1   QUINN EMANUEL URQUHART &
    SULLIVAN, LLP
2   Sean S. Pak (SBN 219032)
    seanpak@quinnemanuel.com
3   50 California Street, 22nd Floor
    San Francisco, CA 94111
4   Telephone: (415) 875-6600
    Facsimile: (415) 875-6700
5
    Andrew H. Schapiro (admitted *pro hac vice*)
6   andrewschapiro@quinnemanuel.com
    191 N. Wacker Drive, Suite 2700
7   Chicago, Illinois 60606
    Telephone: (312) 705-7400
8   Facsimile: (312) 705-4001
9   Alex Spiro (admitted *pro hac vice*)
    alexspiro@quinnemanuel.com
10 295 Fifth Avenue
    New York, NY 10016
11 Telephone: (212) 849-7000
    Facsimile: (212) 849-7100
12
    Rachael L. McCracken (SBN 252660)
13 rachaelmccracken@quinnemanuel.com
    865 South Figueroa Street, 10th Floor
14 Los Angeles, CA 90017
    Telephone: (213) 443-3000
15 Facsimile: (213) 443-3100
16 *Attorneys for Defendant,*
    NVIDIA Corporation
17
18   UNITED STATES DISTRICT COURT

                  NORTHERN DISTRICT OF CALIFORNIA
19
                        OAKLAND DIVISION
20
21 ABDI NAZEMIAN, BRIAN KEENE,   Master File Case No. 4:24-cv-01454-JST (SK)
    STEWART O'NAN, ANDRE DUBUS III and
    SUSAN ORLEAN, individually and on behalf
22 of all others similarly situated,   **DEFENDANT'S NOTICE OF MOTION
                                             AND MOTION TO DISMISS FIRST
23   Plaintiffs,   AMENDED COMPLAINT AND
                                             MEMORANDUM OF POINTS AND
24   v.   AUTHORITIES IN SUPPORT OF
                                             THEREOF**
25 NVIDIA CORPORATION,
                                             Date: April 2, 2026
26   Defendant.   Time: 2:00 p.m.
                                             Location: Courtroom 6
27   Judge: Hon. Jon S. Tigar
28

1

NOTICE OF MOTION AND MOTION TO DISMISS

2   PLEASE TAKE NOTICE THAT on April 2, 2026, at 2:00 p.m., or as soon thereafter as the

3 matter may be heard before the Honorable Jon S. Tigar of the United States District Court, Northern

4 District of California in Courtroom 6, 2nd Floor, United States District Court, 1301 Clay Street,

5 Oakland, California, Defendant NVIDIA Corporation ("NVIDIA"), through its undersigned

6 counsel, will, and hereby does, move to dismiss Plaintiffs' First Amended Consolidated Complaint

7 (ECF No. 235 or "FAC") pursuant to Federal Rule of Civil Procedure 12(b)(6).

8   NVIDIA's Motion to Dismiss ("Motion") is based upon this Notice and Memorandum of

9 Points and Authorities, the concurrently-filed supporting declaration, the complete files and records

10 in this action, and any additional material and arguments considered in connection with the hearing

11 on the Motion.

12

STATEMENT OF RELIEF SOUGHT AND ISSUES TO BE DECIDED

13   NVIDIA seeks an order pursuant to Fed. R. Civ. P. 12(b)(6) dismissing with prejudice:

14   (1) Plaintiffs' direct infringement claim (Count I) for failure to state a claim as to:

15   (A)   allegations on unidentified models, datasets and, and websites,

16   (B)   certain newly added models (the Nemotron-4 15B, Nemotron-4 340B, Megatron

17   345M),

18   (C)   all newly-added datasets and websites (Anna's Archive, SlimPajama, Bibliotik,

19   Z-Library, Libgen, SciHub, Pirate Library Mirror, and unspecified "other shadow

20   libraries"), and use of any "BitTorrent Protocol";

21   (2) Plaintiffs' contributory copyright infringement claim (Count II) for failure to state a claim;

22 and

23   (3) Plaintiffs' vicarious copyright infringement claim (Count III) for failure to state a claim.

24   As to allegations related to models as to which the Court did not grant leave to amend,

25 NVIDIA also moves pursuant to Fed. R. Civ. P. 12(f) to strike on the ground that they are outside

26 the five models for which the Court granted leave to amend.

27

28

TABLE OF CONTENTS

Page

TABLE OF AUTHORITIES

Page(s)

Cases

NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

1

MEMORANDUM OF POINTS AND AUTHORITIES

2

I.   INTRODUCTION

3   The First Amended Consolidated Complaint ("FAC") uses threadbare allegations in an

4   attempt to drastically expand the case to involve new models, new datasets, and two new legal

5   theories that turn on unalleged direct infringement by nonparties. The following claims should be

6   dismissed from Plaintiffs' FAC.

7   *First*, the improper "catch-all" allegations should be dismissed or struck from the FAC. The

8   Court previously denied Plaintiffs discovery beyond NVIDIA's use of a single dataset alleged to

9   contain Plaintiffs' works (Books3) to train a single family of large language models (Nemo

10   Megatron). The Court's recent order found only "that there is good cause to amend to include five

11   other LLMs." ECF No. 232. Yet, the FAC now seeks to sweep in a vast universe of unidentified

12   models, internal research, and training data, including:

13   • Unidentified "NVIDIA[] LLMs" (FAC ¶ 77);

14   • Unidentified "internal models, including checkpoints" never made public and "internal

15   research which did not result in a fully trained LLM" (*id.* ¶¶ 43–44);

16   • Undefined "NextLargeLLM" models (*id.* ¶ 53); and

17   • Unidentified "other shadow libraries" (*id.* ¶¶ 58, 78).

18   Relying on these allegations, Plaintiffs seek to obtain essentially unlimited discovery

19   untethered to any plausible claim. Plaintiffs' bid for limitless discovery is confirmed by the blizzard

20   of discovery requests they served after filing the FAC, which NVIDIA addresses in its concurrently

21   filed Motion to Stay Discovery on Claims and Allegations Subject to NVIDIA's Motion to Dismiss.

22   Plaintiffs' unsupported, vague, and speculative allegations fail to state a claim and cannot open the

23   door to discovery into all of NVIDIA's AI models and products.

