

CANADA
PROVINCE DE QUÉBEC
DISTRICT DE MONTRÉAL

(Chambre des actions collectives)
COUR SUPÉRIEURE

N^o 500-06-001369-251

ANNE ROBILLARD, domiciliée et résidant
au 501-400, rue Serge-Pépin à Beloeil,
province de Québec, J3G 6Z2

Demanderesse

c.

META PLATFORMS, INC., personne morale
dûment constituée, ayant son siège social au
1 Meta Way, Menlo Park, Californie, États-
Unis, 94025

Défenderesse

**DEMANDE POUR AUTORISATION D'EXERCER UNE ACTION COLLECTIVE ET
POUR ÊTRE REPRÉSENTANTE**
(Art. 574 et suivants CPC)

**À L'UN DES HONORABLES JUGES DE LA COUR SUPÉRIEURE, LA
DEMANDERESSE EXPOSE RESPECTUEUSEMENT CE QUI SUIT :**

1. APERÇU

1. La demanderesse sollicite l'autorisation d'exercer une action collective pour le compte des personnes faisant partie du groupe ci-après décrit, dont elle est elle-même membre, à savoir :

Toute personne au Canada titulaire du droit d'auteur sur une œuvre littéraire ou dramatique que la défenderesse Meta Platforms, Inc. a reproduite sans autorisation dans le cadre du développement ou de la formation de grands modèles de langage (en anglais, « *large language models* »).

2. Le droit d'auteur sert l'intérêt public en stimulant la création et la diffusion d'œuvres artistiques et intellectuelles, tout en garantissant aux créateurs une rémunération équitable pour leur travail. La réalisation de ces objectifs est facilitée par l'existence, à l'échelle mondiale, de régimes législatifs et d'une convention internationale qui accordent des droits exclusifs aux créateurs et interdisent à toute personne d'exploiter leurs œuvres sans leur autorisation.

3. Les membres du groupe sont des auteurs et autres titulaires de droit d'auteur sur des œuvres littéraires et dramatiques. Ils n'ont jamais consenti à ce que la défenderesse se serve de leurs œuvres à des fins commerciales ni à toute autre fin.
4. La défenderesse a développé différents grands modèles de langage (en anglais, « *large language models* ») (« **LLM** »), y compris, mais sans s'y limiter, la série de modèles connus sous le nom de « Llama » et le modèle « OneLLM », qu'elle offre en tant que produits autonomes et en tant que composantes intégrées dans ses autres produits commerciaux.
5. Plutôt que d'agir dans le cadre de la loi et dans le respect des droits des membres du groupe, elle a délibérément choisi de former ses LLM à partir d'ensembles de données contenant des copies illicites d'œuvres du monde entier, dont celles des membres du groupe.
6. Ce faisant, la défenderesse a commis ce qui est sans aucun doute l'un des plus grands actes de piratage de l'histoire, exploitant à des fins commerciales des millions d'œuvres dans le but de développer des outils d'intelligence artificielle générative (« **IA générative** ») susceptibles de remplacer les créateurs dont elle a usurpé le travail.
7. Les LLM de la défenderesse n'existeraient pas sans le vaste corpus d'œuvres protégées par le droit d'auteur qui ont servi à les former. Or, ces mêmes LLM mettent aujourd'hui en péril la capacité des membres du groupe à gagner leur vie.
8. En effet, différents outils IA basés sur les LLM de la défenderesse permettent à n'importe qui n'importe où de générer des textes pour lesquels ils paieraient autrement des auteurs.
9. En outre, ces outils IA permettent de produire des œuvres dérivées non autorisées, c.-à-d., du matériel basé sur, imitant, résumant ou paraphasant les œuvres protégées par le droit d'auteur des membres du groupe, et nuisant au marché de ces dernières.
10. La demanderesse, en son nom et pour le compte des membres du groupe, demande donc l'autorisation d'exercer une action collective en dommages-intérêts compensatoires et en restitution de profits ou, au choix, en dommages-intérêts préétablis, de même qu'en dommages-intérêts punitifs et en injonction.

2. LES PARTIES

2.1. La demanderesse

11. La demanderesse est une autrice à succès dont plusieurs livres ont figuré sur les listes de *best-sellers*.
12. Elle est considérée comme une figure incontournable de la littérature fantastique.
13. La demanderesse est connue notamment pour la saga *Les Chevaliers d'Émeraude*, dont le premier tome est paru en 2002. Cette série culte composée de 13 tomes a été vendue à plus de 5 millions d'exemplaires au Québec et en France et traduite en plusieurs langues.
14. Outre cette série, la demanderesse a écrit de nombreux autres livres, dont les séries *Les héritiers d'Enkidiev* (12 tomes), *Les Chevaliers d'Antarès* (12 tomes), *Légendes d'Ashur-Sîn* (7 tomes), *Les magiciens d'Enkidiev* (6 tomes), *Les Wilder* (3 tomes), *A.N.G.E.* (10 tomes), *Les ailes d'Alexanne* (10 tomes), *Les cordes de cristal* (10 épisodes), *Le retour de l'oiseau-tonnerre* (3 tomes), *La malédiction des Dragensblöt* (7 tomes), *Maynard Bennett, enquêteur de l'étrange* (5 tomes) et le roman *Les Chevaliers d'Épées*, tel qu'il appert de la liste des livres d'Anne Robillard publiée sur le site Web anne-robillard.com, **PIÈCE P-1**.
15. En reconnaissance de son succès, la demanderesse a notamment reçu le *Grand Prix littéraire Archambault* pour le cinquième tome de *Les Chevaliers d'Émeraude*, ainsi que le *Prix des lecteurs du Salon du livre de Trois-Rivières* en 2007.
16. Le 26 mars 2025, elle franchira une étape impressionnante avec la publication de son 100^e livre, *Les Magicien d'Enkidiev - Tome 6*.
17. La demanderesse est titulaire des droits d'auteur et moraux sur l'ensemble de son œuvre.

2.2. La défenderesse et ses outils IA

18. La défenderesse Meta Platforms, Inc., anciennement connue sous le nom de Facebook, est une multinationale américaine qui a déclaré un chiffre d'affaires de 164,5 milliards de dollars US en 2024, tel qu'il appert du communiqué de presse intitulé « Meta Reports Fourth Quarter and Full Year 2024 Results » daté du 29 janvier 2025, **PIÈCE P-2**.
19. Elle exploite de nombreuses applications, technologies et services tels que Facebook, Messenger, Instagram, WhatsApp, Ray-Ban Meta glasses, Meta Quest et Meta AI.

20. En février 2023, la défenderesse a lancé sa première version de Llama, Llama 1. Depuis, Llama 2 a été lancée en juillet 2023 et Llama 3 en avril 2024.
21. La défenderesse offre présentement trois sous-versions de Llama 3, Llama 3.1, Llama 3.2 et Llama 3.3, tel qu'il appert d'un extrait du site Web llama.com, **PIÈCE P-3**.
22. La défenderesse met notamment ses modèles Llama à la disposition des développeurs IA et autres entreprises qui souhaitent les intégrer dans leurs propres produits ou services. La licence (*Llama Community License Agreement*) permet une large utilisation commerciale ainsi que la création et la redistribution d'outils IA dérivés de Llama.
23. Selon la défenderesse, Llama serait rapidement devenu le LLM le plus adopté, avec plus de 650 millions de téléchargements de Llama et de ses dérivés :

As Meta Founder & CEO Mark Zuckerberg shared, **Llama has quickly become the most adopted model, with more than 650 million downloads of Llama and its derivatives**, twice as many downloads as we had three months ago. Putting that in perspective, Llama models have now been downloaded an average of one million times a day since our first release in February 2023.

