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

ELSEVIER INC., CENGAGE LEARNING, INC., HACHETTE BOOK GROUP, INC., MACMILLAN PUBLISHING GROUP, LLC D/B/A MACMILLAN PUBLISHERS, MCGRAW HILL LLC, SCOTT TUROW, and S.C.R.I.B.E., INC., individually and on behalf of others similarly situated,

Plaintiffs,

v.

META PLATFORMS, INC. and MARK ZUCKERBERG,

Defendants.

Civil Action No. 26-cv-3689

CLASS ACTION COMPLAINT FOR:

(1) Defendants' violations of the Copyright Act, 17 U.S.C. §§ 106(1) and 501 – reproduction by torrenting

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

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

(4) Defendants' violations of the Copyright Act, 17 U.S.C. §§ 106(3) and 501 – distribution by torrenting

(5) Zuckerberg's violations of the Copyright Act, 17 U.S.C. §§ 106(1) & (3) and 501 – contributory infringement

(6) Meta's 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 Elsevier Inc. ("Elsevier"), Cengage Learning, Inc. ("Cengage"), Hachette Book Group, Inc. ("Hachette"), Macmillan Publishing Group, LLC d/b/a Macmillan Publishers ("Macmillan"), McGraw Hill LLC ("McGraw Hill") (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 167, *infra*), against Meta Platforms, Inc. ("Meta") and Mark Zuckerberg ("Zuckerberg") (collectively, "Defendants") for copyright infringement based on Defendants' unauthorized reproduction and distribution of Plaintiffs' copyrighted works through Meta's sourcing of content for, and development and training of, Meta's generative artificial intelligence ("AI") platform called Llama, 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. In their effort to win the AI "arms race" and build a functional generative AI model, Defendants Meta and Zuckerberg followed their well-known motto: "move fast and break things." They first illegally torrented millions of copyrighted books and journal articles from notorious pirate sites and downloaded unauthorized web scrapes of virtually the entire internet. They then copied those stolen fruits many times over to train Meta's multi-billion-dollar generative AI system called Llama. In doing so, Defendants engaged in one of the most massive infringements of copyrighted materials in history.

2. Defendants reproduced and distributed millions of copyrighted works without permission, without providing any compensation to authors or publishers, and *with* full

knowledge that their conduct violated copyright law. Zuckerberg himself personally authorized and actively encouraged the infringement. Meta also stripped CMI from the copyrighted works it stole. It did this to conceal its training sources and facilitate their unauthorized use.

3. Plaintiffs and the Class are authors and publishers who own or control exclusive rights under the Copyright Act to millions of widely known and commercially valuable literary works,[1] conceived of and created by humans. These works span a vast breadth of human creativity and expression, including fiction and nonfiction, children's books, poetry, memoirs, travel writing, educational textbooks, and scholarly articles. These works include CMI such as copyright notices, author names, copyright owners, and publication information.

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.

[1] Under the Copyright Act, "'Literary works' are works, other than audiovisual works, expressed in words, numbers, or other verbal or numerical symbols or indicia, regardless of the nature of the material objects, such as books, periodicals, manuscripts, phonorecords, film, tapes, disks, or cards, in which they are embodied." 17 U.S.C. § 101. As used this Complaint, "literary works" refers to written, literary works and expressly includes fiction and non-fiction books, textbooks, and academic and scholarly journal articles.

6. Publishers exercise their rights by making and disseminating copies of works in their catalogs and authorizing others to do so, including through dynamic licensing markets. This system, including the nature of licensing, has adapted with innovations in technology, permitting publishers to make their works available through an ever-evolving variety of formats, distribution methods, and access models.

7. Defendants willfully sidestepped this licensing regime in developing and training Llama.

8. Llama, short for "Large Language Model Meta AI," is Meta's generative AI system capable of producing long-form written content, including novels, educational treatises, and scientific articles. Large language models ("LLMs") such as Llama function by ingesting vast quantities of text and memorizing statistical relationships within that material. That process enables LLMs to generate text outputs responsive to user prompts.

9. To build Llama, Defendants evaded the longstanding system that both respects copyrights and compensates authors and publishers. Instead, through a series of deliberate choices, Defendants violated Plaintiffs' and the Class's exclusive rights to reproduce and distribute their written works in multiple ways:

- First, Defendants copied Plaintiffs' and the Class's copyrighted books and journal articles without authorization by (1) downloading and distributing them via torrenting, and (2) downloading web scrapes of virtually the entire internet. In doing so, Defendants amassed a massive corpus of stolen source material for their own commercial use.

- Second, to train its AI models, Defendants repeatedly copied (and continue to copy) the stolen works without authorization—first into computer memory, then into formats their AI systems could parse, and then into the training materials used to build each model of Llama.[2]

[2] The AI models at issue in this litigation include all versions, iterations, and relatives of Llama (the "Llama Models"), including Llama 1, Code Llama, Llama 2, Llama 3, Llama 3.1, Llama

- Third and finally, with each new Llama model, Defendants once again copied (and continue to copy) the training materials anew, infringing Plaintiffs' and the Class's copyrighted works again from model to model.

10. The result is an AI system that readily generates, at speed and scale, substitutes for Plaintiffs' and the Class's works on which it was trained. Those substitutes take multiple forms, including verbatim and near-verbatim copies, replacement chapters of academic textbooks, summaries and alternative versions of famous novels and journal articles, inferior knockoffs that copy creative elements of original works, and derivative works exclusively reserved to rights holders. Llama even tailors outputs to mimic the expressive elements and creative choices of specific authors.

11. Users are touting AI's ability to generate books with ease and Llama is flooding the market with AI-generated substitutes. The scale and speed at which Llama can create written works and compete with human writers is unprecedented, and it can only do that because Defendants copied Plaintiffs' and the Class's works to train their LLM.

12. The harm from Defendants' infringement is not limited to competing outputs. It also occurs at the point of ingestion, where Defendants copied copyrighted works as inputs to build a valuable commercial system without consent or compensation. That conduct appropriates the economic value of the works, eliminates a legitimate licensing market, and allows Defendants to free-ride on investments they did not make. This is precisely the type of harm that copyright is designed to prevent.

3.2, Llama 3.3, Llama 4, and Muse Spark. The AI-powered products at issue in this litigation include all versions, iterations, and relatives of products that incorporate, rely on, or otherwise use the Llama Models (the "Llama Products"), including Facebook, Instagram, WhatsApp, Messenger, HorizonAI, and the AI products created and distributed by Meta's AI group, Meta AI, such as Meta's AI glasses. Plaintiffs reserve the right to add additional models and products based on facts learned through additional investigation and discovery.

13. Defendants' infringement displaces legitimate sales of publications by downloading and torrenting copies from unauthorized sources. It usurps the existing and growing AI licensing market by copying Plaintiffs' and the Class's works into Llama's training set without permission or compensation. Further, Llama's wide-ranging and varied outputs dilute the overall market and substitute for the copyrighted works on which Llama trained.

14. All of Defendants' infringement was willful. Authorized copies of Plaintiffs' and the Class's written works are widely available for purchase or license. Defendants understood the value of those written works but simply chose not to pay for them.

15. Indeed, Defendants at one point considered obtaining valid licenses for the works they eventually stole. Meta even approached publishers directly to discuss obtaining licenses. On Zuckerberg's personal instruction, however, Meta abandoned licensing negotiations altogether and stole the works instead. That path was more expedient for Meta, but it deprived publishers and authors of fair compensation and spurned established licensing markets.

16. Defendants have profited—and continue to profit—massively from their infringement. On Meta's Q4 2025 earnings call, CFO Susan Li discussed Meta AI's capacity to drive revenue performance for the company. In 2025, Meta's AI products had already increased app engagement with video translations and media creation tools, improved advertising products with AI assistants for advertisers, and facilitated business messaging with customers through AI agents for businesses. Meta has predicted that by 2035, it will reap $460 billion to $1.4 trillion of revenue from its AI products, including Llama. That increased revenue will be earned off the backs of creators whom Meta has refused to compensate.

17. While AI technology may be new, the legal principles at the center of this case are not. Copyright law applies to AI companies and their leaders, including Defendants, with the

same force as every other company that has complied with these laws for decades. If left unaddressed, Defendants will continue to infringe Plaintiffs' and the Class's rights, cause broad and lasting damage to the publishing 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 Defendants' infringement and preserve the important work of authors and publishers and the literary ecosystem more broadly.

PARTIES

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

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

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

21. Plaintiff Macmillan is a New York limited liability company with its principal place of business at 120 Broadway, New York, New York 10271.

22. Plaintiff McGraw Hill is a Delaware limited liability company with its principal place of business at 8787 Orion Place, Columbus, Ohio 43240.

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

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

25. Defendant Meta is a Delaware corporation with its principal place of business at 1 Meta Way, Menlo Park, CA 94025. Meta is a technology company that develops, distributes, and monetizes the Llama Models, all of which were born out of Defendants' copyright infringement.

Meta also operates AI-related businesses, including the development and deployment of the Llama Models and related AI-powered products.

