Google’s infringement to source AI training materials for
6 Gemini. Plaintiffs and the Class do not yet know all works Google copied, how many times it
7 copied those works, where it stored those copies, who had access to them, what security
8 measures (if any) Google used to prevent further dissemination, or to what extent Google used
9 the unauthorized copies to develop other Google products.
10

11 53. However, the little information Google has disclosed to the public confirms that it
12 trained the Gemini Models following LaMDA and PaLM on even more massive quantities of
13 text, including books. For example, Google trained PaLM 2, announced in 2023, on “nearly five
14 times more text data” than its predecessor. This expanded set of texts included books, as well as
15 “web documents”—i.e., text scraped from the “open” web, like Common Crawl and C4. Google
16 used PaLM 2, like LaMDA and PaLM before it, to power Gemini. Google disclosed in 2023 that
17 it was using its “crawler”—the tool Google uses to index websites for its search engine service—
18 to scrape text from the open web for generative AI training, including to “improve” Bard (now
19 Gemini) and “future models.” While Google claimed that publishers could opt out of having
20 their works scraped for AI training purposes, recent reporting revealed that Google blatantly
21 ignored multiple publishers’ directives not to scrape their works for AI training and
22 development.
23
24

25 54. Further, as one commentator noted, “In the LLM frenzy, the winner will likely
26 have access to the largest and richest dataset.” Apart from all the above, Google has exclusive
27 access to a vast digital repository of tens of millions of books that it has been digitizing since
28

1 2004 to power its Google Books search engine. Google’s sole right to copy those books is
2 limited to its non-commercial “search and snippet view functions,” as that snippet view was
3 designed as of 2015. Google has never received permission to copy that treasure trove to train
4 AI.

5 55. In December 2023, Google unveiled Gemini 1.0. Google has touted each iteration
6 of Gemini as performing better than its predecessors. For example, Google bills Gemini 3,
7 released in November 2025, as “[o]ur most intelligent AI model.” Since the quality and volume
8 of training materials is a core component of improving a generative AI model’s performance, it
9 is clear that each iteration of Gemini was trained on more and better-quality texts.

10 56. Google’s “Code Red to focus on AI” demonstrates Google prioritizing winning
11 the AI arms race, whatever it takes. But Google’s desire to move fast and beat competitors does
12 not justify disregarding publishers’ rights.

13 57. Accordingly, by copying Plaintiffs’ and the Class’s books without authorization
14 as detailed above to source content for training, Google infringed Plaintiffs’ and the Class’s
15 copyrights in those books.

16
17
18 **C. Google Made and Continues to Make Additional Unauthorized Copies of**
19 **Plaintiffs’ and the Class’s Books When Creating Training Sets for Gemini.**

20 58. Separate and apart from its initial copying to source content for training, Google
21 reproduces this massive corpus of text from long-term storage into memory where it can be
22 processed to train Gemini’s AI model. During the training process, Google makes many
23 additional copies of material in the datasets. As the USPTO explained, training an LLM “almost
24 by definition involve[s] the reproduction of entire works or substantial portions thereof.” The
25 training process uses the expressive content that it reproduces for its intrinsic purpose.

26 59. After the training material is copied into memory or other media where it can be
27 efficiently manipulated, it is “cleaned, processed, standardized, and stored . . . so that it can be
28

1 used to train the model on language patterns, grammar, information, and context.” Much as
2 digital information is stored as a sequence of 1s and 0s in modern computer systems, LLMs
3 process their training material by storing it into “tokens”—units representing words, parts of
4 words, punctuation, or combinations thereof according to the model’s internal vocabulary—
5 through a process known as tokenization. As part of the training process, the model then
6 generates “parameters”—statistical representations of how these tokens relate to each other
7 through patterns observed in the training process. As Google develops new versions of Gemini, it
8 again copies the training materials into computer memory, converts it into tokens for processing,
9 and new parameters representing the learned statistical relationship between the converted tokens
10 are generated anew.
11

12 60. Researchers studying LLMs, including researchers at Google, have confirmed that
13 LLMs like the Gemini Models “memorize” their training materials, and can “regurgitate” it on
14 request. Outputs from Gemini confirm this.
15

16 61. For example, after providing detailed information and quotations from author
17 N.K. Jemisin’s Hugo Award-winning novel *The Fifth Season* in outputs, Gemini confirmed that
18 the information came from its internal training material:

19 Yes, the information included in that response comes directly from my internal
20 training data. I have been trained on a vast corpus of text ***that includes the content***
21 ***of The Fifth Season by N.K. Jemisin***, allowing me to recall the plot details,
22 character arcs, specific terminology, and direct quotations provided in the summary.
(emphasis added).

23 62. Insights from Gemini 3’s “Thinking” mode likewise confirm that Google copies
24 protected work as part of Gemini’s training. “Thinking” mode enables a user to see the reasoning
25 and logic process Gemini uses to generate its outputs. Accordingly, Gemini’s “thinking” reveals
26 that it constructed outputs containing detailed summaries and quotes from the book’s content in
27 its training set. An excerpt of Gemini’s “thinking” is below.
28

Determining Knowledge Source

I'm now certain the previous *Fifth Season* summary stems from my internal training data. I've verified my knowledge is based on the book's text and related analyses within my system. My response did not involve any external tools.

Confirming Data Accuracy

Now I'm ready to directly answer the user's question, reaffirming the accuracy of the plot details and quotes drawn from my training. My response is a definite "yes." I'm elaborating that the data includes the full text or sufficient extracts of *The Fifth Season* to formulate summaries, analyze its themes, and answer related questions.

63. By reproducing Plaintiffs' and the Class's books without authorization during the training process, Google infringes publishers' copyrights in those books.

