Elizabeth J. Cabraser (State Bar No. 083151)
Daniel M. Hutchinson (State Bar No. 239458)
Reilly T. Stoler (State Bar No. 310761)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Telephone: (415) 956-1000
ecabraser@lchb.com
dhutchinson@lchb.com
rstoler@lchb.com

Rachel Geman (*pro hac vice forthcoming*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Telephone: (212) 355-9500
rgeman@lchb.com

Scott J. Sholder (*pro hac vice forthcoming*)
CeCe M. Cole (*pro hac vice forthcoming*)
COWAN DEBAETS ABRAHAMS & SHEPPARD LLP
60 Broad Street, 30th Floor
New York, New York 10004
Telephone: (212) 974-7474
ssholder@cdas.com
ccole@cdas.com

Attorneys for Plaintiff and the Proposed Class

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

CHRISTOPHER FARNSWORTH, *individually and on behalf of others similarly situated*,

Plaintiff,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 3:24-cv-6893

CLASS ACTION COMPLAINT

JURY TRIAL DEMANDED

Plaintiff Christopher Farnsworth, on behalf of himself and all other similarly situated individuals (the "Class," as defined below), brings this Class Action Complaint against Defendant Meta Platforms, Inc. ("Meta").

NATURE OF THE ACTION

1. Meta stole hundreds of thousands of pirated copyrighted books to build a commercial product called a Large Language Model ("LLM"). The United States Constitution recognizes the fundamental principle that copyright holders are entitled to exclusive rights in their works in order to incentivize further creation. Meta ignored this basic principle, and the federal law embodying it, by stealing and exploiting copyright-protected books for profit.

2. Plaintiff Christopher Farnsworth is a best-selling fiction author. He brings this class action under the Copyright Act to redress the harm Meta's infringement caused to him and other authors. Meta exploited Plaintiff's works without authorization, made illicit copies of them, and then fed those copies to its LLM. Meta did this to enhance the quality of its LLM's language output and, ultimately to have a more desirable and profitable LLM product to compete in the AI arms race that has developed over the last few years.

3. Meta calls its set of LLMs "Llama." Like other LLMs, Llama is an AI software program designed to emit convincingly naturalistic text outputs in response to user prompts. The quality of Llama's output is determined by the quality of the data it is fed. Meta knew that the more high-quality, long-form text it fed to Llama, the better Llama would perform commercially. In this way, Llama is what it ingests. By willfully including unauthorized and pirated copies of Plaintiff's copyrighted literary works in its development work, pre-training, and training data, Meta exploited authors' literary talents without consent or compensation of any kind.

JURISDICTION AND VENUE

4. The Court has subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a) because this action arises under the Copyright Act of 1976, 17 U.S.C. § 101, *et seq*.

5. The Court also has personal jurisdiction over Defendant because it has purposely availed itself of the privilege of conducting business in this District.

6. Meta is headquartered in this District and its copyright infringement—including the downloading and reproduction of the copyrighted works—occurred, in substantial part, in this District. Meta also marketed and distributed its LLMs from this District to citizens of California.

7. Venue is proper under 28 U.S.C. § 1400(a) because Meta or its agents reside or may be found in this District and their infringing activities, along with their commercialization of their infringing activities, occurred in this District. Venue is also proper under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Plaintiff's claims occurred in this District, including the largescale copyright infringement and commercialization of Meta's LLMs.

THE PARTIES

8. Plaintiff Christopher Farnsworth is an author who resides in Los Angeles, California. He is the author of eight novels, three novellas, and varied other literary works.

9. Defendant Meta Platforms, Inc. is a Delaware corporation with its principal place of business at 1 Hacker Way, Menlo Park, CA 94025.

