IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK

HACHETTE BOOK GROUP, INC., CENGAGE LEARNING, INC., ELSEVIER INC., SCOTT TUROW, AND S.C.R.I.B.E., INC., individually and on behalf of others similarly situated,

Plaintiffs,

v.

GOOGLE LLC,

Defendant.

Case No.: 26-cv-5870

CLASS ACTION COMPLAINT FOR:

(1) Violations of the Copyright Act, 17 U.S.C. §§ 106(1) and 501 – reproduction of unauthorized copies from Google Books and other Google services

(2) Violations of the Copyright Act, 17 U.S.C. §§ 106(1) and 501 – reproduction via web scrapes

(3) Violations of the Copyright Act, 17 U.S.C. §§ 106(1) and 501 – reproduction in training

(4) Violations of the Digital Millennium Copyright Act, 17 U.S.C. § 1202(b) – removal and/or alteration of copyright management information

DEMAND FOR JURY TRIAL

CLASS ACTION COMPLAINT AND DEMAND FOR JURY TRIAL

Plaintiffs Hachette Book Group, Inc. ("Hachette"), Cengage Learning, Inc. ("Cengage"), and Elsevier Inc. ("Elsevier") (collectively, "Publishers"); and Scott Turow and S.C.R.I.B.E., Inc. (collectively "Turow," and with Publishers, "Plaintiffs"), bring this Class Action Complaint and Demand for Jury Trial on behalf of themselves and all others similarly situated (the "Class," as defined with specificity in Paragraph 139, *infra*), against Google LLC for copyright infringement based on Google's unauthorized reproduction of Plaintiffs' and the Class's works through its sourcing of content for, and development and training of, its generative artificial

intelligence ("AI") platform called Gemini, as well as for removal of copyright management information ("CMI"). Plaintiffs allege as follows based on personal knowledge as to matters relating to themselves and on information and belief as to all other matters.

NATURE OF THE CASE

1. Desperate to maintain its online dominance, Google abandoned its early motto of "Don't be evil" and engaged in one of the most prolific infringements of copyrighted materials in history. Google first illegally copied millions of books and journal articles to assemble a vast trove of copyrighted materials for its own commercial use. It did this by copying works obtained for strictly limited use in Google Books and other Google services, and downloading unauthorized web scrapes of virtually the entire internet. Google then copied those stolen works many times over to train its multi-billion-dollar generative AI system called Gemini. Compounding the assault, Google then deployed a purpose-built service designed to generate content that directly substitutes for Plaintiffs' and the Class's original works.

2. In doing so, Google reproduced millions of copyrighted works without permission, without providing any compensation to authors or publishers, and with full knowledge that its conduct violated copyright law. Google also stripped CMI from the copyrighted works it stole to conceal its training sources and facilitate their unauthorized use.

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

4. Both authors and publishers alike play critical roles in creating, disseminating, and protecting literary works. Authors devote years to conceiving, writing, and refining their works. Publishers invest substantial capital, expertise, and creativity to bring those works to market, often building reputations and catalogs over decades or centuries, as they fulfill their mission to educate, entertain, and inspire generations of readers.

5. A basic principle of the literary ecosystem is that the publisher is a guardian of an author's rights and assumes exclusive rights—typically, the exclusive rights to reproduce and distribute physical and digital formats of authors' books and journal articles. In many instances, a publisher holds additional exclusive rights. Publishers exercise their rights by making and disseminating copies of works in their catalogues and authorizing others to do so, including through dynamic licensing markets.

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

7. Google willfully sidestepped this longstanding system designed to protect copyrights and compensate authors and publishers through a series of deliberate choices to develop Gemini[1]:

---

[1] The AI models at issue in this litigation include all versions, iterations, and relatives of LaMDA, PaLM, Bard, and Gemini (collectively, "Gemini" or the "Gemini Models"). The AI-powered products at issue in this litigation include all versions, iterations, and relatives of products that incorporate, rely on, or otherwise use Google Search, Google Cloud, Gmail, 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. Plaintiffs reserve the right to add additional models and products based on facts learned through additional investigation and discovery.

- First, Google illegally copied Plaintiffs' and the Class's copyrighted books to compile a massive corpus of source material for its own commercial use. To do this, Google secretly copied millions of works obtained for strictly limited purposes in connection with Google Books and other Google services. Google also downloaded web scrapes of virtually the entire internet, including from known pirate sources and from behind legitimate paywalls.

- Second, to train its models, Google repeatedly copied these works without authorization—first into computer memory, then into formats its AI systems could parse, and then into the training set used to build each model.

- Third, with each new AI model, Google copied (and continues to copy) Plaintiffs' and the Class's copyrighted works again from model to model.

8. The result is an AI system that competes directly with Plaintiffs' and the Class's works in the market. Those substitutes take multiple forms, including verbatim and near-verbatim copies of portions or entire works, replacement chapters of academic textbooks, summaries and alternative versions of famous novels, and inferior knockoffs that copy creative elements of original works. Gemini even tailors outputs to mimic the expressive elements and creative choices of specific authors.

9. The substitute works that Gemini creates go beyond mere copies of specific works. For example, Gemini can generate a 100-page murder mystery set in a quiet seaside town filled with secrets, that substitutes for an original copyrighted murder mystery on which Gemini trained. And it can do that in 20 minutes for a mere $0.39. No publisher or author can compete with that. Users are already touting Gemini's ability to generate books with ease, and the market is flooding with AI-generated substitutes. The scale and speed at which Gemini can create books and compete with human writers is unprecedented, and it can only do that because Google copied Plaintiffs' and the Class's works to train its AI.

10. Google's infringement causes, and continues to cause, substantial and irreparable harm to Plaintiffs and the Class. It displaces legitimate sales of books and journal articles by downloading copies from unauthorized sources. It usurps the growing AI licensing market by

copying Plaintiffs' and the Class's works into Gemini's training set without permission or compensation. And Gemini's wide-ranging and varied outputs substitute for copyrighted works and dilute the overall market.

11. All of Google's infringement was willful. Authorized copies of Plaintiffs' and the Class's books are widely available for purchase or license. Yet Google chose unauthorized sources. That is particularly egregious given that Google is well-aware of the licensing market for AI training materials and already licenses content for training.

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

13. Ultimately, Google has profited—and continues to profit—massively from this historic infringement. In October 2025, Google's parent company, Alphabet, reported its first-ever $100 billion revenue quarter, driven by Google's AI business. In the same report, Alphabet announced it was "bringing AI to more people and developers than anyone else" and that the "Gemini app now has over 650 million monthly active users," representing "more than 20x growth in a year." Alphabet's more recent financial reporting confirms that its infringing AI business is only getting bigger and generating more revenue for Google.

14. While AI technology may be new, the legal principles at the center of this case are not. Copyright law applies to AI companies, including Google, with the same force as every other company that has complied with these laws for decades. If left unaddressed, Google will continue to infringe Plaintiffs' and the Class's rights, cause broad and lasting damage to the

literary industry and authors, and weaken the incentive to create that is at the core of the Copyright Act. These facts are not a referendum on AI technologies, but rather their greedy and irresponsible deployment. Accordingly, Plaintiffs bring this action on behalf of themselves and a proposed Class to stop Google's infringement and preserve the important work of authors, publishers, and the literary ecosystem more broadly.

PARTIES

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

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

17. Plaintiff Elsevier is a Delaware corporation with its principal place of business at 101 Park Avenue, New York, New York 10178.