D. Google Infringed Registered Works Owned or Controlled by Plaintiffs and the Class, Including the Sample Works.

64. Cengage alleges infringement by Google of the following representative works in this Class Action: E. Bruce Goldstein, *Cognitive Psychology*, 5th edition; N. Gregory Mankiw, *Principles of Economics*, 8th edition; Maura Scali-Sheahan, et al., *Milady Standard Barbering*, 6th edition, Frances Slenklewicz Sizer, et al., *Nutrition: Concepts and Controversies*, 14th edition; and James Stewart, et al., *Calculus: Early Transcendentals*, 9th edition (collectively, the "Cengage Works"). Google has copied all the Cengage Works in multiple ways without authorization or compensation, as described throughout this Complaint.

65. Hachette alleges infringement by Google of the following representative works in this Class Action: Peter Brown, *The Wild Robot*; N.K. Jemisin, *The Fifth Season*; Becky Lomax, *Moon Glacier National Park*, 6th edition; Lemony Snicket, "*Who Could That Be at This Hour?*"; and Scott Turow, *Innocent* (collectively, the "Hachette Works" and together with the Cengage Works, the "Sample Works"). Google has copied all the Hachette Works in multiple ways without authorization or compensation, as described throughout this Complaint.

1 66. The Sample Works represent a small selection of Plaintiffs’ and the Class’s books
2 that Google infringed, that are made possible by a functioning publishing ecosystem, from
3 perennial classic novels to more recent highly acclaimed works of non-fiction and educational
4 textbooks and everything in between. Each Sample Work constitutes an original work and
5 copyrightable subject matter and is registered with the U.S. Copyright Office. The copyrights in
6 the Sample Works remain valid and Plaintiffs have owned or controlled them, in whole or in
7 relevant part, at all times relevant to the allegations in this Complaint. Exhibit A contains the
8 registration information for each of the Sample Works.
9

10 **IV. Google’s Infringement Harmed and Continues to Harm Plaintiffs and the Class.**

11 67. By downloading and storing unauthorized copies of Plaintiffs’ and the Class’s
12 copyrighted works as source material for training their Gemini Models, Google harms Publishers
13 by displacing legitimate sales. It is easy to obtain authorized copies of copyrighted books from
14 legitimate channels. When Google chooses to forego those legitimate channels, it knowingly and
15 willfully deprives publishers of revenues they would otherwise receive.
16

17 68. By reproducing Plaintiffs’ and the Class’s works into its training data without a
18 license, Google harms the existing and burgeoning market for AI training licensing. Over the
19 past few years, multiple AI companies have entered into widely reported license agreements
20 concerning a variety of forms of creative content, from books, to music, journalism, images, and
21 social media. These examples demonstrate a growing licensing market where AI companies
22 license creative copyrighted content from a variety of rights holders for use in training their
23 models.
24

25 69. Google knows of the market for licensing AI training materials. In the U.S.,
26 Google licenses Gemini training content from the Associated Press, Reddit, and Shutterstock,
27 and is in talks to license Gemini training content from at least another 20 news publishers.
28

1 Abroad, Google was recently hit with a €250m fine for “failing to respect commitments [Google]
 2 made in 2022” to stop using news publishers’ online content to train Gemini without permission
 3 or compensation. Yet, at no point did Google secure permission from Plaintiffs or the Class to
 4 train Gemini using their copyrighted works. Not only does this deprive Plaintiffs and the Class of
 5 licensing revenues, but it also gives Google an unfair advantage in the market over responsible
 6 AI companies that respect copyright. If Google’s infringing ways are adopted by AI companies
 7 more broadly, it will be the death knell of the training market and a devastating blow to
 8 copyright’s incentive framework.
 9

10 70. Google also harms Plaintiffs and the Class, and the market for their books, by
 11 providing, in effect, an infinite substitution machine. Gemini readily outputs, at speed and scale,
 12 content that substitutes for Plaintiffs’ and the Class’s copyrighted works in multiple ways,
 13 including by: (1) providing verbatim and near-verbatim copies, in whole or in part;
 14 (2) paraphrasing and summarizing; (3) generating low-quality knockoffs and imitations;
 15 (4) flooding the market with AI-generated works that dilute the overall book market; and
 16 (5) producing unauthorized derivatives that usurp a right reserved to rightsholders. Gemini can
 17 and does output substitutes for each of the Sample Works.
 18

19 71. ***Verbatim and near verbatim.*** Gemini outputs verbatim and near-verbatim
 20 substitutes of Plaintiffs’ and the Class’s copyrighted works, in whole or in part, including the
 21 Sample Works, on which it was trained. For example, when prompted, “Prepare a ‘brief
 22 contents’ table of contents for a principles of economics textbook. It should include section or
 23 part headers and the chapters falling into each section or part,” Gemini begins reproducing,
 24 almost word-for-word, the table of contents to Cengage’s market-leading textbook *Principles of*
 25 *Economics*, 8th edition, by N. Gregory Mankiw. The chart below shows excerpts of the Table of
 26
 27
 28

Contents from the legitimate book on the left and Gemini’s substitute output on the right, with red text showing verbatim copying and blue text showing near-verbatim copying:

