JEDEDIAH WAKEFIELD (CSB No. 178058)
jwakefield@fenwick.com
RYAN KWOCK (CSB No. 336414)
rkwock@fenwick.com
FENWICK & WEST LLP
555 California Street, 12th Floor
San Francisco, CA 94104
Telephone: 415.875.2300
Facsimile: 650.938.5200

DAVID HAYES (CSB No. 122894)
dhayes@fenwick.com
FENWICK & WEST LLP
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.

Additional counsel listed on signature page

FENWICK & WEST LLP

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

In Re Mosaic LLM Litigation	Master File Case No.: 3:24-cv-01451-CRB Consolidated with Case No. 3:24-cv-02653-CRB **DEFENDANTS' ANSWER TO FIRST** **AMENDED CONSOLIDATED COMPLAINT**

Defendants Databricks, Inc. ("Databricks") and Mosaic ML, LLC, formerly Mosaic ML, Inc. ("MosaicML") (together "Defendants") submit this Answer to Plaintiffs' First Amended Consolidated Complaint ("Complaint") filed by Plaintiffs Stewart O'Nan, Abdi Nazemian, Brian Keene, Rebecca Makkai, and Jason Reynolds on behalf of themselves and others similarly situated (collectively "Plaintiffs"). Unless specifically admitted, Defendants deny each of the allegations in the Complaint.

OVERVIEW[1]

1. Defendants admit that "artificial intelligence" is commonly abbreviated "AI." As Paragraph 1 pertains generally to AI and not any particular AI technologies of Defendants, Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 1, and on that basis deny them.

2. Defendants admit that MosaicML created and that Databricks distributed a large language model ("LLM") series called MosaicML Pretrained Transformer ("MPT"). Defendants admit that Databricks created and distributed an LLM series called DBRX after Databricks' acquisition of MosaicML. Defendants admit that LLMs may generate text outputs in response to user prompts. Defendants deny the remaining allegations in Paragraph 2.

3. As Paragraph 3 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 in Paragraph 3, and on that basis deny them. Defendants deny any suggestion that the LLMs at issue in this litigation have the goal of imitating protected expression.

4. Defendants admit that they have not compensated Plaintiffs to use materials when training the MPT Models, but Defendants further deny that such compensation would be required because the use of such materials to train the MPT Models would constitute fair use under U.S. copyright law. Defendants deny that MosaicML needed Plaintiffs' authorization to use any portions of the RedPajama – Books dataset to train MosaicML's MPT models. Defendants lack

---

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

FENWICK & WEST LLP

knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 4, and on that basis deny them.

5.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 5, and on that basis deny them.

6.   Defendants admit that Databricks is the corporate parent of MosaicML and also distributed DBRX models and certain MPT models. Defendants deny the remaining allegations in Paragraph 6.

7.   Defendants deny the allegations in Paragraph 7.

JURISDICTION AND VENUE

8.   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 8 and on that basis deny them.

9.   The allegations in this Paragraph state a legal conclusion to which no response is required. To the extent a response is deemed to be required, Defendants admit that the Court has personal jurisdiction over it with respect to this litigation and that venue is proper in the Northern District of California. Defendants deny the remaining allegations in Paragraph 9.

10.   The allegations in this Paragraph state a legal conclusion to which no response is required. To the extent a response is deemed to be required, Defendants admit that assignment of this case to the San Francisco Division is proper, and that this action purports to arise under intellectual property laws.

PLAINTIFFS

11.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 11, and on that basis deny them.

12.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 12, and on that basis deny them.

13.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations set forth in Paragraph 13, and on that basis deny them.

14.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations set forth in Paragraph 14, and on that basis deny them.

15.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations set forth in Paragraph 15, and on that basis deny them.

16.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 16 or as to the content of Exhibit A, and on that basis deny the allegations.

DEFENDANTS

17.   Defendants admit the allegations in Paragraph 17.

18.   Defendants admit that MosaicML is a Delaware limited liability company, is located at this address, and is a subsidiary of Databricks.

