



S-255562

No.
Vancouver Registry

In the Supreme Court of British Columbia

Between

James Bernard MacKinnon

Plaintiff

and

Databricks Inc. and MosaicML Inc.

Defendants

Brought under the *Class Proceedings Act*, R.S.B.C. 1996, c. 50

NOTICE OF CIVIL CLAIM

This action has been started by the plaintiff for the relief set out in Part 2 below.

If you intend to respond to this action, you or your lawyer must

- (a) file a response to civil claim in Form 2 in the above-named registry of this court within the time for response to civil claim described below, and
- (b) serve a copy of the filed response to civil claim on the plaintiff.

If you intend to make a counterclaim, you or your lawyer must

- (c) file a response to civil claim in Form 2 and a counterclaim in Form 3 in the above-named registry of this court within the time for response to civil claim described below, and
- (d) serve a copy of the filed response to civil claim and counterclaim on the plaintiff and on any new parties named in the counterclaim.

JUDGMENT MAY BE PRONOUNCED AGAINST YOU IF YOU FAIL to file the response to civil claim within the time for response to civil claim described below.

Time for response to civil claim

A response to civil claim must be filed and served on the plaintiff,

if you were served with the notice of civil claim anywhere in Canada, within 21 days after that service,

(e) if you were served the notice of civil claim anywhere in the United States of America, within 35 days after that service,

(f) if you were served with the notice of civil claim anywhere else, within 49 days after that service, or

(g) if the time for response to civil claim has been set by order of the court, within that time.

PART 1: STATEMENT OF FACTS

Introduction

1. MosaicML Inc. ("Mosaic") and its parent company Databricks Inc. ("Databricks") develop and market artificial intelligence ("AI") large language models. In training their models they relied on a large library of books, believing that using books instead of other lower quality text would give an advantage over its competitors and make for a better product. However, rather than license books to use for this purpose, or rely only on books that were in the public domain, they instead relied on a pirated data set found online that they knew contained unlicensed copyrighted material. They made no effort to pay the owners of these copyrights to obtain or use their works. They made no effort to ensure that their actions did not infringe the moral rights of authors. Instead, they set out to build tools, including a model called "StoryWriter", whose express purposes include competing with authors, and which diminish the market for original human-authored written works.
2. The Plaintiff is an author whose works were used without his consent, license, or permission by the Defendants to train their AI large language models. The Plaintiff

brings a claim in copyright and unjust enrichment on his own behalf and on behalf of a class of persons resident in Canada who also had their copyrighted materials used without their license or permission by the Defendants to train their AI large language models.

The Plaintiff

3. The Plaintiff, James Bernard MacKinnon, is an author, journalist, editor and educator. The Plaintiff is commonly cited in his publications as J.B. MacKinnon. The Plaintiff has an address for service at 400-856 Homer St, Vancouver, British Columbia, and has been a resident of British Columbia throughout the Class Period (defined below).
4. The Plaintiff is the author of, and owns the copyright to, *The Once and Future World*, published in 2013, a national bestselling non-fiction work ("*The Once and Future World*").
5. The Plaintiff is the co-author of, and co-owns the copyright to *The 100-Mile Diet*, published in 2007, a national bestselling non-fiction work ("*The 100-Mile Diet*") (Co-authored and copyright co-owned with Alisa Smith).

Defendants

6. The defendant, Databricks, is a Delaware company with its principal place of business at 160 Spear Street, 15th Floor, San Francisco, California.
7. Databricks was founded in 2013.
8. The defendant, Mosaic, is a Delaware corporation with its principal place of business at 51 2nd Street, Suite 202, San Francisco, California.
9. Databricks acquired Mosaic in July 2023 for approximately \$1.3 billion USD.
10. Mosaic operates as a subsidiary of Databricks.

11. At all material times following the acquisition of Mosaic by Databricks, the Defendants have functioned as an ongoing, organized and continuing business unit sharing common purposes and objectives. The Defendants are and were agents of each other and each is vicariously responsible for the acts and omissions of the other as particularized herein.
12. Databricks operates an office at 1055 West Georgia Street, Suite 2412, in Vancouver, British Columbia.

The Class and the Class Period

13. This action is brought on behalf of all members of a class consisting of the Plaintiff and all holders of Canadian copyright in works used by the Defendants to train one of the various versions of the Defendants' LLMs (defined below) (the "Class" and/or the "Class Members").