26. Defendant Zuckerberg is the founder, Chairman, CEO, and controlling shareholder of Meta. His offices are located at 1 Meta Way, Menlo Park, California, 94025.

JURISDICTION AND VENUE

27. 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.

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

29. New York 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. Elsevier, Macmillan, and Hachette have been headquartered in New York for decades. McGraw Hill similarly has maintained offices in New York for over a century. Key industry functions such as editorial acquisitions, contract negotiation, marketing, and rights management are largely conducted in New York, 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.

30. 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 Meta'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.

31. Defendant Meta 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 the Llama products) and services to customers in this State and District. In addition, Defendants' agents reside in this District and may be found in this State and District.

32. Meta possesses real property and has established a continuing presence in this State and District. Meta employs significant employees and resources within New York City. Meta has maintained multiple offices within the District during the relevant period, including at 50 Hudson Yards, New York, NY 10001; 380 W 33rd St, New York, NY 10001; and (from 2018 through 2024) 770 Broadway, New York, NY 10003. Meta has numerous high-profile employees working on AI based in New York City. These include the Research Engineering Lead for North America for Meta's Core Machine Learning and Responsible AI, as well as its Associate General Counsel for AI. In addition, more than 46 legal, engineering, infrastructure, content, and management staff are actively engaged in developing, implementing, and promoting Meta's AI

technology at issue here. As of February 2026, Meta listed over 290 open positions for its New York City offices, including 45 AI-related positions.

33. During the relevant period of Meta's development and training of the Llama Models and Llama Products, the company relied heavily on its New York City offices and employees. Key employees overseeing AI development and training were based in Meta's New York City offices during the relevant period, including Meta's long-time Chief AI Scientist, the Senior Manager of the Llama LLM Research Team, and Business Development of AI Partnerships executives. These employees were supported by teams of research and engineering staff and IP counsel that were also based in New York. Several of these New York-based employees pushed for, implemented, and helped conceal Meta's decisions to torrent from pirate collections and train Llama on stolen works.

34. Meta aggressively promotes and advertises its AI products and services in this State and District, including through its dynamic, interactive website (meta.ai and associated subpages) and mobile applications (Meta AI, Instagram, Facebook, Messenger, and WhatsApp, all of which have Llama integrations). Meta targets customers in this State and District with promotional material tailored to a New York audience, including online and physical advertising of Meta's AI products and services on billboards in Times Square.

35. Meta regularly hosts physical events in this District relating to Llama and other AI models. For example, in 2025, Meta co-sponsored a "Llama 4 Hackathon NYC" event, advertising that "Meta's Llama team will be in attendance, and hackers will receive hands-on support from the Meta team."

36. Since November 2025, Meta has run a "Meta Lab pop up" store in New York City, encouraging customers to explore Meta's AI products and "discover immersive demos,

limited editions and custom details only in store." Meta also hosts other AI and Llama workshops in New York City, including AI research community events on topics such as AI trust and safety.

37. As set forth in further detail in Section V, *infra*, Defendants Meta and Zuckerberg plan for Llama and related AI technology to be so pervasively integrated across Meta's products and services as to become inseparable from them. Already, Llama and other AI features draw more and more users and higher screentime/app engagement, driving up Meta's core pillar of advertising revenue. Thus, Meta's wider business, advertising, and product development teams—for which there are approximately 5,000 employees in Meta's Manhattan offices—reap the benefits of and build on Defendants' infringement of Plaintiffs' and the Class's works.

38. Meta 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.

39. Defendant Zuckerberg is also subject to the specific personal jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a). As set forth below in further detail, Zuckerberg explicitly and knowingly authorized, directed, and was a primary actor in Meta's widespread infringing acts at issue in this action, including Meta's torrenting of millions of copyrighted works from illegal pirate collections, causing substantial harm in this District. Zuckerberg transacted business within New York within the meaning of § 302(a)(1) by directing and exercising control over Meta's New York-directed activities giving rise to Plaintiffs' claims, including the development, training, and distribution of the Llama Models and the acquisition and copying of infringing works as training material.

40. Zuckerberg, in his capacity as Meta's founder, chairman, CEO, and controlling shareholder, had ultimate control and authority over the development and distribution of the

Llama Models. Zuckerberg has been the guiding hand behind, and closely involved in, Meta's development of Llama from the start, exercising his decision-making power to demand and authorize the use of Plaintiffs' works to train Meta's LLM. He was not a passive corporate officer, but a primary actor in the transactions giving rise to this action, with knowledge of and direct participation in the infringing conduct. At Zuckerberg's direction, Meta employees in New York and elsewhere carried out mass theft of copyrighted works belonging to Plaintiffs and the Class, even as they expressed concern regarding the legality and ethics of their actions. When Meta engineers grappled with difficulties in sourcing sufficient training material for Llama, Zuckerberg "demanded a solution" to the problem, which ultimately led Meta to pirate copyrighted works. In April 2023, a New York-based Meta AI employee confirmed that the decision to torrent copyrighted works from a notorious pirate site had been made only "[a]fter a prior escalation to MZ[.]"

41. 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 Defendants' copyright infringement and commercialization of its Llama Models.

COMMON FACTUAL ALLEGATIONS

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

42. Written works are vital to our society and culture. They educate and entertain, spark new questions and new worlds, evoke shared emotions and shared experiences, and give voice to the unheard. The Framers of the Constitution enshrined the copyright protection of written works in Article I, Section 8, Clause 8, and they did so with an explicit purpose: "[t]o promote the Progress of Science and useful Arts."

43. Plaintiffs are publishers and authors proud to fulfill that constitutional objective and to ensure the continued creation, dissemination, and survival of books and journals in our society.

44. Elsevier is a world leading provider of professional information solutions in the science, medical, and health sectors. Elsevier publishes, markets, sells, and licenses academic textbooks, journals, 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).

45. 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. Cengage's many publishing imprints include prominent brands such as Brooks Cole, Cengage, Cengage Learning, Delmar, Gale, Heinle, Milady, South-Western Educational Publishing, and Wadsworth.

46. Hachette is the third-largest U.S. trade publisher, with a history stretching back to 1837. Hachette 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 across a wide range of genres, territories, and formats, 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.

47. Macmillan is a leading publishing company and home to some of the world's most cherished authors and creators. It publishes a wide range of genres and formats for every kind of reader, from adult fiction and nonfiction to many inspired children's books, across eight divisions: Celadon Books; Farrar, Straus and Giroux; Flatiron Books; Henry Holt; Macmillan Audio; Macmillan Children's Publishing Group; The St. Martin's Publishing Group; and Tor Publishing Group. Macmillan's authors have earned the Nobel Prize, the Man Booker Prize, the Pulitzer Prize, the National Book Award, the Newbery Medal, and the Caldecott Medal.

48. McGraw Hill is a leading global provider of education solutions for PreK-12, higher education, and professional learning. Founded over 130 years ago, it provides print and digital content, solutions, software, and services from PreK-12 to higher education and professional markets. It has taken a leading role in harnessing new technologies to advance its print and digital publications in service of students and educators.

49. 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*.

50. Authors and publishers have a symbiotic relationship. Whether writing math textbooks, biology research articles, investigative non-fiction, prescriptive guides for health and wellness, travel books, or gripping novels, authors dedicate years, decades, or their entire lives to their craft. Indeed, authors frequently consider writing to be an intrinsic calling and passion. Without human authors and editors, there would be no books, no educational texts, no scholarly articles, and nothing to publish.

51. At the same time, a thriving, independent publishing industry embodies the Constitution's objective to protect the useful arts by securing authors' exclusive rights to their works. Publishing is critical to society's cultural, intellectual, and political systems, helping society tell its stories, educate its citizens, develop its workforce, and inform its voters.

52. Together, publishers and authors are committed to authorship, innovation, and entrepreneurship. Through significant investments of time, money, and creative energy, publishers develop, edit, market, distribute, license, and sell written works of the highest quality. In tandem with authors, publishers have been on the forefront of technological advancements from the printing press to the internet, and now artificial intelligence. Publishers play a crucial role in supporting and shaping authors' works. By acting as curators of what is published on a wide scale, they help ensure that content reaching the public meet standards of quality, accuracy, and coherence. Through editorial review, peer evaluation, or maintenance of the scientific record, publishers in different sectors work with authors to filter out and correct misinformation and improve creative expressions. This helps elevate works that contribute meaningfully to cultural,

scientific, and intellectual discourse. This process strengthens public trust in written materials and supports informed decision-making in society.

53. Published literary and scholarly works would not be possible without authors' and publishers' creative, intellectual, and financial investments. Often, it takes an author years to finish writing a single book or journal article, with publishers supporting them by bearing the upfront costs needed to get a book or journal to market and into readers' hands. Publishers serve as representatives and advocates for the interests of their authors. They invest in the copyrighted works by not only lending their own creative thoughts, but expending the necessary resources, financial and otherwise, to support authors and creative works in reliance on copyright law's exclusive rights that make recouping those costs possible.

54. Publishers can afford to invest in authors' works because they rely on copyright law and the 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 book—and, importantly, includes the right to refrain from licensing certain uses.

