

CANADA
PROVINCE OF QUÉBEC
DISTRICT OF MONTRÉAL
NO:

500-06-001406-251

SUPERIOR COURT
(Class Action)

TARAS GRESOE

Applicant

v.

DATABRICKS INC., a legal person, duly constituted under the laws of Delaware, having an office at 1055 West Georgia Street, Suite 2412, in Vancouver, British Columbia.

AND

MOSAICML INC., a legal person, duly constituted under the laws of Delaware, having its principal place of business at 51 2nd Street, Suite 202, San Francisco, California.

Defendants

**APPLICATION FOR AUTHORIZATION TO INSTITUTE A CLASS ACTION
AND TO OBTAIN THE STATUS OF REPRESENTATIVE
(s. 571 CCP and following)**

**TO ONE OF THE HONOURABLE JUSTICES OF THE QUÉBEC SUPERIOR COURT,
SITTING IN AND FOR THE DISTRICT OF MONTRÉAL, THE APPLICANT STATES AS
FOLLOWS:**

PART I: INTRODUCTION

A. Background

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 Applicant is an author whose works were used without his consent, license or permission by the Defendants to train its AI large language models. The Applicant brings a claim in copyright and unjust enrichment on his own behalf and on behalf of a class of persons resident in Quebec who also had their copyrighted materials used without their license or permission by the Defendants to train their AI large language models.

B. The Applicant

3. The Applicant, Taras Grescoe, is a writer and educator. The Applicant has an address for service at 85 rue Saint-Paul Ouest, #410, Montréal (Québec) H2Y 3V4, and is a resident of Quebec.
4. The Applicant is the author of, and owns the copyright to, books including the following titles, as appears from the Applicant's website, attached herein as

Exhibit P-1:

- The Devil's Picnic: Around the World in Pursuit of a Forbidden Fruit (2005) ("Devil's Picnic")
- Bottomfeeder: How to Eat Ethically in a World of Vanishing Seafood (2008) ("Bottomfeeder")

- Straphanger: Saving our Cities and Ourselves from the Automobile (2011) (“Straphanger”)
5. The applicant has won numerous awards for his books, including the Edna Staebler Award for Creative Non-Fiction, the Mavis Gallant Prize for Non-Fiction, the McAulsan First Book Prize, the Hilary Weston Writers' Trust Prize for Nonfiction, and the International Association of Culinary Professionals Award for Literary Food Writing.

C. The 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, as appears in a July 19, 2023 press release by DataBricks, attached herein as **Exhibit P-10** as well as in a June 26, 2023 article in Techcrunch.com, attached herein as **Exhibit P-11**.
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.

D. The Class

12. The Applicant wishes to institute a class action on behalf of the following group, of which they are a member:

all natural persons or corporations, resident in Québec who owned or co-owned the copyright to works used by the Defendants to train one of the various versions of the Defendants' Large Language Models (the "Class" and/or the "Class Members").

or such other class definition as may be approved by the Court.

E. DEFINED TERMS

13. The following definitions apply for the purpose of this application to authorize the bringing of a class action:

- (a) "**CCQ**" means *Civil Code of Québec*, CQLR c CCQ-1991.
- (b) "**CCP**" means Code of Civil Procedure, CQLR c C-25.01.
- (c) "**Copyright Act**" means *Copyright Act*, R.S.C., 1985, c. C-42.
- (d) "**Charter**" means *Charter of Human Rights and Freedoms*, CQLR c C-12.

PART II: STATEMENT OF FACTS

A. Large Language Models

14. Large Language Models ("LLMs") are a form of AI software that are designed to comprehend, and to generate, natural language, as appears in pages 4-13 of US Copyright Report re AI, attached herein as **Exhibit P-2**.
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 programed 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.