AGENTS AND CO-CONSPIRATORS

19.   Defendants deny the allegations in Paragraph 19.

20.   Defendants deny the allegations in Paragraph 20.

FACTUAL ALLEGATIONS

21.   Defendants admit that MosaicML was incorporated in 2020, and one of its goals was to provide tools to facilitate the efficient training of AI models.

22.   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 22.

23.   Defendants admit that the corpus of material used to train an LLM may be referred to as a "training dataset." As Paragraph 23 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 23, and on that basis deny them.

24.   Defendants admit that the MPT-7B LLM contains 6.7 billion parameters but otherwise deny the last sentence of this paragraph. As Paragraph 24 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 24, and on that

FENWICK & WEST LLP

1 basis deny them. Defendants deny any suggestion that the LLMs at issue in this litigation

2 progressively adjusted their output to more closely approximate any protected expression contained

3 in the training dataset.

4   25.   As Paragraph 25 pertains generally to LLMs and not any particular LLM of

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

6 of the allegations set forth in Paragraph 25, and on that basis deny them. Defendants deny any

7 suggestion that the LLMs at issue in this litigation have the goal of imitating any protected

8 expression ingested from the training dataset.

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

10 required. To the extent a response is deemed required, Defendants deny that MosaicML's

11 transformative use of works to train the MPT LLMs required consent, credit, or compensation.

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

13 remaining allegations set forth in Paragraph 26, and on that basis deny them.

14   27.   Defendants admit that MosaicML released the MPT-7B LLM in May 2023 and

15 published a blog post titled "Introducing MPT-7B: A New Standard for Open-Source,

16 Commercially Usable LLMs" available at https://www.databricks.com/blog/mpt-7b. To the extent

17 the allegations set forth in this Paragraph purport to summarize or characterize the contents of the

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

19   28.   To the extent the allegations set forth in this Paragraph purport to summarize or

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

21 document speaks for itself. Defendants deny the remaining allegations in Paragraph 28.

22   29.   To the extent the allegations set forth in this Paragraph purport to summarize or

23 characterize the contents of the website located at

24 https://huggingface.co/datasets/togethercomputer/RedPajama-Data-1T, the document speaks for

25 itself. Defendants lack knowledge and information sufficient to form a belief as to the truth of the

26 remaining allegations in Paragraph 29, and on that basis deny them.

27   30.   To the extent the allegations set forth in this Paragraph purport to summarize or

28 characterize the contents of the document located at https://arxiv.org/pdf/2101.00027.pdf, the

1   document speaks for itself. Defendants lack knowledge and information sufficient to form a belief

2   as to the truth of the remaining allegations in Paragraph 30, and on that basis deny them.

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

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

5   sufficient to form a belief as to the truth of the allegations in Paragraph 31, and on that basis deny

6   them.

7   32.   Defendants lack knowledge and information sufficient to form a belief as to the truth

8   of the allegations in Paragraph 32, and on that basis deny them.

9   33.   Defendants deny that they infringed Plaintiffs' alleged copyrights. Defendants lack

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

11   in Paragraph 33, and on that basis deny them.

12   34.   To the extent the allegations set forth in this Paragraph purport to summarize or

13   characterize the contents of the document located at

14   https://huggingface.co/datasets/the_pile_books3, the document speaks for itself. Defendants lack

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

16   in Paragraph 34, and on that basis deny them.

17   35.   Defendants admit that MosaicML completed training the MPT-7B LLM by May

18   2023. Defendants lack knowledge and information sufficient to form a belief as to the truth of the

19   remaining allegations in Paragraph 35, and on that basis deny them.

20   36.   To the extent the allegations in this Paragraph state a legal conclusion, no response

21   is required. To the extent a response is deemed required, Defendants admit that portions of the

22   RedPajama – Books dataset were used as training data for the MPT-7B LLM. Defendants deny

23   that they infringed Plaintiffs' alleged copyrights. Defendants lack knowledge and information

24   sufficient to form a belief as to the truth of the remaining allegations in Paragraph 36, and on that

25   basis deny them.

