Eugene Y. Turin (SB # 342413)
McGuire Law, P.C.
1089 Willowcreek Road, Suite 200
San Diego, CA 92131
Tel: (312) 893-7002 Ex. 3
Fax: 312-275-7895
eturin@mcgpc.com

Counsel for Plaintiff and the Putative Class Members

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

ELIZABETH LYON, individually and on behalf of similarly situated individuals, Plaintiff, v. ADOBE INC., a Delaware corporation, Defendant.	Case No. **COMPLAINT** **CLASS ACTION** **DEMAND FOR JURY TRIAL**

Plaintiff Elizabeth Lyon ("Plaintiff"), on behalf of herself and all others similarly situated, bring this class action complaint ("Complaint") against Defendant Adobe Inc. ("Adobe" or "Defendant").

OVERVIEW

1. Artificial intelligence ("AI") refers to software engineered to mimic human-like reasoning and inference through algorithmic processes, typically leveraging statistical methods.

2. Small language models ("SLMs") are AI software programs designed to reply to user prompts with natural-sounding text outputs. In contrast to large language models ("LLM's") which often run processing remotely with results transmitted via internet to devices, SLMs are intentionally designed to run efficiently on devices with limited hardware resources, such as smartphones, tablets, and laptops. Adobe's SlimLM is a series of SLM models trained, created, and the released by Adobe that are optimized for document assistance tasks on mobile devices.

3. While the traditional coding process involves human coders inputting explicit instructions, language models are instead trained by processing vast quantities of text from diverse sources (a "pre-training dataset"), learning statistical patterns and associations within that data, and encoding those abstract representations into a vast array of numerical values known as parameters. The goal is to enable the model to learn general language patterns, grammar, factual knowledge, and contextual relationships. When done competently such training results in a versatile base model that can understand and generate human-like text.

4. Creating a high-quality training dataset involves copying an enormous quantity of textual works. Each book or other text in the dataset must be downloaded (or purchased, scanned and OCR'd) copied, stored, and processed (often multiple times) in order to be tokenized, filtered, deduplicated, and ingested in a large-scale pre-training process.

5. The pre-training dataset used by Defendant to train its SlimLM models is called SlimPajama. SlimPajama is a cleaned (deduplicated) version of the RedPajama dataset which is comprised of public domain, licensed, and, crucially, unlicensed copyrighted materials.

6. Plaintiff and Class members are authors. They own registered copyrights in certain books (the "Infringed Works") that were included in the SlimPajama pre-training dataset that Adobe pirated, copied, and used to train its SlimLM models. Plaintiff and Class members never authorized Adobe to download, copy, store, and use their copyrighted works as pre-training materials. Adobe copied, and thus infringed on, these Infringed Works multiple times to train its SlimLM models.

7. Through the acts described in further detail below, Defendant has infringed on Plaintiff's copyrighted works and continues to do so by continuing to store, copy, use, and process the training datasets containing copies of Plaintiff's and the putative Class's Infringed Works.

JURISDICTION AND VENUE

8. 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).

9. Jurisdiction and venue are proper in this judicial district under 28 U.S.C. § 1391(c)(2) because Defendant is headquartered in this District. Defendant copied Plaintiff's and Class members' Infringed Works to train its SlimLM models. Therefore, a substantial part of the events giving rise to the claim occurred in this District. A substantial portion of the affected interstate trade and commerce was carried out in this District. Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendant's conduct has had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in

this District.

10. 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.

PARTIES

11. Plaintiff Elizabeth Lyon is an author who resides in Oregon. She is the author of numerous titles including *Mabel: The Story of One Midwife*, *Nonfiction Book Proposals Anybody Can Write*, *A Writer's Guide to Nonfiction*, and *Manuscript Makeover: Revision Techniques No Fiction Writer Can Afford to Ignore*. Plaintiff does not have an Adobe account.

12. One of Plaintiff's registered copyrights is attached hereto as Exhibit A.

13. Defendant is a Delaware corporation with its principal place of business at 345 Park Avenue, San Jose, CA 95110-2704.

14. The unlawful acts alleged against the Defendant in this Complaint were authorized, ordered, or performed by the Defendant's respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendant's business or affairs. The Defendant's agents operated under the explicit and apparent authority of their principals. Defendant, and its subsidiaries, affiliates, and agents operated as a single unified entity.

15. Various persons or firms not named as defendants may have participated as co-conspirators in the violations alleged herein and may have performed acts and made statements in furtherance thereof. Each acted as the principal, agent, or joint venture of, or for Defendant with respect to the acts, violations, and common course of conduct alleged herein.

FACTUAL ALLEGATIONS

16. Adobe is a computer software company that offers a wide range of programs including web design tools, PDF viewers and editors, photo manipulation, audio/video editing, and importantly AI services.