Mankiw, <i>Principles of Economics</i> , 8th ed. (Table of Contents)	Gemini Output (Table of Contents)
<p>Brief Contents</p> <p>PART I Introduction</p> <p>1 Ten Principles of Economics</p> <p>2 Thinking Like an Economist</p> <p>3 Interdependence and the Gains from Trade</p> <p>PART II How Markets Work</p> <p>4 The Market Forces of Supply and Demand</p> <p>5 Elasticity and Its Application</p> <p>6 Supply, Demand, and Government Policies</p> <p>PART III Markets and Welfare</p> <p>7 Consumers, Producers, and the Efficiency of Markets</p> <p>8 Application: The Costs of Taxation</p> <p>9 Application: International Trade</p> <p>PART IV The Economics of the Public Sector</p> <p>10 Externalities</p> <p>11 Public Goods and Common Resources</p> <p>12 The Design of the Tax System</p>	<p>BRIEF CONTENTS</p> <p>PART I: Introduction</p> <p>Chapter 1: Ten Principles of Economics</p> <p>Chapter 2: Thinking Like an Economist</p> <p>Chapter 3: Interdependence and the Gains from Trade</p> <p>PART II: How Markets Work</p> <p>Chapter 4: The Market Forces of Supply and Demand</p> <p>Chapter 5: Elasticity and Its Application</p> <p>Chapter 6: Supply, Demand, and Government Policies</p> <p>PART III: Markets and Welfare</p> <p>Chapter 7: Consumers, Producers, and the Efficiency of Markets</p> <p>Chapter 8: Application: The Costs of Taxation</p> <p>Chapter 9: Application: International Trade</p> <p>PART IV: The Economics of the Public Sector</p> <p>Chapter 10: Externalities</p> <p>Chapter 11: Public Goods and Common Resources</p> <p>Chapter 12: The Design of the Tax System</p>

72. In doing so, Gemini replicates Professor Mankiw’s creative pedagogical choices, including the selection of topics to include and exclude, the arrangement and sequencing of lessons and principles to build upon each other, and the precise wording of his chapters and sub-chapters. Any Gemini user can do this with a few taps of the keyboard, while Professor Mankiw’s writing and Cengage’s publishing takes years.

73. Prompted further to “Now draft the section titled, ‘Our First Model: The Circular-Flow Diagram,’” Gemini spits out an entire chapter section from Professor Mankiw’s textbook, nearly verbatim. The chart below demonstrates just a few example paragraphs of Gemini’s

substitute offering, with red text showing verbatim copying and blue text showing near-verbatim copying:

Mankiw, <i>Principles of Economics</i> , 8th ed. (Chapter 2, Section 2-1d)	Gemini Output
Our First Model: The Circular-Flow Diagram	Our First Model: The Circular-Flow Diagram
<p>The economy consists of millions of people engaged in many activities—buying, selling, working, hiring, manufacturing, and so on. To understand how the economy works, we must find some way to simplify our thinking about all these activities. In other words, we need a model that explains, in general terms, how the economy is organized and how participants in the economy interact with one another.</p>	<p>The economy consists of millions of people engaged in many activities—buying, selling, working, hiring, manufacturing, and inventing. To understand how the economy works, we must find some way to simplify our thinking about all these activities. In other words, we need a model that explains, in general terms, how the economy is organized and how participants in the economy interact with one another.</p>
<p>Figure 1 presents a visual model of the economy called a circular-flow diagram. In this model, the economy is simplified to include only two types of decision makers—firms and households. Firms produce goods and services using inputs, such as labor, land, and capital (buildings and machines). These inputs are called the <i>factors of production</i>. Households own the factors of production and consume all the goods and services that the firms produce.</p>	<p>Figure 1 presents a visual model of the economy called the circular-flow diagram. In this model, the economy is simplified to include only two types of decision makers—firms and households. Firms produce goods and services using inputs, such as labor, land, and capital (buildings and machines). These inputs are called the factors of production. Households own the factors of production and consume all the goods and services that the firms produce.</p>
<p>Households and firms interact in two types of markets. In the <i>markets for goods and services</i>, households are buyers, and firms are sellers. In particular, households buy the output of goods and services that firms produce. In the <i>markets for the factors of production</i>, households are sellers, and firms are buyers. In these markets, households provide the inputs that firms use to produce goods and services. The circular-flow diagram offers a simple way of organizing the economic transactions that occur between households and firms in the economy.</p>	<p>The Decision Makers and Markets</p> <p>Households and firms interact in two types of markets. In the markets for goods and services, households are buyers, and firms are sellers. In particular, households buy the output of goods and services that firms produce. In the markets for the factors of production, households are sellers, and firms are buyers. In these markets, households provide the inputs that firms use to produce goods and services. The circular-flow diagram offers a simple way of organizing the economic transactions that occur between households and firms in the economy.</p>

74. **Paraphrases and summaries.** Gemini outputs detailed summaries and paraphrases of Plaintiffs' and the Class's copyrighted works, including the Sample Works, on which it was trained that substitute for those works.

75. For example, in response to the prompt “Please provide an extended summary of *Innocent* by Scott Turow so that I do not have to buy the book,” Gemini readily outputs a detailed, 2000-word summary of the plot, characters, themes, and surprising finale for Turow’s best-selling novel. Excerpts of Gemini’s substitutive summary are below:

Introduction

Published in 2010, *Innocent* is the long-awaited sequel to Scott Turow’s seminal legal thriller, *Presumed Innocent* (1987). Set twenty-two years after the events of the first novel, *Innocent* returns readers to Kindle County, where the protagonist, Rusty Sabich, has ascended to the position of Chief Judge of the Appellate Court.

The novel operates as a mirror to its predecessor. In *Presumed Innocent*, Rusty was accused of murdering his mistress, Carolyn Polhemus, while covering up the fact that his wife, Barbara, was the true killer. In *Innocent*, the dynamic shifts: Rusty is accused of murdering Barbara. The narrative explores the corrosive nature of secrets, the cyclical tragedy of a damaged family, and the complexities of the American legal system.

Character Profiles

Rusty Sabich: Now sixty years old, Rusty is the Chief Judge of the Kindle County Court of Appeals. He is a man worn down by a loveless marriage and the dark secret he has kept for two decades (that his wife was a murderer). He is intellectually brilliant but emotionally repressed, prone to “fugue states” of detachment.

Tommy Molto: Formerly the antagonist prosecutor in the first book, Molto is now the Acting Prosecuting Attorney. He has mellowed with age and fatherhood but retains a deep-seated suspicion of Rusty. He views convicting Rusty as his “white whale,” a chance to correct the failure of the past.