24   *Second*, the new allegations relating to named models, datasets, and data sources should be

25   dismissed because they are not plausibly connected to ***Plaintiffs' works***. For example, Plaintiffs

26   merely speculate "[u]pon information and belief, NVIDIA could not obtain the level of books

27   needed for the Nemotron models" without including the handful of books identified as "Plaintiffs'

28   Infringed Works." FAC ¶ 60. As another example, the FAC describes what datasets ***other*** AI

-1-   Master File Case No. 4:24-cv-01454-JST (SK)

NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

companies allegedly downloaded, then asserts without factual support that "NVIDIA followed this industry-wide practice." *Id.* ¶¶ 47, 51. But this guilt-by-association theory fails as a matter of law.

Third, Plaintiffs' new claims for contributory and vicarious infringement should be dismissed because Plaintiffs do not allege specific acts of direct infringement by any of NVIDIA's customers. *See id.* ¶ 71. Moreover, Plaintiffs include only a rote repetition of both claims' legal elements without any factual support. *Id.* ¶¶ 83–90.

Other courts in this District have dismissed similarly deficient allegations in materially identical cases after granting leave to amend. *In re Google Generative AI Copyright Litig.*, 2025 WL 2624885, at *7 (N.D. Cal. Sept. 11, 2025) ("Because Plaintiffs do not allege that any of their works were included in training datasets used to develop these models, Plaintiffs do not plausibly allege copyright infringement."); *In re Mosaic LLM Litig.*, 2025 WL 2402677, at *2 n.2 (N.D. Cal. Aug. 19, 2025) ("It is the lack of factual, rather than conclusory, allegations as to what materials [defendant] actually used to train [its] models that dooms Plaintiffs' direct infringement claim"); *Concord Music Grp., Inc. v. Anthropic PBC*, 2025 WL 1487988, at *4 (N.D. Cal. Mar. 26, 2025) (dismissing secondary infringement claims where "allegations emphasized by [Plaintiffs] in their opposition are conclusory, and provide no facts regarding direct third-party infringement"). This Court should do the same.

II.   FACTUAL BACKGROUND

On March 8, 2024, Plaintiffs filed a class action complaint alleging that NVIDIA infringed their copyrights when it trained the NeMo Megatron family of AI large language models using The Pile, which allegedly contains Plaintiffs' works in a subset called Books3. ECF No. 1 ¶¶ 19–31. Plaintiffs did not allege their asserted works are included in any other dataset or that NVIDIA used any other dataset to train the NeMo Megatron LLMs.

Despite the limited scope of their original complaint, Plaintiffs filed a series of discovery motions seeking to expand the case to additional datasets and models. The Court repeatedly denied Plaintiffs' attempts to seek broader discovery. ECF Nos. 168, 183, 185.

On October 18, 2025, Plaintiffs moved to Modify the Scheduling Order and for Leave to File a First Amended Consolidated Complaint. *See* ECF No. 193. In their reply in support of the

motion, Plaintiffs argued that their proposed amendment was limited to "five discrete models." ECF No. 212 at 12. On January 15, 2026, the Court granted Plaintiffs' motion "to amend to include five other LLMs" listed in the FAC. ECF No. 232 at 6. On January 16, 2026, Plaintiffs filed the FAC, which seeks to make several expansive changes.

First, it broadens Plaintiffs' direct copyright infringement claim from a single family of LLMs (Nemo Megatron) and a single dataset (Books3), to essentially all of NVIDIA's AI models and all datasets. The FAC sweeps in: "[NVIDA's] LLMs" (*id.* ¶ 77); "internal research which did not result in a fully trained LLM" (*id.* ¶ 44); "internal models, including checkpoints," that were never published or even named (*id.* ¶ 43); and other undefined aspects of the "LLM development process" (*id.* ¶ 23). Although Plaintiffs previously represented that their new claims would be limited to "five discrete models," ECF No. 212 at 12, the FAC transgresses that limitation and seeks to shoehorn virtually every NVIDIA model, service, and research project ever developed into this case. *Id.* ¶¶ 23, 43–45, 77. Correspondingly, the FAC expands the class definition to include anything related to any "LLM training, research, or development" by NVIDIA. *Compare* FAC ¶ 92 (new class definition), *with* ECF No. 1 ¶ 39 (prior class definition). As to datasets, the FAC sweeps in an unlimited number of unidentified "other shadow libraries." FAC ¶¶ 5, 23, 51, 57–58, 76–78. Indeed, Plaintiffs recently propounded discovery requests covering "all datasets or files containing books" and "any and all LLMs," confirming the vast breadth of their amendment and the inaccuracy of their representation to the Court that the scope of discovery for the FAC would be "co-extensive" with the current scope of the case. McCracken Decl. Exs. 1–2; ECF No. 193 at 17.

Second, the FAC adds unsupported allegations as to three specific models: Megatron 345 and Nemotron-4 15B and Nemotron-4 340B. However, Plaintiffs do not allege any facts plausibly suggesting their works were used to train those models. NVIDIA does not move to dismiss Plaintiffs' allegations as to the two Retro models because Plaintiffs pled specific facts that those models were trained on Books3.

Third, the FAC adds allegations about several specific websites and datasets, including Bibliotik, LibGen, Sci-Hub, Z-Library, Pirate Library Mirror, Anna's Archive and SlimPajama. But Bibliotik, LibGen, Sci-Hub, Z-Library, Pirate Library Mirror, and Anna's Archive are websites with

1   voluminous content, not datasets, and therefore it is insufficient to allege "on information and belief"

2   that NVIDIA accessed these websites and "downloaded books." Plaintiffs fail to allege that ***their***

3   ***works*** were downloaded from these websites, either as individual works or as part of any dataset.

4   And while Plaintiffs allege that NVIDIA downloaded SlimPajama, they do not allege any facts to

5   establish that this dataset contained their works.

6   *Fourth*, the FAC adds two new counts for contributory and vicarious infringement. FAC

7   ¶¶ 83–90. The new counts are based on conclusory allegations about NVIDIA's "NeMo

8   Framework." *Id.* ¶¶ 70–72. Plaintiffs cite a YouTube clip from March 2023, describing the

9   framework as a set of tools to help developers create their own models. *Id.* ¶ 69 n.10. The FAC

10   alleges that one feature included in the framework was a script for downloading The Pile. *Id.* ¶ 70.