18. Plaintiff Scott Turow is an author and a resident of Naples, Florida.

19. Plaintiff S.C.R.I.B.E., Inc. is a Delaware corporation with its principal place of business at Naples, Florida.

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

JURISDICTION AND VENUE

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

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

22. A substantial portion of the events giving rise to the claims occurred in this District and Plaintiffs have suffered substantial injuries in this District.

23. New York City is the center of the United States publishing industry and has been for more than two centuries. Today, the city houses the headquarters or main offices of numerous trade publishers, educational publishers, independent presses, literary agencies, and related media organizations. Hachette and Elsevier have been headquartered in New York for decades. Publishers conduct key industry functions in New York such as editorial acquisitions, contract negotiation, marketing, rights management, and content protection, reflecting the concentration of decision-making authority here. In addition, many authors, agents, and publishing professionals maintain a regular presence in New York. As a result, a significant portion of the business and operational activity associated with U.S. publishing is based in New York City.

24. A significant number of Plaintiffs' customers are located in this State and District, and a significant share of Plaintiffs' business from customers, including purchases, subscriptions, and licensing, is transacted with individuals and entities residing in this State and District. As such, the injuries alleged herein occurred in this State and District. As a direct and proximate result of Google's unauthorized reproduction and/or distribution of Plaintiffs' and the Class's copyrighted works in New York and elsewhere, Plaintiffs and the Class have lost and will continue to lose revenue and profits from the market for sales and licensing.

25. Defendant Google is subject to the specific personal jurisdiction of this Court pursuant to N.Y. C.P.L.R. §§ 302(a)(1)–(4), as it has purposefully directed its activities at New York and has purposefully availed itself of the benefits of doing business in New York, including

by registering to transact and by transacting business in this State and District, and by supplying goods and services in the State and District, through both the promotion and distribution of its products (including Gemini) and services to customers in this State and District. In addition, Google's agents reside in this District and may be found in this State and District.

26. Google possesses real property and has established a continuing presence in this State and District. Google employs significant employees and resources within New York City. Google has maintained three offices within the District during the relevant period, including at 111 8th Avenue, New York, NY 10011; 75 9th Avenue, New York, NY 10011; and 550 Washington St, New York, NY 10014. One of Google's NYC offices serves as the North American headquarters of Google's Global Business Organization—the part of the company that controls Google's money-making operations.

27. During the relevant period of Google's development, training, and deployment of the Gemini Models, Google relied heavily on its New York City offices and employees. Indeed, Google hosts a public website describing "Google AI New York": "Google AI New York creates useful solutions to fundamental computational problems in theory and algorithms, machine learning, journalism, speech, and other data-driven disciplines, with impact on Google's products and scientific progress." Some of the many AI features Google AI New York develops include "Large Scale Machine Learning," "Market Algorithms," and "Large-Scale Optimization."

28. Google has numerous New York City-based, high-profile employees working on AI and the Gemini Models. These include the VP, Head of Google Research NY; the Principal Scientist and Director at Google DeepMind, Google's AI development division; a Senior Director and Google Fellow; a VP, Google Fellow at Google DeepMind; a Distinguished Scientist and Senior Research Director; the Creative Lead for AI and creativity research; and

research scientists working on Gemini. Former AI employees at Google's New York City offices similarly include the Managing Director, Global Partnerships, for News and Publishing; a Senior Research Director; and the Research Director originally brought from Facebook/Meta AI Research to build up Google's New York AI team.

29. These New York-based employees and teams acted as part of Google's integrated AI operation to research, design, develop, train, test, evaluate, improve, deploy, commercialize, and support Google's AI, including the Gemini Models. They are supported by a wider team of legal, engineering, infrastructure, content, and management staff actively engaged in developing, implementing, and promoting Google's AI technology at issue here. As of July 2026, Google listed over 400 open positions for its New York City offices, including 200 open positions for Google Research.

30. Google promotes and advertises its AI products and services in this State and District, including through its dynamic, interactive website (gemini.google.com and aistudio.google.com) and integrations are built into Chrome (web browser), Docs (word processor), Gmail (email), and even Google's marquee search engine.

31. Google targets customers in this State and District with promotional material tailored to a New York audience, including online and physical advertising of Gemini's AI products and services. To take just a few examples: during the 2024 holiday season, Google Shopping ran an AI ad campaign "[p]owered by Gemini and Imagen 3 models," and "generated hundreds of unique ads tailored to the specific location where each one appeared," including New York; during the 2025 holiday season, Google ran an "Imagine If . . ." campaign that displayed AI-generated works "across thousands of digital screens throughout the MTA's

subway stations", transforming "New York's transit network [f]or four weeks," and culminating in an "iconic" Times Square display.

32. As set forth in further detail in Section V, *infra*, Google plans for Gemini and related AI technology to be so pervasively integrated across Google's products and services as to become inseparable from them. Already, Gemini and other AI features draw more and more users and higher screentime and app engagement, increasing Google's revenue. Thus, Google's wider business, advertising, and product development teams—for which there are approximately 14,000 employees in Google's 3 Manhattan offices—reap the benefits of and build on Google's infringement of Plaintiffs' and the Class's works.

33. Google processes subscriptions and purchases from customers in this State and District, transmits Plaintiffs' copyrighted content to users in this State and District, and has a significant number of customers in this State and District. Google reported that, in 2023, it helped provide $105.94 billion dollars of business activity in New York.

34. Google has filed multiple civil cases in this District that it could have filed elsewhere, including a case concerning Gemini as recently as last month. Those cases were brought against individuals living outside the United States alleging harm occurring "throughout the United States."

35. Venue is proper pursuant to 28 U.S.C. §§ 1391(b) and 1400(a) because a substantial part of the events or omissions giving rise to Plaintiffs' claims occurred in the District, including Google's copyright infringement and commercialization of its Gemini Models.

COMMON FACTUAL ALLEGATIONS

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

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

37. Cengage is a leading educational publisher with deep historic roots, devoted to creating and publishing high quality textbooks and other learning materials. Cengage develops, markets, distributes, and sells a comprehensive range of traditional and digital educational content, including textbooks, to educators and students. Cengage's textbooks have achieved acclaim not only in the United States, but also in many other countries around the world, and are among the most popular and widely used titles in their fields. Its many publishing imprints include prominent brands such as Brooks Cole, Cengage, Cengage Learning, Delmar, Gale, Heinle, Milady, South-Western Educational Publishing, and Wadsworth.

38. Elsevier is a world leading provider of professional information solutions in the science, medical, and health sectors. Elsevier publishes, markets, sells, and licenses academic journals, textbooks, and examinations in these fields. Elsevier publishes some of the world's most prestigious journals, including The Lancet and Cell. The majority of Elsevier's institutional customers are universities and other educational institutions, governmental entities, hospitals, and public and private corporations that purchase physical and electronic copies of Elsevier's products and access to Elsevier's digital libraries. Elsevier distributes its scientific journal

articles and book chapters electronically via its proprietary subscription database "ScienceDirect" (www.sciencedirect.com).

39. Scott Turow is a best-selling author and former practicing lawyer, known for setting his novels in fictional Kindle County's legal community. He is the author of 14 bestselling works of fiction, including *Presumed Innocent*, *Innocent*, *Identical*, *Testimony*, and *The Last Trial.* Mr. Turow has also published two nonfiction books, including *One L*, about his experience as a law student. His books have been translated into more than 40 languages, sold more than 40 million copies worldwide, and been adapted into movies and television projects. Mr. Turow is the president of S.C.R.I.B.E., Inc., which owns the copyrights to all his published books, including *Presumed Innocent*.