Large Language Models

14. Large Language Models ("LLMs") are a form of AI software that are designed to comprehend, and to generate, natural language.
15. In order to construct an LLM it is necessary to "train" it on vast arrays of text. Specifically, the software is fed with huge amounts of text from a variety of sources, from which the machine "learns" how to comprehend and generate language. The text fed to the LLM is called the training data set.
16. An LLM is programmed to copy each text it is fed and to extract information from it. It uses this information to adjust its output so that this output will more closely resemble what it copied from the initial data set. Eventually, an LLM may be able to generate effective simulations of natural written language.

The Defendants' LLMs

17. In May 2023, Mosaic released an LLM called MPT-7B.

18. In training the MPT-7B, Mosaic ML used and copied a database called “RedPajama – Books” (the “RedPajama Dataset”).
19. The RedPajama Dataset at the time it was used by Mosaic to train MPT-7B included within it a database called “Books3” (the “Books3 Dataset”).
20. The Books3 Dataset was accurately described in a paper entitled “The Pile: An 800gb dataset of diverse text for language modeling” by Leo Gao *et al* in the following way:

Books3 is a dataset of books derived from a copy of the contents of the Bibliotik private tracker made available by Shawn Presser (Presser, 2020). Bibliotik consists of a mix of fiction and nonfiction books and is almost an order of magnitude larger than our next largest book dataset (BookCorpus2). We included Bibliotik because books are invaluable for long-range context modeling research and coherent storytelling.

21. Bibliotik contains approximately 196,640 books, including books by Canadian authors with active copyright protection. Mr. Presser has confirmed publicly, and the fact is, that the Books3 dataset contains all of the Bibliotek books.
22. Mosaic downloaded the RedPajama Dataset, including the Books3 Dataset, without paying a licensing fee to the Plaintiff or class members, and used it to train MPT-7B.
23. The Books3 Dataset included *The Once and Future World*, which the Plaintiff authored and which he holds copyright to, and *The 100-Mile Diet*, which the Plaintiff co-authored and which he co-owns copyright to. In respect of *The 100-Mile Diet*, the Books3 dataset included both the Canadian version and the United States version of the book (*The 100-Mile Diet* was published in the United States as “*Plenty: Eating Locally on the 100-Mile Diet*”).
24. Mosaic used *The Once and Future World* and *The 100-Mile Diet*, obtained off of the RedPajama Dataset, to train MPT-7B, which involved making copies of those works. Neither of the Defendants offered or paid a licensing fee to the Plaintiff, or otherwise obtained his consent, to acquire or use *The Once and Future World* and *The 100-Mile Diet* in this way.

25. The RedPajama and Books3 Datasets contained other copyrighted material, including books whose copyright is owned by the Class Members and which is protected under Canadian law.
26. Neither the creators of the RedPajama Dataset nor the creators of the Books3 Dataset licensed the copyrighted work it contained in any way, and did not license *The Once and Future World* or *The 100-Mile Diet*.
27. The Defendants knew at all times that the RedPajama and Books3 Datasets contained copyrighted material.
28. The Defendants did not pay the authors of the copyrighted material contained in the RedPajama and Books3 Datasets in order to access or use their work, and did not otherwise obtain a license or consent from Class Members to use copyrighted material in the datasets.
29. The Defendants knew at all material times that they did not have a right to use the copyrighted material contained in the RedPajama and Books3 Datasets.
30. Mosaic trained and released various successor versions to MPT-7B, including MPT-30B, released in June 2023.
31. Mosaic has released a model called MPT-7B-StoryWriter-65k+ (the “StoryWriter Model”), whose stated purpose is to “read and write stories”.
32. The Defendants released a new general purpose LLM called DBRX in March 2024 (collectively, the LLMs released by the Defendants are the “Defendants’ LLMs”).
33. Training of each of the Defendants’ LLMs relied on and used, or builds off of and incorporates, the copyrighted material taken from the RedPajama and Books3 Datasets.
34. The Defendants decided to train their large language models on books, including newer books still under copyright, because they believed doing so would improve its model and give it an advantage over its competition.