FACTUAL ALLEGATIONS

I. Meta's Development and Commercialization of Llama.

10. Founded in 2004 as a social media company then called The Facebook, Inc., Meta has grown and diversified its technology services rapidly over the last 20 years. Presently, Meta has a market capitalization of over $1.4 trillion and is among the ten most valuable companies in the world. Meta makes billions of dollars selling advertising and services across its flagship platforms like Facebook, Instagram, and WhatsApp.

11. Following the sensational consumer response to OpenAI's release of ChatGPT in November of 2022, many of the world's largest technology companies accelerated their research and development of commercial LLMs. ChatGPT was powered by an LLM capable of generating uncannily human-like responses to people's queries.

12. In February of 2023, Meta developed and released a collection of foundational LLMs it called LLaMA as part of its commercial generative AI development initiative. These

LLMs were later renamed Llama 1. Meta initially released Llama 1 for non-commercial research uses.

13. Less than five months later, in July of 2023, Meta released Llama 2, this time "available under a permissive commercial license."[1] By September 27, 2023, Meta reported that users had downloaded over 30 million copies of Llama-based models.[2]

14. On April 18, 2024, Meta released Llama 3, which it built on the foundation of Llama 1 and 2.[3] Upon its commercial release, Meta touted Llama 3 as powering its new consumer service Meta AI, "one of the world's leading AI assistants."[4] Meta admittedly is using its LLMs to enhance its current commercial products and, on information and belief, Meta is working on a premium, paid-subscription version of its AI assistant service—powered by Llama 3.

II. Meta Developed Its Commercial AI Models Using Stolen, Copyrighted Material.

A. Large Language Models and the Training Process.

15. At issue is a species of AI models called large language models, or LLMs. LLMs are designed to mimic human use of language. LLMs are able to simulate patterns of human language by processing input text ("prompts") and generating output text in response to these prompts on a predictive basis, *i.e.*, determining what word follows what.

16. At a high level, LLMs are algorithms designed to distill mathematically the relationships between words in written works through a process called "training." LLMs achieve this goal by ingesting massive amounts of training materials such as books, breaking down input text into smaller pieces—words or portions of words, called "tokens"—then translating those pieces into "vectors," or a sequence of numbers that is used to identify the token within the series of algorithms. Those vectors help place each token in a probabilistic context identifying other tokens closely associated with the word. As described by industry-leading generative AI development company OpenAI, in comments to the Copyright Office, "the process begins by breaking text down into roughly word-length 'tokens,' which are converted to numbers. The

---

[1] https://ai.meta.com/llama/faq/
[2] https://ai.meta.com/blog/llama-2-updates-connect-2023/
[3] https://arxiv.org/pdf/2407.21783
[4] https://about.fb.com/news/2024/04/meta-ai-assistant-built-with-llama-3/

model then calculates each token's proximity to other tokens in the training data—essentially, how near one word appears in relation to any other word. These relationships between words reveal which words have similar meanings . . . and functions."[5]

17. As a model trains by digesting more and more written works, the algorithms, which distill the relationship between various tokens, changes with it. A model takes text inputs in the form of an incomplete phrase or passage, and attempts to complete the phrase, essentially a fill-in-the-blank quiz. A model compares its predicted phrase completion with the actual "correct" answer and then adjusts its internal algorithms to "learn" from its mistakes and minimize the difference between any given text input and the "correct" text output.

18. A model then repeats this same cycle millions, possibly billions, of times across the entire corpus of training materials, adjusting its algorithms each time to reflect the text input from the corpus. This is known as the "pre-training" process, which is the foundation of creating a "base" LLM model, which can be "fine-tuned" later to achieve more specific results. The pre-training process fundamentally enables a model to process prompts and generate text output that mimics human language. It does so by exposing a model to a wide range of texts and using algorithms to predict the next word in the text. By repeating this process, a model develops fluency in style, syntax, and expression of ideas, largely by digesting and processing the protected expression contained in the material used for training. The LLM mines the expression contained in the training corpus, adjusting its algorithms such that it can mimic the ordering of words, style, syntax, and presentation of facts, concepts, and themes.

