1   JEDEDIAH WAKEFIELD (CSB No. 178058)
    jwakefield@fenwick.com
2   RYAN KWOCK (CSB No. 336414)
    rkwock@fenwick.com
3   FENWICK & WEST LLP
    555 California Street, 12th Floor
4   San Francisco, CA 94104
    Telephone: 415.875.2300
5   Facsimile: 415.281.1350

6   DAVID L. HAYES (CSB No. 122894)
    dhayes@fenwick.com
7   FENWICK & WEST LLP
    801 California Street
8   Mountain View, CA 94041
    Telephone: 650.988.8500
9   Facsimile: 650.938.5200

10 *Attorneys for Defendants*
    DATABRICKS, INC. and
11 MOSAIC ML, LLC, formerly MOSAIC ML, INC.

12

13   **UNITED STATES DISTRICT COURT**
                NORTHERN DISTRICT OF CALIFORNIA
14   **SAN FRANCISCO DIVISION**

15

16

17 REBECCA MAKKAI, an individual; and   Case No.: 3:24-cv-02653-CRB
    JASON REYNOLDS, an individual,   Related Case No. 3:24-cv-01451-CRB

18   Individual and Representative   **DEFENDANTS' ANSWER TO CLASS
                Plaintiffs,   ACTION COMPLAINT**
19
            v.
20
    DATABRICKS, INC., a Delaware corporation;
21 and MOSAIC ML, INC., a Delaware corporation,

22   Defendants.

23

24

25

26

27

28

FENWICK & WEST LLP

FENWICK & WEST LLP

1   Defendants Databricks, Inc. ("Databricks") and Mosaic ML, LLC, formerly Mosaic ML,

2   Inc. ("MosaicML") (together "Defendants") submit this Answer to the Class Action Complaint

3   ("Complaint") filed by Plaintiffs Rebecca Makkai and Jason Reynolds, on behalf of themselves

4   and others similarly situated (collectively "Plaintiffs"). Unless specifically admitted, Defendants

5   deny each of the allegations in the Complaint.

6   **INTRODUCTION**[1]

7   1.   Defendants admit that "artificial intelligence" is commonly abbreviated "AI." As

8   Paragraph 1 pertains generally to AI and not any particular AI technologies of Defendants,

9   Defendants lack knowledge and information sufficient to form a belief as to the truth of the

10   remaining allegations in Paragraph 1, and on that basis deny them.

11   2.   Defendants admit that MosaicML created and that Databricks distributed a large

12   language model ("LLM") series called MosaicML Pretrained Transformer ("MPT"). Defendants

13   admit that LLMs may generate text outputs in response to user prompts.

14   3.   As Paragraph 3 pertains generally to LLMs and not any particular LLM of

15   Defendants, Defendants lack knowledge and information sufficient to form a belief as to the truth

16   of the allegations in Paragraph 3, and on that basis deny them. Defendants deny any suggestion

17   that the LLMs at issue in this litigation have the goal of imitating protected expression.

18   4.   Defendants lack knowledge and information sufficient to form a belief as to the

19   truth of the allegations in Paragraph 4, and on that basis deny them.

20   5.   Defendants lack knowledge and information sufficient to form a belief as to the

21   truth of the allegations in Paragraph 5, and on that basis deny them.

22   6.   Defendants admit that Databricks is the corporate parent of MosaicML and also

23   distributed certain MPT models. Defendants deny the remaining allegations in Paragraph 6.

24   **JURISDICTION AND VENUE**

25   7.   The allegations in this Paragraph state a legal conclusion to which no response is

26   required. To the extent a response is deemed required, Defendants lack knowledge and

27

28
_____

[1] Defendants include the headings as listed in the Complaint without any admission as to the accuracy or appropriateness of the headings.

DEFENDANTS' ANSWER TO CLASS
ACTION COMPLAINT   Case No.: 3:24-cv-02653-CRB

1 information sufficient to form a belief as to the truth of the allegations in Paragraph 7 and on that

2 basis deny them.

