BURSOR & FISHER, P.A.
L. Timothy Fisher (State Bar No. 191626)
1990 North California Blvd., 9th Floor
Walnut Creek, CA 94596
Telephone: (925) 300-4455
Facsimile: (925) 407-2700
E-mail: ltfisher@bursor.com

Attorneys for Plaintiff

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

ARTHUR KLEINER, individually and on behalf of all others similarly situated, Plaintiff, v. ADOBE INC., Defendant.	Case No. **CLASS ACTION COMPLAINT** JURY TRIAL DEMANDED

Plaintiff Arthur Kleiner ("Plaintiff") brings this action on behalf of himself and all others similarly situated against Defendant Adobe Inc. ("Defendant"). Plaintiff seeks to recover injunctive relief and damages as a result of Defendant's unlawful conduct. Plaintiff makes the following allegations pursuant to the investigation of his counsel and are based upon information and belief, except as to the allegations specifically pertaining to himself, which are based on personal knowledge.

NATURE OF THE ACTION

1. This case addresses the surreptitious, non-consensual use and collection of authors' books and written works by Defendant in order to train its SlimLM small language models (SLMs), which is AI software optimized for document assistance tasks on mobile devices. This use violates Adobe's terms of service at the expense of writers who are unknowingly contributing to training models for Defendant's SlimLM software.

2. SlimLM is a series of small language models created by Defendant that are designed and optimized for mobile deployment.[1] SlimLM, according to Defendant's own research, is pretrained on a dataset known as "SlimPajama-627B." It is then fine tuned on DocAssist, a specialized dataset constructed by a team at Adobe.[2]

3. SlimLM, as a language model, is trained by ingesting datasets that are chosen by its developers. It is well-known that developing AI tools necessitates the collection and use of millions of lines of text, including oftencopyrighted materials.[3]

4. The dataset used by Adobe to train its SlimLMs is called SlimPajama, which is a copied, cleaned and deduplicated version of the notorious RedPajama datatset. RedPajama contains the Books3 corpus, which contains hundreds of thousands of copyrighted books that were acquired without the authorization or consent of the authors. The SlimPajama dataset also contains content

[1] Adobe Research, *SlimLM: An Efficient Small Language Model for On-Device Document Assistance* (August 1, 2025).

[2] *Id*.

[3] Ars Technica, *OpenAI says it's "impossible" to create useful AI models without copyrighted material* (Jan. 9, 2024) https://arstechnica.com/information-technology/2024/01/openai-says-its-impossible-to-create-useful-ai-models-without-copyrighted-material/

from the Common Crawl dataset, which is also known to contain copyrighted and unauthorized material.[4]

5. Plaintiff and Class members are authors and copyright holders. They own registered copyrights in books that were included in the SlimPajama (via Common Crawl and RedPajama) datasets that Defendant downloaded, copied, stored, and used without their permission or compensation.

6. Plaintiff and Class members never authorized Defendant to download, copy, store, or use their copyrighted works. Defendant has never compensated Plaintiffs and Class members for downloading, copying, storing, or using their copyrighted works.

7. Adobe has and continues to benefit commercially from its massive acts of copyright infringement. It does so by securing lucrative contracts with enterprise customers for the use of its AI software, including through the SlimLM AI platform, and through deploying such AI software (or tools relying on SlimLM) through its suite of Adobe branded products.

8. Through the above acts, Defendant has infringed Plaintiff's copyrighted works and continues to do so by continuing to store, copy, use, and process the datasets containing copies of Plaintiff's and the putative Class's copyrighted books.

PARTIES

9. Plaintiff Arthur Kleiner is a citizen of New York and resident of Manhattan, New York. Plaintiff Kleiner is a published author of many books, including *The Age of Heretics: Heroes, Outlaws, and the Forerunners of Corporate Change*, which was published in 1996.

10. Plaintiff Kleiner registered his book with the United States Copyright Office in 1995, and has held the ownership of the copyright since.

11. Plaintiff Kleiner's book, *The Age of Heretics*, was included in the datasets that Defendant pirated, copied used, and transcribed to train its SlimLM models.