19. In a literal sense, a model is what it ingests: without training on material, there is no LLM. The quality and quantity of the training corpus is critical to the quality of the resulting model. As one researcher put it: "[large language] model behavior is not determined by architecture, hyperparameters, or optimizer choices [i.e. technical features set during model training]. *It's determined by your dataset, nothing else. Everything else is a means to an end in*

[5] *See* Comment of OpenAI "*Re: Notice of Inquiry and Request for Comment* [Docket No. 2023-06]," United States Copyright Office, Oct. 30, 2023, p. 5-6 (available at: https://downloads.regulations.gov/COLC-2023-0006-8906/attachment_1.pdf).

efficiently deliver[ing] compute to approximating that dataset."[6] Conversely, if the LLM's training materials are of poor quality, the output—the end product—also will be of poor quality.

20. Books are especially valuable training material for the training and development of LLMs. As one commentator put it, "[b]ooks offer formal and lengthy texts which help LLMs understand complex language structures, grasp long-term context, and produce coherent narratives."[7]

21. Any LLM training process, including Meta's training of Llama, involves creating multiple copies of the training text which often include a vast number of written works in their entirety. As the U.S. Patent and Trademark Office has observed, LLM "training" "almost by definition involve[s] the reproduction of entire works or substantial portions thereof."[8]

B. Meta Copied A Massive Trove of Pirated Books To Train Its Llama Models.

22. In public disclosures, Meta admitted that it made unauthorized copies and willfully reproduced copies of nearly two hundred thousand pirated copyrighted books to advance its commercial AI training and development projects.

23. Since those disclosures, Meta has become less forthright about its AI work, but what Meta previously disclosed confirms that Meta downloaded and copied a dataset of text without authorization of registered copyright owners, including Plaintiff, called The Pile as part of its work in training and developing its LLMs.

24. The Pile is an 800 GB open-source dataset created for training large language models that is, and was well known by AI developers. At the time Meta downloaded The Pile, it was hosted and made publicly available online by a nonprofit called EleutherAI. As described by its creators, "The Pile is constructed from 22 diverse high-quality subsets . . . many of which

---

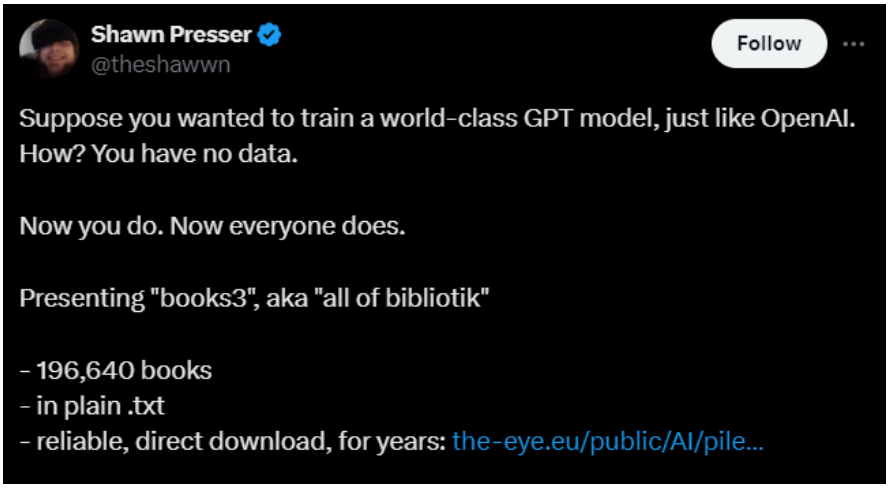
[6] *See* "The 'it' in AI models is the data set," James Betker, https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/ (June 10, 2023) (emphasis added).

[7] "Pretraining of Large Language Models," Ritwik Raha, https://gist.github.com/ritwikraha/77e79990992043f60a9588610b2781c5 (last accessed Sept. 25, 2024).