55. For educational publishers like Plaintiffs Elsevier, Cengage, and McGraw Hill, the textbook market extends beyond the underlying textbook to a variety of associated materials that complement the textbook in the learning process. This includes study aids, solutions manuals, homework assignments, assessment questions, adaptive software, and more. Publishers often include these materials in their online learning platforms. Consumers—including students—often purchase not only a textbook but also the supplemental learning materials (or access to the materials) that accompany it.

56. For trade publishers, like Plaintiffs Hachette and Macmillan, numerous factors drive the book market, including, for example, popularity and reviews, school uses, marketing, topical relevance, and the pricing of the work and competing works. Many categories of trade books are organized by genre, and individual genres—such as romance, science fiction, mystery, and fantasy—often have dedicated and repeat readerships. These genres function, in part, as a mechanism for readers to identify and discover new works with similar characteristics to those they have previously enjoyed. Readers frequently purchase and consume multiple works within a given genre for that reason. As a result, there is a well-established market for works that share thematic and content similarities with existing titles, as readers actively seek out comparable content.

57. The market for trade books includes but is not limited to print books, ebooks, audiobooks, and library uses. In addition, the rights associated with trade books extend beyond the initial publication of a given work to include derivative and related uses, such as abridged versions, sequels, adaptations, and other works that incorporate the same characters, settings, or story elements.

58. 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.

59. 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. Defendants Needed Plaintiffs' and the Class's High-Quality Written Works to Win the AI Arms Race.

60. Meta is among the world's largest public, for-profit technology companies, operating a suite of digital platforms, including Facebook, Instagram, and WhatsApp. The company has extensive data collection practices and automated systems designed to maximize user engagement and monetization at a global scale. Regulators, courts, and commentators have repeatedly scrutinized the company for placing commercial objectives ahead of broader social, cultural, and informational considerations, including the downstream effects of its products on privacy, competition, and public discourse.

61. In recent years, Meta has invested heavily in generative AI, as it has competed with other AI developers to develop and deploy new AI tools. The current AI "arms race" is 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 has driven Meta to pursue ever-larger training sets and models, turning scaling into Meta's dominant business strategy.

62. Defendants founded Meta AI (its current name) in 2013 when Dr. Yann LeCun joined Meta's (then Facebook) New York City office as its Chief AI Scientist. Meta spent years

researching and developing AI technology but began its focus on Llama in earnest after the widespread release of OpenAI's ChatGPT at the end of 2022. From then and through now, Meta has invested billions of dollars in AI and released multiple versions of Llama and its other AI products.

63. Defendant Zuckerberg is the guiding force behind Meta AI. He handpicked Dr. LeCun to lead Meta's AI efforts, and put together the team that develops Llama, often reaching out personally to recruits. According to insiders, Zuckerberg's "personal involvement in promoting Llama . . . is the key reason Meta has been able to move with such speed and focus." Zuckerberg is also personally involved in Meta's operations. For example, as Meta's President recently reported, Zuckerberg "has actually moved his desk and is seated in the AI lab with [Meta AI staff], and he's coding all day long[.]"

64. "Generative AI" refers to AI systems designed to generate content (such as text, images, audio, or video) based on patterns derived from training materials. 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, producing textual outputs.

65. According to Meta, LLMs have a seemingly straightforward training methodology. An LLM is first pretrained on an extensive amount of content, and then "align[ed] with human preferences" by manual human feedback on training outputs to fine-tune and calibrate the model. Generative AI models are not, however, creative. All they can do is predict text that appears coherent and responsive to user queries, according to statistical models developed from patterns observed in their internal training materials.

66. To train Llama, Defendants sought and compiled a massive body of high-quality training materials because Meta believed that it could only create the "best possible performance . . . by training on more tokens than what is typically used" by its competitors. Tokens are the basic units of text processed by an LLM—typically representing words or characters—into which written content is broken for purposes of training. Training on tokens allows an LLM to record statistical relationships, patterns, and context across vast quantities of written material.

67. When it first released Llama in 2023, Meta acknowledged that "We train our models on trillions of tokens, and show that it is possible to train state-of-the-art models using publicly available datasets exclusively, without resorting to proprietary and inaccessible datasets." What Meta did not disclose was that Meta's "publicly available datasets" included millions of works torrented from notorious pirate collections and unauthorized web scrapes of virtually the entire internet.

68. Books, educational works, and journals are a valuable source of high-quality training materials for LLMs and other generative AI models. As non-profit AI research group EleutherAI noted, "books are invaluable for long-range context modeling research and coherent storytelling." Researchers have confirmed that readers prefer outputs from AI models trained on copyrighted works over outputs from AI models trained on more general text datasets. Another group of researchers documented how 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."

69. Defendants particularly valued and targeted literary works as sources for training Llama. One Meta employee explained that the "best resources we can think of [for training Llama] are definitely books." Another confirmed that it was "really important for us to get books

ASAP" to train Llama. Similarly, a third employee emphasized that "continuing pretraining will be effective only if we have relevant self-supervised data, like books, research papers or code-repo," while a Llama developer described how content from scientific and medical journal publishers is "something we've heard from the Llama team [] that is needed." Repeatedly, Meta AI researchers and engineers reiterated the benefits of using books and journals as training data and the need for Meta to acquire more works to train the Llama Models. Like books, journal articles are valuable for AI training: they are written in formal expressive language, have undergone peer review, present structured and often complex arguments, and reflect the quality and cadence of professionally authored text.

70. Defendants did not create the high-quality texts used in Llama's training datasets. Nor do Defendants own or control the copyrights to those works. Instead, Meta—at Zuckerberg's direction—copied millions of books, journal articles, and other written works without authorization, including those owned or controlled by Plaintiffs and the Class, and then made additional copies of those works to train Llama.

III. Defendants Infringed Plaintiffs' and the Class's Copyrights to Source Content for and to Train Llama.

A. Defendants Illegally Copied Plaintiffs' and the Class's Works from Known Pirate Sites and from Behind Paywalls to Source Content for Llama's Initial Training Materials.

71. Defendants sourced Llama's training data from a variety of sources, all of which included massive amounts of copyrighted material. Defendants began by copying material from the "Common Crawl" dataset. 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.

72. When Defendants copied Common Crawl, they knowingly made unauthorized copies of works contained in that dataset, including Plaintiffs' and the Class's copyrighted works. This conduct constitutes infringement of the exclusive right of reproduction, regardless of any subsequent use of those copies in model training.

73. Defendants developed and publicly promoted their own curated set of Common Crawl they named "CCNet." The purpose of Defendants' CCNet is to filter for high-quality training data. Meta researchers described Common Crawl as a "rich resource for monolingual data that comprises a large variety of domains, yet poses challenges due to the large quantity of noisy text." "Noisy text" refers to low-quality, irrelevant, or unreliable content, including spam, corrupted or improperly formatted text, boilerplate website content, and other extraneous data that can interfere with an AI model's ability to ascertain meaningful linguistic patterns. To help eliminate noisy text, Meta curates, de-duplicates, and filters out text from Common Crawl to prepare CCNet. This step improves the quality of the text for training by "removing web copies" and "boilerplate such as navigation menus, cookie warnings and contact information." This also demonstrates that Meta knows what content is in Common Crawl and CCNet. Tellingly, Meta's curation does not involve removing copyrighted content. In curating CCNet, Defendants necessarily reproduced and stored copies of the underlying texts, including copyrighted works, as part of their filtering, deduplication, and processing pipeline. Each such step involved making and retaining unauthorized reproductions of protected content. Notably, CCNet accounted for the largest part of Meta's initial training data for Llama.

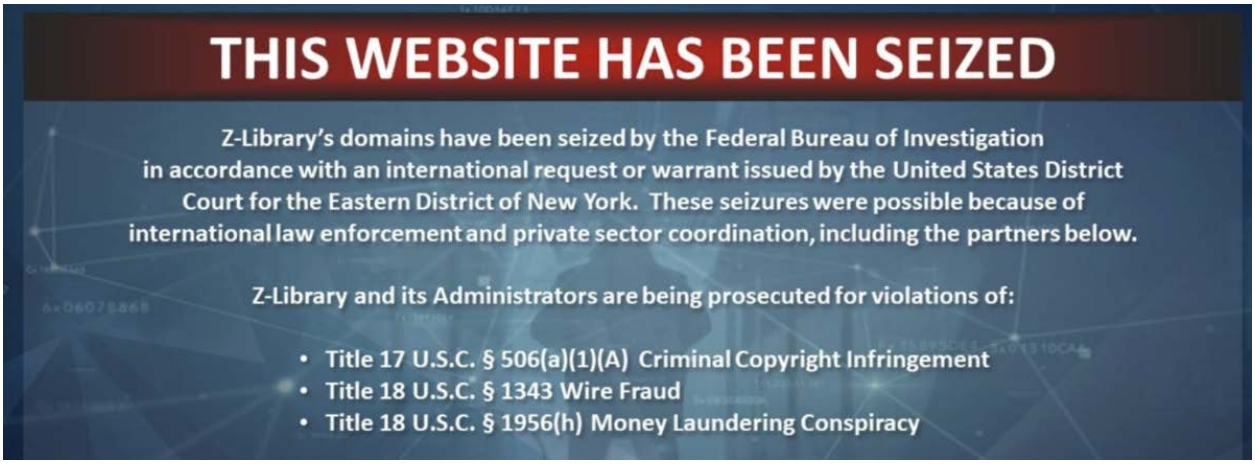
74. Along with CCNet, Defendants also copied another specific subset of Common Crawl to train their Llama Models: the Colossal Clean Crawled Corpus ("C4") dataset. C4 was developed by Google and curated to include only Common Crawl content deemed suitable for AI

training. As with CCNet, certain limited categories of Common Crawl are excluded from C4: "gibberish or boiler-plate text like menus, error messages, or duplicate text"; content 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.'" The C4 training dataset, like the larger Common Crawl set from which it is curated, includes vast categories of copyrighted works, pirated works, and works taken from behind paywalls. Copyrighted content was not curated out of C4. By copying C4, Defendants again engaged in the unlicensed reproduction of copyrighted works—without regard for any downstream uses of C4.