11   The FAC then speculates "on information and belief" that NVIDIA provided the framework to three

12   companies. *Id.* ¶ 71. It further asserts, based solely "on information and belief," that "NVIDIA

13   materially aided numerous other customers in downloading, using, and storing The Pile (and

14   Books3) dataset." *Id.* But there are no factual allegations as to what, if anything, these third parties

15   did with the framework or whether they used the script on which the claims are based, beyond

16   conclusory assertions of "downloading, using, and storing The Pile (and Books3) dataset." *Id.*

17   **III.   LEGAL STANDARD**

18   "The focus of any rule 12(b)(6) dismissal is the complaint." *Cardone v. Cnty. of Alameda*,

19   2019 WL 7816822, at *1 n.1 (N.D. Cal. July 30, 2019) (cleaned up) (citation omitted). To survive

20   a motion to dismiss, "a complaint must contain sufficient factual matter, accepted as true, to 'state

21   a claim to relief that is plausible on its face.'" *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting

22   *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)). The Court must "accept the plaintiffs'

23   allegations as true and construe them in the light most favorable to plaintiffs," but it need not accept

24   as true "allegations that are merely conclusory, unwarranted deductions of fact, or unreasonable

25   inferences." *In re Gilead Scis. Sec. Litig.*, 536 F.3d 1049, 1055 (9th Cir. 2008) (citations omitted).

26   "Threadbare recitals of the elements of a cause of action, supported by mere conclusory

27   statements, do not suffice." *Iqbal*, 556 U.S. at 678 (citations omitted). Importantly, "[f]actual

28   allegations must be enough to raise a right to relief *above the speculative level*." *Twombly*, 550 U.S.

1 at 555 (emphasis added) (citations omitted).   Nor should courts "assume the truth of legal

2 conclusions merely because they are cast in the form of factual allegations." *Fayer v. Vaughn*, 649

3 F.3d 1061, 1064 (9th Cir. 2011) (*per curiam*) (citation omitted).

4 **IV.   ARGUMENT**

5   **A. <u>Plaintiffs' Vague and Conclusory Allegations Regarding Unidentified Models,</u>**

6   **<u>Datasets, and Websites Fail To State a Direct Copyright Infringement Claim</u>**

7   To state a claim for copyright infringement, Plaintiffs must plead "(1) ownership of a valid

8 copyright, and (2) copying of constituent elements of the work that are original." *Aquarian Found.*,

9 *Inc. v. Lowndes*, 127 F.4th 814, 818 (9th Cir. 2025).   As courts in this District have repeatedly

10 confirmed, plaintiffs cannot satisfy the second element without factual allegations that their specific

11 works were included in the relevant training datasets or models. *See Google Generative AI* , 2025

12 WL 2624885, at *7 (dismissing claims where "Plaintiffs do not allege that any of *their works* were

13 included in training datasets used to develop these models" (emphasis added)); *Mosaic*, 2025 WL

14 2402677, at *2 (dismissing claims where plaintiffs "do not allege facts that could establish that the

15 … models are actually trained on any shadow library websites, let alone those that contain

16 *Plaintiffs' works*" (emphasis added)).

17   The FAC is replete with vague references to unidentified "NVIDIA[] LLMs" (FAC ¶ 77),

18 "internal models, including checkpoints" that were never named or published (*id.* ¶ 43), "internal

19 research which did not result in a fully trained LLM" (*id.* ¶ 44), unspecified "NextLargeLLM"

20 models (*id.* ¶ 53), and allegations of unspecified "other shadow libraries" or datasets (*id.* ¶¶ 58, 78).

21 Yet Plaintiffs provide no factual allegations whatsoever connecting these unidentified models,

22 research, or "shadow libraries" to their works. Indeed, Plaintiffs fail to allege which datasets were

23 used to train these unidentified models, let alone that the datasets contained their works. Plaintiffs

24 do not even identify specific models or shadow libraries, instead referring only to a catch-all of

25 undefined "internal models," "research," "development processes," and "other shadow libraries."

26 Without essential factual allegations linking Plaintiffs' copyrighted works to *specific acts of*

27 *copying*, Plaintiffs cannot state a claim for copyright infringement.

28

1   Two courts in this District have recently dismissed claims that rested on similarly vague and

2 speculative allegations. In *Mosaic*, some of these same Plaintiffs alleged that Databricks' "DBRX"

3 AI models "'are trained on open data sets that the community knows,' and that some open data sets

4 'include notorious shadow library websites, which include Plaintiffs' works.'" *Mosaic*, 2025 WL

5 2402677, at *2; *see also Mosaic*, No. 3:24-cv-01451-CRB (N.D. Cal. June 27, 2025), ECF No. 131

6 ¶ 31 (alleging "shadow libraries have long been of interest to the AI-training community because

7 they host and distribute vast quantities of unlicensed copyrighted material"). Judge Breyer held that

8 these "vague and conclusory allegations" were "far too generalized" because plaintiffs did "not

9 allege facts that could establish that the DBRX models are actually trained on any shadow library

10 websites, let alone those that contain Plaintiffs' works." *Mosaic*, 2025 WL 2402677 at *2.

11 Although the *Mosaic* court recently granted the plaintiffs leave to file a further amended complaint,

12 in that decision, the Court did not address the sufficiency of the plaintiffs' new allegations because

13 the defendants did not argue that amendment was futile or that the new allegations failed to state a

14 claim. *Mosaic*, 2026 WL 151109, at *1 (N.D. Cal. Jan. 20, 2026) (noting that "Defendants oppose,

15 citing undue delay and prejudice," but not futility). The amendment there was also different in

16 scope because, unlike here, "the operative complaint already asserts similar claims and allegations

17 against MosaicML," such that the "proposed claims do not inject novel and unrelated legal theories

18 into this case." *Id.* at *2. The court further cautioned that amendment was "not to be interpreted as

19 permitting expansive discovery after amendment," *id.* at *3, exactly what Plaintiffs seek here.

20   Similarly, in *Google Generative AI*, the plaintiffs provided a laundry list of Google models

21 (there, at least identified by name) that the plaintiffs failed to connect to their works. The Court

22 found that "[b]ecause Plaintiffs do not allege that any of their works were included in training

23 datasets used to develop these models, Plaintiffs do not plausibly allege copyright infringement."