[...]

Outside of the US, Llama became a global phenomenon this year, with impressive growth and appetite for our models from developers around the world and a hastened pace of adoption with the launch of our Llama 3 model collection. Llama license approvals have more than doubled in the last six months overall, with notable growth in emerging markets and downloads surging across Latin America, the Asia-Pacific region, and Europe. [Nous soulignons]

- Blog intitulé « The future of AI: Built with Llama » publié sur le site Web ai.meta.com, **PIÈCE P-4** (p. 2 et 3).

24. La défenderesse offre également ses propres outils IA bâtis sur ses LLM, tels que l'assistant IA appelé « Meta AI ».
25. Toujours selon la défenderesse, Meta AI était en voie de devenir l'assistant IA le plus utilisé au monde d'ici la fin de l'année 2024, avec près de 600 millions d'utilisateurs actifs par mois dans 43 pays :

The rapid drumbeat of model innovation we've delivered over the past year is also having a ripple effect in our products. Built with Llama, **Meta AI is on track to be the most used AI assistant in the world by the end of 2024 with almost 600 million monthly active users.**

This year, we expanded access to Meta AI to more countries and in new languages across WhatsApp, Instagram, Facebook, Messenger, and on the

web. By the end of this year, **we anticipate that Meta AI will be available in 43 countries and a dozen languages**, and we look forward to bringing the assistant to more people and places. [Nous soulignons]

- Blog intitulé « The future of AI: Built with Llama » publié sur le site Web ai.meta.com, (PIÈCE P-4, p. 5 et 6).

26. Meta AI est disponible au Québec et dans l'ensemble du Canada.
27. La défenderesse opère notamment le Laboratoire de recherche fondamentale sur l'IA (FAIR), dont un des principaux établissements est situé sur la rue St-Urbain à Montréal, tel qu'il appert de l'article intitulé « Celebrating Five Years of AI Breakthroughs from Our Montreal Research Lab » publié sur le site Web aboutfb.com, **PIÈCE P-5**.
28. L'équipe de FAIR a joué et continue de jouer un rôle clé dans la recherche, le développement et la formation des LLM et autres outils IA de la défenderesse, tel qu'il appert de l'article de *Fortune* intitulé « How Mark Zuckerberg has Fully Rebuilt Meta Around Llama » publié sur le site Web finance.yahoo.com, **PIÈCE P-6**.

3. LA RESPONSABILITÉ DE LA DÉFENDERESSE

3.1. Introduction

29. Les technologies d'IA permettent aux ordinateurs et machines de simuler l'intelligence humaine et ses capacités de résolution.
30. Le terme « IA générative » fait spécifiquement référence à une catégorie d'IA capable de générer différents types de contenu (par ex., texte, image, musique, vidéo, audio ou code logiciel) supposément original en réponse aux instructions données par l'utilisateur.
31. L'IA générative s'appuie sur des modèles d'apprentissage profond (en anglais, « *deep learning* »), c.-à-d., des algorithmes sophistiqués qui identifient et codent les schémas et les relations dans d'énormes quantités de données puis utilisent ces informations pour comprendre et répondre aux requêtes des utilisateurs.
32. Les LLM sont une catégorie de modèles d'apprentissage profond pouvant servir à la génération de texte.
33. Pour développer un LLM, celui-ci doit être « formé » à partir d'énormes quantités de textes et de données provenant d'Internet ou d'autres sources, communément appelées « ensembles de données d'apprentissage » (en anglais, « *training dataset* »).

34. Ce processus est décrit ainsi dans un document de consultation du Gouvernement du Canada sur le droit d'auteur et l'IA générative :

La fouille de textes et de données (FTD) consiste en la reproduction et l'analyse de grandes quantités de données et d'informations, **y compris celles extraites de contenus protégés par le droit d'auteur**, afin d'identifier des tendances et de faire des prédictions. La FTD est une étape essentielle dans la formation de modèles d'apprentissage automatique. Cette technique permet au modèle « d'apprendre » à reconnaître et à reproduire des tendances qui lui permettront d'accomplir certaines tâches, notamment de générer de la poésie, de la musique ou des œuvres d'art. Au-delà de l'IA générative, la FTD peut également faire progresser la science et aider les entreprises à résoudre des problèmes, à innover et à créer davantage de valeur.

Suite à l'essor des outils d'IA générative tels que ChatGPT et DALL-E, **plusieurs titulaires de droits ont exprimé des inquiétudes croissantes concernant l'utilisation de contenus protégés par le droit d'auteur par les développeurs dans la formation de modèles d'apprentissage automatique, sans que les titulaires de droits y consentent ou reçoivent le crédit et la compensation qui leur sont dus**. Certains titulaires de droits ont commencé à proposer des licences pour IA, mais il reste difficile de faire valoir leurs droits et d'obtenir une rémunération pour l'utilisation de contenus protégés par le droit d'auteur dans le cadre des activités FTD. [Nous soulignons]

- Publication du Gouvernement du Canada intitulée « Consultation sur le droit d'auteur à l'ère de l'intelligence artificielle générative », **PIÈCE P-7** (p. 8 et 9).

35. En outre, selon le United States Patent & Trademark Office, le processus implique nécessairement la reproduction d'œuvres protégées par le droit d'auteur :

The ingestion of copyrighted works for purposes of machine learning will almost by definition involve the reproduction of entire works or substantial portions thereof. [...]

[...]

When AI algorithms or processes "learn" their functions by ingesting copyrighted works, reproductions of those works are made in the process as the works are digitized and/or "read" by the AI algorithms or processes. [...]

[...]

- Rapport du United States Patent & Trademark Office intitulé « Public Views on Artificial Intelligence and Intellectual Property Policy », **PIÈCE P-8** (p. 24 et 25).

36. Comme l'explique la défenderesse dans un document soumis au U.S. Copyright Office, la performance d'un modèle LLM dépend largement de la quantité et de la qualité des données utilisées pour son apprentissage :

[...] an AI language model based on a small amount of data will not have enough information from which to derive the highly complex and nuanced rules, meanings, and contradictions of human language. And a model trained on a narrow category of works – like scientific journals – will not have enough information to accurately discern the patterns of modern, everyday human speech. To build a model that can realistically emulate all facets of human language, the developer needs a collection of data that includes a very large number of examples, reflecting a broad range of speech – from casual banter, to literary prose, to scientific jargon.

- Document intitulé « Notice of Inquiry on Artificial Intelligence & Copyright (Dkt. 2023-6): Comments of Meta Platforms, Inc. », **PIÈCE P-9** (p. 2 et 3).

37. Les œuvres écrites par des auteurs constituent une source de données particulièrement précieuse et convoitée pour ces fins. Entre autres, les œuvres littéraires et dramatiques permettent aux LLM de comprendre la complexité de l'expérience humaine et de la reproduire en analysant la manière dont des personnes, réelles ou fictives, communiquent entre elles.
38. Il serait possible d'obtenir de telles œuvres de façon légale. La défenderesse a plutôt choisi de former ses LLM à partir d'ensembles de données d'apprentissage contenant des millions de copies d'œuvres piratées prises de l'Internet, y compris des œuvres des membres du groupe.

3.2. Les contrefaçons liées à Llama 1 et Llama 2

39. En février 2023, la défenderesse a dévoilé l'existence de Llama 1 en publiant un document de recherche sur son site Web. Dans ce document, la défenderesse admet que l'ensemble de données d'apprentissage utilisé pour former Llama 1 incluait le corpus de livres de « Books3 » :

Gutenberg and Books3 [4.5%]. We include two book corpora in our training dataset: the Gutenberg Project, which contains books that are in the public domain, and the Books3 section of ThePile (Gao et al., 2020), a publicly available dataset for training large language models. [...]