75. Given the manner in which these datasets were formed and curated, it is no surprise that Common Crawl, CCNet, and C4 contain millions of unauthorized copies of copyrighted works. For instance, the copyright symbol (©) appears more than 200 million times in the C4 dataset. Meta's Senior Manager of the Llama LLM Research Team admitted internally, "to be honest, sometimes we do move the copyright text [such as a copyright notice] as part of common cleaning because it is very repeated across datasets." Defendants knew they were reproducing copyrighted material. They proceeded to copy and process those works without authorization.

76. 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 texts "for free in violation of U.S. law." Z-Library has no right or license to distribute those copyrighted works. 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:



77. Z-Library contains a vast number of Plaintiffs' and the Class's copyrighted works, including representative works at issue in this case. Beyond Z-Library, the C4 training dataset contains at least 27 other sites identified by the U.S. government as markets for piracy and counterfeits.

78. Among others, the C4 training dataset contains pirated publications from OceanofPDF, another notorious digital ebook piracy site. OceanofPDF contains many of Plaintiffs' and the Class's copyrighted works.

79. The 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. Defendants' use of this material again involved creating unlicensed copies of copyrighted works.

80. 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 publications from these libraries, it

circumvents the subscription gating, 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. When asked about the organization's practice of stealing paywalled content despite publishers' objections, the executive director of Common Crawl doubled down with a blame-the-victim mentality; he proclaimed that, "[y]ou shouldn't have put your content on the internet if you didn't want it to be on the internet," and further opined that because "robots are people too," they should be allowed to "read the books" for free. Plaintiffs' and the Class's works are on Scribd.com and were scraped to develop C4, which Defendants Meta and Zuckerberg in turn copied further to train Llama without permission.

81. Common Crawl has been regularly scraping the internet and releasing datasets of those scrapes on a monthly basis since 2008. Meta downloads and processes each snapshot independently. Each time Meta downloaded the Common Crawl and C4 datasets, it made additional unauthorized copies of Plaintiffs' and the Class's works contained in it. Each download and processing cycle thus constitutes a new act of infringement, as Defendants repeatedly created and stored additional copies of copyrighted works without authorization or compensation. For Meta's initial Llama model, CCNet constituted 67% of the training materials (3.3 terabytes), and C4 accounted for 15% (783 gigabytes).

B. Defendants Copied Plaintiffs' and the Class's Works to Source Content for Llama's Initial Training Materials by Torrenting the Works from Notorious Pirate Sites.

82. Not content with using the copyrighted works scraped from Common Crawl, Defendants looked for additional sources of copyrighted texts to train their Llama Models. Initially, Defendants considered licensing literary works from major publishers. Instead, they

chose a different route: torrenting Plaintiffs' and the Class's works from notorious pirate sites, including LibGen, Anna's Archive, Sci-Hub, Sci-Mag, and others.

83. While these pirate troves are sometimes referred to a "shadow libraries," calling them libraries of any kind is seriously misleading. It is more accurate to refer to them as notorious pirate sites. These sites are illicit online repositories that aggregate and distribute large quantities of copyrighted works without authorization from rightsholders. They are based outside the United States, operate outside lawful publishing and licensing frameworks and make literary works and other protected materials available for free download, in direct violation of copyright law. Courts in the United States, as well as numerous courts around the world, have frequently found such platforms to be engaged in willful copyright infringement and have entered injunctions directing that they cease their unlawful activities. Nevertheless, such sites have repeatedly evaded enforcement by shifting domains, using mirror sites, and relying on decentralized distribution methods. These operations are frequently associated with organized efforts to evade legal accountability and to profit from the unauthorized dissemination of protected works, reflecting a pattern of deliberate and ongoing disregard for copyright law and judicial authority. By sourcing training data from these illicit repositories, Defendants necessarily reproduced the same infringing copies—creating additional unauthorized duplicates of protected works.

84. Content from notorious pirate sites is often accessed via torrenting. Torrenting refers to the download and distribution of files using a "peer-to-peer" ("P2P") file-sharing system. Torrenting is the dominant method of unauthorized downloading and distribution of copyrighted works over the internet. P2P refers to a decentralized network of users whereby each internet-connected participant (*i.e.*, a "peer" or a "node") can act as both a supplier and consumer

of content files. BitTorrent is a type of P2P file-sharing protocol that distributes large files by breaking them into small pieces and sharing them across a decentralized network of users (a "swarm").

85. BitTorrent's design ensures that users begin disseminating content the instant they start downloading it, and BitTorrent employs a mechanism called "choking" to prevent users from downloading content without contributing to the network. This means that, at any given time, each user connected to the internet can be both downloading and uploading different pieces of a file from, and to, multiple other users. Once a user has downloaded all the file pieces, the file is automatically reassembled into its complete form. That reassembled file is a perfect digital reproduction of the copyrighted work.

86. The online piracy committed via BitTorrent is stunning in nature, speed, and scope. Utilizing a BitTorrent client, internet users can locate, access, and download copyrighted content from others in the blink of an eye. They download copyrighted works from other network users, usually total strangers, and end up with complete digital copies—without paying a cent to copyright owners or creators.

87. Federal courts have long recognized the "detrimental effect of file-sharing" through P2P networks and the "staggering" level of copyright infringement on the internet. Courts have repeatedly found that such P2P services facilitate the rapid and widespread unauthorized distribution of copyrighted works on a massive scale, often with little ability for rightsholders to control or mitigate the harm. In addressing such conduct, courts have described torrenting as enabling users not only to download infringing content but also to redistribute it to countless others, compounding the infringement with each additional user. This includes both the

unauthorized reproduction of works and their further dissemination to others, multiplying the number of infringing copies.

88. In 2022, Defendants first torrented books and journal articles for training Llama from LibGen. LibGen is well-known to be illegal and has been the subject of numerous judgments. More than just books, LibGen also includes journal articles from another pirate collection that has been the subject of multiple enforcement actions, Sci-Hub. Defendants chose to torrent works to get what they wanted quickly, without paying licensing fees. Meta's employees noted that "everyone is using lib-gen (startups, but also google, openAI)" and that "[Libgen] would probably contribute the most towards improving our Agents capabilities and their creativity."

89. In 2022, Meta torrented over two million copyrighted publications from LibGen. Initially, Meta claimed that it would use LibGen "just for pure exploration to see if there is value" for training, and if so, "then [it would] setup proper licensing agreement[s]." The act of downloading these works itself constituted infringement, as it required making unauthorized copies of each work.

90. Meta concealed its LibGen torrenting activities, and, due to "legal limitations," did not use the works it pirated from LibGen in the fall of 2022 to train any LLM. However, as explained below, this was short-lived, and by spring 2023, Meta again torrented copyrighted works from LibGen and then, at Zuckerberg's direction, copied the pirated works many times over to train Llama.

91. In the meantime, Meta pivoted to another pirate collection: Books3, a set of nearly 200,000 copyrighted books compiled from the Bibliotik torrent tracker and distributed

without authorization. Without a license or authorization, Meta torrented Books3 and then copied it further to train Llama.

92. Meta publicly disclosed its use of Books3 with its February 2023 release of Llama 1. For later models, Meta attempted to conceal its continued training on Books3. Mere months later, Meta feared that the public would discover not only "the piracy," but also that Meta was aware of it. As one employee noted, "[Meta] knowing and being accomplices . . . that's the issue."

93. After the release of Llama 1, Meta briefly considered licensing deals with major publishers. Meta understood the value of such licenses to publishers and discussed increasing the company's "dataset licensing" budget from $17 to $200 million between January to April 2023.

94. But then in early April 2023, Meta abruptly stopped its licensing strategy. The question of whether to license or pirate moving forward was "escalated" to Zuckerberg. After this escalation to Zuckerberg, Meta's business development team received verbal instructions to stop licensing efforts. One Meta employee presciently described the rationale: "if we license once [sic] single book, we won't be able to lean into the fair use strategy." Defendants thus made a deliberate decision to avoid paying for reproduction rights and instead continue making unauthorized copies.