26   37.   To the extent the allegations in this Paragraph state a legal conclusion, no response

27   is required. To the extent a response is deemed required, Defendants admit that MosaicML released

28   the MPT-30B LLM in June 2023, that portions of the RedPajama – Books dataset were used as

1   training data for the MPT-30B LLM, and that the MPT-30B LLM contains approximately 30 billion

2   parameters. To the extent the allegations set forth in this Paragraph purport to summarize or

3   characterize the contents of a blog post, the document speaks for itself. Defendants deny that they

4   infringed Plaintiffs' alleged copyrights. Defendants lack knowledge and information sufficient to

5   form a belief as to the truth of the remaining allegations set forth in Paragraph 37, and on that basis

6   deny them.

7   38.   Defendants deny the allegations in Paragraph 38.

8   39.   Defendants admit that Databricks announced a new LLM series named DBRX on

9   March 27, 2024. To the extent the allegations in Paragraph 39 purport to summarize or characterize

10   the "Introducing DBRX: A New State-of-the-Art Open LLM" blog post on the Databricks website,

11   the blog post speaks for itself.

12   40.   Defendants admit that Databricks has not publicly identified the specific training

13   data used to train the DBRX models. To the extent the allegations in Paragraph 40 purport to

14   summarize or characterize the contents of the "Databricks spent $10M on new DBRX generative

15   AI model" article on techcrunch.com, the article speaks for itself.

16   41.   Defendants lack knowledge and information to admit or deny that the "shadow

17   library websites" include "Plaintiffs' Works." Defendants otherwise deny the allegations in

18   Paragraph 41.

19   42.   To the extent the allegations in Paragraph 42 refer to the "Introducing DBRX: A

20   New State-of-the-Art Open LLM" blog post on Databricks' website, the blog post speaks for itself.

21   43.   Defendants deny the allegations in Paragraph 43.

22   44.   To the extent the allegations in Paragraph 44 purport to summarize or characterize

23   the contents of a March 27, 2024 social media post on X, the social media post speaks for itself.

24   Defendants deny the remaining allegations in Paragraph 44.

COUNT 1
Direct Copyright Infringement (17 U.S.C. § 501) Against All Defendants

27   45.   Defendants incorporate by reference their responses to Paragraphs 1 through 44

28   above as though fully set forth herein.

FENWICK & WEST LLP

46.   Defendants lack knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 46, and on that basis deny them.

47.   Defendants admit that portions of the RedPajama – Books dataset were used as training data for the MPT-7B and MPT-30B LLMs. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations set forth in Paragraph 47, and on that basis deny them.

48.   Defendants admit that MosaicML released the MPT-7B-StoryWriter-65k+ LLM. 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 48.

49.   The Court's August 19, 2025 Order (Dkt. 162) dismissed Count 1 (Direct Copyright Infringement) of Plaintiffs' Complaint against Databricks. MosaicML lacks knowledge and information sufficient to form a belief as to the truth of the allegations in Paragraph 49, and on that basis denies them.

50.   The Court's August 19, 2025 Order (Dkt. 162) dismissed Count 1 (Direct Copyright Infringement) of Plaintiffs' Complaint against Databricks. Defendants deny that MosaicML needed Plaintiffs' authorization to use any portions of the RedPajama – Books dataset to train MosaicML's MPT models. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 50, and on that basis deny them.

51.   Defendants deny the allegations in Paragraph 51.

52.   Defendants deny the allegations in Paragraph 52.

53.   The Court's August 19, 2025 Order (Dkt. 162) dismissed Count 1 (Direct Copyright Infringement) of Plaintiffs' Complaint against Databricks.   Defendants deny that MosaicML needed Plaintiffs' authorization to use any portion of the RedPajama – Books dataset to train MosaicML's MPT models. Defendants deny that MosaicML violated any exclusive rights under the Copyright Act. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 53, and on that basis deny them.

54.   Defendants deny the allegations in Paragraph 54.

FENWICK & WEST LLP

55.   Defendants deny the allegations in Paragraph 55.

56.   Defendants deny the allegations in Paragraph 56.

COUNT 2
Vicarious Copyright Infringement Against Databricks

57.   Defendants incorporate by reference their responses to Paragraphs 1 through 56 above as though fully set forth herein.

58.   Defendants admit that Databricks acquired MosaicML in July 2023. Defendants deny the remaining allegations in Paragraph 58.