- Document de recherche intitulé « LLaMa: Open and Efficient Foundation Language Models », **PIÈCE P-10** (p. 2).
40. The Pile, d'où la défenderesse confirme avoir pris Books3, est un ensemble de données d'apprentissage créée par EleutherAI, un groupe de recherche sur l'IA.
 41. Dans des publications sur l'application X (anciennement Twitter), le créateur de Books3, Shawn Presser, a confirmé que ce corpus de livres était constitué de copies de tous les livres disponibles sur « Bibliotik » (soit 196 640 livres), tel qu'il appert de publications X de Shawn Presser datées du 25 octobre 2020, **PIÈCE P-11**.

42. Les membres d'EleutherAI ont également admis que les livres de Books3 provenaient de Bibliotik dans un document de recherche qu'ils ont publié pour présenter The Pile :

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.** [Nous soulignons]

- Document de recherche intitulé « The Pile: An 800GB Dataset of Diverse Text for Language Modeling », **PIÈCE P-12** (p. 3 et 4).

43. Or, Bibliotik fait partie de ce qu'on appelle couramment des bibliothèques clandestines ou fantômes (en anglais, « *shadow libraries* »), c.-à-d. des plateformes en ligne qui permettent de télécharger des copies illicites d'œuvres protégées par le droit d'auteur.

44. Il est notoire que Bibliotik recèle de copies d'œuvres protégées par le droit d'auteur qui sont partagées sans l'autorisation des titulaires des droits. D'ailleurs, voici ce que répond l'assistant IA de la défenderesse, Meta AI, lorsqu'on lui demande « Qu'est-ce que Bibliotik? » :

Bibliotik

Bibliotik est une plateforme en ligne qui propose un large éventail de livres électroniques, souvent en format PDF, ainsi que d'autres ressources numériques. Elle est connue pour offrir un accès à des ouvrages qui peuvent être difficiles à trouver ou qui sont soumis à des restrictions de distribution dans certains pays.

Fonctionnement

Bibliotik fonctionne sur le principe du partage de fichiers entre pairs (P2P), où les utilisateurs peuvent télécharger et partager des livres électroniques. **Cependant, il est important de noter que l'utilisation de telles plateformes peut soulever des questions concernant les droits d'auteur et la légalité de la distribution de certains contenus.**

Utilisation responsable

Il est essentiel d'être conscient des lois sur le droit d'auteur et de respecter les créateurs de contenu en utilisant ces plateformes de manière responsable. Dans certains cas, les auteurs ou les éditeurs peuvent offrir leurs œuvres de manière gratuite ou sous licence ouverte, ce qui rend leur partage légal.

Alternatives

Pour les utilisateurs qui cherchent à accéder à des livres électroniques de manière légale, il existe plusieurs alternatives, notamment les bibliothèques numériques publiques, les plateformes de vente de livres électroniques qui proposent des téléchargements légaux, et les initiatives d'édition en ligne qui offrent des contenus sous licences libres.

Si vous avez d'autres questions ou si vous cherchez des recommandations sur les alternatives pour accéder à des livres électroniques, n'hésitez pas à demander. [Nous soulignons]

- Capture d'écran de la réponse donnée par Meta AI à la question « Qu'est-ce que Bibliotik? », **PIÈCE P-13**.
- 45. En août et septembre 2023, le magazine *The Atlantic* a obtenu le contenu de Books3 et développé un moteur de recherche qui permet de vérifier les œuvres qui s'y retrouvent, tel qu'il appert de l'article intitulé « Revealed : The Authors Whose Pirated Books are Powering Generative AI » et de l'article intitulé « These 183,000 Books are Fueling the Biggest Fight in Publishing and Tech » publiés sur le site Web theatlantic.com, **PIÈCE P-14** en liasse.
- 46. Une enquête de CBC/Radio-Canada a révélé qu'au moins 2 500 livres écrits par plus de 1 200 auteurs canadiens se trouveraient dans Books3, tel qu'il appert de l'article intitulé « CBC News Analysis Finds Thousands of Canadian Authors, Books in Controversial Dataset Used to Train AI » publié sur le site Web cbc.ca, **PIÈCE P-15**.
- 47. Un autre reportage de CBC/Radio-Canada rapporte aussi que plusieurs auteurs montréalais ont appris que leurs œuvres se trouvaient sans leur permission dans Books3, tel qu'il appert de l'article intitulé « Montreal Writers Want More Protections as AI Sucks up Their Stories » publié sur le site Web cbc.ca, **PIÈCE P-16**.
- 48. De fait, des recherches aléatoires sur le moteur de recherche développé par *The Atlantic* confirment que Books3 contient de nombreuses œuvres d'auteurs canadiens, tel qu'il appert de captures d'écran des résultats de recherche du contenu de Books3 dans *The Atlantic*, **PIÈCE P-17**.
- 49. La défenderesse s'est aussi servie de Books3 pour former Llama 2, tel que le révèle un échange entre Melanie Kambadur (Senior Manager, Llama Research Team) et Nisha Deo (FAIR Communications Lead), déposé en preuve dans l'affaire *Kadrey et al. v. Meta Platforms, Inc.* (Case No. 3 :23-cv-03417-VC) devant la United States District Court, Northern District of California (« l'affaire **Kadrey** ») :

Nisha Deo: [...]

>Hi Melanie – do you know if we used the Books data set in Llama 2?

Melanie Kambadur:

>yes we did

[...]

Nisha Deo:

>WELP for full context we have an inquiry from the atlantic about the use of books3 in Llama - - so basically wondering if it's at all possible for them to deduce that it was used in llama 2

[...]

Melanie Kambadur:

>yes I went to an AI event yesterday with a bnch [sic] of people from other companies and any time I mention I worked on llama 2 they kept asking me about datasets 😊 I was just like see the paper / no comment lol

[...]

Nisha Deo:

> [...] It's the piracy (and us knowing and being accomplices) that's the issue
- - whether we knew or not at the time

- Courriel de Melanie Kambadur à Melanie Kambadur et Nisha Deo intitulé « Message summary », **PIÈCE P-18**.

50. Lorsqu'elle a pris la décision de télécharger le contenu de Books3 et de s'en servir pour former Llama 1 et Llama 2, la défenderesse savait pertinemment que ce contenu était constitué de copies d'œuvres protégées par le droit d'auteur reproduites sans autorisation.
51. En effet, le fait que Books3 provenait de Bibliotik, une bibliothèque clandestine bien connue, avait déjà été divulgué publiquement par Shawn Presser et EleutherAI.
52. En outre, en novembre 2020, un chercheur de Meta AI, Tim Dettmers, a entamé une conversation sur le serveur public Discord d'EleutherAI afin de s'enquérir sur la possibilité d'utiliser The Pile comme ensemble de données d'apprentissage. Lors de cette conversation, Stella Biderman d'EleutherAI a prévenu Dettmers que « *your legal dept is most likely to be worried about books3 which contains the text of books with active copyrights* ».
53. En août 2023, EleutherAI a retiré cette conversation de la vue du public.
54. Par ailleurs, la défenderesse a aussi admis avoir formé Llama 1 avec des ensembles de données tirées de CommonCrawl et d'un sous-ensemble assemblé par Google appelé C4, tel qu'il appert de la pièce P-10 (p. 2). Or, Common Crawl et C4 contiennent des copies d'œuvres piratées.