24 *Google Generative AI*, 2025 WL 2624885, at *7.

25   Rule 8 prohibits vague and unsupported allegations—like Plaintiffs' allegations about

26 unspecified models trained on unspecified datasets here—as a safeguard against "aimless fishing

27 expeditions." *InfoStream Grp., Inc. v. PayPal, Inc.*, 2012 WL 3731517, at *5 (N.D. Cal. Aug. 28,

28 2012) ("The Rule 8 pleading requirements prevent parties from filing complaints in order to conduct

1 aimless fishing expeditions in the hope that some helpful evidence might possibly be uncovered.")

2 (quotations and citations omitted); *Mosaic*, 2025 WL 2402677, at *3 ("Discovery is not a fishing

3 expedition. . . . Rather, the Federal Rules limit discovery to material that is relevant and proportional

4 to claims that survive past the pleading stage.") (citations omitted). Indeed, Plaintiffs already tried

5 to obtain expansive discovery untethered to their original complaint and the Court repeatedly

6 rejected those efforts (ECF Nos. 168, 183, 185).

7 Plaintiffs' sweeping allegations about unidentified models, websites, and datasets are

8 especially improper because they violate the scope of the Court's leave to amend. In seeking to

9 amend, Plaintiffs claimed that "the FACC is a targeted amendment about five discrete models."

10 ECF No. 212 at 12. This Court's decision granting leave to amend turned critically on Plaintiffs'

11 representation that amendment would be limited to the "five other LLMs" they specifically

12 identified. ECF No. 232 at 6. The FAC violates this limitation by improperly including unbound

13 catch-all allegations that extend to all "NVIDIA's LLMs" and "LLM training, research, or

14 development." FAC ¶¶ 76, 77, 92. The Court should thus strike these allegations under Rule 12(f).

15 *Brown v. Stroud*, 2011 WL 13312050, at *3 (N.D. Cal. Jan. 28, 2011) (striking "actual allegations

16 of an amended complaint that are beyond the permissible scope of a court's prior order granting

17 leave to amend").

18 **B. Plaintiffs' Allegations as to Named Models, Datasets, and Websites Fail To State a**

19 **Direct Copyright Infringement Claim Because They Do Not Plausibly Allege**

20 **Copying of Plaintiffs' Works**

21 The FAC also includes new allegations as to named models, datasets, and websites (or

22 "shadow libraries"), but it does not connect those allegations to ***Plaintiffs' works***. *E.g.*, FAC ¶¶ 46,

23 47, 51, 60. This failure is fatal. As *Mosaic* found, "[i]t is the lack of factual, rather than conclusory,

24 allegations as to what materials [NVIDIA] actually used to train [its] models that dooms Plaintiffs'

25 direct infringement claim against [NVIDIA]." 2025 WL 2402677, at *2 n.2; *Andersen v. Stability*

26 *AI Ltd.*, 700 F. Supp. 3d 853, 868–69 (N.D. Cal. 2023) (concluding that a statement by CEO that

27 "[w]e use the open data sets that are published and train across those," without additional factual

28 allegations, failed to state a claim); *Google Generative AI*, 2025 WL 2624885, at *7 (dismissing

1   where plaintiffs failed to "allege any facts regarding these models" beyond listing them "in a long

2   list of what appears to be every generative AI model that [Defendant] has ever developed").

3   Instead, Plaintiffs' allegations amount to a brazen fishing expedition. Plaintiffs have already

4   propounded discovery requests covering "all datasets or files containing books," and defining "Your

5   LLMs" as "any and all LLMs You developed, trained, or created, or which are currently in

6   development"—an apparent attempt to uncover the very factual basis their complaint lacks. *E.g.*,

7   McCracken Decl. Ex. 1 at 8; Ex. 2 at 3.

8   ***Megatron 345M.*** Plaintiffs allege that "NVIDIA included ***the Pile*** dataset as training data

9   for an LLM known as Megatron 345M." FAC ¶ 41 (emphasis added). But "The Pile" comprises

10   multiple datasets, including Wikipedia, RealNews, OpenWebtext, CC-Stories, and Books3, among

11   others. Only one of those datasets—Books3—allegedly includes Plaintiffs' works. FAC ¶ 32. But

12   Plaintiffs do not allege that Megatron 345M was trained on Books3. *See generally* FAC.[1]

13   In fact, Plaintiffs' conclusory statements about Megatron 345M are directly refuted by public

14   documents. Courts "are not . . . required to accept as true allegations that contradict . . . matters

15   properly subject to judicial notice," including public documents that "neither party disputes the

16   authenticity of . . . or the accuracy of the information displayed therein." *Daniels-Hall v. Nat'l*

17   *Educ. Ass'n*, 629 F.3d 992, 998–99 (9th Cir. 2010). Here, NVIDIA published a model card

18   disclosing that the Megatron 345M "was trained on text sourced from Wikipedia, RealNews,

19   OpenWebtext, and CC-Stories"—not Books3. ECF No. 206-8 at 2. Plaintiffs do not allege that

20   those datasets include their works. Only Books3, which was not used to train Megatron 345M,

21   allegedly contains Plaintiffs' works. FAC ¶ 36.

22   Plaintiffs have not disputed the model card's accuracy and have relied on model cards for

23   other allegations. *E.g.*, *id.* ¶ 30. And Plaintiffs failed to address Megatron 345M's model card

24   during briefing on Plaintiffs' motion for leave to amend, despite having an opportunity to do so.

25   *See* ECF No. 212 at 9 (ignoring NVIDIA's Exhibit 7 when addressing the Megatron 345M model).

26   Because Plaintiffs have not disputed the authenticity and accuracy of these public documents, the

27
_____

28   [1] This motion addresses only the allegations in the FAC. *Khoja*, 899 F.3d at 998.

Master File Case No. 4:24-cv-01454-JST (SK)

NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

1   Court should take judicial notice of them. *Daniels-Hall*, 629 F.3d at 998–99. With no plausible

2   facts to establish that their works were used to train the Megatron 345M model, Plaintiffs cannot

3   state a claim of direct copyright infringement as to this model.