3   8.   The allegations in this Paragraph state a legal conclusion to which no response is

4 required. To the extent a response is deemed to be required, Defendants admit that the Court has

5 personal jurisdiction over it with respect to this litigation and that venue is proper in the Northern

6 District of California. Defendants deny the remaining allegations in Paragraph 8.

7   9.   The allegations in this Paragraph state a legal conclusion to which no response is

8 required. To the extent a response is deemed to be required, Defendants admit that assignment of

9 this case to the San Francisco Division is proper, and that this action purports to arise under

10 intellectual property laws.

PARTIES

12   10.   Defendants lack knowledge and information sufficient to form a belief as to the

13 truth of the allegations in Paragraph 10, and on that basis deny them.

14   11.   Defendants lack knowledge and information sufficient to form a belief as to the

15 truth of the allegations in Paragraph 11, and on that basis deny them.

16   12.   Defendants lack knowledge and information sufficient to form a belief as to the

17 truth of the allegations in Paragraph 12 or as to the content of Exhibit A, and on that basis deny

18 the allegations.

19   13.   Defendants admit the allegations in Paragraph 13.

20   14.   Defendants admit that MosaicML is a Delaware limited liability company, is

21 located at this address, and is a subsidiary of Databricks.

AGENTS AND CO-CONSPIRATORS

23   15.   Defendants deny the allegations in Paragraph 15.

24   16.   Defendants deny the allegations in Paragraph 16.

FACTUAL ALLEGATIONS

26   17.   Defendants admit that MosaicML was incorporated in 2020, and one of its goals

27 was to provide tools to facilitate the efficient training of AI models.

28

FENWICK & WEST LLP

18.   Defendants admit that MosaicML released the MPT-7B LLM in May 2023. Defendants admit that LLMs may generate text outputs in response to user prompts. Defendants deny the remaining allegations in Paragraph 18.

19.   Defendants admit that the corpus of material used to train an LLM may be referred to as a "training dataset." As Paragraph 19 pertains generally to LLMs and not any particular LLM of Defendants, Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 19, and on that basis deny them.

20.   Defendants admit that the MPT-7B LLM contains 6.7 billion parameters but otherwise deny the last sentence of this paragraph. As Paragraph 20 otherwise pertains generally to LLMs and not any particular LLM of Defendants, Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 20, and on that basis deny them. Defendants deny any suggestion that the LLMs at issue in this litigation progressively adjusted their output to more closely approximate any protected expression contained in the training dataset.

21.   As Paragraph 21 pertains generally to LLMs and not any particular LLM of Defendants, Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations set forth in Paragraph 21, and on that basis deny them. Defendants deny any suggestion that the LLMs at issue in this litigation have the goal of imitating any protected expression ingested from the training dataset.

22.   The allegations in this Paragraph state a legal conclusion to which no response is required. To the extent a response is deemed required, Defendants deny that MosaicML's transformative use of works to train the MPT LLMs required consent, credit, or compensation. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations set forth in Paragraph 22, and on that basis deny them.

23.   Defendants admit that MosaicML released the MPT-7B LLM in May 2023 and published a blog post titled "Introducing MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs" available at https://www.databricks.com/blog/mpt-7b. To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents

FENWICK & WEST LLP

of the blog post located at https://www.databricks.com/blog/mpt-7b, the document speaks for itself.

24.   To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents of the blog post located at https://www.databricks.com/blog/mpt-7b, the document speaks for itself. Defendants deny the remaining allegations in Paragraph 24.

25.   To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents of the website located at https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T, the document speaks for itself. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 25, and on that basis deny them.

26.   To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents of the document located at https://arxiv.org/pdf/2101.00027.pdf, the document speaks for itself. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 26, and on that basis deny them.

27.   The allegations in this Paragraph state a legal conclusion to which no response is required. To the extent a response is deemed required, Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 27, and on that basis deny them.

28.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 28, and on that basis deny them.

29.   Defendants deny that they infringed Plaintiffs' alleged copyrights. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 29, and on that basis deny them.