[4] Rights Alliance for the Creative Industries on the Internet, *Report on AI model providers' training data transparency and enforcement of copyrights*, https://rettighedsalliancen.com/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf (Sep. 5, 2024); *see also* The Atlantic, *The Company Quietly Funneling Paywalled Articles to AI Developers* (Nov. 4, 2025) https://www.theatlantic.com/technology/2025/11/common-crawl-ai-training-data/684567/

12. Defendant Adobe, Inc. is a Delaware company with its principal place of business at 345 Park Avenue, San Jose, CA 95110-2704. Adobe is one of the largest companies in the world, with a market value of over $156 billion as of June 2025.

JURISDICTION AND VENUE

13. This Court has subject matter Jurisdiction over this action pursuant to the Class Action Fairness Act ("CAFA"), 28 U.S.C. § 1332(d)(2) because this is a class action in which at least one member of the class is a citizen of a state different from any Defendants, the amount in controversy exceeds $5 million, exclusive of interest and costs, and the proposed class contains more than 100 members. Additionally, this Court has subject-matter jurisdiction under 28 U.S.C. § 1331 because this case arises under the Copyright Act (17 U.S.C. § 501).

14. This Court has personal jurisdiction over the Defendant because Defendant maintains its principal places of business in this District and because a substantial part of the events or omissions giving rise to the claims asserted herein occurred in this District.

15. Venue is proper in this district pursuant to 28 U.S.C. § 1391 because a substantial part of the events or omissions giving rise to the claims asserted herein occurred in this District and because Defendant maintains its principal places of business in this District.

16. Under Civil Local Rule 3-2(c), assignment of this case to the San Jose Division is proper because this case pertains to intellectual-property rights, which is a district-wide case category under General Order No. 44, and therefore venue is proper in any courthouse in this District.

FACTUAL ALLEGATIONS

A. SlimPajama, an Altered Version of "RedPajama," Contains Plaintiff's and Class Members' Works.

17. In October 2020, the Books3 dataset, consisting of approximately 196,640 books sourced from the online pirate "shadow library" Bibliotik, was made available online to provide AI developers with access to high-quality training data. The Books3 dataset included Plaintiff's copyrighted books without permission in violation of the U.S. Copyright Act.

18. The Pile is a dataset curated by a research organization called EleutherAI for use in training AI models. In December 2020, EleutherAI introduced this dataset in a paper called "The Pile: An 800GB Dataset of Diverse Text for Language Modeling."[5] The paper provides a description of the Books3 dataset contained within The Pile:

> 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.[6]

19. Bibliotik is one of a number of notorious pirate websites that also includes Library Genesis (aka LibGen), Z-Library (aka B-ok), Sci-Hub, and Anna's Archive. These illegally sourced libraries have long been of interest to the AI-training community because they contain vast quantities of unauthorized copyrighted material, including books, which are required to train LLM's (large language models) and SLMs (small language models). Using such illegal libraries permits AI companies like Defendant to gain quick access to large quantities of training data rapidly. For that reason, these pirate libraries also violate the U.S. Copyright Act.

20. The person who assembled the Books3 dataset, Shawn Presser, has confirmed in public statements that Books3 represents "all of Bibliotik" and contains approximately 196,640 books.

21. EleutherAI's website encourages the public to download The Pile dataset from a website known as "The-Eye." Anyone who downloads The Pile from The-Eye is therefore also downloading a copy of Books3.

22. The Pile is also available for download from the "Hugging Face" website. Before October 2023, the Books3 subset of The Pile was available for download from Hugging Face as a standalone dataset. But in October 2023, the Books3 dataset was removed with a message that it "is defunct and no longer accessible due to reported copyright infringement."[7]

---

[5] https://arxiv.org/pdf/2101.00027.pdf

[6] *Id.* at 3–4.

[7] https://web.archive.org/web/20231127101818/https://huggingface.co/datasets/the_pile_books3

23. Books3 has also been included in its entirety within another dataset known as RedPajama, which was created by the company Together AI. Released in or around April 2023, the RedPajama dataset contained a subset called "Books" or "RedPajama-Books" that was a direct copy of the "Books3 dataset." The RedPajama dataset is available for download from Hugging Face. Before October 2023, the Books3 subset was "downloaded from Huggingface [sic]" when a user ran the script that automatically assembled the RedPajama dataset.[8] After the "Books3 dataset" was removed from Hugging Face in October 2023, the RedPajama dataset documentation similarly added a message that Books3 is defunct "due to reported copyright infringement."[9]

24. But before October 2023, anyone who downloaded the RedPajama or The Pile datasets from Hugging Face was necessarily downloading a copy of the Books3 dataset.