B. The Defendants' LLMs

17. In May 2023, Mosaic released an LLM called MPT-7B, as appears in a May 5, 2023 DataBricks Blog Entry, attached herein as **Exhibit P-6**.
18. In training the MPT-7B, Mosaic ML used and copied a database called "RedPajama – Books" (the "RedPajama Dataset"), as appears in **Exhibit P-6**.
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"), as appears as of April 17, 2023 on HuggingFace Page, attached herein as **Exhibit P-8**.
20. The Books3 Dataset was 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, as appears in **Exhibit P-3**, attached herein:

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, as appears in page 3 of the US SUMMARY JUDGMENT DECISION, per **Exhibit P-13**.

22. Mr. Presser has confirmed publicly that the Books3 dataset contains all of the Bibliotek books, as appears in the Atlantic article by Alex Reisner “Revealed: The Authors Whose Pirated Books are Powering Generative AI”, attached herein as **Exhibit P-4**.
23. 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.
24. The Books3 Dataset included Bottomfeeder, Straphanger and The Devil's Picnic, all of which the Applicant authored and which he holds copyright to, as appears in the Atlantic's search tool, attached herein as **Exhibit P-5**.
25. Mosaic used Bottomfeeder, Straphanger and The Devil's Picnic obtained off of the RedPajama Dataset (which included the Books3 Dataset) to train MPT-7B, which involved making copies of those works. Neither of the Defendants offered or paid a licensing fee to the Applicant, or otherwise obtained his consent, to obtain or use Bottomfeeder, Straphanger and The Devil's Picnic in this way.
26. 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.
27. Neither the creators of the RedPajama Dataset nor the creators of the Books3 Dataset licensed the copyrighted work they contained in any way, and did not license Bottomfeeder, Straphanger and The Devil's Picnic.
28. The Defendants knew at all times that the RedPajama and Books3 Datasets contained copyrighted material, yet nevertheless downloaded and used the dataset.

29. The Defendants did not pay the authors of the copyrighted material contained in the Books3 Dataset 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 RedPajama and Books3 Datasets.
30. The Defendants knew at all materials times that they did not have a right to use the copyrighted material contained in the RedPajama and Books3 Datasets.
31. Mosaic trained and released various successor versions to MPT-7B, including MPT-30B, released in June 2023, as appears in the DataBricks Blog entry for June 22, 2023, attached herein as **Exhibit P-7**.
32. Mosaic has released a model called MPT-7B-StoryWriter-65k+ (the "StoryWriter Model"), whose stated purpose is to "read and write stories", as appears in **Exhibit P-6**.
33. The Defendants released a new general purpose LLM called DBRX in March 2024 (collectively, the LLMs released by the Defendants are the "Defendants' LLMs"), as appears in the March 27, 2024 DataBricks blog entry, attached herein as **Exhibit P-9**.
34. 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.
35. The Defendants decided to train their large language model on books, including newer books still under copyright, because it believed doing so would improve its model and give it an advantage over its competition.
36. 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.

37. The revenue that Mosaic generates belongs to Databricks, as its parent company.
38. 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 Applicant and other Class Members. They did not have the permission of the copyright holders to alter their works to remove copyright information.
39. 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.
40. 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.
41. 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.
42. The Defendants' LLMs are available in Canada.
43. 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".

C. The Defendants' Concealment

44. 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.

D. The Defendants Benefited

- 45. Training the LLM on copyrighted materials including the Copyrighted Works has improved the performance of the Defendants' LLMs.
- 46. The Defendants has monetized, currently monetizes, and plans in the future to continue to monetize, both directly and indirectly, the Defendants' LLMs.
- 47. Databricks has a valuation of over \$60 billion USD, as appears in the December 17, 2024 DataBricks press release, attached herein as **Exhibit P-12**.

E. The Applicant and the Class Members Suffered a Loss

- 48. The Applicant and the Class Members held copyright in the Copyrighted Works. This copyright was infringed, which inherently causes a loss to a copyright holder.
- 49. Further, the Applicant 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 Applicant and the Class Members may see fit.
- 50. The Applicant 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.
- 51. The Defendants' LLMs pose a threat to the market for the creative written work that the Applicant and the Class Members produce.

52. The Applicant 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 Applicant 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.