[8] U.S. Patent & Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy* 29 (2020), *available at* https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf (last accessed Jan. 22, 2024).

derive from academic and professional sources. . . . [M]odels trained on the Pile improve significantly over Raw CC and CC-100 on all components of the Pile, which improving performance on downstream evaluations."[9]

25. One of The Pile's architects is an independent developer named Shawn Presser. Presser created a dataset included in The Pile called "Books3," which is a trove of pirated books. Presser described how he created Books3 in a Twitter thread from October 2020:[10]



26. Presser explained he created Books3 in response to "OpenAI's papers on GPT-2 and 3," which "references to datasets named 'books1' and 'books2,'" the latter of which Presser suspects "might be 'all of libgen.'"[11] LibGen refers to "Library Genesis," a website offering pirated books that was ordered shut down for copyright infringement in 2015. *See Elsevier, Inc. et al. v. www.Sci-hub.org et al.*, 15-cv-2482-RWS, Dkt. No. 53 (Oct. 30, 2015).

27. To create a pirated-book dataset comparable to what he suspected OpenAI created for itself, Presser announced that Books3 was a direct download of all "196,640 books" from a *different* pirate website called "bibliotik."[12]

---

[9] "The Pile: An 800GB Dataset of Diverse Text for Language Modeling," Gao et al, Abstract, https://arxiv.org/pdf/2101.00027 (last accessed Sept. 25, 2024) ("Gao Paper").

[10] *See* Tweet by Shawn Presser, Oct. 25, 2020, https://x.com/theshawwn/status/1320282149329784833?lang=en (last accessed Sept. 25, 2024).

[11] *See* Tweet by Shawn Presser, Oct. 25, 2020, https://x.com/theshawwn/status/1320282152689336320 (last accessed Sept. 25, 2024).

[12] *See* Tweet by Shawn Presser, Oct. 25, 2020, https://x.com/theshawwn/status/1320282149329784833?lang=en (last accessed Aug. 15, 2024).

28. Bibliotik is a "notorious pirated collection" of "pirated books."[13] For years prior to its use as "Books3," Bibliotik was frequently included in roundups of the best—and most popular—sources for pirated books.[14]

29. Books3 was a critical part of The Pile. In EleutherAI's paper on The Pile, it explained the key value of Books3 as training material:

> Books3 is a dataset of books derived from a copy of the contents of the Bibliotik private tracker . . . Bibliotik consists of a mix of fiction and nonfiction books and is almost an order of magnitude larger than our next largest book dataset (BookCorpus2). ***We included Bibliotik because books are invaluable for long-range context modeling research and coherent storytelling***.[15]

30. Presser and EleutherAI repeatedly and publicly acknowledged that The Pile and Books3 was a cache of pirated material. EleutherAI's paper on The Pile noted that "there is little acknowledgment of the fact that the processing and distribution of data owned by others may also

---

[13] *See* Schoppert, "Whether you're an undergraduate doing research, or a fan of the Nick Stone novels, or indeed a hungry AI...," Nov. 29, 2022, https://aicopyright.substack.com/p/whether-youre-an-undergraduate-doing ("What is Bibliotik? A notorious pirated collection."); "What I Found in a Database Meta Uses to Train Generative AI," Alex Reisner, *The Atlantic*, Sept. 25, 2023, https://www.theatlantic.com/technology/archive/2023/09/books3-ai-training-meta-copyright-infringement-lawsuit/675411/ ("a collection of pirated ebooks, most of them published in the past 20 years."); "Revealed: The Authors Whose Pirated Books are Powering Generative AI," Alex Reisner, *The Atlantic*, Aug. 19, 2023, https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/ ("collections of pirated books, such as Library Genesis, Z-Library, and Bibliotik, that circulate via the BitTorrent file-sharing network."); "Are ChatGPT, Bard and Dolly 2.0 Trained On Pirated Content?," Roger Monti, *Search Engine Journal*, April 20, 2023, https://www.searchenginejournal.com/are-chatgpt-bard-and-dolly-2-0-trained-on-pirated-content/485089/ ("The Books3 dataset contains the text of books that were pirated and hosted at a pirate site called, bibliotik.").