3.3. Les contrefaçons liées à Llama 3 et OneLLM

55. Depuis la publication du document de recherche sur Llama 1 (pièce P-10) et la controverse suscitée par son usage de Books3, la défenderesse ne révèle plus ce que contiennent les ensembles de données d'apprentissage utilisées pour former ses LLM :

One thing everyone WIRED spoke with could agree upon? **All this increased scrutiny on data sets has made AI's big players shy away from transparency. Meta is the prime example.** It openly shared the data sets used to train the first version of its ChatGPT competitor Llama, including Books3. **Now, it's tight-lipped about what is used for newer versions.** "It behooves these companies to be opaque about their sources," McCarthy says. Knowing they're likely to face lawsuits if they fess up to using copyrighted material in their data training sets is a powerful deterrent. **This, in turn, will make it harder for writers to know when their copyright is potentially infringed.** [Nous soulignons]

- Article intitulé « The Battle Over Books3 Could Change AI Forever » publié sur le site Web wired.com, **PIÈCE P-19.**
56. En d'autres mots, la défenderesse a décidé d'envelopper ses activités dans le secret pour mieux se soustraire à ses obligations légales. Par conséquent, elle seule connaît précisément l'ampleur et la provenance des œuvres protégées par le droit d'auteur qu'elle a reproduites sans autorisation dans le cadre du développement et de la formation de ses LLM.
57. Malgré ce secret, il fut récemment révélé que la défenderesse a également choisi de former ses LLM à partir de données téléchargées d'une autre bibliothèque clandestine qu'elle savait contenir des copies d'œuvres piratées, à savoir « Library Genesis » ou « LibGen ».
58. Ce fait a notamment été admis par Melanie Kambadur (Senior Manager, Llama Research Team) dans le cadre d'un interrogatoire hors cour dans l'affaire *Kadrey* :
- Q. Was Libgen used to pre-train the Llama 3 series of models?
A. Yes.
- Extraits de l'interrogatoire hors cour de Melanie Kambadur dans l'affaire *Kadrey et al. v. Meta Platforms, Inc.* (Case No. 3:23-cv-03417-VC) devant la United States District Court, Northern District of California, **PIÈCE P-20** (p. 67).
59. Fondée en Russie, la bibliothèque clandestine LibGen contiendrait plus de 2,8 millions de livres de non-fiction, 2,5 millions de livres de fiction, 84 millions d'articles

scientifiques, et 2 millions de *comics*, tel qu'il appert de la page « Library Genesis » sur le site Web fr.wikipedia.org, **PIÈCE P-21**.

60. Des recherches aléatoires du contenu de LibGen confirment que la bibliothèque contient de nombreuses œuvres d'auteurs canadiens, tel qu'il appert de captures d'écran des résultats de recherche sur LibGen, **PIÈCE P-22**.
61. Il est reconnu qu'une très grande partie sinon la totalité des documents disponibles sur LibGen sont des œuvres partagées sans le consentement des titulaires de droits. Voici d'ailleurs ce que répond l'assistant IA de la défenderesse, Meta IA, lorsqu'on lui demande « What is the LibGen dataset? » :

The LibGen dataset refers to a large collection of books and articles digitized and made available through Library Genesis (LibGen), a shadow library and file-sharing website. **LibGen provides free access to copyrighted materials, including scientific articles, academic books, and fiction, without the permission of copyright holders.**

The dataset contains millions of documents, including:

- Scientific articles from various fields
- Academic books and textbooks
- Fiction books and novels
- Conference proceedings and papers

LibGen has been involved in several copyright infringement lawsuits and has faced opposition from publishers and copyright holders. Despite this, the website remains accessible, and its dataset continues to grow.

It is essential to note that accessing or sharing copyrighted materials without permission may be illegal in some jurisdictions. **Always respect copyright laws and consider using legitimate sources or open-access alternatives when available.** [Nous soulignons]

- Capture d'écran de la réponse donnée par Meta AI à la question « What is the LibGen dataset? », **PIÈCE P-23**.

62. Comme pour Books3, les employés et dirigeants de la défenderesse ont toujours été bien au fait de la nature illicite des copies partagées sur LibGen.
63. Professeure et titulaire d'une bourse William Dawson à l'École d'informatique de l'Université McGill, Joelle Pineau est vice-présidente de la recherche en IA et directrice du Laboratoire de recherche fondamentale sur l'IA (FAIR) de la défenderesse. Elle travaille depuis Montréal où elle dirige notamment l'établissement qui s'y situe.

64. La professeure Pineau est directement impliquée dans les décisions de la défenderesse concernant l'IA, notamment dans les décisions concernant l'usage d'ensembles de données d'apprentissage.
65. En octobre 2022, une employée de la défenderesse, Eleonora « Alex » Presani (Program Manager), a fait part de ses inquiétudes à la professeure Pineau quant à la possibilité d'utiliser LibGen et d'autres bibliothèques clandestines pour former les LLM de la défenderesse. À cette époque, la professeure Pineau convenait que la défenderesse devrait conclure des accords de licences avant de former ses LLM sur des œuvres protégées par le droit d'auteur :

[...]

Eleonora Presani [...]:

>related question: I've been helping with buying content from [CAVIARDAGE] so through that I got in the conversation for Theorem Proving – and Zetta and other teams also need this type of content. One thing that came up is that **they've been looking at dataset such as LibGen and SciHub. Now, I understand that coming from publishing I am biased, but I feel that using pirated material should be beyond our ethical threshold... what do you think?**

[...]

Joelle Pineau [...]:

>Yes, that's part of the conversation. For now we have agreed to use it just for pure exploration to see if there is value. **If we find value, then we will setup proper licensing agreement.** You think it's problematic to use even for this phase? I must say I don't have the full context on that material, so happy to understand better [CAVIARDAGE]

Eleonora Presani [...]:

>I think it's ok to use it for exploratory purposes, but we need to be very careful to use only licensed data for the actual training of the model. **SciHub, ResearchGate, LibGen are basically like PirateBay or something like that, they are distributing content that is protected by copyright and they're infringing it.** [Nous soulignons]

- Courriel de Eleonora Presani à Eleonora Presani et Joelle Pineau intitulé « Message summary », **PIÈCE P-24.**

66. Pourtant, la défenderesse choisira d'inclure les œuvres tirées de LibGen dans ses ensembles de données d'apprentissage pour Llama 3 et OneLLM.
67. Ainsi, en décembre 2023, Sony Theakanath (Director of Product Management) a écrit à la professeure Pineau. Rappelant que « MZ » (Mark Zuckerberg, le fondateur et président-directeur général de la défenderesse) avait déjà approuvé l'utilisation de LibGen pour Llama 3, « with VP sponsor requiring to accept full

risk », M. Theakanath s'adresse maintenant à la professeure Pineau pour qu'elle approuve l'utilisation de LibGen pour former OneLLM. Ce faisant, M. Theakanath insiste sur le fait que, sans le contenu de LibGen, les LLM de la défenderesse ne pourront atteindre les mêmes standards de performance que les modèles concurrents :

Hi Joelle,

The team is writing for a decision on proceeding with using libgen for training OneLLM after guidance from Legal and Policy. **After a prior escalation to MZ, GenAI has been approved to use LibGen for Llama 3 (with VP sponsor requiring to accept full risk)** with a number of agreed upon mitigations. Upon discussion with Legal, **the OneLLM team was asked to escalate to you for use of the Libgen dataset** including implementation of said mitigations (see below).