F. The Defendants Circumvented Technological Protection Measures

53. Some of the Copyrighted Works were made available by their publisher as electronic books (the "Copyrighted E Books")
54. 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").
55. By downloading the Copyrighted E Books from the Books3 data sets, the Defendants circumvented the Technological Protection Measures.
56. 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.
57. The Defendants removed and altered rights management information in electronic form without the consent of the Applicant by using software to remove copyright management information in his works.

G. The Applicant and the Class Members Were Given No Opportunity to Sell or License

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

H. The Defendants Knew

59. The Defendants at all relevant times knew or ought to have known that the Copyrighted Works were copyrighted.
60. The Defendants at all relevant times knew or ought to have known that it 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.
61. 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 Applicant or the Class Members.
62. 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 it was circumventing such measures by acquiring the Copyrighted Works through the RedPajama and Books3 datasets.

PART III: LEGAL BASIS OF APPLICATION

A. Breach of Copyright Act

63. The Applicant and the Class Members are the owners of the copyright and the moral rights in the Copyrighted Works.

64. Without license or permission from the Applicant 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*.
65. The Defendants have infringed the moral rights of the class members, contrary to s. 28.1 of the *Copyright Act*.
66. The Defendants have circumvented technological protection measures contrary to s. 41.1 of the *Copyright Act*.
67. 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*.
68. The Defendants continue to breach the *Copyright Act* as described above.

B. Unjust Enrichment

69. 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 Defendants' LLMs.
70. The Applicant 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.
71. There is no juristic reason for this benefit and corresponding deprivation.

PART IV: Remedies

A. Damages

72. Pursuant to s. 34 of the *Copyright Act*, the Applicant and the Class Members are entitled to all remedies including by way 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 Applicant and the Class Members have suffered losses and damages. These damages include loss of control of their copyrighted works and the loss of licensing revenue.

73. The Applicant 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*.
74. The Applicant and the Class Members claim an accounting of profits and disgorgement and/or restitution in respect of the Defendants' unjust enrichment.

B. Punitive Damages

75. 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.
76. Furthermore, each of the above acts constitutes an unlawful and intentional infringement of the Class Members' rights to integrity and dignity (s. 4) and to the free enjoyment and disposal of their property (s. 6) under the *Charter*.
77. The Applicant asks this Court to award punitive damages against the Defendants in an amount deemed appropriate by the Court.

C. Injunction

78. The Applicant 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 in the Copyrighted Works or circumventing the Technological Protection Measures. This includes publishing or making available in Quebec the Defendants' LLM built off of or integrating infringement of the Applicant and the Class Member's copyright.

PART V: Facts Giving Rise to an Individual Action by the Applicant

79. The Applicant is the author of, and owns the copyright to, the following works:
 - (a) The Devil's Picnic (2005)
 - (b) Bottomfeeder: How to Eat Ethically in a World of Vanishing Seafood (2008)
 - (c) Straphanger: Saving our Cities and Ourselves from the Automobile (2011)
80. The Applicant pleads that the following works he authored are included in the RedPajama and Books3 Datasets used by the Defendants to train their LLMs:
 - Bottomfeeder
 - Straphanger
 - The Devil's Picnic
81. The Applicant learned that his work had apparently been pirated and used by the Defendants to train their LLM in June 2025.
82. The Applicant is concerned that LLMs will affect his ability to maintain a living working as an author.
83. The Applicant never consented to the Defendants using his work to train their LLMs, or to the Defendants acquiring pirated copies of his work.
84. The Applicant was not paid by the Defendants for the use of his works.
85. By making use of the Applicant's works to train its LLM and the Defendants were able to create a marketable product which it sold for profit without providing any of the proceeds to the Applicant.

86. The Applicant pleads that the use the Defendants made of his work to train their LLMs violated his moral rights in his works under s. 14.1 of the *Copyright Act*, that no attempt was made to secure the moral rights, and that this constitutes a clear infringement per s. 28.1 of the *Copyright Act*.
87. Each of the above acts constitutes an unlawful and intentional infringement of the Applicant's rights to integrity and dignity (s. 4) and to the free enjoyment and disposal of his property (s. 6) under the *Charter*.
88. The Applicant therefore claims damages for the violation of his moral rights and the infringement on his works, as well as punitive damages.