17. Adobe's SlimLM models are SLMs for on-device document assistance. Adobe has promoted SlimLM as a part of its AI offerings and as a model intended for integration into mobile devices.

18. Like LLMs, SLMs are trained by ingesting massive training corpora consisting of extremely large volumes of text – often millions or billions of lines. Constructing these corpora is accomplished by acquiring and digitally copying copyrighted works and storing those copies, oftentimes in multiple locations and formats, to support preprocessing, deduplication, tokenization, and training. Sometimes these digital copies are acquired legally, other times, as alleged here, copyrighted works are illegally pirated from the internet.

19. During pre-training, the SLM processes each textual work in the training dataset to learn statistical patterns and associations within it. The SLM adjusts its parameters and weights through optimization techniques to get progressively better at predicting sequences in the data, capturing general linguistic structures rather than specific expressions. The results of this learning process are encoded in a large set of numbers called parameters stored within the model. These parameters are derived from the entire pre-training dataset.

20. The SlimPajama dataset is a training dataset assembled and published by Cerebras Systems Inc. ("Cerebras") that is directly derived from Together Computer, Inc.'s (TogetherAI") RedPajama dataset. Notably, Cerebras describes its SlimPajama dataset as a copied, cleaned, and deduplicated version of Together AI's RedPajama dataset.

21. The RedPajama dataset contained a subset called "Books" or

"RedPajama-Books" that was actually a copy of the Books3 dataset. Specifically, the RedPajama dataset "is a publicly available, fully open, best-effort reproduction of the training data. . . used to train the first iteration of LLaMA family of models." This LLaMA training dataset included the Books3 section of The Pile (a broader publicly available dataset which contained Books3). Books3 was described in a paper by EluetherAI called "*The Pile: An 800GB Dataset of Diverse Text for Language Modeling*" as follows:

> 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.

22. The SlimPajama dataset was created by copying and manipulating the RedPajama dataset (including copying Books3).

23. Thus, because it is a derivative copy of the RedPajama dataset, SlimPajama contains the Books3 dataset, including the copyrighted works of Plaintiff and the Class members. In fact, members of Microsoft's AI team have expressed concern in releasing their own AI models trained on SlimPajama because the dataset contains Books3 and thus copyrighted works.

24. However, Adobe has confirmed that its SlimLM models are pre-trained on the SlimPajama dataset.

25. Thus, in order to train its SlimLM models, Adobe downloaded, copied, stored, and used the SlimPajama dataset that contained Books3 and Plaintiff's Infringed Works. Adobe also repeatedly downloaded, copied, and processed those works during the preprocessing and pretraining of the models.

26. Adobe retained copies of those pretraining datasets that contained copies of the Infringed Works on its servers and continues to store and use them in further

training for new versions of its SlimLM models and other related models at a minimum through retaining the model parameters of its originally trained model.

27. Thus, Defendant directly infringed on Plaintiff's and Class members' copyrighted works on a massive scale. Adobe downloaded and copied copyrighted works and the Infringed Works as contained in the SlimPajama dataset without authorization from, or after providing compensation to, their authors. Adobe then continued copying and storing the datasets and used them to develop and train its SlimLM models and other related models – including retaining such datasets and copyrighted works for undisclosed future uses.

CLASS ALLEGATIONS

28. The "Class Period" as defined in this Complaint begins on at least December 16, 2022, and runs through the present. Because Plaintiff does not yet know when the unlawful conduct alleged herein began, but believes, on information and belief, that the conduct likely began prior to December 16, 2022, Plaintiff reserves the right to amend the Class Period to comport with the facts and evidence uncovered during further investigation or through discovery.

29. Plaintiff seeks certification of the following Class pursuant to Federal Rules of Civil Procedure 23(a), 23(b)(2), and 23(b)(3):

> All persons or entities domiciled in the United States that own a United States copyright in any work that was downloaded, copied, stored, or used as training data by Defendant without authorization during the Class Period according to Defendant's records.

30. Plaintiff will fairly and adequately represent and protect the interests of the other members of the Class. Plaintiff has retained counsel with substantial experience in prosecuting complex litigation and class actions. Plaintiff and her counsel are committed to vigorously prosecuting this action on behalf of the other members of the Class, and have the financial resources to do so. Neither Plaintiff nor

her counsel have any interest adverse to those of the other members of the Class.

31. Absent a class action, most members of the Class would find the cost of litigation their claims to be prohibitive and would have no effective remedy. The class treatment of common questions of law and fact is also superior to multiple individual actions or piecemeal litigation in that it conserves the resources of the courts and the litigants and promotes consistency and efficiency of adjudication.