4   ***Nemotron-4 15B and Nemotron-4 340B.***   Plaintiffs speculate that because the training

5   dataset used to train these models is large and contains books, it must include Plaintiffs' works.

6   FAC ¶ 59 (alleging the Nemotron-4 15B was trained on a dataset "composed of 4.6% books" and

7   "8 trillion tokens"); *id.* ¶ 60 (Nemotron-4 340B's dataset "included the same 8 trillion tokens from

8   the Nemotron-4 15B but added an additional 1 trillion tokens"); *id.* ¶ 61 ("Upon information and

9   belief, NVIDIA could not obtain the level of books needed for the Nemotron models without

10 pirating copyrighted books, including Plaintiffs' Infringed Works.").   But such "[s]peculation

11 cannot serve as the basis of a well-pleaded claim." *Seneca v. Homeaglow Inc.*, 2025 WL 3038592,

12 at *12 (C.D. Cal. Sept. 23, 2025). As the court explained in *Mosaic*, "the mere size of the data sets

13 does not plausibly connect the dots to Plaintiffs' works," because "simply speculating on how

14 widespread the copying may be, without describing any acts or works that infringe fails to state a

15 claim for direct copyright infringement." 2025 WL 2402677, at *2 (cleaned up). Likewise here,

16 the absence of factual allegations that the data used to train Nemotron-4 15B and Nemotron-4 340B

17 included Plaintiffs' works requires dismissal as to those models.

18   ***Anna's Archive.*** Plaintiffs' allegations with respect to Anna's Archive fare no better. The

19 only plausible facts alleged about Anna's Archive are that NVIDIA "contacted Anna's Archive"

20 about unspecified data (FAC ¶ 55), Anna's Archive asked NVIDIA to confirm (*id.* ¶ 56), and

21 NVIDIA gave the "green light" to "proceed" (*id.* ¶ 57). But the mere fact that NVIDIA was in

22 contact with representatives from Anna's Archive does not mean that NVIDIA obtained Plaintiffs'

23 works from Anna's Archive. It's equally plausible NVIDIA did not. *In re Century Aluminum Co.*

24 *Sec. Litig.*, 729 F.3d 1104, 1108 (9th Cir. 2013) (rejecting allegations that were "merely consistent

25 with" their theory, because "[s]omething more is needed, such as facts tending to exclude the

26 possibility that the alternative explanation is true") (citation omitted).

27   ***SlimPajama.*** The FAC alleges that NVIDIA downloaded the "SlimPajama" dataset to train

28 unidentified LLMs, (FAC ¶ 42), but fails to allege that SlimPajama contained their works.

NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

1   Specifically, the FAC alleges that a different dataset, RedPajama, "originally included the Books3

2   dataset," and that SlimPajama was "created by cleaning and deduplicating the 1.2T token

3   RedPajama dataset," *i.e.*, removing some materials. *Id.* In other words, Plaintiffs do not plausibly

4   allege that Books3 was *retained* in SlimPajama. That failure requires dismissal, because Plaintiffs

5   must allege "facts tending to exclude the possibility" of a noninfringing explanation. *In re Century*

6   *Aluminum*, 729 F.3d at 1108 (citation omitted).

7   ***Bibliotik, LibGen, Sci-Hub, Z-Library, and Pirate Library Mirror and BitTorrent***

8   ***Protocols.*** As to the remaining identified datasets, Plaintiffs' allegations rely on only "information

9   and belief" and speculation. The FAC falls short in two ways.

10   *First*, the FAC does not plausibly allege that Plaintiffs' works were downloaded. These

11   websites, by Plaintiffs' own description, host "millions of pirated books." *Id.* ¶ 47. But the FAC

12   contains only conclusory, "on information and belief" allegations that unidentified "books" were

13   downloaded from Bibliotik, LibGen, Sci-Hub, Z-Library, and Pirate Library Mirror, not that

14   Plaintiffs' works were downloaded. *See, e.g.*, FAC ¶ 58 ("On information and belief . . . NVIDIA

15   also downloaded books hosted or sourced from other shadow libraries, including [but not limited

16   to] LibGen, Sci-Hub, and Z-Library"); ¶ 78 (same); ¶ 80 ("On information and belief, NVIDIA has

17   continued to store and make copies" of works downloaded from "shadow libraries"). These

18   allegations fail to state a claim for infringement. Plaintiffs' broad statements that NVIDIA

19   "downloaded *these works*," whether via BitTorrent (for which there are no supporting allegations

20   whatsoever) or otherwise, are "conclusory, unwarranted deductions of fact" that fail to state a claim

21   and should be dismissed, as there are no plausible allegations as to use of a torrent protocol to

22   download. *In re Gilead Scis. Sec. Litig.*, 536 F.3d at 1055 (quotations and citations omitted).

23   *Second*, the FAC does not plausibly allege that NVIDIA itself downloaded anything from

24   these websites. "[E]ven allegations on information and belief must be supported by a factual

25   foundation leading to that belief." *Mosaic*, 2025 WL 2402677, at *2 n.2 (cleaned up); *Menzel v.*

26   *Scholastic, Inc.*, 2018 WL 1400386, at *2 (N.D. Cal. Mar. 19, 2018) (allegations "made on

27   information and belief must still be based on factual information that makes the inference of

28   culpability plausible") (quotations and citation omitted). Here, Plaintiffs point to what other AI

1   companies allegedly did, not to actions taken by NVIDIA. Paragraphs 45–52 describe shadow

2   libraries generally and assert that "[v]irtually every one of the major LLM developers—including

3   OpenAI, Meta, and Anthropic—pirated books from Library Genesis, Z-Library, Sci-Hub, and/or

4   Pirate Library Mirror." FAC ¶ 51. But Plaintiffs must allege facts about "the defendant"—not other

5   companies. *Iqbal*, 556 U.S. at 678; *Synopsys, Inc. v. ATopTech, Inc.*, 2013 WL 5770542, at *4

6   (N.D. Cal. Oct. 24, 2013) ("simply speculating on how widespread the copying may be, without

7   describing any acts or works that infringe," fails to state a claim for direct copyright infringement);

8   *Mosaic*, 2025 WL 2402677, at *2 (noting Rule 8 requires "non-speculative facts about what

9   Defendant infringed and how") (citations omitted).

10   * * * * *

11   In sum, Plaintiffs' allegations about named models, websites, and datasets all fail because

12   they do not connect ***Plaintiffs' works*** to the alleged copying.