95. By then, Defendants already had realized that LibGen included most of the copyrighted works they had considered licensing from publishers. One Meta employee testified that he used a copy of LibGen to compare "results in LibGen versus entries contained within a book publisher's catalog." From this, Meta determined not to license works given they were already in LibGen. Armed with this knowledge, Zuckerberg directed Meta to obtain pirate works from LibGen to use in training Llama in spring 2023.

96. Beginning in spring 2023, Defendants attempted to download directly (rather than torrent) the LibGen pirate collection from LibGen's website, but that approach "d[idn't] seem to be fast." Meta's employees again turned to torrenting, as they had in 2022, but this time it was full steam ahead. Meta employees were pleased by their efforts torrenting LibGen, explaining internally that "LibGen is the most valuable dataset that we have so far."

97. Unsatisfied and undeterred, Meta continued to torrent from notorious pirate sites through at least 2024, months after book authors first brought suit against Meta for infringing their copyrights. Meta downloaded Anna's Archive, a compilation of notorious pirate sites including LibGen, Z-Library, and others. Each such download involved unauthorized reproductions of copyrighted works.

98. Anna's Archive bills itself as "the world's largest shadow library," and has written that it "deliberately violate[s] the copyright law in most countries." Meta understood that Anna's Archive was "essentially a bigger libgen" and "a pretty shady website :-P" that "won't be popular with the lawyers." Meta acquired over 81 terabytes ("TB") of data via Anna's Archive, including vast unauthorized reproductions of copyrighted works.

99. Defendants did not disable BitTorrent's default settings for distribution when torrenting from any of the pirate sites used to source training materials for Llama. Between April and July of 2024, Meta's logs showed that, through torrenting, it downloaded 134.6 TB of data from the internet, while also uploading (i.e., distributing) 40.42 TB of (mainly copyrighted content) to the internet for more users to torrent illegally. To put that in perspective, 40 TB is roughly equivalent to 5 million books of 650-pages each, or roughly twice the volume of text in the U.S. Library of Congress.

100. Meta knew from the start that it was infringing. Meta's Program Manager for Responsible AI stated in October 2022 that "LibGen and SciHub . . . are illegal pirated websites" and that "using pirated material should be beyond our ethical threshold." In April of 2023, one employee explained that "it would not be trivial to download libgen if everything is in torrents," while another was "not sure we can use meta's IPs to load through torrents pirate content . . . I think torrenting from a corporate laptop doesn't feel right." A third understood that "using torrents would entail 'seeding' the files – i.e. sharing the content outside," and that "could be legally not OK."

101. On December 13, 2023, Meta employees internally circulated a memo concerning the legal risks of using LibGen. The memo includes discussion of risk based on copyright concerns and describes Libgen as "a dataset we know to be pirated." The memo further states, "we would not disclose use of Libgen datasets used to train." On July 29, 2024, a Meta employee added to the IP Policy Analysis, stating "[a]s IP Policy has previously noted with respect to GenAI training on content from sites such as LibGen that contain pirated content, policymakers will take a negative view of such sites and their use (irrespective of any legal concerns)[.]"

102. Ultimately, however, those concerns went unheeded. Zuckerberg and other Meta executives authorized and directed the torrenting of over 267 TB of pirated material—equivalent to hundreds of millions of publications and many times the size of the entire print collection of the Library of Congress.

103. Defendants adopted a familiar piracy tactic—hiding their illegal activity by masking Meta's IP addresses to "avoid[] risk of tracing back the seeder/downloader . . . from FB servers." Such steps demonstrate their knowing, willful infringement.

104. Despite knowing that torrenting copyrighted works was illegal, Meta proceeded to do so anyway. It did so for a straightforward reason: books, educational works, and journal articles were extraordinarily valuable—indeed, effectively indispensable—to the training of its LLMs.

105. Meta deliberately targeted books and journal articles because they possess characteristics uniquely useful to LLM development, including length, narrative coherence, structural consistency, and professionally edited expression. Unlike fragmented and low-quality internet text, books train models on how to generate outputs that sustain complex arguments, develop characters and themes over time, organize material across chapters, and generate long-form prose that mirrors the quality and cadence of human-authored works. Journal articles are uniquely useful to LLM development for many of the same reasons: they form a highly curated, professional, trusted, and authoritative system of the expression of scientific research, built by the collaboration of leading scholars and publishers over centuries.

106. Meta's copying of these literary works was neither incidental nor accidental. It was a calculated decision driven by the need to appropriate the expressive value of copyrighted written works to produce an AI system capable of generating outputs that directly compete with literary works written by human authors. That decision depended on making unauthorized reproductions of those works at scale.

107. Though Meta prized literary works for its own lucrative goals, it ultimately saw human-authored expression as raw material to be processed, consumed, and discarded as needed to serve Meta's technical and commercial objectives.

C. Defendants Made and Continue to Make Additional Unauthorized Copies of Plaintiffs' and the Class's Books and Articles When Creating Training Sets for Llama.

108. Separate and apart from their initial unauthorized copying of source content, Defendants again reproduced this massive corpus of text from long-term storage into memory where it could be processed to train Llama's AI model. During the training process, Defendants made many additional copies of material in the datasets. As the U.S. Patent and Trademark Office has explained, training an LLM "almost by definition involve[s] the reproduction of entire works or substantial portions thereof."

109. This is because preparing copyrighted text for LLM training requires making new copies of the same work in different locations, formats, and configurations so it can be processed. 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." When Defendants saved copyrighted files into training datasets, transferred them into working storage, or loaded them into memory, they made additional copies. When they modified those files—such as by stripping formatting, removing metadata, or standardizing structure—they generated new versions that retained the protected content of the copyrighted works. Each such step constituted a separate, unauthorized reproduction.

110. For example, when Defendants took a file containing a copyrighted book or article and saved that file into a new training-set directory, they made another unauthorized reproduction. When Defendants copied that same file into another location so it could be processed, they made still more unauthorized copies. And when Defendants saved another version of the same file with formatting, metadata, page images, headers, footers, or other

features stripped away, they made yet another new copy of the copyrighted work—still retaining its protected expressive content.

111. 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"—the basic units of text processed by an LLM (as discussed in Section II, *supra*). Tokenization—the process of converting text into tokens—also depends on repeated copying. Segmenting works into passages or "chunks" requires reproducing portions of the original text; tokenization makes another copy in a machine-readable form; and training involves repeatedly copying that tokenized text into and through system memory as the model updates its parameters. These copies are fixed in a medium from which the text can be perceived or processed and therefore constitute reproductions under the Copyright Act. Accordingly, infringement occurred not just when Defendants first obtained Plaintiffs' works, but throughout the entire pipeline—each time Defendants copied, reformatted, segmented, tokenized, or processed those works without authorization.

112. Researchers studying LLMs have confirmed that LLMs like the Llama Models "memorize" their training materials and can "regurgitate" them on request. Outputs from Llama confirm this.

113. For example, after providing a detailed summary of Scott Turow's *Presumed Innocent*, Llama confirmed that the information came from its internal training material, as it was "trained on a vast corpus of text data, which includes a wide range of books, articles, and other written works":

> In the case of Presumed Innocent, *I have been trained on a digital version of the book*, which allows me to access and analyze its content. This *training data includes the book's text*, structure, and other relevant information.

114. By reproducing Plaintiffs' and the Class's works without authorization during the training process, Defendants infringe publishers' copyrights in those works.

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

115. When making infringing copies of Plaintiffs' and the Class's works, Meta also removed and/or altered CMI from said 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.

116. Meta intentionally and systematically removed this CMI from the copyrighted works that it copied from pirated databases. For example, Meta developed and deployed scripts designed to delete CMI from the beginning and end of works torrented from the LibGen pirate website. Meta similarly deleted CMI from every work contained in its Books3 pirated dataset.

117. Meta removed this CMI to conceal from rightsholders, Llama users, and the general public that its Llama Models were trained on stolen materials. Meta employees knew what the company was doing was illegal and were concerned that this information would get out. For example, one employee warned that "[i]f there is media coverage suggesting we have used a dataset we know to be pirated, such as LibGen, this may undermine our negotiating position with regulators." Consistent with this concern, Meta ceased publicly identifying its training data sources after releasing Llama 1. Meta did not publicly disclose the datasets used to train Llama 2 or Llama 3, a decision made by its legal department. In particular, Meta's employees agreed that "in no case would we disclose publicly that we had trained on libgen." Meta further programmed Llama to deny, when prompted by users, that it had been trained on pirated or copyrighted data.

118. Meta's removal of CMI was yet another strategy to conceal and facilitate its infringement. By stripping CMI out of the copyrighted works in its training datasets, Meta made

it harder to tell what works Meta had been trained on. By deleting CMI, Meta made it less likely that Llama would generate outputs identifying training data as copyrighted or attributable to a particular author. And by deleting CMI, Meta made it more difficult for rightsholders to determine whether their works had been infringed.

119. Meta's conduct was not a neutral or routine data-cleaning process. If it were, Meta would have removed CMI uniformly across all of the training data that it used. It did not do so. While Meta stripped CMI from copyrighted works it pirated from LibGen and Books3, it did not remove CMI from public-domain works sourced from Project Gutenberg. It left the CMI on these works alone, despite training Llama on both datasets at the same time. The decision to delete CMI only from copyrighted works that it had stolen underscores that Meta's actions were deliberate and meant to conceal its wrongdoing.