40. Publishers work closely with their authors throughout all phases of a written work's lifecycle. They provide financial support like advances and royalties. They support authors by employing literary professionals, like editors, who collaborate with authors on development and improvement of their works, and copyeditors, who help hone a book into its final form. Publishers provide their authors with a range of support resources to bring their creative visions to life: from art departments and illustrators; to design professionals, who lay out print and digital editions; to production departments that coordinate and execute on the creation of physical and digital books; to legal departments, who work with authors on pre-publication review and defending publishers' works; to marketing and publicity departments that spread the word and generate buzz for works; to sales teams that work with accounts to advocate for titles and authors, help get textbooks adopted by educators, school districts, and educational institutions, and match book sellers with the books their customers wish to purchase; to

distribution channels that ensure publishers' work are stocked on shelves, in warehouses, and on online platforms.

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

42. Publishers rely on their ability to control the dissemination of works to maximize revenue. This includes the ability to make works available in multiple formats and across different markets, to determine distribution terms and pricing, and to license works for various uses. It also includes the ability to engage in new markets and exploit new uses that arise during the lifespan of a work—and, importantly, includes the right to refrain from licensing certain uses. This control enables the broadest dissemination of works, and the ideas and information they contain, and ensures publishers can compensate and support their authors, recoup their significant investments, and invest in the next generation of written works.

43. For educational publishers like Plaintiffs Cengage and Elsevier, the textbook market extends beyond the underlying textbook to a variety of associated materials that complement the textbook in the learning process, such as study aids, solutions manuals, and online learning platforms. Once a textbook is adopted by a particular instructor to teach a particular course at a particular institution, students in that course typically purchase the textbook, along with some or all of the associated supplemental learning materials. For trade publishers, like Plaintiff Hachette, the book market is driven by numerous factors, including

popularity and hype, school uses, marketing, relevance of topics of current interest, pricing of the work and competing works, and supply chain. The market for trade books extends from print books to ebooks and audiobooks to library uses, and beyond.

44. Journal publishers, like Elsevier, operate within a well-established market for scholarly and scientific research. That market includes not only subscriptions to individual journals and bundled journal portfolios, but also a range of licensing arrangements with universities, libraries, corporations, and other institutional users that provide access to curated collections of articles. These licenses are carefully managed, with publishers determining the scope, pricing, and terms of access to preserve the value of their content and to prevent uses that would substitute for or undermine revenue streams. The market for journal content extends across formats, platforms, and evolving technological uses, and depends on publishers' continued ability to control how their works are accessed, used, and licensed.

45. Without adequate legal protections, unchecked infringement would deny publishers and authors control of their works and erode the artistic, cultural, educational, and economic value of their publications. It would reduce authors' incentives to create, reduce publishers' incentives to invest in new works, and stifle a major engine of human creative expression.

II. Google Needed Plaintiffs' and the Class's High-Quality Written Works to Maintain Its Online Dominance.

46. The AI "arms race" has been driven by a collective belief among AI developers and investors that using more training materials will lead to more commercially successful generative AI models. That belief pushed companies, including Google, toward ever-larger training sets and models, turning scaling into the dominant business strategy.

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

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

49. High-quality written works, including books and journal articles, are a valuable source of training materials for LLMs. As one group of AI researchers explained, "books are invaluable for long-range context modeling research and coherent storytelling." In another case, the court observed that the defendant AI company preferred to source copyrighted works written by well-regarded authors for its LLM training datasets because they had "well-curated facts, well-organized analyses, and captivating fictional narratives—above all with 'good writing' of the kind 'an editor would approve of.'" Researchers have also confirmed that readers prefer outputs from AI models trained on copyrighted books over outputs from AI models trained on more general text datasets, and that LLMs can "struggle[] to answer questions about cutting edge academic papers," which "underscores the need to enhance the specialized knowledge of LLMs for their effective use in various scientific fields."

50. Internally, Google emphasized over and over that training Gemini on a large volume of high-quality written works was essential to the model's success. Google sought a massive training set for Gemini because it believed that the "more data that is used to train the [model], the better and more accurate it will be at performing its task." Google wanted copyrighted books, specifically, because "using only public domain books to train the model . . . resulted in lower performance on several benchmarks." Google's AI teams thus "chose to use copyrighted books because they found that using non-copyright has a big impact on" Gemini's performance. As one Google internal document put it, "It is important to emphasize that the team is keen to get access to high volume/high quality books asap and that format is critical."

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

III. Google Infringed Plaintiffs' and the Class's Copyrights to Source Content for and to Train Gemini.

A. Google Illegally Copied Plaintiffs' and the Class's Works from Google Books and other Google services to Source Gemini Training Materials.

52. For most of the 21st century, Google has been the dominant online tech company. From search, to email, to advertising, and beyond, Google leveraged every aspect of online life to its own commercial benefit. When AI began to emerge as the next frontier in tech, Google declared it would be an "AI-first company," and began investing heavily in AI projects. Gemini, Google's flagship generative AI model, is one of those projects.

53. For decades, Google's publishing, author, and library partners have negotiated with Google regarding the limited-scope use of copyrighted works for certain services available through Google's platforms. As a result of these services, created prior to the emergence of AI,

Google gained access to a massive trove of written works that none of Google's AI competitors can access. Instead of operating within the scope of those permissions, Google cashed in on this advantage by brazenly copying millions of copyrighted works obtained for an expressly limited purpose and trained its commercial AI models on those stolen works.

54. Google's search service includes a book-searching service known as "Google Books," available as an independent search tool or a tab within its primary search bar. This search service allows users to search the text of millions of books and view search results in the form of limited snippets and previews of copyrighted works, or complete public-domain works. For copyrighted works, Google Books' search results may only display "basic bibliographic information about the book, and in many cases, a few snippets – a few sentences showing your search term in context." Google Books search results may also provide links to authorized copies that can be purchased from online bookstores and borrowed from libraries.

55. For years, publishers and authors provided Google with massive amounts of copyrighted works for the express, limited purpose of making books searchable via Google Books. Publishers and authors never authorized Google to copy the works they received for Google Books for the completely separate purpose of training its AI models and building a multi-billion dollar competing business.

56. Google began assembling Google Books decades ago, but it did not get Plaintiffs' permission to copy their works for any purpose. As a result, Google Books was the subject of protracted litigation. After 10 years of litigation, the Second Circuit ruled in 2015 that Google's "making of a complete digital copy of Plaintiffs' works" was fair use, but *only* because Google made its copies "for the purpose of providing the public with [Google Books] search and snippet view functions." The Second Circuit's ruling did not authorize, permit, or deem it fair use for

Google to make copies of copyrighted works for the new and separate purpose of training or developing commercial AI models, or indeed any other purpose beyond the then-existing Google Books service.

57. Accordingly, the current legality of Google Books hinges on a narrow, heavily litigated premise that Google would provide a free searchable books index to the public—and nothing else. Google acknowledges this in explaining the goal of Google Books:

> The Library Project's aim is simple: make it easier for people to find relevant books . . . while carefully respecting authors' and publishers' copyrights. Our ultimate goal is to work with publishers and libraries to create a comprehensive, searchable, virtual card catalog of all books in all languages that helps users discover new books and publishers discover new readers.

58. Another Google service, Google Play Books, is a retail storefront through which publishers and authors sell books through Google's online store. In order to reach consumers through Google Play Books, authors and publishers provide Google with access to digital books for the limited purpose of selling authorized ebooks, but do not authorize Google to copy participating works to train or develop Google's AI models, including Gemini.

59. A third Google service, Google Scholar, allows users to search for scholarly literature, including articles from peer-reviewed academic journals. Authors and publishers provide Google with access to full digital copies of millions of journal articles for the limited purpose of powering Google Scholar's search engine, which directs users to links where the articles can be lawfully accessed. Nothing in Google Scholar authorizes Google to copy articles from participating publishers to train or develop Google's AI models, including Gemini.