PART VI: FACTS GIVING RISE TO INDIVIDUAL ACTION BY EACH CLASS MEMBER

89. The cause of action and the legal basis of the recourse of each of the Class Members against the Defendants are the same as those of the Applicant.
90. The faults committed by the Defendants with regard to the Class Members are the same as those committed with regard to the Applicant, as detailed above.
91. More particularly, none of the Class Members consented to the Defendants using their work to train its LLMs, or to the Defendants acquiring pirated copies of their work.
92. The Defendants used the Class Members' works to train their LLMs and in so doing violated their moral rights in their works under s. 14.1 of the *Copyright Act*, and no attempt was made to secure the moral rights, constituting a clear infringement per s. 28.1 of the *Copyright Act*.
93. Each of the above acts constitutes an unlawful and intentional infringement of the Class Members' rights to integrity and dignity (s. 4) and to the free enjoyment and disposal of their property (s. 6) under the *Charter*.

94. Each of the Class Members has therefore suffered the same or similar harm as the Applicant and are therefore entitled to damages to compensate for the violation of their moral rights by the Defendants, and the infringement of their works, as well as punitive damages.

PART VII: CONDITIONS REQUIRED FOR AUTHORISATION

A. Common Questions of Fact and Law (s. 575(1) CCP)

95. The claims of the Class Members raise identical, similar or related questions of fact or law namely:
- (a) Did the Defendants download the datasets identified in the Claim as the "RedPajama and Books3 Datasets"?
 - (b) Did the RedPajama and Books3 Datasets contain copyrighted material?
 - (c) Did the Defendants use copyrighted material in the RedPajama and Books3 Datasets in training its LLMs?
 - (d) Did the Defendants keep a library of books for its potential future use of works downloaded from the RedPajama and Books3 Datasets?
 - (e) Did the Defendants attempt to pay or pay class members in order to acquire a copy of copyrighted material contained in the RedPajama and Books3 Datasets?
 - (f) Did the Defendants attempt to obtain or obtain from class members a license or consent to use copyrighted material contained in the RedPajama and Books3 Datasets to train their LLMs?
 - (g) Did the Defendants attempt to obtain or obtain from class members a waiver, or consent in respect, of moral rights held by the class members in

order to use the copyrighted material contained in the RedPajama and Books3 Datasets to train their LLMs?

- (h) Did the Defendants remove or alter rights management information in that were in the RedPajama and Books3 datasets?
- (i) Did the Defendants circumvent technological protection measures when downloading copyrighted electronic books contained in the RedPajama and Books3 datasets?
- (j) Did the Defendants infringe, authorize, or induce the infringement of copyright owned by class members by downloading and/or using to train their LLMs material from the RedPajama and Books3 Datasets, and thereby breach s. 27 of the *Copyright Act*?
- (k) Did the Defendants, in respect of the class members, breach ss. 41 and 41.1, or 41.22 of the *Copyright Act*?
- (l) Did the Defendants unjustly enrich itself at the expense of class members?
- (m) Are the Defendants liable to pay damages to the class members under the *Copyright Act*? If so, is this an appropriate case to award aggregate damages and what is an appropriate award?
- (n) Are the Defendants liable for punitive damages?

B. The Facts Alleged Appear to Justify the Conclusions Sought (s. 676(2) CCP)

96. The Applicant alleges that the facts appear to justify the conclusions sought and refers to paragraphs 1 to 87 of the present application.