Libgen is essential to meet SOTA numbers across all categories, and it is known that OpenAI and Mistral are using the library for their models (through word of mouth). **Without LibGen, research scientist within OneLLM believe it would not be able to achieve the SOTA numbers the industry shows.** We're beating Llama V2 7B across most benchmarks and are close to their 30B benchmarks, but are not able to reach Mistral. Results from experiments in GenAI suggest that adding Libgen would significantly close this gap, and possibly get us to SOTA. [Nous soulignons] [SOTA : state-of-the-art]

- Courriel de Sony Theakanath à Joelle Pineau intitulé « [A/C Priv] Request for decision: LibGen Escalation OneLLM », **PIÈCE P-25**.

68. Le courriel de M. Theakanath révèle également qu'il avait été décidé de dissimuler le fait que la défenderesse se serve du contenu de LibGen pour ses LLM en raison des risques juridiques et politiques :

In **no case** would we disclose publicly that we had trained on libgen, however there is practical risk external parties could deduce our use of this dataset, especially if we release models trained on it (e.g. option #3).

[...]

If there is media coverage suggesting we have used a dataset we know to be pirated, such as LibGen, this may undermine our negotiating position with regulators on these issues. [Soulignements de l'auteur]

69. Compte tenu de l'énorme quantité de données à télécharger de LibGen, la défenderesse a ultimement décidé de procéder par la voie de protocoles de partage de fichiers poste-à-poste (torrent), tel qu'il appert notamment des messages reproduits dans le courriel de Xiaolan Wang à Nikolay Bashlykov et al. intitulé « Message summary » daté du 2 avril 2024, **PIÈCE P-26**, et dans le courriel de David Esiobu à Nikolay Bashlykov et al. intitulé « Message summary » daté du 11 avril 2024, **PIÈCE P-27**.

70. En vertu de tels protocoles, l'utilisateur télécharge les données d'un même fichier depuis les postes de différents autres utilisateurs « pairs » du réseau. Dès que les parties du fichier souhaité sont téléchargées sur son poste, elles deviennent disponibles pour être redistribuées aux autres utilisateurs du réseau qui veulent se procurer le même fichier. On parlera alors de *seeding* :

In computing, and specifically peer-to-peer file sharing, seeding is the uploading of already downloaded content for others to download from. A peer, a computer that is connected to the network, becomes a seed when having acquired the entire set of data, it begins to offer its upload bandwidth to other peers attempting to download the file.

- Page « Seeding (computing) » sur le site Web en.wikipedia.org, **PIÈCE P-28**.

71. Ainsi, en téléchargeant des copies d'œuvres à l'aide de protocoles de partage de fichiers poste-à-poste, la défenderesse n'a pas seulement reproduit des copies piratées pour son propre usage; elle a nécessairement distribué ces copies piratées à d'autres contrevenants.

72. Consciente de l'illégalité de sa conduite, la défenderesse a tenté de camoufler sa participation à un tel réseau de distribution de matériel piraté. Il ressort notamment des messages reproduits à la pièce P-26 qu'il a été demandé aux employés qui ont téléchargé le contenu de LibGen de ne pas utiliser les infrastructures de la défenderesse afin que les partages de fichiers ne puissent pas être retracés jusqu'à elle :

Nikolay Bashlykov [...]:

>for libgen I synced the files to s3 once a chunk was downloaded and removed them from cluster, so this can help save the space

[...]

David Esiobu [...]:

>@Frank can you clarify why we can't use FB infra for this again?

Frank Zhang [...]:

>avoiding risk of tracing back the seeder is from FB server

Frank Zhang [...]:

>avoiding risk of tracing back the seeder/downloader are from FB servers

73. Outre LibGen, il apparaît des messages reproduits à la pièce P-27 que la défenderesse a aussi téléchargé le contenu d'une autre bibliothèque clandestine appelée « Z-Library » à l'aide de protocoles de partage de fichiers poste-à-poste :

Xiolan Wang [...]:

>Hi Everyone, would like to share the downloading progress of AA.

[...]

>- Z-library (25.7 TB out of 54 TB). We may get a few more TB in the next few days, but the remaining would be hard to acquire due to similar reasons above.

74. Or, il est notoire que Z-Library contient des copies d'œuvres piratées. Voici ce que répond l'assistant IA de la défenderesse, Meta AI, lorsqu'on lui demande « Qu'est-ce que Z-Library » :

Z-Library est une bibliothèque numérique en ligne qui propose un accès gratuit à des millions de livres, d'articles et de documents scientifiques. Le site permet aux utilisateurs de télécharger des livres électroniques (e-books) et des articles de revues scientifiques sans frais.

Cependant, **il est important de noter que Z-Library est considéré comme un site Web pirate, car il partage des contenus protégés par le droit d'auteur sans l'autorisation des auteurs ou des éditeurs.** Cela signifie que l'utilisation de Z-Library peut être illégale dans certains pays, y compris le Canada. [...] [Nous soulignons]

- Capture d'écran de la réponse donnée par Meta AI à la question « Qu'est-ce que Z-Library », **PIÈCE P-29.**

75. Des recherches aléatoires du contenu de Z-Library confirment que la bibliothèque contient de nombreuses œuvres d'auteurs canadiens, tel qu'il appert de captures d'écran des résultats de recherche sur Z-Library, **PIÈCE P-30.**
76. En outre, la défenderesse a supprimé les avis de droit d'auteur et autres mentions légales sur les copies des œuvres qu'elle a incorporées dans ses ensembles de données d'apprentissage.

3.4. Les manquements de la défenderesse

77. En vertu de la *Loi sur le droit d'auteur*, L.R.C. (1985), ch. C-42 (la « **LDA** »), les membres du groupe détiennent les droits exclusifs de reproduction sur les œuvres que s'est appropriée la défenderesse pour former et développer ses LLM. Ces droits comprennent en outre le droit d'autoriser telle reproduction, tel qu'il apparaît de l'article 3(1) a) et 3(1) *in fine* LDA.
78. La défenderesse a accompli ou autorisé que soient accomplis à leur insu de nombreux actes réservés aux membres du groupe, portant ainsi de nombreuses atteintes aux droits d'auteur de ceux-ci en vertu de l'article 27(1) LDA.
79. Ainsi, sans restreindre la généralité de ce qui précède, en reproduisant des œuvres des membres du groupe sans leur consentement, la défenderesse a commis des actes réservés aux membres du groupe en vertu du paragraphe 3(1) a) LDA.

80. La défenderesse aurait pu négocier des accords de licences avec les membres du groupe. Elle a plutôt délibérément choisi d'obtenir des copies piratées depuis des bibliothèques clandestines ou autres sources illégitimes.
81. La défenderesse a également participé à la distribution illégale d'œuvres piratées par sa participation à des réseaux de partage de fichiers poste-à-poste. Ce faisant, elle a commis ou autorisé que soient commis des actes réservés aux membres du groupe en vertu du paragraphe 3(1)(a) et 3(1) *in fine* LDA.
82. Par ailleurs, en donnant accès à ses LLM aux développeurs IA et aux détenteurs de licences à travers le monde, la défenderesse a mis et continue de mettre en circulation les œuvres des membres du groupe de façon à leur porter préjudice, commettant ainsi des violations à une étape ultérieure en vertu du paragraphe 27(2) b) LDA.
83. De même, la défenderesse a commis et continue de commettre des violations à une étape ultérieure en vertu du paragraphe 27(2) LDA en développant et en commercialisant des outils IA, tels que Meta AI, qui permettent à leurs utilisateurs de générer du contenu en violation du droit d'auteur des membres du groupe.
84. De plus, en incorporant les œuvres des membres du groupe dans ses LLM et ses outils IA sans aucune attribution, la défenderesse porte atteinte aux droits moraux des auteurs membres du groupe en vertu des art. 14.1 et 28.1 LDA.
85. En outre, en s'appropriant sans droit des œuvres protégées par le droit d'auteur pour développer des outils IA qui peuvent générer du contenu qui fait concurrence au travail des auteurs et met en péril leur capacité à gagner leur vie, la défenderesse se livre à une concurrence parasitaire et déloyale qui engage sa responsabilité extracontractuelle.
86. Par ces actes, la défenderesse s'est enrichie sans justification aux dépens des membres.
87. Finalement, chacun des actes précités constitue une atteinte illicite et intentionnelle aux droits des membres québécois à l'intégrité et à la dignité (art. 4) et à la libre jouissance et à la libre disposition de leurs biens (art. 6) en vertu de la *Charte des droits et libertés de la personne*, L.R.Q., ch. C-12 (la « **Charte** »).