32. Defendant has acted and failed to act on grounds generally applicable to Plaintiff and the other members of the Class, requiring the Court's imposition of uniform relief to ensure compatible standards of conduct toward the members of the Class, and making injunctive or corresponding declaratory relief appropriate for the Class as a whole.

33. The factual and legal bases of Defendant's liability to Plaintiff and to the other members of the Class are the same, resulting in injury to Plaintiff and to all of the other members of the Class. Plaintiff and the other members of the Class have all suffered harm and damages as a result of Defendant's unlawful and wrongful conduct.

34. There are many questions of law and fact common to the claims of Plaintiff and the other members of the 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 limited to, the following:

a. Whether Defendant violated the copyrights of Plaintiff and the Class by obtaining and creating copies of Plaintiff's Infringed Works with the intent to use the Infringed Works for commercial benefit;

b. Whether Defendant did use the Infringed Works of Plaintiff and the Class for commercial benefit;

c. Whether Defendant violated the copyrights of Plaintiff and the Class by using illicitly obtained copies of Plaintiff's Infringed Works to train Defendant's AI models.

FIRST CAUSE OF ACTION
Direct Copyright Infringement,
(17 U.S.C. § 501)
(On Behalf of Plaintiff and the Class)

35. Plaintiff repeats the allegations contained in the foregoing paragraphs as if fully set forth herein.

36. Plaintiff, as the owner of registered copyrights, holds the exclusive rights to those books under 17 U.S.C. § 106.

37. In order to supply enough data for pre-training of the SlimLM models and other related models, Defendant downloaded, copied, stored, optimized and used copies of the SlimPajama dataset which included Plaintiff's copyrighted works. Defendant made multiple copies of the dataset (and thus Plaintiff and the Class Members' copyrighted works) for pre-training its models.

38. Neither Plaintiff nor Class Members authorized Defendant to make copies of, make derivative works, publicly display copies (or derivative works), or distribute copies (or derivative works) of their copyrighted works. The U.S. Copyright Act bestows all the aforementioned rights only on Plaintiff and the Class Members.

39. By downloading, copying, storing, processing, reproducing, and using the datasets containing copies of Plaintiff and the Class Members' Infringed Works, Defendant has directly infringed on their exclusive rights in their copyrighted works.

40. Defendant repeatedly copied, stored, and used the Infringed Works without Plaintiff's and members of the Class's permission in violation of their exclusive rights under the Copyright Act.

41. By and through the actions alleged above, Defendant has infringed and will continue to infringe on Plaintiff and the Class Members' copyrights.

42. Plaintiff has been injured by Defendant's acts of direct copyright infringement. 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 all others similarly situated, seeks judgment against Defendant, as follows:

a. For an order certifying the Class, naming Plaintiff as Class Representative, and naming Plaintiff's attorneys as Class Counsel to represent the Class;

b. For an order declaring that Defendant's conduct violates 17 U.S.C. § 501;

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 awards to Plaintiff and the Class, and that such interest be awarded at the highest legal rate from and after the date this class action complaint is first served on Defendant; and

h. Further relief for Plaintiff and the Class as the Court deems may be appropriate.

JURY TRIAL DEMAND

Plaintiff demands a trial by jury on all causes of action and issues so triable.

DATED: December 16, 2025

Respectfully submitted,

ELIZABETH LYON, individually and on behalf of similarly situated individuals

By: /s/ *Eugene Y. Turin*

Plaintiff's Attorney
Eugene Y. Turin (SB # 342413)
David L. Gerbie (*pro hac vice* forthcoming)
Jordan R. Frysinger (*pro hac vice* forthcoming)
MCGUIRE LAW, P.C.
1089 Willowcreek Road, Suite 200
San Diego, CA 92131
Tel: (312) 893-7002 Ex. 3
Fax: 312-275-7895
eturin@mcgpc.com
dgerbie@mcgpc.com
jfrysinger@mcgpc.com

Counsel for Plaintiff and the Putative Class

EXHIBIT A

Registration Number
TX 7-004-354

Effective date of registration:
June 20, 2008

Title

Title of Work: MANUSCRIPT MAKEOVER: Revision Techniques No Fiction Writer Can Afford to Ignore

Completion/Publication

Year of Completion: 2007

Date of 1st Publication: April 1, 2008 **Nation of 1st Publication:** United States

Author

- **Author:** Elizabeth Lyon

Author Created: entire text (excluding brief excerpts from other sources)

Work made for hire: No

Citizen of: United States

Anonymous: No **Pseudonymous:** No

Copyright claimant

Copyright Claimant: Elizabeth Lyon

c/o Penguin Group (USA) Inc., 375 Hudson Street, New York, NY, 10014

Limitation of copyright claim

Previously registered: No

Certification

Name: Benjamin Hojem, authorized agent of Elizabeth Lyon

Date: April 8, 2008

Correspondence: Yes