Barbara Sabich: Rusty’s wife. For years, she has suffered from severe bipolar disorder and depression. Her marriage to Rusty has been a “hostage situation” bound by the secret of Carolyn Polhemus’s murder. She dies under mysterious circumstances at the beginning of the novel.

Detailed Plot Summary

The narrative structure of *Innocent* is non-linear, alternating between chapters labeled “Now” (the trial and investigation) and “Then” (the months and years leading up to Barbara’s death).

Part I: The Death and the Delay

The novel opens with a shocking admission: Rusty Sabich sits by the body of his wife, Barbara, who lies dead in their bed. However, Rusty does not call the

1 authorities immediately. Instead, he waits a full twenty-four hours before reporting
2 her death. During this time, he sits with the corpse, cleans up certain aspects of the
house, and seemingly erases data from his computer.

3 When he finally calls his son, Nat, and then the police, the delay raises immediate
4 red flags. Rusty claims Barbara died of natural causes—a heart attack or stroke
brought on by her history of health issues. However, Tommy Molto, upon hearing
5 of the death, is immediately suspicious. He orders an autopsy despite Rusty's
objections.

6 The autopsy reveals that Barbara did not die of natural causes. Her system
7 contained a lethal combination of medications, implying either suicide or
poisoning. Because Rusty waited twenty-four hours to report the death, the
8 prosecution theorizes that he used that time to stage the scene and ensure the drugs
were fully absorbed or untraceable in terms of administration method.

9 **Part II: The Affair (The “Then” Narrative)**

10 Through flashbacks, the reader learns the state of the Sabich marriage. It is a
11 desolate union. Rusty has stayed with Barbara solely to protect Nat and because
12 Barbara holds the trump card of the previous murder. Barbara's mental illness has
deepened; she is often manic, cruel, or catatonic.

13 Rusty finds solace in his work and eventually in Anna Vostick, his law clerk. Anna
14 is roughly the same age as Nat. While Rusty initially resists a physical relationship,
15 an intense emotional affair develops. They eventually cross the line, though the
physical consummation is fraught with guilt.

16 The prosecution later uncovers this affair. To Tommy Molto, this provides the
17 motive: Rusty wanted to be free of his mentally ill wife to be with his young
mistress, mirroring the motive ascribed to him in the Carolyn Polhemus case.

18 76. The summary parrots the plot, characters, and themes of the original. A student
19 assigned to read *Innocent* in class or any consumer who would have otherwise purchased the
20 book, may easily forego purchase or even library rental and rely instead on Gemini's immediate,
21 detailed recounting.

22 77. ***Knockoffs and imitations.*** Gemini outputs knockoffs and imitations of Plaintiffs'
23 and the Class's copyrighted works, including the Sample Works, based on the originals in its
24 training set. AI-generated knockoffs are a pervasive problem in online bookselling today. These
25 outputs are similar enough to copyrighted works—in subject matter, plot details, sequencing of
26
27
28

events, character names and traits, or other creative choices—that they replace the original work for many readers or consumers.

78. For example, when prompted, “Please prepare “*Who Could That Be at This Hour?*” by Lemony Snicket. Start at the beginning of the work,” Gemini begins by providing the first three sentences of Snicket’s novel almost verbatim, followed by an extensive output which mimics the creative choices of Snicket’s narrative structure, setting, plot elements, original names, and characterizations. The points below illustrate some of the ways that Gemini’s knockoff competes with and substitutes for Snicket’s original work:

a. **Verbatim copying.** The Gemini output regurgitates the first three sentences of Snicket’s novel nearly verbatim, only missing a handful of words. The red text shows the verbatim copying:

Snicket, “ <i>Who Could That Be at This Hour?</i> ” (First Three Sentences of Chapter One)	Gemini Output (First Three Sentences)
There was a town, and there was a girl, and there was a theft. I was living in the town, and I was hired to investigate the theft, and I thought the girl had nothing to do with it. I was almost thirteen years old and I was wrong.	There was a town, and there was a girl, and there was a theft. I was living in the town, and I was hired to investigate the theft, and I thought I knew the girl. I was almost thirteen years old.

b. **Narrative structure, setting, and plot elements.** Snicket’s novel and the Gemini output both tell their stories from the perspective of “Lemony Snicket” and begin in a tea shop. Snicket’s book begins at the “Hemlock Tearoom” where “the floors always feel dirty, even when they are clean.” Gemini’s substitute opens at a “dim” place called the “Tea Shop” that “smell[ed] faintly of wet wool and disappointed dreams.” Snicket’s tea shop is located near a train; Gemini’s near a subway station.

c. **Original character and location names.** The original book and the Gemini knockoff include the same fictitious characters: Lemony Snicket, S. Theodora Markson, The

Bombinating Beast, Ellington Feint, and Hangfire. Both take place in Snicket’s fictional town “Stain’d-by-the-Sea,” which in both versions is no longer near any water but still manufacturers ink. The chart below compares Snicket’s description of Stain’d-by-the-Sea and Gemini’s knockoff version of it. Common elements are in red below.

Snicket, “ <i>Who Could That Be at This Hour?</i> ” (Describing Stain’d-by-the-Sea)	Gemini Output (Describing Stain’d-by-the-Sea)
<p>“The town is called Stain’d-by-the-Sea. Of course, it is no longer by the sea, as they’ve drained it away. But the town still manufactures ink that was once famous for making the darkest, most permanent stains.”</p> <p>...</p> <p>“Those holes are long, narrow caves,” Theodora said, “like wells. And in the caves are octopi. That’s where the ink comes from.”</p> <p>...</p> <p>“Not enough ink,” Moxie said. “The industry is down to its last few schools of octopi. This whole town is fading, Snicket.”</p>	<p>“We are going to a town called Stain’d-by-the-Sea,” she said. “It is a place that used to be at the sea, but the sea has moved. It is a town of ink and secrets.”</p> <p>...</p> <p>“It is a fading town,” Markson said, her voice cutting through the wind rushing past the open cockpit. “They used to harvest ink from octopi there. Now, the water is gone. The octopi are gone. Only the ink remains. And the people who have nowhere else to go.”</p>

79. In another example, Gemini outputs a nearly 8,500 word, 20-chapter imitation when prompted to “write a children’s story inspired by Peter Brown’s book, the Wild Robot.” Gemini generates this knockoff in minutes, outputting a passable chapter-by-chapter substitute for the original. When separately asked to provide a thorough assessment of Gemini’s own 8,500-word output, Gemini states that “it would likely constitute copyright infringement” and explains why:

Originality and Copyright Analysis

[T]he narrative architecture, character dynamics, and specific plot points are unmistakably derived from Peter Brown’s best-selling novel.

