

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

E. MOLLY TANZER,
Plaintiff,

v.

ADOBE INC.,
Defendant.

Case No. 26-cv-04712-JSC

**ORDER GRANTING DEFENDANT’S
MOTION TO DISMISS**

Re: Dkt. No. 30

Plaintiff E. Molly Tanzer alleges Defendant Adobe Inc. unlawfully used her copyrighted books in developing two generative artificial intelligence products. (Dkt. No. 8.)¹ Defendant moves to dismiss the portion of Plaintiff’s sole cause of action which pertains to NVIDIA’s Nemotron product. (Dkt. No. 30.) Having carefully considered the parties’ arguments, and having had the benefit of oral argument on August 20, 2026, the Court GRANTS Defendant’s motion. Plaintiff does not sufficiently allege Article III standing or a direct copyright infringement claim arising from Adobe’s development of the Nemotron product because, drawing inferences in Plaintiff’s favor, Plaintiff does not allege facts sufficient to support an inference Adobe trained the Nemotron model by acquiring Plaintiff’s copyrighted works from Anna’s Archive.

PLAINTIFF’S ALLEGATIONS

Plaintiff is an author who owns registered copyrights on at least three books. (Dkt. No. 8 ¶ 10.) Defendant Adobe Inc. is a company which “provides services and tools for image, audio, video, and text editing and viewing.” (*Id.* ¶ 27.) “Adobe has in recent years added artificial

¹ Record citations are to material in the Electronic Case File (“ECF”); pinpoint citations are to the ECF-generated page numbers at the top of the documents.

intelligence [(‘AI’)] capabilities into its products.” (*Id.* ¶ 28.)

To develop its AI products, Adobe acquired datasets from so-called “shadow libraries.” (*Id.* ¶ 42.) “Shadow libraries are illicit online repositories of copyrighted content that openly defy copyright laws.” (*Id.* ¶ 43.) Plaintiff alleges six examples of shadow libraries: “(1) Library Genesis (‘LibGen’) [...] (2) Z-Library [...] (3) Sci-hub,” (4) “Internet Archive,” (5) “Bibliotok,” and (6) “Anna’s Archive.” (*Id.* ¶¶ 43, 44, 51, 52.) Anna’s Archive is “[t]he most active and notorious shadow library.” (*Id.* ¶ 44.) In 2002, Anna’s Archive “made available *all* of the content of other shadow libraries,” and therefore “makes *millions* of pirated books available for download to the public.” (*Id.*)

The products at issue are generative AI models. “Generative [AI ...] models are designed to generate output—such as text, images, video, or audio—in response to user prompts.” (*Id.* ¶ 15.) “[A]n AI model is trained by copying an enormous quantity of data and then feeding these copies into the model. This corpus of input material is called the training dataset.” (*Id.*) “A common type of generative AI model is known as a large language model (‘LLM’)[.]” (*Id.* ¶ 17.) One type of LLM is a “Small Language Model” (“SLM”). (*Id.* ¶ 47.)

“The first step” of “training” an AI model is “acquiring and curating the data that goes into the model.” (*Id.* ¶ 19.) “Copyrighted books tend to be high-quality data for training LLMs” and “models [...] perform best when their training data comprises books,” so “the inclusion of books in the training mixture of LLMs is essential.” (*Id.* ¶ 20.) “During training, the AI model copies and ingests each work in the training dataset and extracts protected expression from it.” (*Id.* ¶ 21.)

At issue in this case are two series of AI models. The first is Adobe’s “SlimLM series.” (*Id.* ¶ 47.) “These models were trained on the ‘SlimPajama’ dataset,” which itself contains a “Books3 dataset.” (*Id.*) The creator of Books3 has “confirmed in public statements that Books3 represents ‘all of Bibliotik,’” a “shadow library” of books. (*Id.* ¶¶ 51-52.) The Books3 dataset “contains approximately 196,640 books,” including Plaintiff’s copyrighted books. (*Id.* ¶ 55.) So, Plaintiff alleges Adobe unlawfully trained the SlimLM models on Plaintiff’s copyrighted books.