C. The Composition of the Class (s. 575(3) CCP)

97. The composition of the Class makes it difficult or impracticable to apply the rules for mandates to take part in judicial proceedings on behalf of others or for

consolidation of proceedings, with respect to provision 575(3) of the *CCP*, for the following reasons:

- (a) Class Members are numerous and are scattered across Québec estimated to be in the thousands;
- (b) The applicant is unaware of how many persons throughout Québec have published works contained in the RedPajama and Books3 datasets;
- (c) The names and addresses of the Class Members are not known to the Applicant;
- (d) Given the costs and risks inherent in an action before the courts, many people will hesitate to institute an individual action against the Defendants. Even if the Class Members themselves could afford such individual litigation, the Court system could not as it would be overloaded;
- (e) Further, individual litigation of the factual and legal issues raised by the conduct of the Defendants would increase delay and expense to all parties and to the Court system;
- (f) A multitude of actions risks having contradictory judgments on questions of fact and law that are similar or related to all Class Members;
- (g) These facts demonstrate that it would be impractical, if not impossible, to contact each and every Class Member to obtain mandates and to join them in one action; and
- (h) In these circumstances, a class action is the only appropriate procedure for all of the Class Members to effectively pursue their respective rights and have access to justice.

D. The Applicant is in a Position to Adequately Represent the Class members (s. 575(4) CCP)

98. The Applicant requests that they be appointed the status of representative Applicant for the following main reasons:

- (a) They are a member of the Class and have a personal interest in seeking the conclusions that they propose herein;
- (b) They are competent, in that they have the potential to be the mandatary of the action if it had proceeded under article 91 of the *CCP*; and
- (c) Their interests are not antagonistic to those of other Class Members.

99. Additionally, the Applicant respectfully adds that:

- (a) They contacted and mandated their attorneys to file the present application for the sole purpose of having their rights, as well as the rights of other Class Members, recognized and protected so that they may be compensated for the damages that they have suffered as a consequence of the Defendants' failures in their legal obligations and so that the Defendants can be held accountable;
- (b) They are aware of several other Class Members in the same situation as them;
- (c) They have the time, energy, will and determination to assume all the responsibilities incumbent upon them in order to diligently carry out the action;
- (d) They cooperate and will continue to fully cooperate with their attorneys, who have experience in intellectual property-related class actions;
- (e) They understand the nature of the action;

PART VIII: NATURE OF THE ACTION AND CONCLUSIONS SOUGHT

100. The nature of the class action that the Applicant intends to bring on behalf of the Class Members is an action damages, in restitution of profits and in injunction .

101. The conclusions that the Applicant seeks are the following:

GRANT the Applicant's application to institute proceedings;

ORDER the Defendants to pay the Applicant damages to be assessed by the Court with interest from the date of service of this application as well as the additional indemnity provided for in Article 1619 CCQ and with full costs and expenses including expert fees and notice fees and fees relating to administering the plan of distribution of the recovery in this action .

ORDER the Defendants to pay the Applicant punitive and exemplary damages to be assessed by the Court;

DECLARE that the Defendants have been unjustly enriched by the amount of revenue earned as a result of the use of Class Members' works to train the Defendants' LLM and **ORDER** the indemnification of the Applicant for all such revenue;

CONDEMN the Defendants to pay the Applicant punitive and exemplary damages to be assessed by the Court;

ORDER the Defendants to pay to each of the Members damages and interest to be assessed by the Court, with interest from the date of service of the present application , as well as the additional indemnity provided for in Article 1619 C.C.Q;

CONDEMN the Defendants to pay to each of the Members punitive and exemplary damages to be assessed by the Court;

ORDER that the aforementioned damages be subject to collective compensation, in accordance with the provisions of sections 595 to 598 of the *CCP*;

CONDEMN the Defendants to any other appropriate remedy deemed fair and reasonable.

PART IX: JURISDICTION

102. The Applicant suggests that the Superior Court of Quebec has jurisdiction since the Class Members reside in the Province of Québec and because the application relates to the Defendants' activities in Quebec, the faults were committed in whole or in part in Quebec, and the Class Members have suffered and are suffering harm in Quebec.
103. Moreover, the Defendants' faults, in particular the collection and reproduction of texts and data from the Internet, including copies of Quebec works, necessarily took place at least in part in Quebec.
104. The Applicant proposes that this class action be brought in the judicial district of Montréal since the Applicant and a significant number of the Class Members are domiciled there, and the undersigned lawyers whose services have been retained by the Applicant have a principal place of business in the judicial district of Montreal.
105. The present application for authorization to institute a class action is well founded in fact and in law.
106. The interests of justice weigh in favour of this application being granted in accordance with its conclusions.