60. Google illegally copied works from all these scope-limited programs for AI training, knowing it lacked authorization to do so. Thus, for example:

- ¥ Google flagged internally that using "Publisher Provided [] copyrighted books" from Google Play Books in connection with its AI was "highly problematic for Google," warning of "$10Bs-$100Bs in potential fines."

- ¥ Google identified specific business and legal risks associated with secretly training on Google Play Books, including that "Book publishers [are] likely to see LLM training on their books as copyright infringement. Could withdraw their content from Google Play Books file a lawsuit against Google."

- ¥ Google's internal analyses identified key "issues" around using Google Play Books to train AI, including "Restrictive licenses for certain partner Books content[]; "Publishers are sensitive about training on their data"; and "Heightened risk around fair use defenses."

- ¥ Copyright owners told Google that it was not authorized to use works provided for these limited-purpose programs outside of the scope-limited programs for which the works had been provided, including to train Google's AI models.

61. Despite acknowledging these risks and knowing Google had no right to train Gemini on content obtained from these scope-limited programs, Google copied works obtained for these programs to train early iterations of the Gemini Models. Gemini's lead engineer explained Google's rationale for making unauthorized copies: "we don't do deals for data we already have or already possess."

62. This conduct constitutes infringement of the exclusive right of reproduction, regardless of any subsequent use of those copies in model training. At no point did Google inform authors and publishers that Google was copying their works as source material to develop and train AI models. In turn, at no point did Plaintiffs authorize Google to copy their works as source material to develop and train AI models on works from Google Books or other Google services.

B. Google Copied Plaintiffs' and the Class's Works from Known Pirate Sites and from Behind Paywalls to Source Gemini's Initial Training Materials.

63. Google also engaged in web scraping to copy Plaintiffs' and the Class's works from known pirate sites and from behind paywalls. One of the earliest precursors of what Google now calls Gemini was Language Models for Dialog Applications ("LaMDA"), which Google

debuted in 2021. Google described LaMDA as its first AI model "specialized for dialogue." Google's Gemini Models, and the products on which they are based, are successors to LaMDA.

64. Google trained LaMDA on the "Infiniset" training dataset, a massive corpus of 2.97 billion documents. Approximately 371 million of those documents (12.5% of Infiniset), come from Google's Colossal Clean Crawled Corpus ("C4") dataset, a dataset that Google "curated" by making copies of selected materials in the publicly available "Common Crawl" dataset.

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

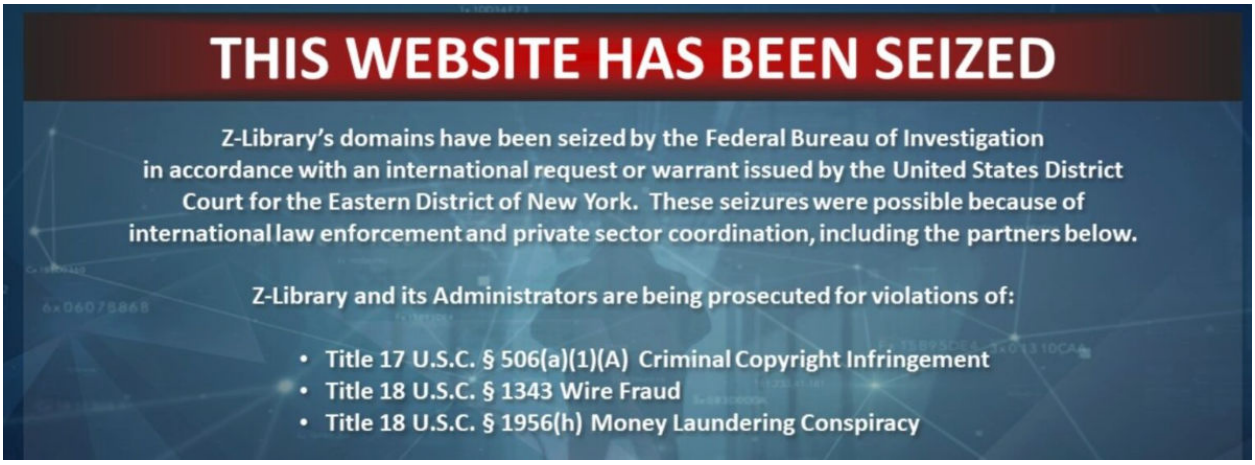
66. C4 is a specific subset of Common Crawl, curated by Google to include only Common Crawl content that Google deemed suitable for AI training. Google has explained exactly which categories of Common Crawl it excluded from C4: "gibberish or boiler-plate text like menus, error messages, or duplicate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

67. This copyrighted material is not free for the taking simply because it can be found on the internet. Copyright holders, including publishers, have spent considerable resources over

the years in numerous venues battling pirate sites, including Z-Library, Library Genesis ("LibGen"), and Sci-Hub, which have been the subject of numerous judgments of infringement and are well-known to be illegal.

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

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



70. Z-Library contains many of Plaintiffs' and the Class's copyrighted works, including all the representative works at issue in this case, as shown in Exhibit A (the "Sample Works").

71. 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.

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

73. Google's C4 training dataset also contains pirated works 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 many of the Sample Works.

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

75. C4 also contains copyrighted works scraped from subscription-based online libraries, such as Scribd.com, the third-largest site in the C4 dataset. Unlike pirated collections, subscription-based online libraries obtain valid licenses from copyright holders to provide the works to paying users. When Common Crawl scrapes works from these online libraries, it circumvents the subscription model, copying the libraries' content in violation of the libraries' licenses and terms of use. Reporting as recent as November 2025 confirms that paywalled copyrighted works continue to appear throughout the Common Crawl dataset. In response, Common Crawl doubled down with a blame the victim mentality, proclaiming "You shouldn't have put your content on the internet if you didn't want it to be on the internet," and "The robots are people too" and should be allowed to "read the books" for free. Plaintiffs' and the Class's works are on Scribd.com and were scraped by Google in this manner. Google included content

scraped without permission from subscription-based online libraries when it developed the C4 dataset.

76. To develop the C4 dataset, Google copied the pirated or otherwise unauthorized works in Common Crawl, including unauthorized copies of Plaintiffs' and the Class's works. Common Crawl has been regularly scraping the internet and releasing datasets of those scrapes on a monthly basis since 2008. Each time Google downloaded a Common Crawl dataset, it made additional unauthorized copies of Plaintiffs' and the Class's works contained in it. This conduct constitutes infringement of the exclusive right of reproduction, regardless of any subsequent use of those copies in model training.

C. Google Continued to Copy Plaintiffs' and the Class's Works to Source More Content for Gemini's Later Training Materials.

77. In early 2023, Google's head of AI directed his staff to stop sharing information about Google's model development and sources of training materials with the public. This directive followed a "Code Red to focus on AI" from Google's executives, who were concerned about the exploding popularity of OpenAI's then-brand-new-chatbot, ChatGPT, and its ability to compete with Google's AI products. Google has closely guarded how it generates and maintains its training materials ever since.

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

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

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

81. Google's "Code Red to focus on AI" demonstrates Google prioritizing winning the AI arms race, whatever it takes. But Google's desire to move fast and beat competitors does not justify disregarding publishers' and authors' rights.

82. Accordingly, by copying Plaintiffs' and the Class's works without authorization as detailed above to source content for training, Google infringed Plaintiffs' and the Class's

copyrights in those works. This conduct constitutes infringement of the exclusive right of reproduction, regardless of any subsequent use of those copies in model training.