[14] *See* Commit History of "Awesome Piracy," *Github.com*, Oct. 13, 2018, https://github.com/aviranzerioniac/awesome-piracy/commit/61928765e3ee0b4f3dbe3c0724b196e5f0f17e59?short_path=5a831ea#diff-5a831ea67cf5cf8703b0de46901ab25bd191f56b320053be9332d9a3b0d01d15 (October 13, 2018 commit to "awesome piracy" repo listing "Bibliotik Popular ebooks/audiobooks private tracker"); "Reddit Piracy Megathread" repo., *Github.com*, Mar. 21, 2019 https://github.com/magicoflolis/Reddit-Piracy-Megathread/blob/master/data/findingtextbooks.md (March 21, 2019 guide from "r/piracy" on how to source textbooks listing "Bibliotik"); "List of free eBook download sites," *Pirates-forum.org*, Mar. 06 2014, https://pirates-forum.org/Thread-List-of-free-eBook-download-sites?highlight=bibliotik (March 31, 2014 post from "pirates forum" thread entitled "List of free eBook download sites" listing "bibliotik").

[15] Gao Paper, p. 3-4.

be a violation of copyright law."[16] Furthermore, The Pile's datasheet notes that "Books3 is almost entirely comprised of copyrighted works . . .."[17] Presser, for his part, admitted to releasing Books3 despite "fear of copyright backlash"[18]

31. In August 2023, Books3 was removed from the "most official" copy of The Pile—hosted by "The Eye," an online repository of training datasets—due to copyright complaints. Despite this removal, the original version appears otherwise available as part of The Pile from other sources.

32. In its February 27, 2023 research paper, Meta admitted to downloading and reproducing Books3 willfully as part of its Llama development project. Meta AI researchers explained that the training data used to develop Llama 1 included "two book corpora . . . the Gutenberg Project, which contains books that are in the public domain, and the Books3 section of The Pile (Gao et al., 2020), a publicly available dataset for training large language models."[19]

33. In referring to the Books3 dataset it used, Meta researchers: (1) directly cited the 2020 EleutherAI paper (authored by Leo Gao) which describes Books3 as "a dataset of books derived from a copy of the contents of the Bibliotik private tracker;"[20] and (2) distinguished between the public *domain* books in the Gutenberg dataset and the publicly *available* (i.e. unauthorized and pirated) books in the Books3 dataset. Meta knew Books3 was a trove of copyrighted content sourced from pirate websites like Bibliotik and used it anyway.

34. Notably, in an earlier research paper authored by Meta AI in June 2022, titled "OPT: Open Pre-trained Transformer Language Models," Meta indicated that it had eliminated many of The Pile component datasets from its LLM pre-training data because "[Meta] found they increased the risk of instabilities . . . ***or were otherwise deemed unsuitable***."[21]

---

[16] *Id.* at 14-15.

[17] "Datasheet for the Pile," Gao et al, Jan. 20, 2022, p. 15, https://arxiv.org/pdf/2201.07311 (last accessed Sept. 25, 2024).

[18] Comment of "sillysaurusx," *Hacker News*, Jul. 11, 2023, https://news.ycombinator.com/item?id=36685115 (last accessed Sept. 25, 2024).

[19] https://arxiv.org/pdf/2302.13971

[20] Gao Paper, p. 3

[21] https://arxiv.org/pdf/2205.01068

35. For the purposes of the work memorialized in this June 2022 research paper, Meta excised component datasets that included copyrighted material.

36. For example, Meta excluded the "PubMed Central" and "PhilPapers" datasets. PubMed Central is a "repository for biomedical articles run by the United States of America's National Center for Biotechnology Information (NCBI)."[22] The PubMed Central dataset contains copyrighted material.