25. SlimPajama is a "deduplicated, multi-corpora" dataset that was created by "cleaning and deduplicating the 1.21T token RedPajama dataset."[10]

26. In this context, "cleaning" data does not mean removing copyrighted or protected works. It means removing errors, corrupted data, or duplicated/repeated data, with the result of making a dataset more accurate and streamlined.

27. This means that the SlimPajama dataset contains copyrighted material that has not been licensed and authorized by its creators.

---

[8] https://web.archive.org/web/20230420075601/https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T

[9] https://web.archive.org/web/20240510231649/https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T

[10] Cerebras.ai, *SlimPajama: A 627B token, cleaned and deduplicated version of RedPajam*, https://www.cerebras.ai/blog/slimpajama-a-627b-token-cleaned-and-deduplicated-version-of-redpajama (Jun. 9, 2023).

28. The SlimPajama dataset is also based in part on—or overlaps with—the Common Crawl dataset,[11] which is known to contain copyrighted material, including "articles published on news websites and in digital newspapers."[12]

29. Plaintiff's copyrighted book, listed in **Exhibit A**, is among the works in the Books3 dataset—and therefore the RedPajama, and SlimPajama— dataset. Below, this book is referred to as the **Infringed Work**.

B. AI Models Require Datasets.

30. A large language model ("LLM") is Artificial Intelligence software designed to emit convincingly naturalistic text outputs in response to user prompts. Small language models ("SLM") are similar to LLMs, but trained "on data from specific domains," meaning their training utilizes smaller or more targeted sets of data.[13]

31. Though SLMs and LLMs are software programs, they are not created the way most software programs are—that is, by human software programmers writing code. Rather, an SLM is *trained by* copying extremely large quantities of textual works and then feeding these copies into the model. This corpus of input material is called the *training dataset*.

32. Training consists of a multi-stage process (known as the training pipeline) that includes the acquisition and curation of the dataset, processing of the dataset, feeding the dataset into the model so that the model can extract the patterns and relationships from the protected expression contained therein; and further fine-tuning the model for more specialized uses with even more data. This process also involves experimentation to improve the data mixture (i.e., the final

---

[11] Rights Alliance for the Creative Industries on the Internet, *Report on AI model providers' training data transparency and enforcement of copyrights*, https://rettighedsalliancen.com/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf (Sep. 5, 2024); *see also* The Atlantic, *The Company Quietly Funneling Paywalled Articles to AI Developers* (Nov. 4, 2025); *see also SlimPajama-DC: Understanding Data Combinations for LLM Training*, https://ar5iv.labs.arxiv.org/html/2309.10818#S3.SS1

[12] *Netherlands: Common Crawl removes 2 million scraped news articles, complying with request*, *Piracy Mon*https://piracymonitor.org/netherlands-common-crawl-removes-2-million-scraped-news-articles-complying-with-request/

[13] *LLMs vs. SLMs: The Differences in Large & Small Language Models*, Splunk, https://www.splunk.com/en_us/blog/learn/language-models-slm-vs-llm.html (Feb. 17, 2025).

training dataset and the proportion of each component dataset) of the model. Each experiment is also known as an "ablation" in technical parlance.

33. The first step in training the model is acquiring and curating the data that goes in to the model. This acquisition necessarily includes copying and downloading data, usually onto persistent storage. This is because training an SLM or any other generative AI model is expensive and resource intensive, so retaining persistent copies of datasets avoids the need to reacquire the data when developing new models, or altering the data mixtures during the development process.

34. Training an SLM is not only a function of quantity of data, but also of quality. The selection and curation of training data is therefore an important first step in training. Copyrighted books are well known among developers of AI models to be high-quality data for training SLMs.