35. The Defendants make their LLM products available to purchase, including on a “pay as you go” model where users pay based on their usage of the products.
36. The revenue that Mosaic generates belongs to Databricks, as its parent company.
37. The Defendants used software to remove copyright management information from the books downloaded to train the Defendants’ LLMs, including books covered by copyright held by the Plaintiff and other Class Members. They did not have the permission of the copyright holders to alter their works to remove copyright information.
38. The Defendants knew at all material times that copyrighted e-books are sold with technological protection measures that prevent unauthorized copying of the books, including copying those works as part of training an LLM.
39. The Defendants knew at all material times that most if not all pirated books such as those contained in the RedPajama and Books3 Datasets have had the technological protection measures that preclude unauthorized copying removed.
40. By acquiring and using versions of copyrighted e-books that had the technological protection measures removed, the Defendants avoided the technological protection measure that would have otherwise prevented it from using copyrighted e-books to train the Defendants’ LLMs.
41. The Defendants’ LLMs are available in Canada.
42. Collectively, the material copyrighted under Canadian law owned by authors resident in Canada used by the Defendants to train the Defendants’ LLMs, including material in the RedPajama and Books3 Datasets, is “the Copyrighted Works”.

The Defendants’ Concealment

43. The Defendants took steps to conceal the fact that it had infringed copyright in training the Defendants’ LLMs. These steps include:

- a. Training the Defendants' LLMs to respond in a misleading way if a user asked if it was built off of copyrighted material;
- b. Removing copyright management information from material before it was fed into the Defendants' LLMs so that the LLM did not itself learn that it was built off of copyrighted material.

The Defendants Benefited

44. Training the LLM on copyrighted materials including the Copyrighted Works has improved the performance of the Defendants' LLMs.
45. The Defendants have monetized, currently monetize, and plan in the future to continue to monetize, both directly and indirectly, the Defendants' LLMs.
46. Databricks has a valuation of over \$60 billion USD.

The Plaintiff and the Class Members Suffered a Loss

47. The Plaintiff and the Class Members held copyright in the Copyrighted Works. This copyright was infringed, which inherently causes a loss to a copyright holder.
48. Further, the Plaintiff and the Class Members lost the opportunity to choose to license their copyright in the Copyrighted Works to the Defendants, for some or any purpose that the Plaintiff and the Class Members may see fit.
49. The Plaintiff and the Class Members lost the opportunity to choose not to contribute to the development of the Defendants' LLMs, which pose a threat to people who do creative work, including authors.
50. The Defendants' LLMs pose a threat to the market for the creative written work that the Plaintiff and the Class Members produce.
51. The Plaintiff and the Class Members hold moral rights to the integrity of their work and the right to be associated with the work or remain anonymous. The Defendants' use of the works to train the Defendants' LLMs harmed the Plaintiff and the Class

Member's moral rights by undermining the integrity of the works and associating the works with a project that ultimately poses a threat to the business and livelihood of authors, and to the literacy of society generally.

The Defendants Circumvented Technological Protection Measures

52. Some of the Copyrighted Works were made available by their publisher as electronic books (the "Copyrighted E Books")
53. The Copyrighted E Books were protected from unauthorized copying by encryption and/or digital rights management ("DRM") tools such as access control technology (collectively, encryption and DRM are the "Technological Protection Measures").
54. By downloading the Copyrighted E Books from the Books3 data sets, the Defendants circumvented the Technological Protection Measures.
55. The Defendants removed and altered rights management information in electronic form without the consent of the owner of the copyright in the work by using software to remove copyright management information in the Copyrighted Works, including the Copyrighted E Books.
56. The Defendants removed and altered rights management information in electronic form without the consent of the Plaintiff by using software to remove copyright management information in his works.

The Plaintiff and the Class Members Were Given No Opportunity to Sell or License

57. The Defendants provided copyright holders in Canada, including the Plaintiff and other Class Members, no opportunity to license their books to the Defendants for the purposes of training the Defendants' LLMs.

The Defendants Knew

58. The Defendants at all relevant times knew or ought to have known that the Copyrighted Works were copyrighted.

59. The Defendants at all relevant times knew or ought to have known that they did not have a license or otherwise have permission to download the Copyrighted Works, or to use the Copyrighted Works to train the Defendants' LLMs.
60. The Defendants determined that, rather than attempt to obtain a license, they would violate copyright and face whatever consequences arose. In so doing they acted in a high handed manner without regard for the rights of the Plaintiff or the Class Members.
61. The Defendants at all relevant times knew or ought to have known that the Copyrighted E Books were protected by Technological Protection Measures, and that they were circumventing such measures by acquiring the Copyrighted Works through the RedPajama and Books3 Datasets.