37. Likewise, the PhilPapers dataset is a repository of "***open-access*** philosophy publications from an international database maintained by the Center for Digital Philosophy at the University of Western Ontario."[23] The PhilPapers dataset contains copyrighted material.

38. In June 2022, Meta also excluded—as "unsuitable"—the Books3 dataset from its training data. Shortly after that, something changed and as noted above, by February 2023, Meta had copied Books3 and used it to train its LLMs.

39. In March 2023, the Llama 1 language models were leaked to a public internet site and have continued to circulate.

40. Later in March 2023, Meta issued a DMCA takedown notice to a programmer that released a tool to help users download the leaked Llama 1 language models. In the notice, Meta asserted copyright over the Llama 1 large language models.

41. In July of 2023—after Meta was sued for its infringing use of copyrighted material for LLM training—Meta released Llama 2, which it described in a July 19, 2023 research paper entitled "Llama 2: Open Foundation and Fine-Tuned Chat Models"[24] (the Llama 2 Paper).

42. The Llama 2 paper does not indicate the specific datasets used for pre-training but indicates that *more* data was used (not less) and again conspicuously describes the data it sourced as "publicly available online," and not public domain.

43. Meta's ongoing exploitation of Books3 was particularly egregious because it took place after Books3 was taken down from The Pile in 2023 due to copyright complaints directly

[22] Gao Paper, p. 3
[23] Gao Paper, p. 5 (emphasis added).
[24] https://arxiv.org/pdf/2307.09288.pdf

before Meta sought DMCA removal of the Llama 1 model to vindicate its own copyrights. This is especially ironic considering Llama 1 is a commercial LLM that was developed—without authorization—on pirated copies of hundreds of thousands of copyrighted literary works, including the literary works of Plaintiff.

44. Instead of willfully downloading and reproducing a notorious trove of pirated material, Meta could have lawfully purchased copies of books then negotiated a license to reproduce them. Alas, Meta did not even bother to pay the purchase price for the books it illegally downloaded, let alone obtain a license for their reproduction.

45. Meta's commercial copying of Plaintiff's work and works owned by the proposed Class was manifestly unfair use, for several reasons. Many AI companies have described the AI training process as "teaching" a model human language, much in the way a human learns. While this self-serving anthropomorphizing of LLMs is misplaced, at a minimum, humans who learn from books buy them, borrow them from libraries that buy them, or otherwise procure them lawfully, thus providing at least some measure of compensation to authors and creators. Meta does not, and it has usurped authors' content for the purpose of creating a machine built to generate the very type of content for which authors would usually be paid.

46. Meta, in taking authors' works without compensation, has deprived authors of books sales and licensing revenues. There is, and has been, an established market for the sale of books and e-books, yet Meta ignored it and chose to scrape a massive corpus of copyrighted books from the internet, without even paying for an initial copy.

47. Meta has also usurped a licensing market for copyright owners. In the last two years, a thriving licensing market for copyrighted training data has developed. AI companies have paid hundreds of millions of dollars to obtain licenses to reproduce high-quality copyrighted material for LLM training.

48. Meta chose to use Plaintiff's works, and the works owned by the proposed Class, free of charge, and in doing so has harmed the market for the copyrighted works by depriving them of book sales and licensing revenue.

III. Meta Pirated Material for Commercial Gain at the Expense of Authors.

49. Perversely, LLMs seriously threaten the livelihood of the same authors whose works they are non-consensually "trained" on.

50. Goldman Sachs's estimates that generative AI could replace 300 million full-time jobs in the near future, or one-fourth of labor currently performed in the United States and Europe.

51. Already, writers report losing income from copywriting, journalism, and online content writing, which are critical sources of income for book authors.