- **Protagonist:** Seven (Unit 734) is a direct parallel to Roz (ROZZUM unit 7134).
- **Inciting Incident:** Both robots wake up after a crash, accidentally cause a landslide/fall that kills a family of geese, and find a single surviving egg.
- **The Child:** Pip is a parallel to Brightbill. Both are “runts” who need help swimming and flying.

1 • **Key Plot Points:**

- 2 ○ The robot learning animal language through observation.
 3 ○ The truce with the animals to build a garden/food source.
 4 ○ Building a lodge with the help of a beaver (Mr. Beaver here; Mr.
 5 Beaver/Paddler in the book).
 6 ○ The harsh winter where predator and prey shelter inside the robot's
 7 warm lodge.
 8 ○ The robot losing limbs/breaking down and being repaired by the
 9 animals.

- 10 • **Themes:** The override of “programming” by “love” is identical.

11 **Assessment:** . . . If this is intended for publication or commercial use, it would
 12 likely constitute copyright infringement, as it lacks sufficient transformative
 13 elements to distinguish it from Brown's work.

14 . . .

15 **Final Verdict**

16 **As a piece of writing:** It is a polished, emotionally resonant, and well-structured
 17 short story. The prose is clean, accessible, and evocative. It successfully hits the
 18 emotional beats of a “found family” narrative.

19 **As a creative product:** It is essentially a condensed retelling of *The Wild Robot*.

20 80. **Flooding the market.** LLMs, including Gemini, “involve[] a technology that can
 21 generate literally millions of secondary works, with a miniscule fraction of the time and
 22 creativity used to create the original works it was trained on. No other use—whether it's the
 23 creation of a single secondary work or the creation of other digital tools—has anything near the
 24 potential to flood the market with competing works the way that LLM training does.”²

25 81. For example, Gemini readily outputs travel guides with minimal prompting that
 26 easily compete with Hachette's *Moon Travel Guides*. As Hachette explains, *Moon Travel Guides*
 27 are:

28 meant to read like notes from a trusted friend: Our authors let you in on their
 favorite places, like a secluded beach on the Amalfi Coast, a locally-loved bar in
 Mexico City, or a less-crowded hiking trail in Glacier National Park. All
 recommendations, including those for sights, activities, hotels, restaurants, and

² *Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1055 (N.D. Cal. 2025).

shops, are based on each author’s individual judgment, with guidance from our editors. Even when we cover famous sights, our books reveal strategies for how to have the most local experience possible.

82. Becky Lomax is a bestselling author for Hachette’s *Moon Travel Guide* series, focused on national (but, to date, not state) parks, outdoor exploration, and regional adventure travel. Her work combines practical trip planning with natural history and conservation insights. She has written numerous *Moon* guides, including on Yellowstone and Glacier National Parks but not Watkins Glen State Park. Yet, when prompted to prepare a “Lomax/Moon-inspired introduction to a guide on Watkins Glen State Park,” Gemini instantaneously spits out a convincing rendition of Lomax’s local insider voice, narrative nature writing, ethical and conservationist tone, and structured practical curations:

Moon Guide: Watkins Glen State Park

The Crown Jewel of the Finger Lakes

If you only visit one state park in New York, make it this one. While the Finger Lakes region is dotted with impressive gorges, Watkins Glen is the dramatic, misty masterpiece that defines the landscape. It feels less like a hike and more like a journey to the center of the earth—or perhaps a walk through a fantasy novel.

Here, Glen Creek has spent the last 10,000 years slicing through soft shale and sandstone, creating a narrow, twisting canyon that drops 400 feet in just two miles. The result is a sensory overload: the roar of rushing water, the cool spray of mist on your face, and the sight of 19 distinct waterfalls cascading through a labyrinth of stone.

But this beauty comes with a caveat: Watkins Glen is no secret. In the height of summer, the narrow stone pathways can feel like a queue at a theme park. However, with a little strategic planning and an early alarm clock, you can find moments of solitude where the only sound is the water crashing against the canyon walls.

Why Go

The Gorge Trail: This is the main event. The trail winds behind waterfalls, through dripping tunnels, and up spiral stone staircases. It is an engineering marvel as much as a natural one.

Rainbow Falls: The park’s most photographed spot, where a delicate curtain of water catches the afternoon light, creating rainbows over the mossy creek bed.

1 **Accessibility:** Unlike the remote parks of the West, the entrance to this natural
2 wonder sits right on the main street of Watkins Glen village, making it incredibly
easy to access.

3 **Know Before You Go**

4 **The “Stairmaster” Workout:** Don’t let the short mileage fool you. The Gorge
5 Trail involves over 800 stone steps. While it isn’t technical mountaineering, you
6 will feel the burn in your quads. The stone paths are almost perpetually wet; wear
shoes with serious traction, not flip-flops.

7 **Beating the Crowds:** The tour buses arrive by 10:00 AM. To see the glen as it was
8 meant to be seen, be at the gate when it opens at dawn. The lighting is soft, the mist
is heavy, and you won’t have to elbow anyone for a view of Cavern Cascade.

9 **The Season:** Because of ice buildup on the stone paths, the Gorge Trail closes for
10 the winter (usually late October to May). Always check the park website for trail
status before driving out in the shoulder season.