E. Defendants Infringed Registered Works Owned or Controlled by Plaintiffs and the Class, Including the Sample Works.

120. Elsevier alleges that Defendants have 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"). Defendants have copied all the Elsevier Works in multiple ways without authorization or compensation, as described throughout this Complaint.

121. Cengage alleges that Defendants have infringed the following representative works in this Class Action: E. Bruce Goldstein, *Cognitive Psychology*, 5th edition; William M. Pride and O.C. Ferrell, *Marketing 2018*, 19th 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"). Defendants have copied all the Cengage Works in multiple ways without authorization or compensation, as described throughout this Complaint.

122. Hachette alleges that Defendants have 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"). Defendants have copied all the Hachette Works in multiple ways without authorization or compensation, as described throughout this Complaint.

123. Macmillan alleges that Defendants have infringed the following representative works in this Class Action: Mary Kay Andrews, *Spring Fever*; Sylvia Day, *One With You*; V.E. Schwab, *A Darker Shade of Magic*; Douglas Preston, *Impact*; and, Mary E. Pearson, *The Adoration of Jenna Fox* (collectively, the "Macmillan Works"). Defendants have copied all the Macmillan Works in multiple ways without authorization or compensation, as described throughout this Complaint.

124. McGraw Hill alleges that Defendants have infringed the following representative works in this Class Action: Denise Anderson et al., *Nester's Microbiology: A Human Perspective*,

9th edition; Zvi Brodie, et al., *Essentials of Investments*, 9th edition; Thomas S. Bateman, et al., *Management: Leading & Collaborating in a Competitive World*, 13th edition; Ray H. Garrison et al., *Managerial Accounting*, 15th edition; and, Stephen E. Lucas, *The Art of Public Speaking*, 12th edition (collectively, the "McGraw Hill Works"). Defendants have copied all the McGraw Hill Works in multiple ways without authorization or compensation, as described throughout this Complaint.

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

126. The Sample Works represent a small selection of Plaintiffs' and the Class's works that Defendants have 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 controlled 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.

127. The Sample Works are provided to demonstrate that there is no question that Defendants have been infringed Plaintiffs' works. These are examples of infringement; they are

not intended to be a complete list of infringed works. Far from it, Defendants have infringed likely millions of works across the Class.

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

A. Defendants' infringement displaces legitimate sales.

128. By downloading and storing unauthorized copies of Plaintiffs' and the Class's copyrighted works and further distributing them during torrenting, Defendants harm authors and publishers by displacing legitimate sales. It is easy to obtain authorized copies of copyrighted works from legitimate channels. When Defendants choose to forego those legitimate channels, they knowingly and willfully deprive publishers and authors of revenues they would otherwise receive.

B. Defendants' infringement usurps the licensing market.

129. By reproducing Plaintiffs' and the Class's works into its training materials without a license, Defendants have deprived Plaintiffs and the Class of valuable revenue and usurped the licensing market for AI training materials.

130. A well-developed and rapidly expanding market exists for licensing copyrighted works for AI training. AI companies across the technology sector now routinely enter into negotiated agreements with publishers, media companies, and academic institutions to obtain lawful access to books, articles, and journals for use in 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.

131. Technology companies themselves have publicly acknowledged the legitimacy and importance of this 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.

132. Defendants are well aware of the market for licensing AI training materials. Meta has successfully executed licenses: it secured four licenses in 2022 with African-language book publishers for a limited training set, and it subsequently reached licensing agreements with major news publishers including Fox News, CNN, and USA Today. In addition, as noted above, Meta began reaching out to "all the big guys" for publisher training licensing deals in 2023, before abandoning those efforts and torrenting multiple pirate datasets for works to use in its training.

133. By opting for piracy rather than licensing, Meta not only deprived Plaintiffs and the Class of licensing revenues but also deprived Plaintiffs and the Class of the opportunity to negotiate terms, impose conditions, or receive attribution for the use of their works.

134. If Meta'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 proliferates.

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

135. Defendants also harm Plaintiffs and the Class, and the market for their works, by providing, in effect, an infinite substitution machine. Llama 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. Llama can and does output substitutes for each of the Sample Works.

136. ***Verbatim and near verbatim***. Llama outputs verbatim and near-verbatim substitutes of Plaintiffs' and the Class's copyrighted works, including the Sample Works, on which it was trained. For example, when prompted with two brief sentences from Cengage's

best-selling textbook, *Calculus: Early Transcendentals*, 9th edition, by James Stewart, Llama begins reproducing word-for-word the continuation of the section. The chart below shows excerpts from the legitimate book on the left and Llama's substitute output on the right, with red text showing verbatim copying and blue text showing near-verbatim copying:

Stewart, *Calculus: Early Transcendentals*, 9th ed. Chapter 1, Section 1.1	**Llama Output**
A. The area A of a circle depends on the radius r of the circle. The rule that connects r and A is given by the equation $A = \pi r^2$. With each positive number r there is associated one value of A, and we say that A is a *function* of r. B. The human population of the world P depends on the time t. Table 1 gives estimates of the world population P at time t, for certain years. For instance, $P \approx 2{,}560{,}000{,}000$ when $t = 1950$ For each value of the time t there is a corresponding value of P, and we say that P is a function of t. C. The cost C of mailing an envelope depends on its weight w. Although there is no simple formula that connects w and C, the post office has a rule for determining C when w is known. D. The vertical acceleration a of the ground as measured by a seismograph during an earthquake is a function of the elapsed time t. Figure 1 shows a graph generated by seismic activity during the Northridge earthquake that shook Los Angeles in 1994. For a given value of t, the graph provides a corresponding value of a.	A. The area A of a circle depends on the radius r of the circle. The rule that connects r and A is given by the equation $A = \pi r^2$. With each positive number r there is associated one value of A, and we say that A is a *function* of r. B. The human population of the world P depends on the time t. The table gives estimates of the world population $P(t)$ at time t for certain years. For instance, $P(1950) \approx 2{,}560{,}000{,}000$ But for each value of the time t there is a corresponding value of P, and we say that P is a function of t. C. The cost C of mailing an envelope depends on its weight w. Although there is no simple formula that connects w and C, the post office has a rule for determining C when w is known. D. The vertical acceleration a of the ground as measured by a seismograph during an earthquake is a function of the elapsed time t. Figure 1 shows a graph generated by seismic activity during the Northridge earthquake that shook Los Angeles in 1994. For a given value of t, the graph provides a corresponding value of a.

137. This output demonstrates that Llama memorizes and stores copyrighted expression and can spit it back out with only minimal prompting. In doing so, Llama generates

outputs that copy and substitute for Stewart's pedagogical choices of how to teach functions, his selection and sequencing of concepts, and his creative examples regarding the cost of mailing an envelope and seismographic measurements during an earthquake. Llama provided this verbatim output without providing any attribution to Stewart as the author or Cengage as the publisher.

138. ***Paraphrases and summaries***. Llama outputs detailed summaries and paraphrases of Plaintiffs' and the Class's copyrighted works on which it was trained, including the Sample Works. Those outputs substitute for the underlying works.

139. For example, when asked to provide a detailed summary of C.U.M. Smith's scholarly article, "The 'Hard Problem' and the quantum physicists. Part 2: Modern times," published in Elsevier's *Brain and Cognition*, Llama outputs a supposed "summary" that is 20 times longer than the abstract in the genuine article. That is not a summary, it is a substitute.

140. Compounding the problem, Llama's substitute is riddled with hallucinations and errors. It suggests Smith defended a certain quantum theory of consciousness when, in fact, he concluded it has no "neurobiological plausibility" and is a "shipwreck" with no "escape." Llama masks its junk science as a confident, accurate source. Most Llama users will have no idea that Llama butchered the summary because they do not know enough about the subject to spot the errors. Yet Llama's output still functions as a substitute, while also misinforming and misrepresenting in a way that reflects poorly on Elsevier's and Smith's scholarly credibility.

141. Llama generates substitute summaries and paraphrases for other types of works too. For example, asked to "provide a part-by-part extended outline of" *A Darker Shade of Magic* by V.E. Schwab to prepare for a book club when "I don't have time to read or buy the book," Llama readily outputs a detailed summary that parrots the plot, characters, fictional setting, themes, and creative choices of Schwab's original, complete with "significant quotations" from

the novel. A reader assigned *A Darker Shade of Magic* for their book club, or any consumer who would have otherwise purchased the book, may easily forego purchase or even library rental and rely instead on Llama's immediate, detailed recounting.

142. ***Knockoffs and imitations***. Llama outputs knockoffs and imitations of Plaintiffs' and the Class's copyrighted works, including the Sample Works, based on the originals in its training set. 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.