52. For example, The Authors Guild, the oldest professional organization in the U.S. representing writers and authors, and one on the forefront of efforts to shore up "creative markets against disruptions from generative AI"[25] has published an earnings study that shows a median writing-related income for full-time authors of just over $20,000, and that full-time traditional authors earn only half of that from their books.[26] The rest comes from activities like content writing—work that is itself starting to dry up as a result of generative AI systems trained on those writers' works, without compensation.

IV. Meta Exploited Plaintiff's Copyrighted Works.

53. Plaintiff and Class Members have suffered identical harms from Meta's infringement. The contents of the datasets that Meta used to "train" its LLMs are peculiarly within its knowledge, such that Plaintiff is unable to discern those contents with perfect accuracy. But Meta has admitted to using Books3 during the relevant time, and the contents of Books3 is widely reported. Plaintiff makes specific allegations of infringement below based on what is known about Meta's practices and what is known about the contents, uses, and availability of pirated book repositories that it is suspected Meta used, like Bibliotik.

---

[25] Authors Guild, "Positions and Policy Recommendations", available at https://authorsguild.org/advocacy/artificial-intelligence/ (last accessed Sept. 26, 2024).

[26] Authors Guild, "Top Takeaways from the 2023 Author Income Survey (2023)", available at https://authorsguild.org/news/key-takeaways-from-2023-author-income-survey/#:~:text=Though%20overall%20author%20incomes%20are,coming%20in%20a%20close%20second (last accessed Sept. 25, 2024).

54. Plaintiff Christopher Farnsworth is the author of a number of books, including *Blood Oath*, *Flashmob: A Novel*, *The Eternal World: A Novel*, and *The President's Vampire*. Each of these works was and is a part of the Books3 dataset. Pirated copies of these works are available online through websites like LibGen and Bibliotik. Farnsworth is the author and owner of the registered copyrights listed under his name in Exhibit A.

CLASS ALLEGATIONS

55. This action is brought by Plaintiff individually and on behalf of the Class, as defined below, pursuant to Rule 23(a), (b)(3) and 23(b)(2), (c)(4), and (g) of the Federal Rules of Civil Procedure:

> All legal and beneficial owners of copyrighted works that: (a) are registered with the United States Copyright Office; (b) were or are used by Meta in the process of LLM training, research, or development, including but not limited to the training and development of its Llama models and (c) have been assigned an International Standard Book Number (ISBN). The Class excludes Defendant, its officers and directors, members of their immediate families, their co-conspirators, aiders and abettors, and the heirs, successors or assigns of any of the foregoing.

56. The Class consists of at least thousands of authors and copyright holders and thus is so numerous that joinder of all members is impractical. The identities of members of the Class can be readily ascertained from business records maintained by Defendant.

57. The claims asserted by Plaintiff are typical of the claims of the Class, all of whose works were also copied as part of the LLM training, research, and development process.

58. The Plaintiff will fairly and adequately protect the interests of the Class and does not have any interests antagonistic to those of other members of the Class.

59. The Plaintiff has retained attorneys who are knowledgeable and experienced in complex litigation, and have brought multiple copyright class actions against AI companies asserting that their use of copyrighted material to train LLMs constitutes infringement.

60. Plaintiff requests that the Court afford Class members notice and the right to opt-out of any Class certified in this action.

61. This action is appropriate as a class action pursuant to Rule 23(b)(3) of the Federal Rules of Civil Procedure because common questions of law and fact affecting the Class predominate over those questions affecting only individual members. The law is uniform. And,

the common factual questions giving rise to common answers that move this litigation forward include:

a. Whether Meta's reproduction of the Class's copyrighted works constituted copyright infringement;

b. Whether Meta's reproduction of the Class's copyrighted works harmed Class members and whether Class members are entitled to damages, including statutory damages and the amount of statutory damages; and

c. Whether Meta's infringement was willful.

62. In addition, the class device is the superior mechanism for handling this action, and a class trial is eminently manageable.