35. After the initial copying and processing of data, the SLM copies each textual work in the training dataset and extracts protected expression from it. During what is known as *pretraining*, the SLM progressively adjusts its output to more closely approximate the protected expression copied from the training dataset. The SLM records the results of this process in a large set of numbers called weights (also known as *parameters*) that are stored within the model. These weights are entirely and uniquely derived from the protected expression in the training dataset. Once a model is pretrained, the result is a trained model known as a *base* or *foundational* model.

36. During the development process of an SLM, engineers may also conduct experiments known as ablations or "ablation studies" that test the effect of certain data on the model. This can include, for example, determining whether there is a difference in the quality of a model's output if it is trained with certain books or without. A dataset may be used to run such experiments but ultimately be excluded from the final training mixture of the model. Importantly, these datasets used for ablation studies often also consist of copyrighted works, including books.

37. Once an SLM has copied the textual works in the training dataset and extracted the protected expression into stored weights, an SLM is able to emit convincing simulations of natural written language in response to user prompts. Whenever an SLM generates text output in response to a user prompt, it is performing a computation that relies on these stored weights which represent the protected expression ingested from the training dataset.

38. Throughout each step of the training pipeline, the same dataset is typically used (i.e., copied) multiple times. Indeed, given the cost of developing an SLM, it is a ubiquitous practice to retain datasets for future use, whether that use is to pretrain future models, to perform ablations on a model, or to fine-tune an already trained base model. Each step involves making additional copies of the underlying data. The implication is that if a dataset contains unlawfully obtained copyrighted material, each step of the training pipeline may result in an unauthorized use (i.e., infringement) of that copyrighted work.

C. Adobe Used Datasets Known to Contain Pirated Books to Develop its Small Language Models

39. Adobe is a computer software company that "empowers everyone, everywhere to imagine, create, and bring any digital experience to life. From creators and students to small businesses, global enterprises, and nonprofit organizations — customers choose Adobe products to ideate, collaborate, be more productive, drive business growth, and build remarkable experiences."[14]

40. Adobe's software includes the popular PDF editing suite known as Adobe Acrobat, the image editor software known as Photoshop, the video editor software known as Adobe Premiere, and the graphic design tool known as Adobe Illustrator.

41. As part of Adobe's improvement of its "consumer technology" products, Defendant has developed SlimLM, "a series of small language models specifically designed and optimized for mobile deployment."[15]

42. Adobe says that its SlimLM SLM is pretrained on SlimPajama-627B. As detailed above, SlimPajama is known to contain the contents of RedPajama, which in turns contains part of the Books3 corpus, comprised of thousands of copyrighted works.

[14] Adobe, *Changing the world through personalized digital experiences*, https://www.adobe.com/about-adobe.html (last accessed Jan. 22, 2026).

[15] Adobe Research, *SlimLM: An Efficient Small Language Model for On-Device Document Assistance* (August 1, 2025).

43. SlimPajama is also known to contain parts of, if not all, of the Common Crawl dataset. Specifically, SlimPajama's composition (by token percentage) is 52.2% Common Crawl.[16]

44. The Common Crawl dataset was created and curated by Common Crawl. "Common Crawl is a non-profit organization that crawls the internet to provide an extensive archive or dataset to the public. Tech companies use this dataset as training data for their generative AI models, including Apple's openELM, Microsoft's Phi, OpenAI's ChatGPT, NVIDIA's Nemo Megatron, Deepseek's Deepseek V3, and Anthropic's Claude."[17]

45. BREIN, a Dutch non-profit foundation, asked the creators of Common Crawl to remove over "two million news articles belonging to popular Dutch news outlets from its AI training dataset." According to BREIN, "these articles were copied without permission."[18]

46. Common Crawl is known to contain other copyrighted works beyond the millions of unlicensed Dutch news articles.[19]

47. In other words, Defendant's SlimLM is trained on potentially multiple datasets known to contain copyrighted works: SlimPajama and Common Crawl.

48. This runs contrary to Adobe's outspoken stance on ethical AI development and use.

49. In numerous places on its website, Adobe touts its Adobe Firefly product, an AI image generator that was built through "thoughtful, responsible development." Adobe claims it "developed Adobe Firefly to prevent it from creating content that infringes copyright or intellectual property rights, and it is designed to be commercially safe," and that they "do not mine content from the web to train Adobe Firefly."[20]

---

[16] Emergent Mind, *SlimPajama Dataset*, https://www.emergentmind.com/topics/slimpajama-dataset

[17] CyberNews, *Common Crawl removes AI dataset containing over 2M news articles*, https://cybernews.com/ai-news/common-crawl-ai-dataset-2m-news-articles/ (Nov. 5, 2025).