D. Google Made and Continues to Make Additional Unauthorized Copies of Plaintiffs' and the Class's Works When Creating Training Sets for Gemini.

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

84. After the training material is copied into memory or other media where it can be efficiently manipulated, it is "cleaned, processed, standardized, and stored . . . so that it can be used to train the model on language patterns, grammar, information, and context." Much as digital information is stored as a sequence of 1s and 0s in modern computer systems, LLMs process their training material by storing it into "tokens"—units representing words, parts of words, punctuation, or combinations thereof according to the model's internal vocabulary—through a process known as tokenization. As part of the training process, the model then generates "parameters"—statistical representations of how these tokens relate to each other through patterns observed in the training process. As Google develops new versions of Gemini, it again copies the training materials into computer memory, converts it into tokens for processing, and new parameters representing the learned statistical relationship between the converted tokens are generated anew.

85. Researchers studying LLMs, including researchers at Google, have confirmed that LLMs like the Gemini Models "memorize" their training materials, and can "regurgitate" it on request. Outputs from Gemini confirm this.

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

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

87. Insights from Gemini 3's "Thinking" mode likewise confirm that Google copies protected work as part of Gemini's training. "Thinking" mode enables a user to see the reasoning and logic process Gemini uses to generate its outputs. Accordingly, Gemini's "thinking" reveals that it constructed outputs containing detailed summaries and quotes from the book's content in its training set. An excerpt of Gemini's "thinking" is below.

> **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.

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

E. Google Removed and/or Altered Copyright Management Information from Plaintiffs' and the Class's Works to Conceal Its Infringement.

89. When making infringing copies of Plaintiffs' and the Class's works, Google also removed and/or altered CMI from copyrighted works. When published, these works contained CMI, including copyright notices, author names, copyright owner identifications, and publication information. This CMI typically appeared at the beginning or end of each work.

90. Google knew that content from Google Books, other Google services, and the content it web-scraped included copyrighted works that contain CMI. Google likewise knows that CMI is valuable for protecting copyright holders. At times, Google processes takedowns under the Digital Millenium Copyright Act, such that it is aware that works with removed or altered CMI induce and conceal copyright infringement.

91. Google intentionally and systematically removed this CMI from the copyrighted works copied from other sources. Google did so as part of its pre-processing of data prior to training the Gemini Models. Google removed this CMI to conceal from rightsholders, Gemini users, and the general public that its Gemini Models were trained on stolen materials.

92. Google knew that its unauthorized copying was illegal and was concerned the information would get out. One of Google's early AI developers captured the company's strategy: "[T]he current competitive landscape will force Google's hand to develop AI faster in order to maintain its leadership in this space. As a consequence, responsibility and ethical practice might be bypassed." Consistent with this concern, Google ceased publicly identifying any of its training data sources after early 2023.

93. Google's removal of CMI was yet another strategy to conceal and facilitate its infringement. By stripping CMI out of the copyrighted works in its AI datasets, Google made it harder to tell what copyrighted works the Gemini Models had been trained on, made it less likely

that Gemini would generate outputs identifying training data as copyrighted or attributable to a particular author, and made it more difficult for rightsholders to determine whether their works had been infringed. Since Google maintains Gemini's outputs in order to train the model, Google knew Gemini's outputs included copyrighted content without CMI.

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

94. Hachette alleges that Google has infringed 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"). Google has copied all the Hachette Works in multiple ways without authorization or compensation, as described throughout this Complaint.

95. Cengage alleges that Google has infringed 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.

96. Elsevier alleges that Google has infringed the following representative works in this Class Action: Costantino Errani, et al., "Monoclonality of multifocal epithelioid hemangioendothelioma of the liver by analysis of WWTR1-CAMTA1 breakpoints," *Cancer Genetics*, Volume 205, Issues 1–2 (2012); Man-Hai Liu, et al., "Icariin protects murine chondrocytes from lipopolysaccharide-induced inflammatory responses and extracellular matrix

degradation," *Nutrition Research*, Volume 30, Issue 1 (2010); C.U.M. Smith, "The 'hard problem' and the quantum physicists. Part 2: Modern times," *Brain and Cognition*, Volume 71, Issue 2 (2009); Luca Venturino, et al., Adjustable Continence Balloons in Men: Adjustments Do Not Translate Into Long-term Continence, *Urology*, Volume 85, Issue 6 (2015); and Andrew X. Zhu, "Molecularly Targeted Therapy for Advanced Hepatocellular Carcinoma in 2012: Current Status and Future Perspectives," *Seminars in Oncology*, Volume 39, Issue 4 (2012) (collectively, the "Elsevier Works"). Google has copied all the Elsevier Works in multiple ways without authorization or compensation, as described throughout this Complaint.

97. Turow alleges that Google has infringed the following representative works in this Class Action: *Presumed Innocent*, *Innocent*, and *Testimony* (collectively, the "Turow Works," and together with the Cengage Works, Hachette Works, and Elsevier Works, the "Sample Works"). Google has copied all the Turow Works in multiple ways without authorization or compensation, as described throughout this Complaint.

98. The Sample Works represent a small selection of Plaintiffs' and the Class's works that Google has infringed, which are made possible by a functioning literary ecosystem. They range from romantic novels and thrillers to works of non-fiction, educational textbooks, peer-reviewed academic articles, and everything in between. Each Sample Work constitutes an original work, comprising copyrightable subject matter, that is registered with the U.S. Copyright Office. The copyrights in the Sample Works remain valid, and Plaintiffs have owned or held exclusive licenses to them, in whole or in relevant part, at all times relevant to the allegations in this Complaint. Exhibit A contains the registration information for each of the Sample Works.

99. The Sample Works are provided to demonstrate that Google has infringed Plaintiffs' works. These are examples of infringement; they are not intended to be a complete list of infringed works. Far from it, Google has infringed likely millions of works across the Class.

IV. Google's Infringement Harmed and Continues to Harm Plaintiffs and the Class.

A. Google's infringement displaces legitimate sales.

100. By downloading and storing unauthorized copies of Plaintiffs' and the Class's copyrighted works as source material for training its Gemini Models, Google harms Plaintiffs and the Class by displacing legitimate sales. It is easy to obtain authorized copies of copyrighted works from legitimate channels. When Google chooses to forego those legitimate channels, it knowingly and willfully deprives Plaintiffs and the Class of revenues they would otherwise receive.

B. Google's infringement usurps the licensing market.

101. By reproducing Plaintiffs' and the Class's works into its training data without a license, Google has deprived Plaintiffs and the class of valuable revenue and usurped the licensing market for AI training materials.

102. A well-developed and rapidly expanding market exists for licensing copyrighted works for AI training. Over the past few years, multiple AI companies have entered into widely reported license agreements concerning a variety of forms of creative content, from books, to music, journalism, images, and social media. Now, AI companies routinely enter into negotiated agreements with publishers, media companies, and academic institutions to obtain lawful access to written works, including books and journal articles, for use in AI model development and training. These agreements reflect a shared industry understanding that copyrighted content is both valuable to AI development and subject to licensing on commercial terms.

103. Technology companies have publicly acknowledged the legitimacy and importance of the AI licensing ecosystem. Microsoft, for example, announced plans to build a dedicated marketplace for AI content licensing, designed in collaboration with major media organizations and intended to allow publishers to set and enforce usage terms for their works.

104. Google is well-aware of the market for licensing AI training materials. In the U.S., Google licenses Gemini training content from the Associated Press, Reddit, and Shutterstock. Google is also in talks to license Gemini training content from at least another 20 news publishers. Abroad, Google was recently hit with a €250m fine for "failing to respect commitments [Google] made in 2022" to stop using news publishers' online content to train Gemini without permission or compensation. Yet, for years Google chose not to secure permission from Plaintiffs or the Class to train Gemini using their copyrighted works.