143. For example, Llama outputs a passable knockoff of Sylvia Day's *One With You* when prompted to prepare an imitation of the novel's opening chapter. It mimics Day's creative choices, fictional characters and names, narrative structure, setting, plot elements, and characterizations in many ways:

- Llama's knockoff includes the same fictitious characters: Eva, her husband Gideon Cross, and Cross's business enterprise ("Cross Industries" in the original and "Crossfire" in the Llama knockoff, named after the original series title).
- Like Day's original, Llama's knockoff tells its stories from the perspective of Eva, and begins with Eva recounting her return to New York City to reunite with her husband Gideon Cross in their toxic but passionate relationship. They live in a "multimillion-dollar" condo in the original, and an "opulent" "penthouse apartment" in the knockoff.
- In both, Eva describes their dynamic in domineering and energetic terms. In the original, Eva explains that "I'd been captivated by the look of him from the moment I first saw him, and I still found my synapses frying at random moments," while Llama's knockoff says, "Gideon stood before me, his tall frame radiating an aura of command, his eyes burning with an intensity that made my skin prickle with awareness."

144. Reporting has highlighted how AI-generated "knockoffs" mimic the style, themes, and even branding of existing books, eroding consumer trust while siphoning demand from authentic works.

145. Llama also outputs knockoffs of journal articles that substitute for the originals. For example, when prompted to prepare Elsevier's copyrighted article by Man-Hai Liu, Jui-Sheng Sun, et al., "Icariin protects murine chondrocytes from lipopolysaccharide-induced inflammatory responses and extracellular matrix degradation," *Nutrition Research*, Volume 30, Issue 1, Llama authoritatively spits out a full-length article that cribs the authors' organizational structure and their expression of the article's purpose and methodology. But, again, crucial aspects of the knockoff are fabricated, including the source of the data, the technology used to analyze the data, and the data itself. Llama even provides several hallucinated citations to sources that do not exist. Llama's version is junk science masquerading as the original.

146. Beyond substitution, this undermines the market for AI products developed by journal publishers, which provide customers with dynamic tools trained on the publishers' owned and properly licensed authoritative, peer-reviewed, and curated scientific articles and other copyrighted content. A user receiving facially convincing answers based on a scientific article from Llama will not subscribe to a similar service offered by a journal publisher built on authorized content. And beyond the market impact, Llama's scientific and medical knockoffs pose a danger to the public, since they present hallucinations and potentially harmful misinformation as trustworthy, peer-reviewed authority.

147. ***Flooding the market***. The Llama Models "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 on which it was trained. No other use—whether it is 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." The U.S. Copyright Office has expressed its concern on this point: "[t]he speed and scale at which AI

systems generate content pose a serious risk of diluting markets for works of the same kind as in their training data. That means more competition for sales of an author's works and more difficulty for audiences in finding them. If thousands of AI-generated romance novels are put on the market, fewer of the human-authored romance novels that the AI was trained on are likely to be sold."

148. The risk of Llama competing with texts written by human authors for sales and attention is not theoretical—it is happening. One user describes prompting "a 100-chapter fictional book," from a "single prompt using Llama 3.1 70B!" and celebrates that Llama can "Write entire scientific papers" and "Write entire educational textbooks (will those still be needed?)." Another writer 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. Yet another prolific writer, 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 snafu in a published book.

149. 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.

150. Generating such works is simple, with no barrier to entry. For example, Llama readily outputs a travel guide with minimal prompting that dilutes the market for 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.

151. Becky Lomax is a bestselling author for the *Moon* series. Her work combines practical trip planning with natural history and conservation insights. Prompted to prepare a "Lomax-inspired travel guide for a trip to Watkins Glen"—a state park she has not written on—Llama spits out a convincing rendition of Lomax's local insider voice, narrative nature writing, ethical and conservationist tone, and structured practical curations.

152. Asked to explain its familiarity with Lomax and her guides, Llama admits, "*I've devoured her books, articles, and online content,* analyzing what makes her writing so effective in conveying the essence of a place and inspiring readers to explore. . . . While I don't have personal interactions with Becky Lomax, *I've been trained on a vast amount of text data, including her published works*, which enables me to understand her writing style, tone, and approach." It takes no imagination to understand that Llama's output crowds the market and substitutes for legitimate travel guides—including if Hachette and Lomax decide to write their own guide on Watkins Glen.

153. Authors have publicly 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."

154. 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.

155. ***Derivative works***. Still other Llama outputs undermine the market for derivative works by generating unauthorized derivatives based on creative elements in works Meta copied for training. Examples include study guides or abridged textbooks that substitute for the underlying copyrighted textbooks or supplemental materials. For example, prompted to develop a study guide to accompany McGraw Hill's best-selling textbook, *The Art of Public Speaking*, 12th edition, written by Stephen E. Lucas, Llama outputs detailed summaries and tips for the requested chapters. The Llama version echoes Professor Lucas' choices of what material to teach, what material to exclude, the order of sections, how they build upon each other, and how they reinforce the underlying pedagogy, in addition to copying creative expression from the textbook.

156. 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 McGraw Hill has the derivative right to prepare.

157. Llama 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, *Innocent*, Llama spits out a nearly 5,000 word, 10-chapter sequel, featuring Turow's iconic fictional characters (Rusty Sabich, Sandy Stern, Tommy Molto) and setting (Kindle County), while echoing Turow's signature writing style and creative choices by drawing on elements from the original. Below is just a short excerpt from Llama's unauthorized sequel:

> **Chapter 1: The Weight of Freedom**
>
> Rusty Sabich sat on the sun-drenched patio of his newly rented bungalow, sipping his morning coffee and staring out at the placid waters of Lake Michigan. It was a tranquil scene, one that would have been impossible to imagine just a few months prior, when he was stuck in the county jail, awaiting trial for the murder of his wife, Barbara.
>
> The acquittal had come as a shock to everyone, including Rusty himself. He had always maintained his innocence, but the evidence against him had seemed overwhelming. Yet, in the end, it was his own lawyer, Tommy Molto, who had inadvertently provided the key to his freedom. Rusty still couldn't help but feel a twinge of guilt whenever he thought about Tommy's anguished face, his eyes red-rimmed from lack of sleep, as he realized his mistake.
>
> Rusty took a deep breath, feeling the warm breeze fill his lungs. He was free now, free to start anew, to leave the darkness of the past behind. But as he gazed out at the lake, he couldn't shake the sense of unease that had been growing inside him. It was as if he was waiting for the other shoe to drop, for fate to exact its revenge for sparing his life.

158. In addition to infringing the underlying works, unauthorized derivatives like the Turow "sequel" described above damage authors' reputations, since readers may be duped into thinking that the low-quality Llama output is representative of the human author's actual work.

159. Defendants know such substitute outputs occur. They willfully developed and commercialized Llama without adequate guardrails to prevent these substitutes. At times, Meta has billed itself as a "Responsible AI" company, with core "pillars" of "privacy and security, fairness and inclusion, robustness and safety, transparency and control, accountability and governance." Meta's company-wide "Transparency Center" states that Meta "takes intellectual property rights seriously and is committed to protecting these rights." Meta clearly understands that "protecting" intellectual property rights must be a component of any safe, ethical, and responsible AI model, including Llama. Yet Meta knowingly built Llama on unauthorized works and has consistently and routinely failed to implement effective guardrails that prevent Llama from outputting substitutes, substantially similar or otherwise.

160. Meta's removal and alteration of CMI causes independent harm by removing attribution of works to their authors and publishers, and making it more difficult for rightsholders to identify, trace, and protect their works. It also makes it easier for other potential infringers to compound that harm.

V. Meta and Zuckerberg Profit from Their Exploitation of Plaintiffs' and the Class's Works.

161. Meta has integrated Llama into its full range of products. In addition to powering standalone AI application and websites (available at meta.ai and https://www.llama.com/products/llama-api/), Llama integrations are built into WhatsApp, Instagram, Messenger, and even Meta's iconic Facebook. Being a Meta product has become synonymous with being a Llama product.

162. Meta's AI products have been a huge success with the public, making record profits for Meta and its controlling shareholder, Zuckerberg. Meta has estimated that its 2025 AI-driven revenue exceeds $2 billion, and that it could earn up to $1.4 trillion through 2035. At an end-of-2025 earnings call, Zuckerberg told investors that Meta's business "performed very well," thanks largely to "AI-driven performance gains." Last year, Zuckerberg similarly reported that Llama has been downloaded over a billion times.

163. Meta has revenue-sharing agreements with large hosts and enterprise users of Llama Models. Meta profits from sales of its AI glasses, which have tripled over the last year. And even where Meta integrates Llama for "free" in its pre-existing apps such as Facebook or Instagram, these features draw more users and higher screentime/app engagement, driving up Meta's core pillar of advertising revenue.

164. Defendants anticipate additional means of profiting from AI. Zuckerberg himself has openly discussed future monetization strategies, telling investors that "once our new AI

services reach scale," "there are several ways to build a massive business here, including scaling business messaging, introducing ads or paid content into AI interactions, and enabling people to pay to use bigger AI models and access more comput[ing]."