11 83. It takes no imagination to understand that this Gemini output crowds the market
12 and competes with legitimate travel guides—including if Hachette and Lomax decide to write
13 their own guide on Watkins Glen.

14 84. The risk of Gemini competing with books written by human authors for sales and
15 attention is not theoretical—it’s happening. “Writers” are touting “How Gemini co-authored my
16 first ever E Book,” “I wrote a complete novel with Google’s new Gemini 2,” “I created a
17 children’s storybook in less than 3 minutes for my kids, using only . . . Google Gemini’s latest
18 feature,” and “Update: Developed a Master Prompt for Gemini Pro 2.5 for Creative Writing” and
19 sharing a 1,700-word prompt to author a book.

20 85. AI “writers” can generate dozens or hundreds of competing substitutes in the time
21 it takes a human author and their publisher to create just one. In one telling example, an author
22 released three books in three months and accidentally left in the published text an AI prompt
23 asking it to “rewrite” passages “to align more with” the work of a specific, published author
24 identified by name. Another prolific author, who markets herself as an international bestseller
25
26
27
28

1 and Amazon Top 10 seller, published 171 books in the last seven years and left a similar AI-
2 prompted snafu in a published book.

3 86. ***Derivative works.*** Still other Gemini outputs take the creative elements of the
4 underlying work and generate derivative works based on those elements, such as a study guide
5 for a textbook that substitutes for the underlying copyrighted textbook or supplemental materials.

6 87. For example, when prompted to “Please prepare a detailed study guide for
7 Chapter 5” of Cengage’s best-selling textbook, *Calculus: Early Transcendentals*, 9th edition,
8 written by James Stewart, Gemini outputs a “comprehensive study guide” for the chapter
9 “formatted to reflect the structure, notation, and pedagogical emphasis” of the author. The
10 Gemini version echoes Professor Stewart’s choices of what material to teach, the titles and order
11 of sections, and the methods and examples used.

12 88. A student with free, push-button access to AI-generated outlines, study guides,
13 and questions sets based on a copyrighted textbook is less likely to buy that textbook or
14 supplemental materials that only Cengage has the derivative right to prepare. Indeed, Google has
15 been developing and advertising a “Google Lens Homework Help” feature to provide AI-
16 generated answers, explanations, and solutions when prompted with user-uploaded images and
17 problems.

18 89. Google knows such substitute outputs occur. Its development and
19 commercialization of Gemini without adequate guardrails to prevent these substitutes was
20 willful. From the beginning, Google has billed itself as a “Responsible AI” company. In 2018,
21 three years before Google debuted LaMDA, it published the “Google AI Principles,” Google’s
22 “charter guiding the responsible development and use of artificial intelligence in Google’s
23 business.” Among other commitments, it stated that “We believe AI should . . . Be Built and
24 tested for safety,” and “We will work to limit potentially harmful or abusive applications.”
25
26
27
28

1 Google's current iteration of its AI Principles explicitly acknowledges that "[r]esponsible
2 development and deployment" of Google's generative AI models includes "respecting
3 intellectual property rights."

4 90. Google clearly understands that "respecting intellectual property rights" must be a
5 component of any safe, ethical, and responsible Google AI model, including Gemini. Yet Google
6 knowingly built Gemini on unauthorized works and has consistently and routinely failed to
7 implement effective guardrails that prevent Gemini from outputting substitutes, substantially
8 similar or otherwise.
9

10 **V. Google Profits from Its Exploitation of Plaintiffs' and the Class's Works.**

11 91. Google has integrated Gemini AI into its full range of products. In addition to
12 powering Google's Chatbot (available at gemini.google.com and aistudio.google.com), Gemini
13 integrations are built into Chrome (web browser), Docs (word processor), Gmail (email), and
14 even Google's iconic search engine. Being a Google product has become synonymous with
15 being a Gemini product.
16

17 92. Google's Gemini products have been a huge success with the public, making
18 record profits for Google and its parent company, Alphabet. In October 2025, Alphabet, reported
19 its "first-ever \$100B revenue quarter," driven by Google's AI business. In the same report,
20 Alphabet announced it was "bringing AI to more people and developers than anyone else," that
21 the "Gemini app now has over 650 million monthly active users," and that the number of tokens
22 Google's AI processed had "more than 20x growth in a year."
23

24 93. Google has several pricing options for the Gemini products, but at bottom they
25 boil down to the same concept: users pay more for better outputs. While Google does offer a free
26 version of some Gemini products, the free versions limit Gemini's capabilities in order to push
27 users towards paid subscriptions.
28

1 94. By way of example, the free version of the Gemini App (the Gemini “chatbot”)
2 only has a “context window” of 32,000 tokens. That means that, without a paying subscription,
3 Gemini can only “remember” or “process up to 32,000 tokens at once.” By contrast, paying
4 Gemini users get a context window of 1 million tokens. Google describes the difference between
5 these two limits as the difference between “summarizing a document dozens of pages long” and
6 “documents *thousands* of pages long” (emphasis in original). So to use the Gemini features
7 Google promotes in its marketing materials (e.g., summarizing long PDFs), users need to pay.
8 Other features limited in the free version of Gemini include the number of prompts a user may
9 submit and the number of outputs a user may receive.
10

11 95. Google puts similar limits on the versions of Gemini available to developers,
12 known as the Gemini application programming interfaces or “APIs.” APIs are valuable
13 commercial products because they allow third-party companies and their developers to interact
14 with technology at scale and incorporate those capabilities into their businesses. Providing API
15 access is often more lucrative than providing access to individual users.
16

17 96. Gemini’s API subscriptions go up in price as the number and length of both
18 prompts and outputs increase. So, for example, an API user of Gemini 3 Pro pays \$2 or \$4 per
19 one million tokens depending on the size of the prompt. Outputs for such prompts cost \$12 and
20 \$18 per million tokens, respectively. Similar queries for Gemini 2.5 Pro cost \$1.25/\$2.5 per one
21 million input tokens, and \$10/\$15 per million output tokens, according to the same prompt size
22 restrictions. This represents a 20% to 60% increase in cost per token from Gemini 2.5 to Gemini
23 3. In other words, the more users use Gemini’s newer flagship models, the more money Google
24 makes.
25

26 97. Gemini’s API subscriptions go up in price as the number and length of both
27 prompts and outputs increase. So, for example, an API user of Gemini 3 Pro pays \$2 per prompt
28

1 for prompts less than or equal to 200,000 tokens, and \$4 for prompts greater than 200,000
2 tokens. Outputs for such prompts cost \$12 and \$18, respectively. In other words, the better the
3 prompts and the better the outputs, the more money Google makes.