PART 2: RELIEF SOUGHT

62. The Plaintiff claims on their own behalf and on behalf of the proposed Class Members:
 - a. An order pursuant to s. 4 of the *Class Proceedings Act*, [RSBC 1996] c. 50, ("*CPA*") certifying this action as a class proceeding;
 - b. An order pursuant to s. 2 of the *CPA* appointing the Plaintiff as the representative Plaintiff;
 - c. A declaration that the Defendants are liable for infringing, authorizing, and or inducing the infringement of the copyright and moral rights owned by the Plaintiff and the class members in the Copyrighted Works;
 - d. A declaration that the Defendants are liable for circumventing a technological protection measure contrary to ss. 41 and 41.1 of the *Copyright Act*, R.S.C. 1985, c. 42;

- e. A declaration that the Defendants are liable for knowingly removing or altering rights management information in electronic form contrary to s. 41.22(1) of the *Copyright Act*;
- f. A declaration that the Defendants are liable for unjustly enriching themselves at the expense of the Plaintiff and the Class Members;
- g. Damages, including damages under ss. 35, 41.1 and 41.22 of the *Copyright Act*, and/or statutory damages under s. 38.1 of the *Copyright Act*;
- h. An order pursuant to ss. 35 and/or 41.1 of the *Copyright Act* for, in addition to damages, such part of the profits that the Defendants have made from the infringement that were not taken into account in calculating the damages that the court considers just;
- i. An accounting and disgorgement of profits and/or restitution in respect of the Defendants' unjust enrichment;
- j. Aggregate damages;
- k. Punitive damages for the Defendants' repeated, wilful, and knowing infringement of the Plaintiff and the Class Members' copyright and moral rights in the Copyrighted Works;
- l. A permanent injunction pursuant to s. 34, 41.1 and 41.22 of the *Copyright Act* restraining the Defendants, their directors, officers, employees, agents, licensees, successors, assigns, related or affiliated companies, and all those under the control of the Defendants from, directly or indirectly, infringing copyright or moral rights in the Copyrighted Works or circumventing the Technological Protection Measures;
- m. Pre-judgement and post-judgement interest pursuant to the *Court Order Interest Act*;

- n. Costs of all notices and of administering the plan of distribution of the judgment in this action, together with applicable taxes; and
- o. Such further and other relief as this Honourable Court may deem just.

PART 3: LEGAL BASIS

63. The Plaintiff plead and relies upon:

- a. The *Class Proceedings Act*;
- b. The *Copyright Act*;
- c. The *Court Order Interest Act*, [R.S.B.C. 1996], c. 79; and
- d. The Supreme Court Civil Rules, B.C. Reg. 168/2009.

Causes of Action

Breach of Copyright Act

- 64. The Plaintiff and the Class Members are the owners of the copyright and the moral rights in the Copyrighted Works.
- 65. Without license or permission from the Plaintiff and the Class Members, the Defendants have reproduced and continues to reproduce and exploit the Copyrighted Works (or a substantial part thereof), contrary to ss. 3 and 27 of the *Copyright Act*.
- 66. The Defendants have infringed the moral rights of the class members, contrary to s. 28.1 of the *Copyright Act*.
- 67. The Defendants have circumvented technological protection measures contrary to s. 41.1 of the *Copyright Act*.
- 68. The Defendants removed and altered rights management information in electronic form without the consent of the owner of the copyright contrary to s. 41.22 of the *Copyright Act*.

69. The Defendants continue to breach the *Copyright Act* as described above.

Unjust Enrichment

70. The Defendants have benefited from its use of the Copyrighted Work without license or permission as that work allowed it to improve the potentially extremely lucrative Defendant's LLMs.

71. The Plaintiff and the Class Members have suffered a corresponding deprivation as they lost the ability to either license their work to the Defendants or choose to not participate in the development of the Defendants' LLMs.

72. There is no juristic reason for this benefit and corresponding deprivation.

Remedies

Damages

73. Pursuant to s. 34 of the *Copyright Act*, the Plaintiff and the Class Members are entitled to all remedies including by way of damages, accounts, delivery up and otherwise that are or may be conferred by law for the infringement of a right. As a result of the Defendants' acts and omissions, as particularized above, the Plaintiff and the Class Members have suffered losses and damages. These damages include loss of control of their copyrighted works and moral rights, and the loss of licensing revenue.

