1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6

7   ABDI NAZEMIAN, et al.,   Case No. 24-cv-01454-JST

8   Plaintiffs,

9   v.   **ORDER DENYING PLAINTIFFS'
                                                MOTION FOR RELIEF FROM NON-
10   NVIDIA CORPORATION,   DISPOSITIVE PRETRIAL ORDER**

                     Defendant.   Re: ECF No. 174
11

12

13   Before the Court is Plaintiffs' motion for relief from a non-dispositive pretrial order

14   concerning the scope of discovery. ECF No. 175. The Court will deny the motion.

15   **I.   BACKGROUND**

16   This is a suit for copyright infringement brought by Plaintiffs Abdi Nazemian, Brian

17   Keene, Stewart O'Nan, and other proposed class members who are authors of books that were

18   pirated and part of a training dataset that Defendant NVIDIA Corporation [NVIDIA] used for its

19   series of large language models named NEMO Megatron-GPT. ECF No. 1 ¶¶ 1–4.

20   On August 6, 2025, the parties asked Judge Kim to decide whether Plaintiffs were entitled

21   to seek discovery into NVIDIA's potential use of pirated datasets other than "The Pile," which is a

22   copy of "Books3"—the database that Plaintiffs had knowledge of and made allegations about in

23   the complaint. ECF No. 152 at 2; ECF No. 1 ¶ 31. Plaintiffs argued that the proposed searches

24   terms on six databases containing pirated materials are highly relevant and posed no undue burden

25   to NVIDIA. ECF No. 152 at 3, 5–6. Plaintiffs also argued that NVIDIA limited their ability to

26   identify other pirated databases they used to train their LLMs by rejecting requests either to

27   identify what other datasets their LLMs are trained or to run search terms intended to identify

28   potential other datasets. *Id.* at 6. NVIDIA argued that the complaint only alleges that their LLMs

United States District Court
Northern District of California

1   were **trained** on Books3, and discovery on other pirated databases is an improper attempt to seek

2   discovery on datasets beyond those alleged to contain putative class members' works, exceeding

3   the permissible scope of discovery. *Id.* at 7.

4   Judge Kim denied Plaintiffs' request, ruling that since the Plaintiffs are only alleging that

5   the LLMs were trained on the Books3 database, discovery was appropriately limited to that

6   database. ECF No. 168 at 3. Judge Kim found that Plaintiffs were not entitled to information

7   about the use of any dataset by NVIDIA, and their discovery request lacked sufficient bounds. *Id*

8   at 3–4. On September 4, 2025, Plaintiffs filed a timely motion for relief from Judge Kim's order,

9   which NVIDIA opposed on September 19, 2025. ECF Nos. 174, 178.

10   **II.   JURISDICTION**

11   The Court has jurisdiction under 28 U.S.C. § 1331.

12   **III.   LEGAL STANDARD**

13   "A non-dispositive order entered by a magistrate must be deferred to unless it is 'clearly

14   erroneous or contrary to law.'" *Grimes v. City & Cnty. of San Francisco*, 951 F.2d 236, 241 (9th

15   Cir. 1991) (quoting Fed. R. Civ. P. 72(a); 28 U.S.C. § 636(b)(1)(A)). "A decision is 'contrary to

16   law' if it applies an incorrect legal standard or fails to consider an element of the applicable

17   standard." *Ingram v. Pac. Gas & Elec. Co.*, No. 12-cv-02777-JST, 2013 WL 6174487, at *2

18   (N.D. Cal. Nov. 25, 2013) (citation omitted). While the "'clearly erroneous' standard applies to

19   the magistrate judge's findings of fact," the magistrate's "legal conclusions are [reviewed] de

20   novo." *Wolpin v. Philip Morris Inc.*, 189 F.R.D. 418, 422 (C.D. Cal. 1999).