3.5. Le préjudice subi par les membres du groupe et les remèdes

88. Les membres du groupe détiennent des droits exclusifs sur leurs œuvres.
89. Non seulement investissent-ils énormément de temps, d'énergie et de ressources dans leur travail; en tant qu'artistes, ils investissent tout leur être dans leur expression créative.
90. Il existe un marché bien établi leur permettant d'être rémunéré équitablement pour ce travail.
91. Au lieu de respecter les droits exclusifs des membres du groupe, la défenderesse a délibérément choisi de reproduire leurs œuvres sans leur consentement et sans rémunération et de s'en servir pour développer et former ses LLM. Elle a de plus tout mis en œuvre pour tenter de camoufler ce fait.
92. En outre, les LLM qu'elle a formés grâce à cette violation intentionnelle des droits des membres du groupe servent à développer des produits IA qui peuvent générer des textes substantiellement équivalents et pouvant faire concurrence aux œuvres créées par les membres du groupe, mettant du même coup en péril leur capacité à gagner leur vie.
93. En conséquence directe de ses gestes, la défenderesse a profité et profite indûment de la propriété intellectuelle des membres du groupe au mépris de leurs droits, leur causant des préjudices matériels et moraux.
94. Cette exploitation massive des œuvres des membres du groupe pour des fins commerciales est manifestement inéquitable. Ayant absolument besoin de ces œuvres pour développer des LLM atteignant les standards de l'industrie, la défenderesse a conclu qu'il valait mieux se tourner vers des sources pirates plutôt que de négocier de bonne foi l'utilisation des œuvres avec les détenteurs de droits.
95. Tout le cynisme de cette approche est bien résumé par un échange entre deux employés de la défenderesse, Sergey Edunov (Director of Engineering, Generative AI) et Laurens van der Maaten (Senior Research Director) :

Laurens van der Maaten [...]:

>Flagged to me offline: Some people on the team have mentioned they think we should spend more on licensing data for pre-training. Things like school/university text books. Any thoughts on that? Melanie was pretty clear that she wasn't planning to license data right now. And it may be less relevant given that we are using libgen..?

Sergey Edunov [...]:

>the problem is that people don't realize that **if we license once [sic] single book, we won't be able to lean into fair use strategy. So we will have to drop all of the libgen and books datasets**

Laurens van der Maaten [...]:

>Ah interesting. Makes sense!

Sergey Edunov [...]:

>there are also many issues with book licensing

[...]

>particularly that publishers don't have the rights

[...]

>you have to license books with individual authors

[...]

>and thats a shit ton of work

[...]

- Courriel de Laurens van der Maaten à Sergey Edunov et Laurens van der Maaten intitulé « Message summary » daté du 23 août 2023, **PIÈCE P-31** [Nous soulignons].

96. Le développement des technologies d'IA générative est au cœur des ambitions commerciales de la défenderesse, tel que le rapportait récemment l'agence Bloomberg :

Meta Platforms Inc. Chief Executive Officer Mark Zuckerberg exuded confidence in his company's artificial intelligence strategy, saying 2025 will be a "really big year" in which its AI assistant will become the most widely used in the industry.

"In AI, I expect that this is going to be the year when a highly intelligent and personalized AI assistant reaches more than 1 billion people," Zuckerberg said Wednesday on a call with investors following the company's fourth-quarter earnings report. "And I expect Meta AI to be that leading AI assistant."

Zuckerberg told investors he believes Meta will eventually spend hundreds of billions on AI infrastructure, imitating and improving on the best technology presented by competitors to ensure its dominance.

The remarks sent Meta shares up as much as 3.2% after markets opened in New York on Thursday. [...]

- Article intitulé « Meta Gains After Zuckerberg Predicts 'Really Big Year' in AI » publié sur le site finance.yahoo.com, **PIÈCE P-32**.

97. Or, le succès des opérations de la défenderesse en matière d'IA générative repose en grande partie sur le pillage délibéré de la propriété intellectuelle des membres du groupe.
98. Sans restreindre la généralité de ce qui précède, les membres du groupe ont subi un préjudice matériel découlant des fautes de la défenderesse, y compris, mais sans s'y limiter, des pertes de revenus.
99. Ils ont également subi un préjudice moral en raison de la douleur, de la tristesse et de l'anxiété de voir les fruits de leur expression créative usurpés et instrumentalisés par une multinationale multimilliardaire qui développe des produits susceptibles de les concurrencer, et ce, sans leur accorder le moindre crédit.
100. Par conséquent, les membres du groupe ont droit à des dommages-intérêts et à la restitution des profits en vertu de l'article 35 LDA ou, au choix, à des dommages-intérêts préétablis au montant 20 000 \$ par œuvre en vertu de l'article 38.1 LDA.
101. Les membres du groupe ont également droit à des dommages-intérêts punitifs en vertu de la Charte et de la common law.
102. Finalement, les membres du groupe ont droit à une injonction pour faire cesser les violations de leurs droits et toute activité résultant de telles violations.

4. LES CONDITIONS REQUISES POUR L'AUTORISATION

4.1. Les faits donnant ouverture au recours individuel de la demanderesse

103. La demanderesse n'est pas parvenue à son succès facilement ou par hasard.
104. Diplômée de l'Université de Montréal en français et en traduction, elle a exercé le métier d'adjointe juridique durant 27 ans puis travaillé trois ans comme traductrice à la pige.
105. Pendant ces 30 années, elle a inlassablement soumis ses romans à diverses maisons d'édition qui, bien qu'appréciant leur qualité, ne croyaient pas suffisamment à l'existence d'un marché pour la littérature de genre fantastique au Québec.
106. Son destin a changé en 2002 avec la parution de son roman *Le feu dans le ciel*, premier tome de la série *Les Chevaliers d'Émeraude*.
107. L'univers des *Chevaliers d'Émeraude* a rapidement rejoint son lectorat, jeunes et jeunes de cœur, et est devenu un véritable phénomène littéraire.

108. Après des années d'acharnement, la demanderesse pouvait finalement vivre de sa plume.
109. Elle a depuis écrit plus de 100 livres, chacun étant le fruit de son imagination fertile et du travail acharné qu'elle consacre sans relâche à son art.
110. LibGen contient plus d'une centaine de copies piratées de livres de la demanderesse, tel qu'il appert de captures d'écran des résultats de recherche pour « Anne Robillard » dans le contenu de LibGen, **PIÈCE P-33**.
111. Z-Library contient plus d'une centaine de copies piratées de livres de la demanderesse, tel qu'il appert de captures d'écran des résultats de recherche pour « Anne Robillard » dans le contenu de Z-Library, **PIÈCE P-34**.
112. La demanderesse, qui détient les droits d'auteur et moraux sur l'ensemble de son œuvre, n'a jamais consenti à ce que des copies de ses livres soient mis à disposition pour téléchargement sur LibGen ou Z-Library, ni sur aucune autre bibliothèque clandestine.
113. Elle n'a jamais été approchée par la défenderesse et n'a jamais consenti à ce que celle-ci reproduise ses œuvres pour quelque fin que ce soit, et encore moins, pour développer des LLM et outils IA.
114. La demanderesse a récemment appris que la défenderesse avait téléchargé le contenu de bibliothèques clandestines, y compris celles où se retrouvent illégalement ses œuvres, dans le cadre du développement et de la formation de ses LLM.
115. Elle est outragée qu'une multinationale comme la défenderesse ait fait le choix d'usurper ainsi sa propriété intellectuelle et celle de centaines de milliers d'auteurs et de créateurs dans une course effrénée à la domination du marché de l'IA.