59.   Defendants deny the allegations in Paragraph 59.

60.   Defendants deny the allegations in Paragraph 60.

CLASS ALLEGATIONS

61.   The allegations in Paragraph 61 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 61.

62.   The allegations in Paragraph 62 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 62.

63.   The allegations in Paragraph 63 state legal conclusions or arguments to which no response is required.

64.   The allegations in Paragraph 64 state legal conclusions or arguments to which no response is required. To the extent a response is deemed required, Defendants deny that they possess information concerning the exact number of members of Plaintiffs' putative class. Defendants lack knowledge and information sufficient to form a belief as to the truth of the remaining allegations in Paragraph 64, and on that basis deny them.

65.   The allegations in Paragraph 65 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 65.

66.   The allegations in Paragraph 66 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 66.

67.   The allegations in Paragraph 67 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 67.

68.   The allegations in Paragraph 68 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 68.

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.

<u>DEFENSES</u>

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.

FENWICK & WEST LLP

FENWICK & WEST LLP

FIRST DEFENSE
(Failure to State a Claim)

Plaintiffs' and 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, the copying constituted fair use, 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 class members' claims fail in whole or in part because the models created by Defendants 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 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 class members' claims are barred in whole or in part by the doctrines of waiver, abandonment, and/or forfeiture.

SEVENTH DEFENSE
(License)

Plaintiffs' and 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 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 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 and Innocent Conduct)

Plaintiffs' and class members' claims for statutory damages must be reduced because Defendants' conduct was innocent, not willful.

ELEVENTH DEFENSE
(Lack of Injury)

Plaintiffs' and 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.

TWELFTH DEFENSE
(Failure to Mitigate Damages)

Plaintiffs and 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.

FOURTEENTH DEFENSE
(Class Cannot be Certified)

The purported class cannot be certified under Federal Rule of Civil Procedure 23 because, inter alia, the purported class, class representatives and/or class counsel fail to meet the typicality, commonality, adequacy, superiority, and predominance requirements for class actions.

RESERVATION OF ADDITIONAL DEFENSES

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.

DEFENDANTS' DEMAND FOR JUDGMENT

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;

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

2 and

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

4

5 Dated: September 2, 2025   FENWICK & WEST LLP

6

7   By: */s/ Jedediah Wakefield*
                                                 Jedediah Wakefield (CSB No. 178058)

8   jwakefield@fenwick.com
                                                 Ryan Kwock (CSB No. 336414)

9   rkwock@fenwick.com
                                                 555 California Street, 12th Floor

10   San Francisco, CA 94104
                                                 Telephone: 415.875.2300

11   Facsimile: 650.938.5200

12   David Hayes (CSB No. 122894)
                                                 dhayes@fenwick.com

13   Diana C. Buck (CSB No. 339314)
                                                 dbuck@fenwick.com

14   801 California Street
                                                 Mountain View, CA 94041

15   Telephone: 650.988.8500
                                                 Facsimile: 650.938.5200

16

17   Brian D. Buckley (admitted *pro have vice*)
                                                 bbuckley@fenwick.com

18   Deena Feit (admitted *pro hac vice*)
                                                 dfeit@fenwick.com

19   401 Union Street, 5th Floor
                                                 Seattle, WA 98101

20   Telephone: 206.389.4510
                                                 Facsimile: 650.938.5200

21

22   Charles Moulins (admitted *pro hac vice*)
                                                 cmoulins@fenwick.com

23   Justine Vandermel (admitted *pro hac vice*)
                                                 justine.vandermel@fenwick.com

24   902 Broadway, 18th Floor
                                                 New York, NY 10010

25   Telephone: 212.430.2600
                                                 Facsimile: 650.938.5200

26

27   Zachary Harned (CSB No. 335898)
                                                 zharned@fenwick.com

28

DEFENDANTS' ANSWER TO FIRST   13   Case No.: 3:24-cv-01451-CRB
AMENDED CONSOLIDATED COMPLAINT

FENWICK & WEST LLP

FENWICK & WEST LLP

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

730 Arizona Avenue, 1st Floor
Santa Monica, CA 90401
Telephone: 310.434.5400
Facsimile: 650.938.5200

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