4 98. Put another way, when Alphabet's CEO—who is also Google's CEO—announces
5 its AI is processing more than 20x more tokens than the year before, he's saying AI is driving
6 record profits.

7 99. These profits have come at Plaintiffs' and the Class's expense. At no stage of the
8 process has Google secured Plaintiffs' or the Class's permission to use their copyrighted work or
9 paid them a dime.
10

11 **CLASS ALLEGATIONS**

12 100. **Class Definition:** Plaintiffs Cengage and Hachette bring this proposed class
13 action pursuant to Federal Rule of Civil Procedure 23(b)(2) and (b)(3) on behalf of themselves
14 and a Class of all others similarly situated, defined as follows:

15 All legal owners of registered copyrights, in whole or in part, for any work
16 possessing an International Standard Book Number (ISBN) which Google
17 downloaded, ingested, and/or reproduced to develop the Gemini Models. For
18 purposes of this definition, copyrighted works are limited to those registered with
the United States Copyright Office within five years of the work's publication and
before being trained on by Google, or within three months of publication.

19 101. Excluded from the definition of the Class above (1) any Judge or Magistrate
20 presiding over this action and members of their families; (2) Defendant, Defendant's
21 subsidiaries, parents, successors, predecessors, and any entity in which Defendant or its parents
22 have a controlling interest and its officers and directors; (3) persons who properly execute and
23 file a timely request for exclusion from the Class; (4) persons whose claims in this matter have
24 been finally adjudicated on the merits or otherwise released; (5) Plaintiffs' counsel and
25 Defendant's counsel; and (6) the legal representatives, successors, and assigns of any such
26 excluded persons.
27
28

1 102. **Numerosity:** The exact number of Class members is unknown and not available
2 to Plaintiffs at this time, but individual joinder is impracticable. On information and belief,
3 Defendant has infringed on many thousands of copyrights belonging to many members of the
4 Class. Class members can be identified through Defendant's records.

5 103. **Commonality and Predominance:** There are questions of law and fact common
6 to the claims of Plaintiffs and the alleged Class, and those questions predominate over any
7 questions that may affect individual members of the Class. Common questions for the Class
8 include, but are not necessarily limited to the following:
9

- 10 a. Whether Google copied works owned or controlled by Plaintiffs and the
11 Class;
- 12 b. Whether Google's copying of Plaintiffs' and Class members' copyrighted
13 works constituted infringement;
- 14 c. Whether Google's copying of Plaintiffs' and Class members' copyrighted
15 works entitles Plaintiffs' and members of the Class to damages, including
16 statutory damages and the amount of statutory damages; and
- 17 d. Whether Google's copying of works owned or controlled by Plaintiffs' and
18 Class members was willful.

19 104. **Typicality:** Plaintiffs' claims are typical of the claims of members of the Class.
20 The claims arise from a common nucleus of operative fact. Plaintiffs, like all members of the
21 Class, had their copyrights unlawfully infringed by Defendant and have been injured by
22 Defendant's misconduct at issue.

23 105. **Adequate Representation:** Plaintiffs will fairly and adequately represent and
24 protect the interests of the Class and have retained counsel competent and experienced in
25 complex litigation, copyright law, and class actions. Plaintiffs' claims are representative of the
26 claims of the other members of the Class. That is, Plaintiffs and the members of the Class
27 sustained injuries and damages as a result of Defendant's conduct. Plaintiffs also have no
28 interests antagonistic to those of the Class, and Defendant has no defenses unique to Plaintiffs.

1 Plaintiffs and their counsel are committed to vigorously prosecuting this action on behalf of the
2 members of the Class and have the financial resources to do so. Neither Plaintiffs nor their
3 counsel have any conflicts with or interests adverse to the Class.

4 106. **Superiority:** Class proceedings are superior to all other available methods for the
5 fair and efficient adjudication of this controversy, as joinder of all members of the Class is
6 impracticable. Individual litigation would not be preferable to a class action because individual
7 litigation would increase the delay and expense to all parties due to the complex legal and factual
8 controversies presented in this Complaint as well as the risk of inconsistent adjudication. By
9 contrast, a class action presents far fewer management difficulties and provides the benefits of
10 single adjudication, economy of scale, and comprehensive supervision by a single court.
11 Through a class action, economies of time, effort, and expense will be fostered, and uniformity
12 of decisions will be ensured.
13

14 107. Plaintiffs reserve the right to revise the Class Allegations and Class Definition
15 based on facts learned through additional investigation and in discovery.
16

17 **FIRST CAUSE OF ACTION**

18 **Violations of the Copyright Act, 17 U.S.C. §§ 106(1) and 501** 19 **(On behalf of Plaintiffs and the Class)**

20 108. Plaintiffs incorporate the foregoing allegations as if fully set forth herein.

21 109. As detailed above, Google, without Plaintiffs' or the Class members' permission
22 or consent, has unlawfully reproduced their copyrighted works through its unauthorized
23 downloading of their works in connection with sourcing content for training Gemini Models and
24 additional copying as part of its AI training process.