FOR THESE REASONS, MAY IT PLEASE THE COURT:

GRANT the present application;

AUTHORISE the class action, the nature of which is an action in damages, in restitution of profits and in injunction;

APPOINT the Applicant the status of representative Applicant of the persons included in the class herein described as:

all natural persons or corporations, resident in Québec who owned or co-owned the copyright to works used by the Defendants to train one of the various versions of the Defendants' Large Language Models.

IDENTIFY the principal questions of fact and law to be treated collectively as the following:

- (a) Did the Defendants download the datasets identified in the Claim as the "RedPajama and Books3 Datasets"?
- (b) Did the RedPajama and Books3 Datasets contain copyrighted material?
- (c) Did the Defendants use copyrighted material in the RedPajama and Books3 Datasets in training its LLMs?
- (d) Did the Defendants keep a library of books for its potential future use of works downloaded from the RedPajama and Books3 Datasets?
- (e) Did the Defendants attempt to pay or pay class members in order to acquire a copy of copyrighted material contained in the RedPajama and Books3 Datasets?
- (f) Did the Defendants attempt to obtain or obtain from class members a license or consent to use copyrighted material contained in the RedPajama and Books3 Datasets to train their LLMs?
- (g) Did the Defendants attempt to obtain or obtain from class members a waiver, or consent in respect, of moral rights held by the class members in

order to use the copyrighted material contained in the RedPajama and Books3 Datasets to train their LLMs?

- (h) Did the Defendants remove or alter rights management information in that were in the RedPajama and Books3 datasets?
- (i) Did the Defendants circumvent technological protection measures when downloading copyrighted electronic books contained in the RedPajama and Books3 datasets?
- (j) Did the Defendants infringe, authorize, or induce the infringement of copyright owned by class members by downloading and/or using to train their LLMs material from the RedPajama and Books3 Datasets, and thereby breach s. 27 of the *Copyright Act*?
- (k) Did the Defendants, in respect of the class members, breach ss. 41 and 41.1, or 41.22 of the *Copyright Act*?
- (l) Did the Defendants unjustly enrich itself at the expense of class members?
- (m) Are the Defendants liable to pay damages to the class members under the *Copyright Act*? If so, is this an appropriate case to award aggregate damages and what is an appropriate award?
- (n) Are the Defendants liable for punitive damages?

IDENTIFY the conclusions sought by the class action to be instituted as being the following:

GRANT the Applicant's application to institute proceedings;

ORDER the Defendants to pay the Applicant damages to be assessed by the Court with interest from the date of service of this application as well as the additional indemnity provided for in Article 1619 CCQ and with full costs

and expenses including expert fees and notice fees and fees relating to administering the plan of distribution of the recovery in this action .

ORDER the Defendants to pay the Applicant punitive and exemplary damages to be assessed by the Court;

DECLARE that the Defendants have been unjustly enriched by the amount of revenue earned as a result of the use of Class Members' works to train the Defendants' LLM and **ORDER** the indemnification of the Applicant for all such revenue;

CONDEMN the Defendants to pay the Applicant punitive and exemplary damages to be assessed by the Court;

ORDER the Defendants to pay to each of the Members damages and interest to be assessed by the Court, with interest from the date of service of the present application, as well as the additional indemnity provided for in Article 1619 C.C.Q;

CONDEMN the Defendants to pay to each of the Members punitive and exemplary damages to be assessed by the Court;

ORDER that the aforementioned damages be subject to collective compensation, in accordance with the provisions of sections 595 to 598 of the *CCP*;

CONDEMN the Defendants to any other appropriate remedy deemed fair and reasonable.