[18] *Id*.

[19] The Business Standard, *New York Times successfully removes copyrighted content from AI training dataset*, https://www.tbsnews.net/tech/new-york-times-successfully-removes-copyrighted-content-ai-training-dataset-741266 (Nov. 18, 2023).

[20] https://www.adobe.com/ai/overview/ethics.html

50. Despite Adobe's claims of ethical AI development for its Firefly product, Adobe has nevertheless celebrated the development of its SlimLM SLM, which was trained on multiple datasets of copyrighted material, including Plaintiff's. At no point in Adobe's paper describing its SlimLM model did it discuss compensation for the writers and creators of the works on which SlimLM was trained.[21]

51. Adobe infringed on Plaintiff and Class members' copyrighted works on a massive scale. Adobe downloaded these books from the SlimPajama dataset—which includes copyrighted works via RedPajama and Common Crawl—without authorization from or compensation to their authors. Adobe then continued copying and storing the datasets, and used them to develop and train the SlimLM series of SLMs.

52. Adobe stored and used Plaintiff's and Class members' copyrighted works to train other models, including non-public models, whether sourced from SlimPajama, Common Crawl or other pirated and/or unauthorized, copyrighted sources.

53. Plaintiff and Class members have retained ownership rights in their works. Plaintiff and Class members did not consent to the use of their works as training material for SlimLM. Nonetheless, their works were downloaded, copied, stored, and used for training SlimLM to be a critical support system for current or future Adobe products.

D. Adobe's Use of Copyrighted Works Is Not Fair Use.

54. Adobe's downloading, copying, storage, and use of Plaintiff's and Class members' works was not fair use. Adobe could have—but chose not to—lawfully obtain the Infringed Work. It is near impossible that "any accused infringer could ever meet its burden of explaining why downloading source copies from pirate sites *that it could have purchased or otherwise accessed lawfully* was itself reasonably necessary to any subsequent fair use." *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025). Adobe used the SlimPajama dataset (comprised of data from the RedPajama and Common Crawl datasets), which includes materials sourced from shadow libraries known to indiscriminately hoover up countless copyrighted works, for its centralized

[21] Adobe Research, *SlimLM: An Efficient Small Language Model for On-Device Document Assistance* (August 1, 2025)

commercial database, all with complete disregard to the rights of copyright holders. Adobe used these datasets containing Plaintiff's and Class members' copyrighted works as a substitute for purchasing or compensating creators for authorized copies of those works. In doing so, Adobe displaced or diluted the market for Plaintiffs' and Class members' works. Plaintiff's works were copied and maintained for future commercial purposes, including to further develop its SlimLM series of SLMs. Pirating otherwise purchasable works is copyright infringement.

55. The exploitation of Plaintiff's Infringed Work was not indirect. It was and is a direct and illegal download, use and scraping of the entirety of Plaintiff's' Infringed Work (as part of the SlimPajama dataset) with no transformation of form. Plaintiff's Infringed Work was copied, used and maintained for future commercial purposes. Defendant's use was not transitory as Plaintiff's Infringed Work was not immediately destroyed, but was retained to train Adobe's SLMs. Integration of this technology into Adobe's existing document editing and creation suite could generate billions of dollars for Adobe. Adobe's "piracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for [a] transformative use and immediately discarded." *Id.* at 1025.

56. Adobe's use of these copyrighted works to train its SlimLM SLMs are not transformative. SLM outputs are the result of statistical outputs based on the training corpora. Broadly, SLMs operate by probabilistically predicting the next "token." Tokens are units of language which can be words, combinations of letters, or even a punctuation or space. Adobe's SlimLM models are no different. Recent studies demonstrate that the LLMs created by the leading AI developers are capable of regurgitating substantial portions of the training data for those models, even with the implementation of safeguards designed to prevent the regurgitation of training data.[22]

57. Even if the court applies the statutory fair use factors to Adobe's unmitigated copying through downloading datasets including SlimPajama, and its use of those datasets to train

[22] Ahmed Ahmed et. a., "Extracting Books from Production Language Models" (Jan. 6, 2026), available at https://arxiv.org/pdf/2601.02671. This study analyzed the output of LLMs created by Anthropic, Google, OpenAI, and xAI—models which use the same or similar architecture to the xGen models.

its SLMs, the application of the factors weighs against fair use. The four factors of fair use are: (1) "the purpose and character of the use," (2) "the nature of the copyrighted work," (3) "the amount and substantiality of the portion used," and (4) "the effect of the use upon [Plaintiff's] potential market." 17 U.S.C. § 107.