30.   To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents of the document located at https://huggingface.co/datasets/the_pile_books3, the document speaks for itself. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 30, and on that basis deny them.

DEFENDANTS' ANSWER TO CLASS
ACTION COMPLAINT   5   Case No.: 3:24-cv-02653-CRB

31.   Defendants admit that MosaicML completed training the MPT-7B LLM by May 2023. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 31, and on that basis deny them.

32.   To the extent the allegations in this Paragraph state a legal conclusion, no response is required. To the extent a response is deemed required, Defendants admit that portions of the RedPajama – Books dataset were used as training data for the MPT-7B LLM. Defendants deny that they infringed Plaintiffs' alleged copyrights. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 32, and on that basis deny them.

33.   To the extent the allegations in this Paragraph state a legal conclusion, no response is required. To the extent a response is deemed required, Defendants admit that MosaicML released the MPT-30B LLM in June 2023, that portions of the RedPajama – Books dataset were used as training data for the MPT-30B LLM, and that the MPT-30B LLM contains approximately 30 billion parameters. To the extent the allegations set forth in this Paragraph purport to summarize or characterize the contents of a blog post, the document speaks for itself. Defendants deny that they infringed Plaintiffs' alleged copyrights. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations set forth in Paragraph 33, and on that basis deny them.

34.   Defendants deny the allegations in Paragraph 34.

COUNT 1

Direct Copyright Infringement (17 U.S.C. § 501) against MosaicML

35.   Defendants incorporate by reference their responses to Paragraphs 1 through 34 above as though fully set forth herein.

36.   Defendants admit that Plaintiffs purport to bring a claim for direct copyright infringement on behalf of themselves individually and on behalf of a putative class. To the extent the allegations in Paragraph 36 are premised upon the appropriateness of Plaintiffs as class representatives or upon the ability of a class to be certified, those allegations state legal

FENWICK & WEST LLP

1 conclusions and arguments to which no response is required. To the extent a further response is

2 deemed required, Defendants deny those allegations.

3   37. Defendants lack knowledge and information sufficient to form a belief as to the

4 truth of the allegations in Paragraph 37, and on that basis deny them.

5   38. Defendants admit that portions of the RedPajama – Books dataset were used as

6 training data for the MPT-7B and MPT-30B LLMs. Defendants lack knowledge and information

7 sufficient to form a belief as to the truth of the remaining allegations set forth in Paragraph 38,

8 and on that basis deny them.

9   39. Defendants admit that MosaicML released the MPT-7B-StoryWriter-65k+ LLM.

10 To the extent the allegations set forth in this Paragraph purport to summarize or characterize the

11 contents of the blog post located at https://www.databricks.com/blog/mpt-7b, the document

12 speaks for itself. Defendants deny the remaining allegations in Paragraph 39.

13   40. Defendants deny that MosaicML needed Plaintiffs' authorization to use any

14 portions of the RedPajama – Books dataset to train MosaicML's MPT models. Defendants lack

15 knowledge and information sufficient to form a belief as to the truth of the remaining allegations

16 in Paragraph 40, and on that basis deny them.

17   41. Defendants deny that MosaicML needed Plaintiffs' authorization to use any

18 portion of the RedPajama – Books dataset to train MosaicML's MPT models. Defendants deny

19 that MosaicML violated any exclusive rights under the Copyright Act. Defendants lack

20 knowledge and information sufficient to form a belief as to the truth of the remaining allegations

21 in Paragraph 41, and on that basis deny them.

22   42. Defendants deny the allegations in Paragraph 42.

23 **COUNT 2**

24 **Vicarious Copyright Infringement against Databricks**

25   43. Defendants incorporate by reference their responses to Paragraphs 1 through 42

26 above as though fully set forth herein.

27   44. Defendants admit that Plaintiffs purport to bring a claim for vicarious copyright

28 infringement on behalf of themselves individually and on behalf of a putative class. To the extent

FENWICK & WEST LLP

DEFENDANTS' ANSWER TO CLASS
ACTION COMPLAINT   7   Case No.: 3:24-cv-02653-CRB

1   the allegations in Paragraph 44 are premised upon the appropriateness of Plaintiffs as named

2   representatives or upon the ability of a class to be certified, those allegations state legal

3   conclusions and arguments to which no response is required. To the extent a further response is

4   deemed required, Defendants deny those allegations.