DECLARE that all class members that have not requested their exclusion from the class in the prescribed delay to be bound by any judgment to be rendered on the class action to be instituted ;

FIX the delay of exclusion at thirty (30) days from the date of the publication of the notice to the class members, date upon which the members of the class that have not exercised their means of exclusion will be bound by any judgment to be rendered herein;

ORDER the publication of a notice to the class members in accordance with section 579 of the *CCP*, within sixty (60) days from the judgement to be rendered herein pursuant to a further Order of the Court, and **ORDER** Defendants to pay for said publication costs;

RENDER any other order that this Honourable Court shall determine;

THE WHOLE with costs, including the costs of all publications of notices.

Montréal, July 30, 2025



SIMKIN LÉGAL

Maitre Michael Simkin

85 rue Saint-Paul O., #410
Montréal (Québec) H2Y 3V4

Phone: 1-514-582-9236

Fax: 1-438-800-2363

Attorney for the Applicant

SUMMONS
(Art. 145 and following C.C.P.)

Filing of a judicial application

Take notice that the Applicant has filed this Application for Authorization to Institute a Class Action and to Appoint the Status of Representative Applicant in the office of the Superior Court in the judicial district of Montréal.

Defendants' answer

You must answer the application in writing, personally or through a lawyer, at the courthouse of Montréal situated at 1 Rue Notre-Dame Est, Montréal, Québec, H2Y 186, within 15 days of service of the Application or, if you have no domicile, residence or establishment in Québec, within 30 days. The answer must be notified to the Applicant's lawyer or, if the Applicant is not represented, to the Applicant.

Failure to answer

If you fail to answer within the time limit of 15 or 30 days, as applicable, a default judgement may be rendered against you without further notice and you may, according to the circumstances, be required to pay the legal costs.

Content of answer

In your answer, you must state your intention to:

- negotiate a settlement;
- propose mediation to resolve the dispute;
- defend the application and, in the case required by the Code, cooperate with the Applicant in preparing the case protocol that is to govern the conduct of the proceeding. The protocol must be filed with the court office in the district specified above within 45 days after service of the summons or, in family matters or if you have no domicile, residence or establishment in Québec, within 3 months after service;
- propose a settlement conference.

The answer to the summons must include your contact information and, if you are represented by a lawyer, the lawyer's name and contact information.

Change of judicial district

You may ask the court to refer the originating Application to the district of your domicile or residence, or of your elected domicile or the district designated by an agreement with the Applicant.

If the application pertains to an employment contract, consumer contract or insurance contract, or to the exercise of a hypothecary right on an immovable serving as your main

residence, and if you are the employee, consumer, insured person, beneficiary of the insurance contract or hypothecary debtor, you may ask for a referral to the district of your domicile or residence or the district where the immovable is situated or the loss occurred. The request must be filed with the special clerk of the district of territorial jurisdiction after it has been notified to the other parties and to the office of the court already seized of the originating application.

Transfer of application to Small Claims Division

If you qualify to act as a Applicant under the rules governing the recovery of small claims, you may also contact the clerk of the court to request that the application be processed according to those rules. If you make this request, the Applicant's legal costs will not exceed those prescribed for the recovery of small claims.

Calling to a case management conference

Within 20 days after the case protocol mentioned above is filed, the court may call you to a case management conference to ensure the orderly progress of the proceeding. Failing this, the protocol is presumed to be accepted.

Exhibits supporting the application

- Exhibit P-1:** Applicant's website list of published books
- Exhibit P-2:** US Copyright Report re AI
- Exhibit P-3:** "The Pile: An 800gb dataset of diverse text for language modeling" by Leo Gao et al.
- Exhibit P-4:** "Revealed: The Authors Whose Pirated Books are Powering Generative AI", by Alex Reisner.
- Exhibit P-5:** The Atlantic Search Tool.
- Exhibit P-6:** DataBricks Blog Entry Re MPT-7B
- Exhibit P-7:** DataBricks Blog Entry re MPT-30B
- Exhibit P-8:** HuggingFace Website re RedPajama Dataset
- Exhibit P-9:** DataBricks Blog Entry re DBRX
- Exhibit P-10:** DataBricks announcement re acquisition of Mosaic
- Exhibit P-11:** Article re DataBricks acquisition of Mosaic for 1.3 billion USD
- Exhibit P-12:** DataBricks Announcement re 62 Billion USD Valuation:

Exhibit P-13 United States District Court summary judgement ruling of June 23, 2025 in
Bartz et al v. Anthropic PBC

The exhibits in support of the application are available upon request.