58. **Factor one.** The purpose and character of Adobe's use is strictly commercial and not transformative. The factor one analysis includes (1) whether the purpose of copying is commercial or non-commercial and (2) whether the purpose is transformative. Adobe's use was and is not for educational or commentary purposes, but furthered a commercial endeavor to utilize a massive library of valuable copyrighted works that helped train and develop financially lucrative Adobe products. Further, there is nothing transformative about using unauthorized copies of Plaintiff's Infringed Work to train Adobe's SlimLM models in place of sourcing free market alternatives. Piracy is not transformative and is not fair use.

59. **Factor two.** The factor two analysis primarily questions whether the nature of the copied work is more creative or factual, and secondarily questions whether the copied works are accessible or out of print. Here, Plaintiff's work is highly creative and is therefore entitled to broader protections under copyright. Further, inaccessible works are provided less protection because there is no market for inaccessible works that a defendant's copying could displace. However, "this is not a case where source copies were unavailable," *Anthropic PBC*, 787 F. Supp. 3d at 1027, and Adobe chose not to purchase or license this copyrighted material – even though that option was readily available. The nature of Plaintiff's Infringed Work ensures Plaintiff the greatest copyright protection and narrows Defendant's argument for fair use.

60. **Factor three.** The third factor asks whether a qualitatively and/or quantitatively substantial portion of Plaintiff's Infringed Work was stolen. The entirety of Plaintiff's Infringed Work was compiled in the SlimPajama dataset, and copied and stored for Adobe's use on an ongoing basis. These works were not copied in a modified form (for example, as an image thumbnail) into the dataset, such that the substance was not extracted. Plaintiff's Infringed Work was copied in its entirety to take both the whole and the heart of the work. The third factor weighs strongly in favor of infringement.

61. **Factor four.** The harm to Plaintiff's market is clear as "[t]he copies used to build a central [training] library *and* that were obtained from pirated sources plainly displaced demand for Authors' books — copy for copy." *Id.* at 1033. Factor four considers whether Adobe's actions displaced or diluted the market for Plaintiff's and the Class Members' Infringed Works. More accurately, market displacement considers both whether the infringer's conduct itself causes harm or whether widespread replication of the infringer's conduct would harm Plaintiff's market. At every turn, Adobe's conduct injures Plaintiff. Adobe injures Plaintiff's market when downloading a pirated library or corpus of unlicensed data, and when training its various SLMs.

62. Adobe's unauthorized use of the copyrighted content in SlimPajama—via RedPajama and Common Crawl—displaced Plaintiff's market through Defendant's own piracy. Widespread piracy inevitably leads to market harm as it is difficult for creators that charge for their work, *i.e.*, through book sales and licensing, to compete with free versions of their works. Moreover, while merely pointing towards a theoretical licensing market is insufficient, Adobe has demonstrated that there is a thriving licensing market for Plaintiff's Infringed Work—and that Adobe itself knows the value of that market—as Adobe has compensated some copyright holders whose work it chose not to infringe upon.[23] Adobe's individual actions have injured Plaintiffs' market and widespread adoption of Adobe's tactics will only exacerbate this harm.

63. The purpose of copyright, since the beginning of this country, is to incentivize original creativity. Adobe's displacement of the market for the works of Plaintiff and the Class severely disincentivizes their future creativity. Therefore, considering copyright's historic purpose, Adobe's market displacement requires an adjudication that Adobe's conduct is not fair use.

CLASS ALLEGATIONS

64. Plaintiff seeks to represent a class defined as all legal or beneficial owners of copyrighted works registered with the United States Copyrighted Office that Adobe downloaded, copied, stored, or otherwise used without authorization to develop any of Adobe's products during the Class Period (the "Nationwide Class").