13   **C.   Plaintiffs' Secondary Copyright Infringement Claims Fail**

14   Both contributory and vicarious infringement claims require, as a threshold matter, an

15   underlying act of direct infringement by a third party. *Perfect 10, Inc. v. Yandex N.V.*, 962 F. Supp.

16   2d 1146, 1157–58 (N.D. Cal July 12, 2013) (noting "secondary liability otherwise cannot exist"

17   without "an underlying act of direct infringement"). To state a claim for contributory infringement,

18   Plaintiffs must further allege that NVIDIA had "knowledge of the infringing activity" and that

19   NVIDIA either (a) "materially contribute[d] to" or (b) "induce[d]" that infringement. *Cobbler*

20   *Nevada, LLC v. Gonzales*, 901 F.3d 1142, 1147 (9th Cir. 2018). For vicarious infringement,

21   Plaintiffs must allege that NVIDIA had "(1) the right and ability to supervise the infringing conduct

22   and (2) a direct financial interest in the infringing activity." *Visa*, 494 F.3d at 802 (citation omitted).

23   In other words, a defendant "infringes vicariously by profiting from direct infringement while

24   declining to exercise a right to stop or limit it." *Id.* (quoting *MetroGoldwyn-Mayer Studios Inc. v.*

25   *Grokster, Ltd.*, 545 U.S. 913, 930 (2005)).

26   As to both contributory and vicarious liability, Plaintiffs' allegations fall far short.

27

28

1. Plaintiffs Have Not Alleged Any Predicate Act of Direct Third-Party Infringement for Either Claim

The FAC lacks the foundational requirement of "underlying direct infringement." *Yandex*, 962 F. Supp. 2d at 1157. This deficiency is fatal both to Plaintiffs' contributory and vicarious infringement claims. Plaintiffs allege only that NVIDIA helped customers "to create and build their own AI models," FAC ¶ 69, including by providing a script to download The Pile dataset *from third-party Eleuther AI*. *Id.* ¶ 70 (emphasis added). Plaintiffs then speculate, based on nothing more than "information and belief," *id.* ¶ 71, that NVIDIA's customers actually downloaded, stored, and used The Pile. But "while facts may be alleged upon information and belief, that does not mean that conclusory allegations are permitted." *Menzel v. Scholastic, Inc.*, 2018 WL 1400386, at *2 (N.D. Cal. Mar. 19, 2018). Allegations "made on information and belief must still be based on factual information that makes the inference of culpability plausible." *Id.* (citation modified). Plaintiffs provide no such factual foundation. Although Plaintiffs name three purported NVIDIA customers, they do not identify any customer alleged to have downloaded or used The Pile. The only allegation *not* couched on "information and belief" is that NVIDIA "provided resources, guidance, and tools for its customer[s]" (FAC ¶ 71)—but providing business assistance to customers is not secondary infringement if those customers are not alleged to infringe. Plaintiffs' remaining allegations rest on sheer speculation unsupported by any cited documents, statements, or other evidence.

Moreover, the FAC does not identify what "tools and scripts" NVIDIA allegedly provided those customers. If Plaintiffs are referring to the URL identified in their Motion for Leave to Amend, ECF No. 193 at 5, the cited configuration file provides *optional* functionality. The end user must make the independent choice to use The Pile to train (rather than their own datasets or other public datasets). Plaintiffs conspicuously stop short of alleging that any NVIDIA customer made the independent choice to use that functionality for downloading and using The Pile, and instead simply hypothesize about what *could* be done. Allegations concerning what unidentified third parties *could have* done is not a substitute for "clearly alleg[ing] a predicate act of direct third-party

1   infringement." *Concord Music Grp., Inc. v. Anthropic PBC*, 2025 WL 1487988, at *4–5 (N.D. Cal.

2   Mar. 26, 2025).

2.   Plaintiffs Fail to Sufficiently Allege the Remaining Elements of Contributory Infringement

5   In addition to Plaintiffs' failure to allege an act of direct infringement, they also fail to allege

6   that NVIDIA (1) knew of any such infringement, and (2) materially contributed to, or (3) induced

7   third-party infringement, as required to state a claim for contributory infringement.

8   ***No Knowledge of Specific Infringing Acts.*** Contributory infringement "requires more than

9   a generalized knowledge" of "the possibility of infringement"—it demands "actual knowledge of

10 specific acts of infringement." *Luvdarts, LLC v. AT & T Mobility*, LLC, 710 F.3d 1068, 1072 (9th

11 Cir. 2013) (quotations and citations omitted). This requirement is rigorous: even a defendant's

12 general knowledge that its product "is in fact used for infringement" will not suffice absent

13 knowledge of particular infringements of a plaintiff's work. *Grokster*, 545 U.S. at 933. Courts

14 therefore dismiss contributory infringement claims where plaintiff's allegations "merely suggest

15 that Defendant operated [its website] indifferent to the risk of copyright infringement generally[.]"

16 *Davis v. Pinterest, Inc.*, 2021 WL 879798, at *3–4 (N.D. Cal. Mar. 9, 2021).

17   The FAC alleges only that "NVIDIA knew or had reason to know of the direct infringement

18 by others using the NeMo Megatron framework, because NVIDIA is ***fully aware of the capabilities***

19 ***of its own product***, platforms and tools upon which third parties downloaded and acquired at least

20 The Pile dataset, and potentially other datasets including copyrighted books as well." FAC ¶ 85

21 (emphasis added). This generalized awareness of product capabilities—*optional* capabilities that

22 require independent choices by nonparties no less—is precisely the type of "generalized knowledge"

23 that *Luvdarts* holds insufficient. 710 F.3d at 1072; *see also YZ Prods. v. Redbubble*, 545 F. Supp.