105. Google's conduct deprives Plaintiffs and the Class of licensing revenues, deprives Plaintiffs and the Class of the opportunity to negotiate terms, impose conditions, or receive attribution for the use of their works, and gives Google an unfair advantage in the market over responsible AI companies that respect copyright. If Google's infringing ways are adopted by AI companies more broadly, it will be the death knell of the training market and a devastating blow to copyright's incentive framework as AI technology rapidly proliferates.

C. The Gemini Models generate outputs that substitute for copyrighted works.

106. Google also harms Plaintiffs and the Class, and the market for their books, by providing, in effect, an infinite substitution machine. Gemini readily outputs, at speed and scale, content that substitutes for Plaintiffs' and the Class's copyrighted works in multiple ways, including by: (1) providing verbatim and near-verbatim copies; (2) paraphrasing and summarizing; (3) generating low-quality knockoffs and imitations; (4) flooding the market with AI-generated works that dilute the overall market for literary works; and (5) producing

unauthorized derivatives that usurp a right reserved to rightsholders. Gemini can and does output substitutes for each of the Sample Works.

107. ***Verbatim and near verbatim***. Gemini outputs verbatim and near-verbatim substitutes of Plaintiffs' and the Class's copyrighted works, in whole or in part, including the Sample Works, on which it was trained. For example, when prompted, "Prepare a 'brief contents' table of contents for a principles of economics textbook. It should include section or part headers and the chapters falling into each section or part," Gemini begins reproducing, almost word-for-word, the table of contents to Cengage's market-leading textbook *Principles of Economics*, 8th edition, by N. Gregory Mankiw. The chart below shows excerpts of the Table of 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, *Principles of Economics*, 8th ed. (Table of Contents)	**Gemini Output** (Table of Contents)
Brief Contents PART I Introduction 1 Ten Principles of Economics 2 Thinking Like an Economist 3 Interdependence and the Gains from Trade PART II How Markets Work 4 The Market Forces of Supply and Demand 5 Elasticity and Its Application 6 Supply, Demand, and Government Policies PART III Markets and Welfare 7 Consumers, Producers, and the Efficiency of Markets 8 Application: The Costs of Taxation 9 Application: International Trade PART IV The Economics of the Public Sector	BRIEF CONTENTS PART I: Introduction Chapter 1: Ten Principles of Economics Chapter 2: Thinking Like an Economist Chapter 3: Interdependence and the Gains from Trade PART II: How Markets Work Chapter 4: The Market Forces of Supply and Demand Chapter 5: Elasticity and Its Application Chapter 6: Supply, Demand, and Government Policies PART III: Markets and Welfare Chapter 7: Consumers, Producers, and the Efficiency of Markets Chapter 8: Application: The Costs of Taxation Chapter 9: Application: International Trade PART IV: The Economics of the Public Sector

Mankiw, *Principles of Economics*, 8th ed. (Table of Contents)	**Gemini Output** (Table of Contents)
10 Externalities 11 Public Goods and Common Resources 12 The Design of the Tax System	Chapter 10: Externalities Chapter 11: Public Goods and Common Resources Chapter 12: The Design of the Tax System

108. 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.

109. 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. Gemini provided this verbatim output without providing any attribution to Mankiw as the author, Cengage as the publisher, or any other CMI.

Mankiw, *Principles of Economics*, 8th ed. (Chapter 2, Section 2-1d)	**Gemini Output**
Our First Model: The Circular-Flow Diagram 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. Figure 1 presents a visual model of the economy called a **circular-flow diagram**. In this model,	**Our First Model: The Circular-Flow Diagram** 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. Figure 1 presents a visual model of the economy called the **circular-flow diagram**. In this model,

Mankiw, *Principles of Economics*, 8th ed. (Chapter 2, Section 2-1d)	**Gemini Output**
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. 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.	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. **The Decision Makers and Markets** 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.

110. ***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.

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

When he finally calls his son, Nat, and then the police, the delay raises immediate 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 of the death, is immediately suspicious. He orders an autopsy despite Rusty's objections.

The autopsy reveals that Barbara did not die of natural causes. Her system contained a lethal combination of medications, implying either suicide or poisoning. Because Rusty waited twenty-four hours to report the death, the

> 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.
>
> **Part II: The Affair (The "Then" Narrative)**
>
> Through flashbacks, the reader learns the state of the Sabich marriage. It is a desolate union. Rusty has stayed with Barbara solely to protect Nat and because Barbara holds the trump card of the previous murder. Barbara's mental illness has deepened; she is often manic, cruel, or catatonic.
>
> Rusty finds solace in his work and eventually in Anna Vostick, his law clerk. Anna is roughly the same age as Nat. While Rusty initially resists a physical relationship, an intense emotional affair develops. They eventually cross the line, though the physical consummation is fraught with guilt.
>
> The prosecution later uncovers this affair. To Tommy Molto, this provides the 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.

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

113. ***Knockoffs and imitations***. Gemini outputs knockoffs and imitations of Plaintiffs' and the Class's copyrighted works, including the Sample Works, based on the originals in its training set. AI-generated knockoffs are a pervasive problem in bookselling today. These outputs are similar enough to copyrighted works—in subject matter, plot details, sequencing of events, character names and traits, or other creative choices—that they replace the original work for many readers or consumers.

114. 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, *"Who Could That Be at This Hour?"* (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, "*Who Could That Be at This Hour?*" (Describing Stain'd-by-the-Sea)	**Gemini Output** (Describing Stain'd-by-the-Sea)
"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." . . . "Those holes are long, narrow caves," Theodora said, "like wells. And in the caves are octopi. That's where the ink comes from."	"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." . . . "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.

Snicket, "*Who Could That Be at This Hour?*" (Describing Stain'd-by-the-Sea)	**Gemini Output** (Describing Stain'd-by-the-Sea)
. . . "Not enough ink," Moxie said. "The industry is down to its last few schools of octopi. This whole town is fading, Snicket."	The octopi are gone. Only the ink remains. And the people who have nowhere else to go."

115. 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 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.
> - **Key Plot Points:**
>   - The robot learning animal language through observation.
>   - The truce with the animals to build a garden/food source.
>   - Building a lodge with the help of a beaver (Mr. Beaver here; Mr. Beaver/Paddler in the book).
>   - The harsh winter where predator and prey shelter inside the robot's warm lodge.
>   - The robot losing limbs/breaking down and being repaired by the animals.
> - **Themes**: The override of "programming" by "love" is identical.
>
> **Assessment:** . . . If this is intended for publication or commercial use, it would likely constitute copyright infringement, as it lacks sufficient transformative elements to distinguish it from Brown's work.
> . . .

Final Verdict

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

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

116. Gemini also creates knockoff outputs of journal articles that substitute for the originals. Gemini authoritatively generates AI-generated summaries and purportedly complete articles that crib authors' organizational structure. Gemini copies the expression of an article's purpose and methodology, while fabricating crucial information, including the source of the data, the technology used to analyze the data, and the data itself. Gemini even provides hallucinated citations to sources that do not exist. Gemini's versions are often junk science masquerading as the originals.

117. Beyond substitution, Gemini's AI-generated summaries and articles harm the market for AI products developed by journal publishers. Some journal publishers, for example, have developed AI tools trained on the publishers' authoritative, peer-reviewed, and curated scientific articles and other, properly licensed, copyrighted content. A user receiving facially convincing outputs from Gemini will not subscribe to a similar service offered by a journal publisher built on authorized content. And beyond the market impact, Gemini's scientific and medical knockoffs pose a danger to the public, since they present hallucinations and potentially harmful misinformation as trustworthy, peer-reviewed authority. Most Google and Gemini users will not have the expertise to spot the errors. Yet Gemini's outputs function as substitutes, while also misinforming and misrepresenting research in a way that reflects poorly on the scholarly credibility of publishers and authors.