Notice of presentation of an application

If the application is an application in the course of a proceeding or an application under Book III, V, excepting an application in family matters mentioned in article 409, or VI of the Code, the establishment of a case protocol is not required; however, the application must be accompanied by a notice stating the date and time it is to be presented.

**NOTICE OF PRESENTATION
(Articles 146 and 574 CCP)**

TO:

DATABRICKS INC.

1055 West Georgia Street, Suite 2412,
Vancouver, British Columbia.

AND

MOSAICML INC.

51 2nd Street, Suite 202,
San Francisco, California
USA,

Defendants

TAKE NOTICE that Applicant's *Application for Authorization to Institute a Class Action and to Obtain the Status of Representative* will be presented before the Superior Court at 1 Rue Notre-Dame E, Montréal, Québec, H2Y 1B6, on the date set by the coordinator of the Class Action chamber.

GOVERN YOURSELF ACCORDINGLY.

Montréal, July 30, 2025



SIMKIN LÉGAL

Maître Michael Simkin

85 rue Saint-Paul O., #410
Montréal (Québec) H2Y 3V4

Phone: 1-514-582-9236

Fax: 1-438-800-2363

Attorney for the Applicant

CANADA
PROVINCE OF QUÉBEC
DISTRICT OF MONTRÉAL
NO:

SUPERIOR COURT
(Class Action)

TARAS GRESCOE

Applicant

v.

DATABRICKS INC.

AND

MOSAICML INC.

Defendants

LIST OF EXHIBITS

- Exhibit P-1:** Applicant's website list of published books
- Exhibit P-2:** US Copyright Report re AI
- Exhibit P-3:** "The Pile: An 800gb dataset of diverse text for language modeling" by Leo Gao et al.
- Exhibit P-4:** "Revealed: The Authors Whose Pirated Books are Powering Generative AI", by Alex Reisner.
- Exhibit P-5:** The Atlantic Search Tool.
- Exhibit P-6:** DataBricks Blog Entry Re MPT-7B
- Exhibit P-7:** DataBricks Blog Entry re MPT-30B
- Exhibit P-8:** HuggingFace Page re RedPajama Dataset
- Exhibit P-9:** DataBricks Blog Entry re DBRX
- Exhibit P-10:** DataBricks announcement re acquisition of Mosaic
- Exhibit P-11:** Article re DataBricks acquisition of Mosaid for 1.3 billion USD

Exhibit P-12: DataBricks Announcement re 62 Billion USD Valuation

Exhibit P-13 United States District Court summary judgement ruling of June 23, 2025
in *Bartz et al v. Anthropic PBC*

Montréal, July 30, 2025



SIMKIN LÉGAL

Maître Michael Simkin

85 rue Saint-Paul O., #410

Montréal (Québec) H2Y 3V4

Phone: 1-514-582-9236

Fax: 1-438-800-2363

Attorney for the Applicant

500-06-001406-251

NO:

**SUPERIOR COURT
DISTRICT OF MONTRÉAL
(Class Action)**

TARAS GRESCOE*Applicant*

V.

DATABRICKS INC.

AND

MOSAICML INC.*Defendants*

APPLICATION FOR AUTHORIZATION TO
INSTITUTE A CLASS ACTION AND TO
OBTAIN THE STATUS OF
REPRESENTATIVE,
LIST OF EXHIBITS P-1 TO P-13

ORIGINAL

Nature : Class Action

Mon dossier : 0150-003

AS0NB7



**SIMKIN
LEGAL**

info@simkinlegal.com
85 rue Saint-Paul O., #410
Montréal (Québec) H2Y 3V4
T : 1-514-582-9236
Fax: 1-438-800-2363