4.2. Les questions de fait ou de droit identiques, similaires ou connexes

116. Les demandes des membres soulèvent les questions de droit ou de fait identiques, similaires ou connexes suivantes :
 - (a) La défenderesse a-t-elle reproduit sans autorisation des œuvres protégées par le droit d'auteur dont les membres du groupe sont les titulaires?
 - (b) La défenderesse a-t-elle formé ses LLM avec des ensembles de données d'apprentissage reproduisant des œuvres protégées par le droit d'auteur dont les membres du groupe sont les titulaires?

- (c) La défenderesse a-t-elle violé de quelque façon le droit d'auteur des membres du groupe?
- (d) La défenderesse a-t-elle violé de quelque façon les droits moraux des membres du groupe sur leurs œuvres?
- (e) La défenderesse a-t-elle commis une faute envers les membres du groupe en se livrant à une concurrence parasitaire et déloyale?
- (f) La défenderesse s'est-t-elle enrichie sans justification aux dépens des membres?
- (g) La défenderesse a-t-elle violé les droits des membres en vertu de la *Charte des droits et libertés de la personne*?
- (h) Quels sont les remèdes et dommages auxquels ont droit les membres du groupe?

4.3. La composition du groupe

- 117. La composition du groupe rend difficile ou peu pratique, voire impossible, l'application des règles sur le mandat d'ester en justice pour le compte d'autrui ou sur la jonction d'instance.
- 118. Le secret entretenu par la défenderesse sur les œuvres qui ont servis et servent à développer et former ses LLM fait en sorte qu'il est impossible d'identifier et de retracer tous les membres du groupe.
- 119. Toutefois, on peut raisonnablement estimer que le groupe comprend des dizaines de milliers de personnes, incluant des auteurs, éditeurs, cessionnaires et ayant droits.
- 120. Ainsi, selon Statistique Canada entre 6 800 et 8 485 nouveaux ouvrages d'auteurs canadiens sont publiés annuellement au Canada, tel qu'il appert du « Tableau 21-10-0207-01 Éditeurs de livres, nouveaux ouvrages édités selon la catégorie commerciale et l'origine des auteurs » de Statistique Canada, **PIÈCE P-35**.
- 121. Dans ces circonstances, l'action collective est une voie appropriée pour que les membres du groupe puissent faire valoir leurs droits et aient accès à la justice.

4.4. La représentation adéquate

- 122. La demanderesse est en mesure d'assurer une représentation adéquate des membres du groupe.
- 123. Elle est prête à consacrer le temps nécessaire au dossier.

124. Elle a donné mandat à ses avocats d'obtenir tous les renseignements pertinents au dossier et se tiendra informée des développements.
125. Elle est de bonne foi et entreprend ces procédures dans l'unique but de faire en sorte que ses droits et ceux des membres du groupe soient reconnus et respectés et qu'il soit remédié aux préjudices que chacun a subis.

5. LA COMPÉTENCE DU TRIBUNAL

126. Les autorités québécoises sont compétentes pour entendre l'action collective du fait que la défenderesse a un établissement au Québec et que la contestation est relative à son activité au Québec, que des fautes ont été commises en tout ou en partie au Québec, et que les membres du groupe ont subi et subissent un préjudice au Québec.
127. Joelle Pineau et le Laboratoire FAIR qu'elle dirige depuis Montréal ont joué un rôle central dans la recherche, le développement et la formation des LLM de la défenderesse, y compris dans la prise de décisions relativement à la reproduction d'œuvres protégées par le droit d'auteur pour ces fins.
128. En outre, les fautes de la défenderesse, notamment la collecte et la reproduction de textes et de données provenant de l'Internet, dont des copies d'œuvres québécoises, ont nécessairement eu lieu au moins en partie au Québec.
129. Les membres du groupe qui résident ou sont domiciliés au Québec y ont tous subi ou subissent un préjudice résultant de la violation intentionnelle de leurs droits par la défenderesse.
130. Sans limiter la généralité de ce qui précède, en s'accaparant et en reproduisant sans autorisation les œuvres dont les droits d'auteurs sont détenus par les membres québécois, la défenderesse a privé ces derniers de leur droit d'obtenir, au Québec, une juste compensation pour leur travail.
131. En outre, la défenderesse savait que ses ensembles de données d'apprentissage contenaient des œuvres piratées du monde entier, y compris d'auteurs québécois. Elle pouvait donc prévoir que des préjudices découlant de sa décision d'exploiter ces œuvres sans autorisation pour ses fins commerciales se manifesteraient au Québec.
132. La demanderesse propose que la présente action collective soit exercée dans le district judiciaire de Montréal puisqu'un nombre important des membres du groupe et les avocats soussignés y sont domiciliés, et que la défenderesse y a un établissement.

POUR CES MOTIFS, PLAISE À LA COUR

- A. **ACCUEILLIR** la présente demande;
- B. **AUTORISER** l'exercice de l'action collective, dont la nature est celle d'une action en dommages-intérêts, restitution de profits et injonction;
- C. **DÉSIGNER** Anne Robillard à titre de représentante du groupe des personnes ci-après décrit :

Toute personne au Canada titulaire du droit d'auteur sur une œuvre littéraire ou dramatique que la défenderesse Meta Platforms, Inc. a reproduite sans autorisation dans le cadre du développement ou de la formation de grands modèles de langage (en anglais, « *large language models* »).

- D. **IDENTIFIER** comme suit les principales questions de droit et de fait qui seront traitées collectivement :

- (a) La défenderesse a-t-elle reproduit sans autorisation des œuvres protégées par le droit d'auteur dont les membres du groupe sont les titulaires?
- (b) La défenderesse a-t-elle formé ses LLM avec des ensembles de données d'apprentissage reproduisant des œuvres protégées par le droit d'auteur dont les membres du groupe sont les titulaires?
- (c) La défenderesse a-t-elle violé de quelque façon le droit d'auteur des membres du groupe?
- (d) La défenderesse a-t-elle violé de quelque façon les droits moraux des membres du groupe sur leurs œuvres?
- (e) La défenderesse a-t-elle commis une faute envers les membres du groupe en se livrant à une concurrence parasitaire et déloyale?
- (f) La défenderesse s'est-t-elle enrichie sans justification aux dépens des membres?
- (g) La défenderesse a-t-elle violé les droits des membres en vertu de la *Charte des droits et libertés de la personne*?
- (h) Quels sont les remèdes et dommages auxquels ont droit les membres du groupe?