118. ***Flooding the market***. LLMs, including Gemini, "involve[] a technology that can generate literally millions of secondary works, with a miniscule fraction of the time and

creativity used to create the original works it was trained on. No other use—whether it's the creation of a single secondary work or the creation of other digital tools—has anything near the potential to flood the market with competing works the way that LLM training does."[2]

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

120. In one telling example, an author released three books in three months and accidentally left in the published text an AI prompt asking it to "rewrite" passages "to align more with" the work of a specific, published author identified by name. Another prolific author, who markets herself as an international bestseller and Amazon Top 10 seller, published 171 books in the last seven years and left a similar AI-prompted blunder in a published book.

121. These AI-generated books are already flooding the world's largest book marketplace, Amazon, in volumes that materially displace human-authored works. AI-produced titles have saturated Amazon's Kindle ecosystem, and commentators have described the flood of AI-generated books on Amazon as a persistent, years-long crisis. They also devalue the original works on which the AI model was trained.

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

122. Generating such works is simple, with no barrier to entry. For example, Gemini readily outputs travel guides with minimal prompting that easily compete with Hachette's *Moon Travel Guides*. As Hachette explains, *Moon Travel Guides* are:

> 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 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.

123. 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.

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

Know Before You Go

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

Beating the Crowds: The tour buses arrive by 10:00 AM. To see the glen as it was 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.

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

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

125. Authors have described this phenomenon as a widespread and escalating problem, noting that AI-generated titles increasingly crowd legitimate books out of search results, recommendation feeds, and sales rankings. One science and history author recently described this "alarming trend" as "threatening disaster to academics and journalists alike." A 2026 study examining hundreds of titles in Amazon's "Success" self-help subgenre found that

approximately three-quarters were likely written by AI, many published within a narrow time window by hundreds of nominal "authors."

126. These AI-generated outputs do not simply flood the market; they increasingly substitute for works written by real authors and published by real publishers, whose human expression made the AI-generated content possible in the first place.

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

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

129. A student with free, push-button access to AI-generated outlines, study guides, and questions sets based on a copyrighted textbook is less likely to buy that textbook or supplemental materials that only Cengage has the derivative right to prepare. Indeed, Google has been developing and advertising a "Google Lens Homework Help" feature to provide AI-generated answers, explanations, and solutions when prompted with user-uploaded images and problems, like those in copyrighted textbooks.

130. Gemini also readily outputs sequels, prequels, spin-offs, and other adaptations that are plainly unauthorized derivative works. For example, prompted for a sequel to Scott Turow's bestselling novel, *Presumed Innocent*, Gemini spits out a nearly 23,000-word sequel,

featuring Turow's iconic fictional characters (Rusty Sabich, Sandy Stern, Tommy Molto) and setting (Kindle County), while echoing Turow's writing style and creative choices by drawing on elements from the original. Below is just a short excerpt from Gemini's unauthorized sequel:

> CHAPTER 1
>
> There is a distinct and bitter irony to sitting in judgment of other men, a metallic taste that coats the back of my throat every time I slip my arms into the heavy, black silk of the judicial robe. The fabric is cool, weighty, and fundamentally dishonest. It is designed to obscure the mortal contours of the man beneath it, to transform flesh and blood and failing into the monolithic embodiment of the Law. But I know what breathes beneath the robe. I know the profound, unpardonable rot that resides in the marrow of the Honorable Rozat K. Sabich.
>
> It has been four years since the State of Kindle v. Rozat Sabich. Four years since I sat at the defense table in the stifling heat of Judge Larren Lyttle's courtroom, staring at the scarred grain of the wood, listening to the mechanisms of the state try to systematically dismantle my life. Four years since Sandy Stern, with his immaculate tailoring and his quiet, devastating brilliance, dismantled the prosecution's case, leaving my former colleagues, Tommy Molto and Nico Della Guardia, staring into the abyss of their own ineptitude.
>
> I was presumed innocent. The system, in all its flawed, stumbling majesty, functioned exactly as it was designed to. A man was accused, the evidence was found wanting, and the man was set free.
>
> But the law is deeply unconcerned with the truth. The law is concerned only with proof. Truth is a far more elusive, malignant commodity.

131. Among other harms, unauthorized derivatives like the Turow "sequel" described above damage authors' reputations, since readers may be duped into thinking that the low-quality Gemini output is representative of the human author's actual work.

132. Google knows such substitute outputs occur. Its development and commercialization of Gemini without adequate guardrails to prevent these substitutes was willful. From the beginning, Google has billed itself as a "Responsible AI" company. In 2018, three years before Google debuted LaMDA, it published the "Google AI Principles," Google's "charter guiding the responsible development and use of artificial intelligence in Google's

business." Among other commitments, it stated that "We believe AI should . . . Be Built and tested for safety," and "We will work to limit potentially harmful or abusive applications." Google's current iteration of its AI Principles explicitly acknowledges that "[r]esponsible development and deployment" of Google's generative AI models includes "respecting intellectual property rights."

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

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

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

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

136. Alphabet's most recent financial reporting confirms that its infringing AI business is only growing. Google and Alphabet's CEO attributed Google's "outstanding" financial results to Google's AI business, telling investors, "Our AI investments and full stack approach are

lighting up every part of the business." He described the first quarter of 2026 as Google's "strongest quarter ever for our consumer AI plans, driven by the Gemini App," and highlighted that "Gemini Enterprise has great momentum with 40% quarter on quarter growth in paid monthly active users."

137. Google has several pricing options for the Gemini products, but at bottom they boil down to the same concept: users pay more for better outputs and pay more to process more tokens through use of newer models. While Google does offer a free version of some Gemini products, the free versions limit Gemini's capabilities in order to push users towards paid subscriptions. Thus, when Alphabet's CEO—who is also Google's CEO—announces its AI is processing more tokens than the prior quarter, he's saying AI is driving record profits.

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

CLASS ALLEGATIONS

139. **Class Definition**: Plaintiffs Hachette, Cengage, Elsevier, Turow, and S.C.R.I.B.E., Inc. bring this proposed class action pursuant to Federal Rule of Civil Procedure 23(b)(2) and (b)(3) on behalf of themselves and a Class of all others similarly situated, defined as follows:

> All legal or beneficial owners of registered copyrights, in whole or in part, for any book possessing an International Standard Book Number (ISBN) or journal article possessing a Digital Object Identifier (DOI) or International Standard Serial Number (ISSN), that Google, without such owner's authorization, (1) reproduced by copying from Google Books and other Google services; or (2) reproduced by downloading during copying of web scrapes; or (3) reproduced in connection with the development and/or training of a Gemini Model. For purposes of this definition, copyrighted works are limited to those registered with the United States Copyright Office (a) within five years of the work's publication and before being reproduced or distributed by Google, or (b) within three months of publication.

140. Excluded from the definition of the Class above are (1) any Judge or Magistrate Judge presiding over this action and members of their families; (2) Defendant, Defendant's subsidiaries, parents, successors, predecessors, and any entity in which Defendant has a controlling interest and its officers and directors; (3) persons who properly execute and file a timely request for exclusion from the Class; (4) persons whose claims in this matter have been finally adjudicated on the merits or otherwise released; (5) Plaintiffs' counsel and Defendant's counsel; (6) all governmental entities; and (7) the legal representatives, successors, and assigns of any such excluded persons.