21   Federal Rule of Civil Procedure 26(b)(1) provides that "[p]arties may obtain discovery

22   regarding any nonprivileged matter that is relevant to any party's claim or defense and

23   proportional to the needs of the case, considering the importance of the issues at stake in the

24   action, the amount in controversy, the parties' relative access to relevant information, the parties'

25   resources, the importance of the discovery in resolving the issues, and whether the burden or

26   expense of the proposed discovery outweighs its likely benefit." Relevancy, for purposes of

27   discovery, is broadly defined to encompass "any matter that bears on, or that reasonably could

28   lead to other matter that could bear on, any issue that is or may be in the case." *Doe v. Kaiser*

United States District Court
Northern District of California

1   *Found. Health Plan, Inc.*, 2024 WL 3225904, at *1 (N.D. Cal. June 28, 2024) (*citing In re*

2   *Williams-Sonoma, Inc.*, 947 F.3d 535, 539 (9th Cir. 2020)). "Though a magistrate judge's

3   determination of relevance is a legal conclusion, district courts have tended not to disturb such

4   conclusions absent abuse of discretion." *AMEC Env't & Infrastructure, Inc. v. Integral*

5   *Consulting, Inc.*, No. 12-CV-01735-SC, 2014 WL 6601960, at *2 (N.D. Cal. Nov. 19, 2014).

6   **IV.   DISCUSSION**

7   Plaintiffs seek relief from Judge Kim's order declining to compel discovery to run searches

8   on the use of six pirated databases. They argue that the order committed (1) legal error by

9   restricting discovery of evidence only to information known to Plaintiffs at the pleading stage, and

10   (2) abuse of discretion by failing to address Plaintiffs' relevance argument in the absence of any

11   burden to NVIDIA. ECF No. 174 at 5. Plaintiffs argue that limiting discovery only to databases

12   that they already knew NVIDIA was using ignores precedent in copyright infringement cases and

13   undermines the utility and stated purpose of discovery. *Id*. at 4. NVIDIA opposed the motion by

14   arguing that there was no evidence that the order was clearly erroneous or contrary to the law.
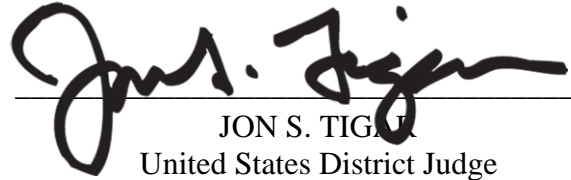
15   ECF No. 178 at 2.

16   Judge Kim's order limited the scope of the databases that the Plaintiffs could examine to

17   only the Books3 dataset that Plaintiffs specifically alleged as a source of the pirated training

18   material. ECF No. 168 at 3. The Court does not find that Judge Kim committed legal error or an

19   abuse of discretion here by refusing to grant discovery on the additional databases based on

20   theories of infringement that exceed the scope of the allegations in the complaint. Courts in this

21   district have routinely denied discovery regarding theories that exceed the scope of the operative

22   complaint. *See Anderson v. Stability AI Ltd.*, 23-cv-00201-WHO, ECF No. 14 at 2 (N.D. Cal. Jun.

23   27, 2025); *In re Mosaic LLM Litig*, 2025 WL 2402677, at *2–3 (N.D. Cal. Aug. 19, 2025). It was

24   well within the magistrate court's discretion to limit discovery to the specific allegations of

25   infringement made in the complaint. Further, Plaintiffs incorrectly argue that Judge Kim's order

26   improperly limits the proposed class of authors whose work may exist in the other pirated

27   databases, when the limitations on the proposed class stem from the Plaintiffs' own operative

28   complaint. ECF No. 174 at 6.

United States District Court
Northern District of California

3

1

CONCLUSION

2   For the above reasons, the Court denies Plaintiffs' motion for relief from non-dispositive

3   pretrial order.

4   **IT IS SO ORDERED.**

5   Dated: September 25, 2025

6   _____

7   JON S. TIGAR
                                                                     United States District Judge

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28