24 3d 756, 764 (N.D. Cal. 2021) (dismissing contributory infringement claim because bare allegation

25 that defendant had "specific knowledge of" infringement "through Defendant's system" was too

26 conclusory); *Grokster*, 545 U.S. at 932–33 (courts may not "presum[e] or imput[e] intent . . . solely

27 from the design or distribution of a product" used for infringement); *Nat'l Photo Grp., LLC v.*

28 *Allvoices, Inc.*, 2014 WL 280391, at *6 (N.D. Cal. Jan. 24, 2014) ("That Defendant operates a

website where infringing content may be posted is not, by itself, enough to state a claim for contributory infringement.").[2]

No Material Contribution. "Material contribution turns on whether the activity in question 'substantially assists' direct infringement." *Louis Vuitton Malletier, S.A. v. Akanoc Sols., Inc.*, 658 F.3d 936, 943 (9th Cir. 2011) (citation omitted) (setting out standard and applying it to facts where, unlike here, there was support for substantial assistance because the defendant "had direct control" over the site with the infringing materials); *Perfect 10, Inc. v. Visa Int'l Serv. Ass'n* ("*Visa*"), 494 F.3d 788, 795 (9th Cir. 2007) (requiring "steps that are substantially certain to result in such direct infringement"). In the technology context, material contribution typically requires that the defendant's own infrastructure or systems be "an essential step in the infringement process"—not that they simply facilitate access to readily available third-party content, as alleged here. *Louis Vuitton*, 658 F.3d at 944 (citation omitted); *Visa*, 494 F.3d at 799–800 (holding defendants did not materially contribute in part because they "d[id] not operate the servers on which [the infringing images] reside[d]"); *Napster*, 239 F.3d. at 1022 (holding that "Napster materially contribute[d] to the infringing activity" because, "[w]ithout the support services defendant provides, Napster users could not find and download the music they want with the ease of which defendant boasts" (internal quotation marks removed) (alteration in original)).

Here, Plaintiffs allege only that NVIDIA provided an unidentified script within a larger framework of optional tools that could allow its customers to access The Pile dataset through a publicly hosted third-party website, if they chose to do so. Plaintiffs do not—and cannot—allege that NVIDIA hosted, stored, promoted, or operated the website where The Pile resided. Nor do Plaintiffs allege that one optional choice within NVIDIA's larger framework of tools was an essential step in the infringement process—to the contrary, individuals could just as easily download and use The Pile from Eleuther AI's website. Therefore, Plaintiffs' allegations are precisely the type that *Visa* held insufficient as a matter of law.

---

[2]   Nor does the FAC even attempt to allege willful blindness, as it does not claim that NVIDIA "subjectively believed specific infringement was likely occurring" and "took deliberate steps to avoid learning about [it]." *Luvdarts*, 710 F.3d at 1073.

1   ***No Inducement.*** Inducement liability requires "an ***object*** of promoting the product's use to

2   infringe copyright." *VHT, Inc. v. Zillow Grp., Inc.*, 918 F.3d 723, 746 (9th Cir. 2019) (cleaned up)

3   (emphasis added). Plaintiffs admit that the purpose of the framework at issue in this claim was to

4   assist customers "to create and build their own AI Models," FAC ¶ 69, not to promote infringement

5   of works in The Pile—a publicly available dataset that Plaintiffs themselves acknowledge was well

6   known within the AI community, *id.* ¶ 49.

7   ***Substantial Non-Infringing Uses Bar Liability.*** Independently, Plaintiffs' contributory

8   infringement claim fails because the NeMo framework is capable of substantial non-infringing uses

9   (e.g., pre-processing other datasets like public-domain or licensed datasets, and training, optimizing,

10   or fine-tuning LLMs or working with other modalities like automatic speech recognition, text-to-

11   speech generation, speaker recognition, natural language processing, biomolecular data, and other

12   functionalities), which bars liability. *See Sony Corp. of America v. Universal City Studios, Inc.*, 464

13   U.S. 417 (1984); *see also Cobbler Nevada,* 901 F.3d at 1148 (noting that, under *Sony*, "a court

14   would be unable to find contributory infringement liability merely based on a failure to take

15   affirmative steps to prevent infringement, if the device otherwise was capable of substantial

16   noninfringing uses.") (citation omitted).

17   **3.   Plaintiffs Fail to Sufficiently Allege the Remaining Elements of Vicarious**

18   **Infringement**

19   Plaintiffs have also failed to allege that NVIDIA had (1) the "the right and ability to control"

20   the purported—but wholly unidentified—third-party infringement or (2) any "direct financial

21   interest" in such activity. *Visa*, 494 F.3d at 802.

22   ***No Right and Ability to Control.*** Plaintiffs proclaim that "NVIDIA had the right and ability

23   to control the direct infringements of [its] customers," FAC ¶ 88, but offer no factual substantiation.

24   A defendant's "ability to control its own [] webpages does not give [that defendant] the right to

25   control the infringing acts of third parties even though that ability would allow [the defendant] to

26   affect those infringing acts to some degree." *Visa*, 494 F.3d at 803 (citation omitted). Rather, the

27   defendant must have "***both*** a legal right to stop or limit the directly infringing conduct, as well as

28   the practical ability to do so." *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1173 (9th Cir.

1   2007) (emphasis added). NVIDIA's ability to modify its own training script does not translate into

2   the **legal right** to control independent actions by users or customers, including their independent

3   decisions to download The Pile from a third-party website and use it to train their LLMs. Plaintiffs

4   identify no mechanism giving NVIDIA such control—because none exists.

5   ***No Direct Financial Interest.***   Plaintiffs assert that "NVIDIA has directly benefitted

6   financially from the direct infringement of its customers because NVIDIA generated revenue from

7   customers using the NeMo Megatron Framework to download the Pile." FAC ¶ 89; *see also id.*

8   ¶ 73 (alleging "[t]he Pile (and Books3) was key to NVIDIA attracting customers"). This conclusory

9   speculation lacks any factual substantiation and, regardless, fails to establish the requisite "causal

10   relationship between the infringing activity and any financial benefit a defendant reaps." *Perfect*

11   *10, Inc. v. Giganews, Inc.*, 847 F.3d 657, 673 (9th Cir. 2017). The scripts identified by Plaintiffs,

12   along with the NeMo Megatron Framework, were available to the public for free—NVIDIA

13   received no direct financial benefit. *Perfect 10, Inc. v. Google, Inc.*, 2010 WL 9479060, at *9 (C.D.