The second series of AI models is Nemotron, which are LLMs. In February 2024,

1 NVIDIA released the Nemotron-4 15B model. (*Id.* ¶ 37.) NVIDIA “did not identify the specific
2 datasets used to train the model” but “indicated generally that the model was trained with
3 ‘books.’” (*Id.*) “The entire training corpus” for the model “consisted of 8 trillion tokens.” (*Id.*)
4 “One token is about four characters, or 3/4 of a word.” (*Id.*) “240.6 billion tokens” of the training
5 corpus were “English-language books.” (*Id.*) With these numbers, and some back-of-the-napkin
6 estimates regarding “[a]n average-length novel[s],” Plaintiff alleges the Nemotron-4 15B model
7 was trained on a dataset containing “1,881,250” English-language books. (*Id.*)

8 In March 2023, nearly a year before NVIDIA released Nemotron-4 15B, Adobe announced
9 a partnership with NVIDIA to ‘create the next generation of generative AI models.’” (*Id.* ¶ 36.)
10 Then, in March 2024, a month after the release of Nemotron-4 15B, “Adobe again announced it
11 was working together with NVIDIA to ‘train new NVIDIA LLMs.’” (*Id.* ¶ 36.) In the March
12 2024 announcement, Adobe stated it “was ‘building datasets to train and tune next-generation
13 NVIDIA AI foundation models, including NVIDIA Nemotron LLMs.’” (*Id.*) Citing a March
14 2024 blog post on Adobe’s website, Plaintiff alleges “Adobe itself admitted to training and fine-
15 tuning Nemotron models, and building datasets to train the Nemotron models using its Adobe PDF
16 Extract tool (an AI-powered tool allowing the extraction of content from PDFs into a format
17 suitable for LLM training).” (*Id.* ¶ 40.)

18 “Upon information and belief, to develop the Nemotron language models, Adobe copied
19 datasets originating from Anna’s Archive to develop these models and incorporate the datasets
20 into the models’ training data.” (*Id.* ¶ 75.) Given Plaintiff estimates Nemotron-4 15B was trained
21 on “1,881,250” English-language books, “[u]pon information and belief, obtaining such a large
22 number of English-language books to train the Nemotron models could only be possible by
23 pirating copyrighted books from sources such as Anna’s Archive.” (*Id.* ¶¶ 37, 38, 43.) “Plaintiff’s
24 copyrighted books [...] are among the works available at Anna’s Archive[.]” (*Id.* ¶ 55.) “The use
25 of pirated books from Anna’s Archive, LibGen, Z-Library, and other shadow libraries has been
26 commonplace within the AI development community.” (*Id.* ¶ 45.) “Shadow libraries
27 themselves,” including the creators of Anna’s Archive, “acknowledge that the explosion in piracy
28 and patronage by LLM companies has saved shadow libraries from extinction.” (*Id.* ¶ 46.)

DISCUSSION

Adobe moves to dismiss the portion of Plaintiff's direct copyright infringement claim arising from the Nemotron model under Federal Rules of Civil Procedure 12(b)(1) and (b)(6).

A. Adobe's Motion is Procedurally Proper

As an initial matter, Defendant's request to dismiss part of Plaintiff's claim is proper. Plaintiff's sole cause of action alleges copyright infringement based upon two different AI products released by different companies and trained on different datasets. The allegations regarding the two AI models therefore "must be stated in a separate count" because the copyright infringement claims arising from those models are "founded on [] separate transaction[s] or occurrence[s]." Fed. R. Civ. P. 10(b). Plaintiff may not defeat dismissal by improperly including separate claims in a single count. Moreover, the Ninth Circuit has explained it "frequently dismiss[es] those portions of a claim that are barred while permitting the remaining portion [...] to go forward." *Benson v. JPMorgan Chase Bank, N.A.*, 673 F.3d 1207, 1216 (9th Cir. 2012). *Benson* relied on a Ninth Circuit case, *Guerrero v. Gates*, 442 F.3d 697 (9th Cir. 2006), which affirmed a district court's grant of a motion to dismiss "as to all but one half of" a plaintiff's claims because "half of" a claim was barred by a statute of limitations, but the other half was not. *Id.* at 701, 705-07, 709. Plaintiff's cited district-court cases, none of which cite a Ninth Circuit case, are therefore unpersuasive. See *Franklin v. Midwest Recovery Sys., LLC*, 2020 WL 3213676, at *1 (C.D. Cal. Mar. 9, 2020); *Redwind v. W. Union, LLC*, 2019 WL 3069864, at *4 (D. Or. June 21, 2019), report and recommendation adopted, 2019 WL 3069841, at *1 (D. Or. July 12, 2019); *Snell v. G4S Secure Solutions (USA) Inc.*, 424 F. Supp. 3d 892, 903-04 (E.D. Cal. 2019). So, the Court considers Defendant's motion to dismiss the portion of Plaintiff's copyright infringement claim arising from the Nemotron model.