[23] Yahoo Finance, *Adobe Is Buying Videos for $3 Per Minute to Build AI Model* https://finance.yahoo.com/news/adobe-buying-videos-3-per-213215719.html (Apr. 10, 2024).

65. Specifically excluded from the Class is Defendant, Defendant's officers, directors, agents, trustees, parents, children, corporations, trusts, representatives, employees, principals, servants, partners, joint ventures, or entities controlled by Defendant, and its heirs, successors, assigns, or other persons or entities related to or affiliated with Defendant and/or Defendant's officers and/or directors, the judge assigned to this action, and any member of the judge's immediate family.

66. Plaintiff reserves the right to expand, limit, modify, or amend the class definition, including the addition of one or more subclasses, in connection with his motion for class certification, or at any other time, based on, *inter alia*, changing circumstances and/or new facts obtained.

67. **Numerosity**. On information and belief, hundreds of thousands of copyright holdersfall into the definition of the Class. Members of the Class can be identified through Defendant's records, discovery, and third-party sources.

68. **Commonality and Predominance**. Common questions of law and fact exist as to all members of the Class and predominate over any questions affecting only individual members of the Class. These common legal and factual questions include, but are not limited to, the following:

(a) Whether Defendant violated the copyrights of Plaintiff and the Class when it obtained copies of Plaintiff's Infringed Work, stored the Infringed Works, copied the Infringed Works, and used the Infringed Works to develop Defendant's SLMs;

(b) Whether any affirmative defense excuses Defendant's conduct; and

(c) Whether any statutes of limitation limit the potential for recovery for Plaintiffs and the Class.

69. **Typicality.** Plaintiff's claims are typical of the claims of the other members of the Class in that, among other things, all Class members were similarly situated and were comparably injured through Defendant's wrongful conduct as set forth herein. Further, there are no defenses available to Defendant that are unique to Plaintiff.

70. **Adequacy of Representation**. Plaintiff will fairly and adequately protect the interests of the Class. Plaintiff has retained counsel that is highly experienced in complex consumer class action litigation, and Plaintiff intends to vigorously prosecute this action on behalf of the Class. Furthermore, Plaintiff has no interests that are antagonistic to those of the Class.

71. **Superiority.** A class action is superior to all other available means for the fair and efficient adjudication of this controversy. The damages or other financial detriment suffered by individual Class members are relatively small compared to the burden and expense of individual litigation of their claims against Defendant. It would thus be virtually impossible for the Class to obtain effective redress for the wrongs committed against the members on an individual basis. Furthermore, even if Class members could afford such individualized litigation, the court system could not. Individualized litigation would create the danger of inconsistent or contradictory judgments arising from the same set of facts. Individualized litigation would also increase the delay and expense to all parties and the court system from the issues raised by this action. By contrast, the class action device provides the benefits of adjudication of these issues in a single proceeding, economies of scale, and comprehensive supervision by a single court, and presents no unusual management difficulties under the circumstances.

72. Further, Defendant has acted and refused to act on grounds generally applicable to the proposed Class, thereby making appropriate final injunctive and declaratory relief with respect to the Class as a whole.

CAUSES OF ACTION

COUNT I
Direct Copyright Infringement (17 U.S.C. § 501) against Defendant
(On Behalf of Plaintiff and the Class)

73. Plaintiff incorporates by reference and re-alleges each and every allegation set forth above as though fully set forth herein.

74. As the owner of the registered copyrights in the Infringed Work, Plaintiff holds the exclusive rights to those books under 17 U.S.C. § 106.

75. Adobe downloaded, ingested, or otherwise acquired copies of the SlimPajama dataset containing, in part, Common Crawl and Redpajama, which includes the Infringed Work. The initial downloading constitutes the first unauthorized copying of Plaintiff's works by Adobe in the training pipeline.

76. Upon information and belief, Adobe created and stored numerous copies of SlimPajama in its internal servers.

77. Upon information and belief, to develop the SlimLM SLMs, Adobe copied the SlimPajama dataset containing Common Crawl and RedPajama (comprised in part by Books3) to develop these models and incorporate the dataset (and various iterations thereof) into the models' training data.