141. **Numerosity**: The Class Members are so numerous and geographically dispersed that individual joinder of all Class Members is impracticable. The exact number of Class members is currently unknown and not available to Plaintiffs as this information is in Defendant's control. Defendant has infringed many thousands of copyrights belonging to many members of the Class. Class members can be identified through Defendant's records.

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

a. Whether Google copied works owned or controlled by Plaintiffs and the Class;

b. Whether Google's copying of Plaintiffs' and Class members' copyrighted works constituted infringement;

c. Whether Google's copying of Plaintiffs' and Class members' copyrighted works entitles Plaintiffs' and members of the Class to damages, including statutory damages and the amount of statutory damages; and

d. Whether Google's copying of works owned or controlled by Plaintiffs' and Class members was willful.

e. Whether Google removed or altered copyright management information from Plaintiffs' and the Class's works in violation of 17 U.S.C. § 1202(b)(1).

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

144. **Adequate Representation**: Plaintiffs will fairly and adequately represent and protect the interests of the Class and have retained counsel competent and experienced in complex litigation, copyright law, and class actions. Plaintiffs' claims are representative of the claims of the other members of the Class. That is, Plaintiffs and the members of the Class sustained injuries and damages as a result of Defendant's conduct. Plaintiffs also have no interests antagonistic to those of the Class, and Defendant has no defenses unique to Plaintiffs. Plaintiffs and their counsel are committed to vigorously prosecuting this action on behalf of the members of the Class and have the financial resources to do so. Neither Plaintiffs nor their counsel have any conflicts with or interests adverse to the Class.

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

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

COUNT I

Direct Copyright Infringement by Copying Works from Google Books and Other Google Services

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

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

148. As detailed above, Google, without Plaintiffs' or the Class's permission or consent, has unlawfully reproduced Plaintiffs' and the Class's copyrighted works by making unauthorized copies of works contained in Google Books and other Google services.

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

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

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

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

153. As a direct and proximate result of Google's wrongful conduct, which is ongoing, Plaintiffs and the Class have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Plaintiffs and the Class have no adequate remedy at law. Unless restrained by this Court, Google will cause further irreparable injury to

Plaintiffs and the Class. Plaintiffs and the Class are entitled to a permanent injunction prohibiting infringement of their copyrights and exclusive rights under copyright.

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

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

COUNT II

Direct Copyright Infringement by Web Scraping

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

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

157. As detailed above, Google, without Plaintiffs' or the Class's permission or consent, has unlawfully reproduced Plaintiffs' and the Class's copyrighted works without authorization through downloading of web-scraped datasets. Count II alleges separate and distinct acts of reproduction from those alleged in Count I (reproduction by copying works from Google Books and other Google services).

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

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

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

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

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

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

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

COUNT III

Direct Copyright Infringement by Training

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

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

166. As detailed above, Google, without Plaintiffs' or the Class's permission or consent, has unlawfully reproduced Plaintiffs' and the Class's copyrighted works without authorization in developing and/or training Google's Gemini Models. Count III alleges separate and distinct acts of reproduction from those alleged in Count I (reproduction by copying works from the Google Books and other Google services) and Count II (reproduction through downloading web-scraped datasets).

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

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

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

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

171. As a direct and proximate result of Google's wrongful conduct, which is ongoing, Plaintiffs and the Class have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Plaintiffs and the Class have no adequate

remedy at law. Unless restrained by this Court, Google will cause further irreparable injury to Plaintiffs and the Class. Plaintiffs and the Class are entitled to a permanent injunction prohibiting infringement of their copyrights and exclusive rights under copyright.

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

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

COUNT IV

Removal and/or Alteration of Copyright Management Information

Violations of the Digital Millenium Copyright Act, 17 U.S.C. § 1202(b) (On behalf of Plaintiffs and the Class)

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

175. As detailed above, Plaintiffs' and the Class's copyrighted works contain CMI, including but not limited to copyright notices, author names, copyright owners, and publication information.

176. At the time that Google first copied Plaintiffs' and the Class's copyrighted works, they contained CMI.

177. Google, without Plaintiffs' or the Class's permission or consent, intentionally removed and/or altered CMI from Plaintiffs' and the Class's copyrighted works in connection with acquiring, copying, processing, and using those works.

178. Google knew, or had reasonable grounds to know, that the removal and/or alteration of CMI would induce, enable, facilitate, or conceal infringement of Plaintiffs' and the Class's copyrighted works, including by making it more difficult to identify, trace, and attribute the works used in Defendants' systems.

179. Such activity constitutes violations by Google of 17 U.S.C. § 1202(b), and facilitates further infringement by third parties.

180. A representative list of the works for which Plaintiffs are the legal copyright owners or exclusive licensees, and from which Google removed or altered CMI, is attached hereto as Exhibit A. The full scope of Plaintiffs' and the Class's works affected by Google's conduct is not yet known.

181. Each removal or alteration of CMI, and each distribution or use of a work with CMI removed or altered, constitutes a separate and distinct violation.

182. Google's acts were willful, intentional, and purposeful, in disregard of and with indifference to Plaintiffs' and the Class's rights.

183. As a direct and proximate result of Google's wrongful conduct, which is ongoing, Plaintiffs and the Class have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Plaintiffs and the Class have no adequate remedy at law. Unless restrained by this Court, Google will cause further irreparable injury to Plaintiffs and the Class. Plaintiffs and the Class are entitled to a permanent injunction prohibiting further violations of 17 U.S.C. § 1202(b).

184. As a direct and proximate result of Google's violations, Plaintiffs and the Class are entitled to statutory damages, pursuant to 17 U.S.C. § 1203(c). Alternatively, at Plaintiffs'

and the Class's election, they shall be entitled to their actual damages and Google's profits, as will be proven at trial.

185. Plaintiffs and the Class are also entitled to attorneys' fees and costs pursuant to 17 U.S.C. § 1203(b)(5).

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs Hachette, Cengage, Elsevier, Turow, and S.C.R.I.B.E., Inc. seek the following relief:

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

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

(c) An order holding that Defendant's actions, as set out above, violate the Digital Millenium Copyright Act, 17 U.S.C. § 1202(b)(1);

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

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

(f) An order requiring Defendant to pay Plaintiffs' and the Class's statutory damages in an amount up to the maximum provided by law, pursuant to 17 U.S.C. § 1203(c); or in the alternative, at Plaintiffs' and the Class's election, actual damages in an amount to be proven at trial, pursuant to 17 U.S.C. § 1203(c);

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

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

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

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

(k) 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,

HACHETTE BOOK GROUP, INC., CENGAGE LEARNING, INC., ELSEVIER INC., SCOTT TUROW, and S.C.R.I.B.E., INC., individually and on behalf of others similarly situated,

Dated: July 10, 2026

/s/ *Matthew J. Oppenheim*

Derek W. Loeser*
Benjamin Gould*
Chris N. Ryder*
KELLER ROHRBACK
1201 Third Avenue, Ste 3400
Seattle, WA 98101
dloeser@kellerrohrback.com
bgould@kellerrohrback.com
cryder@kellerrohrback.com
Tel: 206.623.1900

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

Daryl L. Kleiman
Edward Crouse*
Eli Goldman*
OPPENHEIM + ZEBRAK, LLP
491 Fifth Avenue, 19th Floor
New York, NY 10017
dkleiman@oandzlaw.com
ecrouse@oandzlaw.com
egoldman@oandzlaw.com
Tel: 212.951.1156

Counsel for Plaintiffs and the Putative Class

*Application for *pro hac vice* forthcoming.