74. The Plaintiff and the Class Members claim damages under ss. 35, 41.1 and 41.22 of the *Copyright Act*, and may elect at any time prior to final judgement being rendered to pursue statutory damages under s. 38.1 of the *Copyright Act*.

75. The Plaintiff and the Class Members claim an accounting of profits and disgorgement and/or restitution in respect of the Defendants' unjust enrichment.

Punitive Damages

76. The Defendants' conduct in acquiring, sharing, and using the Copyrighted Works was high handed, arrogant, and displayed a reckless disregard for the rights of the

Class Members. The Plaintiff asks this Court to award punitive damages against the Defendants in an amount deemed appropriate by the Court.

Injunction

77. The Plaintiff asks this Court to issue an injunction pursuant to s. 34, 41.1 and 41.22 of the *Copyright Act* restraining the Defendants, their directors, officers, employees, agents, licensees, successors, assigns, related or affiliated companies, and all those under the control of the Defendants from, directly or indirectly, infringing copyright or moral rights in the Copyrighted Works or circumventing the Technological Protection Measures. This includes publishing or making available in British Columbia any LLM built off of or integrating infringement of the Plaintiff and the Class Members' copyright.

Jurisdiction

78. This action has a real and substantial connection with British Columbia because, among other things, it concerns a business carried on in British Columbia, property located in British Columbia, a tort that caused damage to the Plaintiff in British Columbia, restitutionary obligations arising in British Columbia, is a claim for an injunction ordering a party to do or refrain from doing anything in British Columbia.

Plaintiff's address for service:

CFM LAWYERS LLP
#400 – 856 Homer Street
Vancouver, BC V6B 2W5
Tel: (604) 689-7555
Fax: (604) 689-7554
Email: service@cfmlawyers.ca

Place of trial: Vancouver Law Courts

Address of the registry: 800 Smithe Street, Vancouver, BC V6Z 2E1

Date: 24/July/2025



Signature of lawyers
for Plaintiff

Reidar Mogerman and Oliver Pulleyblank

**ENDORSEMENT ON ORIGINATING PLEADING OR PETITION FOR SERVICE
OUTSIDE BRITISH COLUMBIA**

The Plaintiff claims the right to serve this pleading/petition on the Defendants outside British Columbia on the ground that:

The Plaintiff has at all material times been a resident of British Columbia and has suffered loss in British Columbia. This Action concerns a business carried on in British Columbia, property in British Columbia, a tort that was committed in British Columbia, is a claim for an injunction ordering a party to do or refrain from doing anything in British Columbia, and concerns restitutionary obligations that, to a substantial extent, arose in British Columbia. The Supreme Court of British Columbia has jurisdiction with respect to this matter and the Plaintiff pleads the *Court Jurisdiction and Proceedings Transfer Act*, 2003, SBC Chapter 28 and amendments thereto.

Rule 7-1 (1) of the Supreme Court Civil Rules states:

- (1) Unless all parties of record consent or the court otherwise orders, each party of record to an action must, within 35 days after the end of the pleading period,
 - (a) prepare a list of documents in Form 22 that lists
 - (i) all documents that are or have been in the party's possession or control and that could, if available, be used by any party at trial to prove or disprove a material fact, and
 - (ii) all other documents to which the party intends to refer at trial, and
 - (b) serve the list on all parties of record.

APPENDIX

CONCISE SUMMARY OF NATURE OF CLAIM:

A claim on behalf of persons resident in Canada who, during the Class Period, had their copyrighted works used by the Defendants without license of permission to train its AI Large Language Models.

THIS CLAIM ARISES FROM THE FOLLOWING:

A personal injury arising out of:

- ☐ a motor vehicle accident
- ☐ medical malpractice
- ☐ another cause

A dispute concerning:

- ☐ contaminated sites
- ☐ construction defects
- ☐ real property (real estate)
- ☐ personal property
- ☐ the provision of goods or services or other general commercial matters

- ☐ investment losses
- ☐ the lending of money
- ☐ an employment relationship
- ☐ a will or other issues concerning the probate of an estate
- ☒ a matter not listed here

THIS CLAIM INVOLVES:

- ☒ a class action
- ☐ maritime law
- ☐ aboriginal law
- ☐ constitutional law
- ☐ conflict of laws
- ☐ none of the above
- ☐ do not know