14   Cal. July 30, 2010), *aff'd,* 653 F.3d 976 (9th Cir. 2011) (holding plaintiff had "failed to provide

15   evidence that it is likely to be able to establish that [defendant] derives a direct financial benefit"

16   where plaintiff "does not dispute that [defendant] provides [the at-issue product] as a free service to

17   users, without charging a fee"). Moreover, the script is one optional component among many in a

18   general-purpose framework offered free of charge and designed, as Plaintiffs concede, to help

19   customers build their own AI models on any platform using any dataset or training materials. FAC

20   ¶¶ 69–70. Any alleged financial benefit NVIDIA received cannot be causally tied to the alleged

21   infringement, as opposed to the framework's legitimate, non-infringing functionality. As the Ninth

22   Circuit emphasized, "the direct financial benefit prong of the vicarious infringement test 'demands

23   more than evidence that customers were 'drawn' to [a defendant] to obtain access to infringing

24   material in general." *Giganews*, 847 F.3d at 673 (citation omitted).

25   **D. <u>Dismissal Should Be with Prejudice and Without Leave to Amend</u>**

26   The deficiencies in the FAC cannot be cured by amendment but are instead rooted in the

27   absence of any plausible factual basis to satisfy the elements of the claims Plaintiffs allege.

28   As to the direct infringement claims, Plaintiffs' catch-all allegations are unsupported by any

1 factual allegations whatsoever, and their allegations about the databases and models they name are

2 doomed by the absence of any basis to contend that Plaintiffs' works were included in these datasets

3 or were used to develop these models. Plaintiffs' conclusory allegations are insufficient as a matter

4 of law, and these defects cannot be remedied through amendment.

5   As to the secondary liability claims, Plaintiffs rely on conclusory allegations and abstract

6 descriptions of NVIDIA's training script because Plaintiffs lack any plausible or concrete

7 allegations of any third party infringing any of Plaintiffs' works, let alone of NVIDIA's knowledge

8 of such an act or of conduct by NVIDIA that materially contributed to or induced the act. *See*

9 *Luvdarts*, 710 F.3d at 1073 (district court properly dismissed complaint, including contributory

10 infringement claim, with prejudice where plaintiff failed to adequately plead specific knowledge of

11 third-party infringement). Similarly, Plaintiffs fail to adequately allege NVIDIA's right and ability

12 to supervise that conduct or direct financial interest in the infringing activity. *See Visa*, 494 F.3d at

13 802. Amendment would be futile because Plaintiffs lack the specific facts required to allege viable

14 claims.

15   Further amendment would also unduly prejudice NVIDIA. This case has been pending for

16 nearly two years, and fact discovery is set to close in five weeks. Substantial judicial and party

17 resources have been expended. Plaintiffs represented to the Court that their amended claims would

18 require only "co-extensive" discovery (*i.e*, of the same extent), yet immediately served a barrage of

19 requests demanding new discovery on "all datasets or files containing books," defining "Your

20 LLMs" as "any and all LLMs [NVIDIA] developed, trained, or created, or which are currently in

21 development," and explicitly noting requests were "not limited to" the five models the Court allowed

22 them to add—a dramatic expansion that exponentially increases costs and burdens on both NVIDIA

23 and the Court, as detailed in NVIDIA's concurrently filed Motion to Stay Discovery on Claims and

24 Allegations Subject to NVIDIA's Motion to Dismiss. *E.g.*, McCracken Decl. Ex. 1 at 5, 8; Ex. 2 at

25 3. Plaintiffs' bait-and-switch should be rejected. "The pleadings must be settled *now* to focus the

26 parties' efforts in this massive and complex case." *Google Generative AI*, 2025 WL 2624885, at

27 *13 (emphasis in original).

28   These factors support denying leave to amend.

1 **V.   CONCLUSION**

2   For the foregoing reasons, NVIDIA respectfully requests that the Court dismiss with

3 prejudice the FAC's (1) direct infringement claim (Count I) as to (A) allegations on unidentified

4 models, datasets and, and websites, (B) certain newly-added models (the Nemotron-4 15B,

5 Nemotron-4 340B, Megatron 345M), (C) all newly-added datasets and websites (Anna's Archive,

6 SlimPajama, Bibliotik, Z-Library, Libgen, SciHub, Pirate Library Mirror, and unspecified "other

7 shadow libraries") and use of any "BitTorrent Protocol"; (2) contributory copyright infringement

8 claim (Count II); and (3) vicarious copyright infringement claim (Count III). NVIDIA also requests

9 that the Court strike allegations related to models other than the five models for which the Court

10 granted leave to amend.

11

12

13   Dated: January 29, 2026   Respectfully Submitted,

14
                                        By: */s/ Sean S. Pak*
15   QUINN EMANUEL URQUHART & SULLIVAN, LLP

16   Sean S. Pak (SBN 219032)
                                             seanpak@quinnemanuel.com
17   50 California Street, 22nd Floor
                                             San Francisco, CA 94111
18   Telephone: (415) 875-6600
                                             Facsimile: (415) 875-6700
19
                                             Andrew H. Schapiro (admitted *pro hac vice*)
20   andrewschapiro@quinnemanuel.com
                                             191 N. Wacker Drive, Suite 2700
21   Chicago, Illinois 60606
                                             Telephone: (312) 705-7400
22   Facsimile: (312) 705-4001

23   Alex Spiro (admitted *pro hac vice*)
                                             alexspiro@quinnemanuel.com
24   295 Fifth Avenue
                                             New York, NY 10016
25   Telephone: (212) 849-7000
                                             Facsimile: (212) 849-7100
26
                                             Rachael L. McCracken (SBN 252660)
27   rachaelmccracken@quinnemanuel.com
                                             865 South Figueroa Street, 10th Floor
28   Los Angeles, CA 90017

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

Telephone: (213) 443-3000
Facsimile: (213) 443-3100

Attorneys for Defendant,
NVIDIA Corporation

Master File Case No. 4:24-cv-01454-JST (SK)
NVIDIA'S MOTION TO DISMISS FIRST AMENDED COMPLAINT

1   **ATTESTATION PURSUANT TO CIVIL L.R. 5-1(i)(3)**

2   I attest that the other signatory listed is a registered CM/ECF filer, concurs in the content of

3   the filing, and has authorized the filing.

4

5   Dated: January 29, 2026   By: */s/ Rachael L. McCracken*

6   Rachael L. McCracken

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28