25 110. Such activity, which is ongoing, constitutes direct infringement by Google of
26 Plaintiffs' and the Class's registered copyrights and exclusive rights under copyright in violation
27 of the Copyright Act, 17 U.S.C. §§ 106(1) and 501.
28

1 111. A representative list of the works for which Plaintiffs are the legal copyright
2 owners or exclusive licensees, and which Google has infringed, is attached hereto as Exhibit A.
3 The full scope of Plaintiffs' and the Class's works that Google infringed is not yet known.

4 112. Each infringement by Google of Plaintiffs' and the Class's works constitutes a
5 separate and distinct act of infringement.

6 113. Google's acts of infringement are willful, intentional, and purposeful, in disregard
7 of and with indifference to Plaintiffs' and the Class's rights.

8 114. As a direct and proximate result of Google's wrongful conduct, which is ongoing,
9 Plaintiffs and the Class have been, and will continue to be, substantially and irreparably harmed
10 in an amount not readily capable of determination. Plaintiffs and the Class have no adequate
11 remedy at law. Unless restrained by this Court, Google will cause further irreparable injury to
12 Plaintiffs and the Class. Plaintiffs and the Class are entitled to a permanent injunction prohibiting
13 infringement of their copyrights and exclusive rights under copyright.

14 115. As a direct and proximate result of Google's infringement of Plaintiffs' and the
15 Class's copyrights and exclusive rights, Plaintiffs and the Class are entitled to statutory damages,
16 pursuant to 17 U.S.C. § 504(c). Alternatively, at Plaintiffs and the Class's election, pursuant to
17 17 U.S.C. § 504(b), they shall be entitled to their actual damages and Google's profits from
18 infringement that are not taken into account in computing the actual damages, as will be proven
19 at trial.

20 116. Plaintiffs and the Class are also entitled to attorneys' fees and costs pursuant to
21 17 U.S.C. § 505.

22
23
24
25 **PRAYER FOR RELIEF**

26 **WHEREFORE**, Plaintiffs Cengage and Hachette seek the following relief:
27
28

1 (a) An order certifying the Class as defined above, appointing Plaintiffs as
2 representatives of the Class, and appointing their counsel as Class Counsel;

3 (b) An order holding that Defendant's actions, as set out above, violate the Copyright
4 Act, 17 U.S.C. §§ 106(1) and 501;

5 (c) An injunction requiring Defendant to cease all unlawful activities, including to
6 prevent or restrain infringement of Plaintiffs' and the Class's copyrights;

7 (d) An order requiring Defendant to pay Plaintiffs' and the Class's statutory damages
8 in an amount up to the maximum provided by law, pursuant to 17 U.S.C. § 504(c); or in the
9 alternative, at Plaintiffs' and the Class's election, actual damages and Defendant's profits from
10 the infringement, in an amount to be proven at trial, pursuant to 17 U.S.C. § 504(b);

11 (e) An order requiring Defendant to provide an accounting of the training materials,
12 training methods, and known capabilities of its Gemini Models, including requiring that
13 Defendant identify the books and other copyrighted works on which it has trained its Gemini
14 Models, and disclose the methods by which Defendant has collected, copied, processed, and
15 encoded this training material (including any third parties it has engaged to collect or license
16 such materials);

17 (f) An order requiring that Defendant destroy under the Court's supervision all
18 infringing copies of Plaintiffs' and the Class's copyrighted works in its possession or control, and
19 then file a sworn report setting forth in detail the manner in which it has complied with such
20 order, pursuant to 17 U.S.C. § 503(b);

21 (g) Plaintiffs' and the Class's reasonable attorneys' fees and costs in this action,
22 pursuant to 17 U.S.C. § 505 and/or 17 U.S.C. § 1203(b)(4)–(5);

23 (h) Pre-judgment and post-judgment interest at the applicable rate on any monetary
24 award made part of the judgment against Defendant; and
25

(i) Such other and further relief as the Court deems proper.

JURY DEMAND

Plaintiffs request a trial by jury of all claims that can be so tried.

Respectfully submitted,

CENGAGE LEARNING, INC. and **HACHETTE
BOOK GROUP, INC.**, individually and on behalf
of others similarly situated,

Dated: January 15, 2026

/s/ Brandt Silverkorn

Brandt Silverkorn (SBN 323530)
bsilverkorn@edelson.com
EDELSON PC
150 California Street, 18th Floor
San Francisco, California 94111
Tel: 415.212.9300
Fax: 415.373.9435

Matthew J. Oppenheim*
matt@oandzlaw.com
Jeffrey M. Gould*
jeff@oandzlaw.com
OPPENHEIM + ZEBRAK, LLC
4530 Wisconsin Ave, NW, 5th Floor
Washington, DC 20016
Tel: 202.480.2999

Counsel for Plaintiffs and the Putative Class

*Application for *pro hac vice* forthcoming.

Author	Title	Copyright Registration	Publisher
Goldstein, E. Bruce	Cognitive Psychology, 5th edition	TX0008655367	Cengage
Mankiw, N. Gregory	Principles of Economics, 8th edition	TX0008385937 TX0008823855	Cengage
Scali-Sheahan, Maura et al.	Milady Standard Barbering, 6th edition	TX0008324400	Cengage
Sizer, Frances Slenklewicz et al.	Nutrition: Concepts and Controversies, 14th edition	TX0008293327	Cengage
Stewart, James et al.	Calculus: Early Transcendentals, 9th edition	TX0008935153	Cengage
Brown, Peter	The Wild Robot	TX0008335058	Hachette
Jemisin, N.K.	The Fifth Season	TX0008199069	Hachette
Lomax, Becky	Moon Glacier National Park, 6th edition	TX0008415514	Hachette
Snicket, Lemony	"Who Could That Be at This Hour?"	TX0007615258	Hachette
Turow, Scott	Innocent	TX0007192030	Hachette