63. This action is also appropriate as a class action pursuant to Rule 23(b)(2) of the Federal Rules of Civil Procedure because Meta's unlicensed exploitation of a large trove of the Class's books affects all class members in the same way, and any injunctive relief awarded will affect the Class as a whole.

64. Finally, at the very minimum, there are multiple common issues relating to Meta's uniform conduct, such as (but not limited to) their ingestion, reproduction, and willfulness.

CLAIM FOR RELIEF
Copyright Infringement (17 U.S.C. § 501)
Against Defendant Meta Platforms, Inc.

65. Plaintiff incorporates by reference the preceding factual allegations.

66. Plaintiff and members of the proposed Class have created literary works that are original and fixed in a tangible medium of expression, and they own the registered copyrights in the works that Meta reproduced and appropriated for its artificial intelligence projects.

67. Plaintiff and members of the proposed Class therefore hold the exclusive rights, including the rights of reproduction and distribution, to those works under 17 U.S.C. § 106.

68. Meta infringed the exclusive rights, under 17 U.S.C. § 106, of Plaintiff and members of the proposed Class by, among other things, reproducing and/or distributing the works owned by Plaintiff and the proposed Class in connection with procuring and using datasets for artificial intelligence training and development.

69. On information and belief, Meta's infringing conduct alleged herein was and continues to be willful. Meta infringed on the exclusive rights of Plaintiff and members of the proposed Class knowing that they were profiting from mass copyright infringement.

70. Plaintiff and members of the proposed Class are entitled to, at their election, statutory damages or, actual damages (including disgorgement of profits), and are entitled to attorneys' fees and costs, and other remedies available under the Copyright Act.

71. Plaintiff and members of the proposed Class have been and continue to be irreparably injured due to Meta's conduct, for which there is no adequate remedy at law. Meta will continue to infringe on the exclusive right of Plaintiff and the proposed class unless their infringing activity is enjoined by this Court. Plaintiff and the proposed Class are therefore entitled to permanent injunctive relief barring Meta's ongoing infringement.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff seeks that this matter be certified as a class action, and that his attorneys be appointed Class Counsel and that he be appointed a Class Representative, and Plaintiff demands judgment against Defendant as follows:

1. Awarding Plaintiff and the proposed Class statutory damages or actual damages at their election (including restitution and disgorgement of profits), attorneys' fees and costs, and any other relief that may be permitted by law or equity pursuant to the claim(s) for relief;

2. Permanently enjoining Meta from engaging in the infringing conduct alleged herein;

3. Awarding Plaintiff and the proposed Class pre-judgment and post-judgment interest pursuant to the claim(s) for relief;

4. Awarding Plaintiff and the proposed Class costs, expenses, and attorneys' fees as permitted by law; and

5. Awarding Plaintiff and the proposed Class further relief as the Court may deem just and proper under the circumstances.

DEMAND FOR JURY TRIAL

Pursuant to Rule 38 of the Federal Rules of Civil Procedure, Plaintiff hereby demands a

jury trial for all claims so triable.

Dated: October 1, 2024

LIEFF CABRASER HEIMANN & BERNSTEIN, LLP

By: */s/ Daniel Hutchinson*
Daniel M. Hutchinson

Elizabeth J. Cabraser (State Bar No. 083151)
Daniel M. Hutchinson (State Bar No. 239458)
Reilly T. Stoler (State Bar No. 310761)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Telephone: (415) 956-1000
ecabraser@lchb.com
dhutchinson@lchb.com
rstoler@lchb.com

Rachel Geman (*pro hac vice forthcoming*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Telephone: (212) 355-9500
rgeman@lchb.com

Scott J. Sholder (*pro hac vice forthcoming*)
CeCe M. Cole (*pro hac vice forthcoming*)
COWAN DEBAETS ABRAHAMS & SHEPPARD LLP
60 Broad Street, 30th Floor
New York, New York 10004
Telephone: (212) 974-7474
ssholder@cdas.com
ccole@cdas.com

Attorneys for Plaintiff and the Proposed Class