5   45.   Defendants admit that Databricks acquired MosaicML in July 2023.

6   46.   Defendants deny the allegations in Paragraph 46.

7   47.   Defendants deny the allegations in Paragraph 47.

8   48.   Defendants deny the allegations in Paragraph 48.

9   **CLASS ALLEGATIONS**

10   49.   The allegations in Paragraph 49 state legal conclusions or arguments to which no

11   response is required. To the extent a response is deemed required, Defendants deny the

12   allegations in Paragraph 49.

13   50.   The allegations in Paragraph 50 state legal conclusions or arguments to which no

14   response is required.

15   51.   The allegations in Paragraph 51 state legal conclusions or arguments to which no

16   response is required.

17   52.   The allegations in Paragraph 52 state legal conclusions or arguments to which no

18   response is required. To the extent a response is deemed required, Defendants deny that they

19   possess information concerning the exact number of members of Plaintiffs' putative class.

20   Defendants lack knowledge and information sufficient to form a belief as to the truth of the

21   remaining allegations in Paragraph 52, and on that basis deny them.

22   53.   The allegations in Paragraph 53 state legal conclusions or arguments to which no

23   response is required. To the extent a response is deemed required, Defendants deny the

24   allegations in Paragraph 53.

25   54.   The allegations in Paragraph 54 state legal conclusions or arguments to which no

26   response is required. To the extent a response is deemed required, Defendants deny the

27   allegations in Paragraph 54.

28

55.   The allegations in Paragraph 55 state legal conclusions or arguments to which no response is required. To the extent a response is deemed required, Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 55, and on that basis deny them.

56.   The allegations in Paragraph 56 state legal conclusions or arguments to which no response is required. To the extent a response is deemed required, Defendants deny the allegations in Paragraph 56.

DEMAND FOR JUDGMENT

Defendants deny that Plaintiffs and the proposed Class are entitled to any relief, including but not limited to the relief sought in the section of the Complaint titled "Demand for Judgment." To the extent that this section contains any allegations, Defendants deny them.

JURY TRIAL DEMANDED

Defendants hereby demand a trial by jury on all claims, defenses, and issues in this action so triable.

GENERAL DENIAL

Unless expressly admitted above, Defendants deny each and every allegation set forth in Plaintiffs' Complaint.

DEFENSES

Defendants further plead the following separate and additional defenses to the Complaint. By pleading these defenses, Defendants do not in any way agree or concede that they have the burden of proof or persuasion on any claims or defenses. Defendants reserve the right to plead any and all defenses that may be evident or revealed after investigation and discovery in this matter.

FIRST DEFENSE

(Failure to State a Claim)

Plaintiffs' and putative class members' claims fail, in whole or in part, because the Complaint, and each purported cause of action therein, fails to state a claim upon which relief may be granted and/or to state facts sufficient to constitute a claim for relief against Defendants.

SECOND DEFENSE

(Fair Use)

To the extent that Defendants made any unauthorized copies of any of Plaintiffs' or putative class members' registered copyrighted works, such copying constituted fair use under 17 U.S.C. § 107, given, among other factors, the purpose and transformative character of the use and the effect of the use upon the potential market for or value of the allegedly copyrighted works.

THIRD DEFENSE

(Noninfringing Use)

Plaintiffs' and putative class members' claims fail in whole or in part because the models that are the subject of this action, and Defendants' products, services, or actions in connection with those models, have and are capable of substantial noninfringing uses and commercially significant noninfringing uses, and are widely used for legitimate, unobjectionable purposes.

FOURTH DEFENSE

(Invalidity or Unenforceability of Copyright)

Plaintiffs' and putative class members' claims fail in whole or in part to the extent the accused datasets include works in the public domain, unregistered works, works to which copyright protection has been abandoned or expired, works that lack requisite originality, works not subject to copyright protection under 17 U.S.C. § 102(b), or otherwise unprotectable under applicable law.