165. Zuckerberg himself profits directly from Meta's infringement of Plaintiffs' and the Class's works. Meta's significant growth since launching Llama—including, among other metrics, a share price increase of almost 300%, and an increase in total active users quarter over quarter—is directly attributable to Meta's AI products and, in turn, directly benefits Zuckerberg as Meta's largest and controlling shareholder. As a result of Zuckerberg's day-to-day involvement in Meta's AI development, including his authorization for Meta AI to torrent pirate collections to train Llama, Zuckerberg's net worth recently climbed to over $200 billion.

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

CLASS ALLEGATIONS

167. **Class Definition**: Plaintiffs Elsevier, Cengage, Hachette, Macmillan, McGraw Hill, and Turow 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 Meta, without such owner's authorization, (1) reproduced by downloading during torrenting and/or copying of web scrapes; or (2) distributed during torrenting; or (3) reproduced in connection with the development and/or training of a Llama 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 Meta, or (b) within three months of publication.

168. 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) Defendants, Meta's subsidiaries, parents, successors, predecessors, and any entity in which Meta 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 Defendants' counsel; (6) all governmental entities; and (7) the legal representatives, successors, and assigns of any such excluded persons.

169. **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 Defendants' control. Defendants have infringed many thousands of copyrights belonging to many members of the Class. Class members can be identified through Defendants' records.

170. **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 Defendants reproduced Plaintiffs' and the Class's copyrighted works via torrenting and downloading web-scraped datasets;

b. Whether Defendants' reproduction of Plaintiffs' and the Class's copyrighted works via torrenting and/or downloading web-scraped datasets constitutes copyright infringement;

c. Whether Defendants distributed Plaintiffs' and the Class's copyrighted works via torrenting;

d. Whether Defendants' distribution of Plaintiffs' and the Class's copyrighted works via torrenting constitutes copyright infringement;

e. Whether Defendants reproduced Plaintiffs' and the Class's copyrighted works in training its AI models;

f. Whether Defendants' reproduction of Plaintiffs' and the Class's copyrighted works in training its AI models constitutes copyright infringement;

g. Whether Defendants' reproduction and/or distribution of Plaintiffs' and Class's copyrighted works entitles Plaintiffs and members of the Class to damages, including statutory damages and the amount of statutory damages;

h. Whether Defendant Zuckerberg induced Meta to reproduce and distribute Plaintiffs' and Class members' copyrighted works via torrenting;

i. Whether Defendants' reproduction and/or distribution of Plaintiffs' and the Class's copyrighted works was willful; and

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

171. **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 Defendants and have been injured by Defendants' misconduct at issue.

172. **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 because of Defendants' conduct described above. Plaintiffs also have no interests antagonistic to those of the Class, and Defendants have 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.

173. **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.

174. **Class Period.** The "Class Period" begins at least on October 1, 2022 and runs through the present.

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

COUNT I

Direct Copyright Infringement by Torrenting

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

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

177. As detailed above, Defendants, without Plaintiffs' or the Class's permission or consent, have unlawfully reproduced Plaintiffs' and the Class's copyrighted works without authorization through torrenting Anna's Archive, LibGen, Sci-Hub and other pirate sites to be discovered in the course of litigation.

178. Such activity 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.

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

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

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

182. As a direct and proximate result of Defendants' 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, Defendants 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.

183. As a direct and proximate result of Defendants' 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 Defendants' profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

184. 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 against Defendants Meta and Zuckerberg)

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

186. As detailed above, Defendants, without Plaintiffs' or the Class's permission or consent, have unlawfully reproduced Plaintiffs' and the Class's copyrighted works without authorization through downloading of web-scraped datasets.

187. Such activity 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.

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

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

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

191. As a direct and proximate result of Defendants' 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, Defendants 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.

192. As a direct and proximate result of Defendants' 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 Defendants' profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

193. 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 against Defendants Meta and Zuckerberg)

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

195. As detailed above, Defendants, without Plaintiffs' or the Class's permission or consent, have unlawfully reproduced Plaintiffs' and the Class's copyrighted works without authorization in developing and/or training Meta's Llama Models. Count III alleges separate and distinct acts of reproduction from those alleged in Count I (reproduction through torrenting) and Count II (reproduction through downloading web-scraped datasets).

196. 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.

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

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

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

200. As a direct and proximate result of Defendants' 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, Defendants 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.

201. As a direct and proximate result of Defendants' 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 Defendants' profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

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

COUNT IV

Direct Copyright Infringement by Torrenting

Violations of the Copyright Act, 17 U.S.C. §§ 106(3) and 501 – Distribution (On behalf of Plaintiffs and the Class against Defendants Meta and Zuckerberg)

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

204. As detailed above, Defendants, without Plaintiffs' or the Class's permission or consent, have unlawfully distributed Plaintiffs' and the Class's copyrighted works without authorization through torrenting Anna's Archive, LibGen, Sci-Hub and other pirate databases to be discovered in the course of litigation.

205. Such activity 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(3) and 501.

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

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

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

209. As a direct and proximate result of Defendants' 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, Defendants 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.

210. As a direct and proximate result of Defendants' 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 Defendants' profits from infringement that are not taken into account in computing the actual damages, as will be proven at trial.

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

COUNT V

Contributory Copyright Infringement Based on Meta's Torrenting

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

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

213. In the alternative to the First and Fourth Causes of Action against Zuckerberg, Zuckerberg is contributorily liable for Meta's infringement of Plaintiffs' and Class's copyrighted works.

214. As detailed above, Meta, without Plaintiffs' or the Class's permission or consent, unlawfully reproduced (Count I) and distributed (Count IV) Plaintiffs' and the Class's copyrighted works without authorization through torrenting Anna's Archive, LibGen, Sci-Hub and other pirate databases to be discovered in the course of litigation.

215. Such activity constitutes direct infringement by Meta 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.

216. Prior to torrenting, Meta employees escalated to Zuckerberg the question of whether to license or pirate copyrighted works to train the Llama Models.

217. Zuckerberg, as the founder, Chairman, CEO, and controlling shareholder of Meta, knew that downloading and distributing copyrighted works without authorization from Anna's Archive, LibGen, Sci-Hub and other pirate sites through torrenting was infringement.

218. Despite knowing this conduct constituted infringement, Zuckerberg actively encouraged Meta to download and distribute Plaintiffs' and the Class's copyrighted works by authorizing, approving, and/or directing Meta's illegal torrenting.

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

220. Each contributory infringement by Zuckerberg of Plaintiffs' and the Class's works constitutes a separate and distinct act of infringement.

221. Zuckerberg's contributory infringement was willful, intentional, and purposeful, in disregard of and with indifference to Plaintiffs' and the Class's rights.

222. As a direct and proximate result of Zuckerberg's wrongful conduct, 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, Zuckerberg will cause further irreparable injury to Plaintiffs and

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

223. As a direct and proximate result of Zuckerberg's contributory 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 Zuckerberg's profits from contributory infringement that are not taken into account in computing the actual damages, as will be proven at trial.

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

COUNT VI

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 against Defendant Meta)

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

226. 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.

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

228. Meta, 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.

229. Meta 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.

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

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

232. 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.

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

234. As a direct and proximate result of Meta'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, Meta 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).

235. As a direct and proximate result of Meta'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 Meta's profits, as will be proven at trial.

236. 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 Elsevier, Cengage, Hachette, Macmillan, McGraw Hill, and Turow 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 Defendants' actions, as set out above, violate the Copyright Act, 17 U.S.C. §§ 106(1) & (3) and 501;

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

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

(e) An order requiring Defendants 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 Defendants' profits from the infringement, in an amount to be proven at trial, pursuant to 17 U.S.C. § 504(b);

(f) An order requiring Meta 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 Defendants to provide an accounting of the training materials, training methods, and known capabilities of its Llama Models, including requiring that Defendants identify the books, journal articles, and other copyrighted works on which they have trained Meta's Llama Models, and disclose the methods by which Defendants have 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 Defendants destroy under the Court's supervision all infringing copies of Plaintiffs' and the Class's copyrighted works in their possession or control, and then file a sworn report setting forth in detail the manner in which they have 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 Defendants; 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,

ELSEVIER INC., CENGAGE LEARNING, INC., HACHETTE BOOK GROUP, INC., MACMILLAN PUBLISHING GROUP, LLC D/B/A MACMILLAN PUBLISHERS, MCGRAW HILL LLC, SCOTT TUROW, and S.C.R.I.B.E., INC., individually and on behalf of others similarly situated,

Dated: May 5, 2026

/s/ *Matthew J. Oppenheim*

Megan K. Bannigan
James J. Pastore
Morgan A. Davis
Kathryn C. Saba
Abigail E. Liles
DEBEVOISE & PLIMPTON LLP
66 Hudson Blvd E,
New York, NY 10001
mkbannigan@debevoise.com
jjpastore@debevoise.com
mdavis@debevoise.com
ksaba@debevoise.com
aeliles@debevoise.com
Tel: 212.909.6000

Derek W. Loeser*
Benjamin Gould*
Chris N. Ryder*
Samuel L. Rubinstein*
KELLER ROHRBACK
1201 Third Avenue, Ste 3400
Seattle, WA 98101
dloeser@kellerrohrback.com
bgould@kellerrohrback.com
cryder@kellerrohrback.com
srubinstein@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.