78. Plaintiff and the Class members never authorized Defendant to make copies of their Infringed Works, make derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works). All those rights belong exclusively to Plaintiff and the Class members under the U.S. Copyright Act.

79. By copying, storing, processing, reproducing, and using the dataset containing pirated copies of Plaintiff's Infringed Works, Defendant has directly infringed Plaintiff's exclusive rights in their copyrighted works.

80. By copying, storing, processing, and reproducing the SlimLM models trained on Plaintiff's Infringed Work, Adobe has directly infringed Plaintiff's exclusive rights in their copyrighted works.

81. Defendant repeatedly copied, stored, and used the Infringed Work without Plaintiff's permission. Defendant made these copies without Plaintiff's permission and in violation of their exclusive rights under the Copyright Act.

82. Because the complete training mixture for the SlimLM models has not been made public, other unauthorized repositories of pirated copies of Plaintiff's and Class members' copyrighted works may have been used to develop Adobe's SlimLM and other models.

83. Defendant's infringing conduct alleged herein was and continues to be willful and carried out with full knowledge of Plaintiff's rights in the copyrighted works. As a direct result of their conduct, Defendant has wrongfully profited from copyrighted works that they do not own.

84. By and through the actions alleged above, Defendant has infringed and will continue to infringe Plaintiff's copyrights.

85. Plaintiff has been and will continue to be injured by Defendant's acts of direct copyright infringement. Adobe's infringement has directly resulted and is directly causing ongoing harm, including, but not limited to (1) loss of value to Plaintiff's and Class Members' works due to Adobe's mass infringement that has replaced and depressed the overall market for the Plaintiff's and Class Members' respective commercial markets, and (2) lost licensing revenue that Plaintiff and Class members would have received had Adobe properly obtained authorization, consent, or license to acquire and use their works.

86. Plaintiff is entitled to statutory damages, actual damages, restitution of profits, and all appropriate legal and equitable relief.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff, individually and on behalf of the all others similarly situated seeks judgment against Defendant as follows:

(a) For an order certifying the Class under Rule 23 of the Federal Rules of Civil Procedure and naming Plaintiff as the representative for the Class and Plaintiff's attorneys as Class Counsel;

(b) Judgment in favor of Plaintiff and the Class and against Defendant.

(c) An award of statutory and other damages under 17 U.S.C. § 504 for violations of the copyrights of Plaintiff and the Class by Defendant.

(d) Reasonable attorneys' fees and reimbursement of costs under 17 U.S.C. § 505 or otherwise.

(e) A declaration that such infringement is willful.

(f) Destruction or other reasonable disposition of all copies Defendant made or used in violation of the exclusive rights of Plaintiff and the Class, under 17 U.S.C. § 503(b).

(g) Pre- and post-judgment interest on the damages awarded to Plaintiff and the Class, and that such interest be awarded at the highest legal rate from and after the date this Complaint is first served on Defendant.

(h) For an order establishing that Defendant is responsible financially for the costs and expenses of a Court-approved notice program through post and media designed to give immediate notification to the Class.

(i) For injunctive relief as the Court may deem proper; and

(j) Further relief for Plaintiff and the Class as may be appropriate.

JURY TRIAL DEMANDED

Plaintiff demands a trial by jury on all claims so triable.

Dated: February 9, 2026

BURSOR & FISHER, P.A.

By: */s/ L. Timothy Fisher*

L. Timothy Fisher (State Bar No. 191626)
1990 North California Blvd., 9th Floor
Walnut Creek, CA 94596
Telephone: (925) 300-4455
Facsimile: (925) 407-2700
E-mail: ltfisher@bursor.com

Attorneys for Plaintiff

EXHIBIT A

Copyright

Registration Number / Date:
TX0004280899 / 1996-05-31

Type of Work:
Text

Title:
The age of heretics :heroes, outlaws, and the forerunners of corporate change /Art Kleiner.

Date of Creation:
1995

Date of Publication:
1996-05-01

Copyright Claimant:
Art Kleiner

Description:
414 p.

Imprint:
New York : Currency, c1996.

Edition:
1st ed.

Names:
Kleiner, Art

USCO Catalog Link:
https://publicrecords.copyright.gov/detailed-record/voyager_15216775