B. Motion to Dismiss

Article III standing requires (1) Plaintiffs have suffered an "injury in fact" that is "actual or imminent," not "conjectural or hypothetical," (2) the injury is "fairly traceable" to the defendant, and (3) the injury will "likely [...] be redressed by a favorable decision." *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560-61 (1992) (cleaned up). Additionally, Plaintiff's copyright

1 infringement claim requires her to allege (1) she “owns the copyright and” (2) “the defendant
2 [it]self violated” Plaintiff’s rights. *Cobbler Nevada, LLC v. Gonzales*, 901 F.3d 1142, 1147 (9th
3 Cir. 2018).

4 Adobe contends Plaintiff’s allegations do not satisfy the second element for copyright
5 infringement or the first two requirements for Article III standing because, among other things, (1)
6 Plaintiff does not allege facts supporting an inference the Nemotron dataset was trained on
7 Plaintiff’s copyrighted works from Anna’s Archive, and (2) even if she did, her allegations do not
8 permit an inference Adobe, as opposed to NVIDIA, acquired Plaintiff’s books. The Court agrees
9 Plaintiff does not sufficiently allege a copyright infringement claim or an injury in fact, and
10 therefore does not reach Adobe’s second argument.² Given the Article III standing inquiry
11 overlaps with the merits, the Court addresses Adobe’s 12(b)(1) and (b)(6) motions together.

12 For the reasons discussed at oral argument, the Complaint does not include factual
13 allegations that support a plausible inference Adobe “acquired” Plaintiff’s copyrighted works
14 “from Anna’s Archive.” (Dkt. No. 8 ¶¶ 57, 75.) The allegation—made “[u]pon information and
15 belief”—is a bare conclusion unsupported by factual allegations that make it plausible rather than
16 merely possible. *See In re Century Aluminum Co. Sec. Litig.*, 729 F.3d 1104, 1105 (9th Cir. 2013)
17 (holding an allegation that is merely possible rather than plausible is not sufficient to defeat a
18 12(b)(6) motion); *see also Quinones v. Cnty. of Orange*, No. 20-56177, 2021 WL 5855113, at *1
19 (9th Cir. Dec. 9, 2021) (“At the motion to dismiss stage, the court must take all factual allegations
20 as true, strip out all conclusory statements, and ask whether the plaintiff has stated a plausible, not
21 just possible, claim for relief”) (citing *Bell Atl. Corp. v. Twombly*, 550 U.S. 544 (2007); *Aschroft*
22 *v. Iqbal*, 556 U.S. 662 (2009)).

23 For example, at oral argument, when asked what Complaint allegations support the use of
24 Anna’s Archive to train Nemotron, Plaintiff identified paragraphs 42 and 45. Those paragraphs
25 allege:

26
27 ² The Court therefore does not consider the portions of the alleged blog posts which were not
28 alleged in the complaint and which Defendant attaches to counsel’s declaration. (Dkt. No. 31.)

NVIDIA acquired millions of unauthorized copies of copyrighted books from the shadow library Anna's Archive in order to develop its LLMs. Upon information and belief, Adobe used those millions of pirated books to train and fine-tune Nemotron models, including through the use of its PDF Extract tool.

The use of pirated books from Anna's Archive, LibGen, Z-Library, and other shadow libraries has been commonplace with the AI development community. Major AI companies, including OpenAI, Meta, Anthropic, and NVIDIA pirated books from Anna's Archive, Library Genesis, Z-Library, Sci-Hub, and/or Pirate Library Mirror.

(Dkt. No. 8 ¶¶ 42, 45.) But the paragraphs' allegations as to use of Anna's Archive are conclusory and again do not allege facts that support a plausible inference NVIDIA utilized Anna's Archive. Repeating a conclusory allegation several times in a complaint does not make it any less conclusory.

In sum, the Complaint's allegations do not support a plausible inference Nemotron models were trained using Anna's Archive (and therefore Plaintiff's copyrighted works), regardless of whether Adobe or NVIDIA performed that training. *See In re Mosaic LLM Litig.*, No. 24-CV-01451-CRB, 2025 WL 2402677, at *2 (N.D. Cal. Aug. 19, 2025) (holding the plaintiff's allegations regarding the underlying data sets were too generalized to support a plausible inference the defendant infringed the plaintiff's copyrighted works). The district court cases Plaintiff cites fail to persuade the Court otherwise.

Accordingly, the Court grants Defendant's motion to dismiss Plaintiff's claim arising from the Nemotron model.

CONCLUSION

For the reasons explained above and at oral argument, the Court grants Defendant's motion to dismiss with leave to amend. Any amended complaint must be filed by October 2, 2026.

This Order disposes of Docket No. 30.

IT IS SO ORDERED.

Dated: August 20, 2026


JACQUELINE SCOTT CORLEY
United States District Judge