FIFTH DEFENSE

(Misuse, Unclean Hands, Laches, Estoppel)

Plaintiffs' and putative class members' claims fail in whole or in part by the doctrines of misuse, unclean hands, laches, estoppel, and/or other equitable defenses.

SIXTH DEFENSE

(Waiver, Abandonment, Forfeiture)

Plaintiffs' and putative class members' claims are barred in whole or in part by the doctrines of waiver, abandonment, and/or forfeiture.

DEFENDANTS' ANSWER TO CLASS
ACTION COMPLAINT

10

Case No.: 3:24-cv-02653-CRB

SEVENTH DEFENSE

(License)

Plaintiffs' and putative class members' claims are barred or limited to the extent that the works over which they assert copyright and copyright infringement were subject to an express or implied license or permission given to Defendants or their agents.

EIGHTH DEFENSE

(Lack of Standing)

Plaintiffs' and putative class members' claims are barred to the extent they do not own a valid copyright for some or all of the works, do not hold a valid copyright registration for some or all of the works, and/or otherwise lack standing or fail to meet statutory requirements to assert their claims.

NINTH DEFENSE

(Unavailability of Injunctive Relief)

Plaintiffs' and putative class members' claims fail, in whole or in part, because Plaintiffs are not entitled to injunctive relief (temporarily, preliminarily, or permanently), including because any injury to them is not immediate or irreparable, Plaintiffs would have an adequate remedy at law, the balance of hardships favors no injunction, and the public interest is best served by no injunction.

TENTH DEFENSE

(Lack of Willfulness)

Plaintiffs' and putative class members' claims fail in whole or in part because Defendants' conduct was innocent, not willful.

ELEVENTH DEFENSE

(Lack of Injury)

Plaintiffs' and putative class members' claims fail in whole or in part because they have not suffered injury as a result of the conduct alleged in the Complaint.

FENWICK & WEST LLP

TWELFTH DEFENSE

(Failure to Mitigate Damages)

Plaintiffs and putative class members are barred from recovery of damages because of and to the extent of their failure to mitigate their alleged damages (to which, in any event, they are not entitled).

THIRTEENTH DEFENSE

(Attorneys' Fees Improper)

The Complaint fails to state a cause of action or allege sufficient facts to support a claim for attorneys' fees.

<u>RESERVATION OF ADDITIONAL DEFENSES</u>

Defendants do not presently know all facts concerning Plaintiffs' claims and the putative class sufficient to state all affirmative defenses at this time. Defendants will seek leave to amend this Answer should they later discover facts demonstrating the existence of additional affirmative defenses. Defendants reserve any and all additional affirmative defenses available to them.

<u>DEFENDANTS' DEMAND FOR JUDGMENT</u>

WHEREFORE, Defendants pray for judgment with respect to Plaintiffs' Complaint and Defendants' defenses as follows:

a)   A judgment in Defendants' favor denying Plaintiffs all relief requested in the Complaint and dismissing the Complaint with prejudice;

b)   For an award to Defendants of their attorneys' fees, costs, and expenses of litigation; and

c)   Such other relief as the Court shall deem just and proper.

FENWICK & WEST LLP

Dated: May 29, 2024   FENWICK & WEST LLP

By: */s/ Jedediah Wakefield*
    Jedediah Wakefield (CSB No. 178058)
    jwakefield@fenwick.com
    Ryan Kwock (CSB No. 336414)
    rkwock@fenwick.com
    555 California Street, 12th Floor
    San Francisco, CA 94104
    Telephone: 415.875.2300
    Facsimile: 415.281.1350

    David Hayes (CSB No. 122894)
    dhayes@fenwick.com
    801 California Street
    Mountain View, CA 94041
    Telephone: 650.988.8500
    Facsimile: 650.938.5200

    Attorneys for Defendants
    DATABRICKS, INC. and
    MOSAIC ML, LLC, formerly MOSAIC ML,
    INC.