- E. **IDENTIFIER** comme suit les conclusions recherchées qui s'y rattachent :

- A. ACCUEILLIR la présente demande;
- B. CONDAMNER la défenderesse à payer une somme à être déterminée à titre de dommages-intérêts punitifs, avec intérêt au taux légal et l'indemnité additionnelle à compter du jugement;

- C. CONDAMNER la défenderesse à payer à chaque membre du groupe un montant de VINGT MILLE DOLLARS (20 000 \$), sauf à parfaire, par œuvre dont elle s'est servie sans autorisation à titre de dommages-intérêts compensatoires, avec intérêt au taux légal et l'indemnité additionnelle à compter de l'assignation;
 - D. CONDAMNER la défenderesse à payer, en sus des dommages-intérêts, une proportion équitable des profits réalisés en commettant les violations du droit d'auteur des membres du groupe, avec intérêt au taux légal et l'indemnité additionnelle à compter de l'assignation;
 - E. AU CHOIX, EN REMPLACEMENT DES CONCLUSIONS C ET D : CONDAMNER la défenderesse à payer à chaque membre du groupe un montant de VINGT MILLE DOLLARS (20 000 \$) par œuvre dont elle s'est servie sans autorisation à titre de dommages-intérêts préétablis, avec intérêt au taux légal et l'indemnité additionnelle à compter de l'assignation;
 - F. ENJOINDRE à la défenderesse de cesser toute commercialisation, distribution, ou publication de LLM ou de tout autre outil d'IA formé sur des œuvres des membres du groupe sans leur consentement et de détruire toute copie de ces œuvres ou données qu'elle a en sa possession;
 - G. ORDONNER le recouvrement collectif des réclamations;
 - H. ORDONNER à la défenderesse de déposer au greffe de cette cour la totalité des sommes susdites, ainsi que les intérêts et l'indemnité additionnelle;
 - I. LE TOUT avec frais de justice incluant les frais d'experts, de publication d'avis et d'administration des réclamations;
- F. **DÉCLARER** qu'à moins d'exclusion, les membres du groupe seront liés par tout jugement à intervenir sur l'action collective de la manière prévue par la loi;
 - G. **FIXER** le délai d'exclusion à trente (30) jours après la date de l'avis aux membres, délai à l'expiration duquel les membres du groupe qui ne se seront pas exclus seront liés par tout jugement à intervenir;
 - H. **ORDONNER** la publication d'un avis aux membres dans les termes et par les moyens à déterminer;
 - I. **RÉFÉRER** le dossier au juge en chef pour détermination du district dans lequel l'action collective devra être exercée et désignation du juge pour l'entendre;
 - J. **ORDONNER** au greffier de cette cour, pour le cas où l'action doit être exercée dans un autre district, de transmettre le dossier, dès décision du juge en chef, au greffier de cet autre district;

K. **LE TOUT** avec frais de justice incluant les frais de publication de l'avis.

MONTREAL, le 21 mars 2025

A handwritten signature in blue ink that reads "Audren Rolland". The signature is written in a cursive style with a large initial 'A'.

Audren Rolland s.e.n.c.r.l.

Avocats de la demanderesse Anne Robillard

AVIS D'ASSIGNATION
(Art. 145 et suivants CPC)

Dépôt d'une demande en justice

Prenez avis que le demandeur a déposé au greffe de la Cour supérieure du Québec, district judiciaire de Montréal, la présente Demande pour autorisation d'exercer une action collective et pour être représentante.

Pièces au soutien de la demande

Au soutien de sa Demande pour autorisation d'exercer une action collective et pour être représentante, la demanderesse invoque les pièces suivantes :

- P-1** Liste des livres d'Anne Robillard publiée sur le site Web anne-robillard.com
- P-2** Communiqué de presse intitulé « Meta Reports Fourth Quarter and Full Year 2024 Results » daté du 29 janvier 2025
- P-3** Extrait du site Web llama.com
- P-4** Blog intitulé « The future of AI: Built with Llama » publié sur le site Web ai.meta.com
- P-5** Article intitulé « Celebrating Five Years of AI Breakthroughs from Our Montreal Research Lab » publié sur le site Web aboutfb.com
- P-6** Article de Fortune intitulé « How Mark Zuckerberg has Fully Rebuilt Meta Around Llama » publié sur le site Web finance.yahoo.com
- P-7** Publication du Gouvernement du Canada intitulée « Consultation sur le droit d'auteur à l'ère de l'intelligence artificielle générative »
- P-8** Rapport du United States Patent & Trademark Office intitulé « Public Views on Artificial Intelligence and Intellectual Property Policy »
- P-9** Document intitulé « Notice of Inquiry on Artificial Intelligence & Copyright (Dkt. 2023-6): Comments of Meta Platforms, Inc. »
- P-10** Document de recherche intitulé « LLaMa: Open and Efficient Foundation Language Models »
- P-11** Publications X de Shawn Presser datées du 25 octobre 2020
- P-12** Document de recherche intitulé « The Pile: An 800GB Dataset of Diverse Text for Language Modeling »

- P-13** Capture d'écran de la réponse donnée par Meta AI à la question « Qu'est-ce que Bibliotik? »
- P-14** Articles intitulés « Revealed : The Authors Whose Pirated Books are Powering Generative AI » et « These 183,000 Books are Fueling the Biggest Fight in Publishing and Tech » publiés sur le site Web theatlantic.com, en liasse
- P-15** Article intitulé « CBC News Analysis Finds Thousands of Canadian Authors, Books in Controversial Dataset Used to Train AI » publié sur le site Web cbc.ca
- P-16** Article intitulé « Montreal Writers Want More Protections as AI Sucks up Their Stories » publié sur le site Web cbc.ca
- P-17** Captures d'écran des résultats de recherche du contenu de Books3 dans The Atlantic
- P-18** Courriel de Melanie Kambadur à Melanie Kambadur et Nisha Deo intitulé « Message summary »
- P-19** Article intitulé « The Battle Over Books3 Could Change AI Forever » publié sur le site Web wired.com
- P-20** Extraits de l'interrogatoire hors cour de Melanie Kambadur dans l'affaire Kadrey et al. v. Meta Platforms, Inc. (Case No. 3:23-cv-03417-VC) devant la United States District Court, Northern District of California
- P-21** Page « Library Genesis » sur le site Web fr.wikipedia.org
- P-22** Captures d'écran des résultats de recherche sur LibGen
- P-23** Capture d'écran de la réponse donnée par Meta AI à la question « What is the LibGen dataset? »
- P-24** Courriel de Eleonora Presani à Eleonora Presani et Joelle Pineau intitulé « Message summary »
- P-25** Courriel de Sony Theakanath à Joelle Pineau intitulé « [A/C Priv] Request for decision: LibGen Escalation OneLLM »
- P-26** Courriel de Xiaolan Wang à Nikolay Bashlykov et al. intitulé « Message summary » daté du 2 avril 2024
- P-27** Courriel de David Esiobu à Nikolay Bashlykov et al. intitulé « Message summary » daté du 11 avril 2024
- P-28** Page « Seeding (computing) » sur le site Web en.wikipedia.org
- P-29** Capture d'écran de la réponse donnée par Meta AI à la question « Qu'est-ce que Z-Library »
- P-30** Captures d'écran des résultats de recherche sur Z-Library

- P-31** Courriel de Laurens van der Maaten à Sergey Edunov et Laurens van der Maaten intitulé « Message summary » daté du 23 août 2023
- P-32** Article intitulé « Meta Gains After Zuckerberg Predicts 'Really Big Year' in AI » publié sur le site finance.yahoo.com
- P-33** Captures d'écran des résultats de recherche pour « Anne Robillard » dans le contenu de LibGen
- P-34** Captures d'écran des résultats de recherche pour « Anne Robillard » dans le contenu de Z-Library
- P-35** « Tableau 21-10-0207-01 Éditeurs de livres, nouveaux ouvrages édités selon la catégorie commerciale et l'origine des auteurs » de Statistique Canada

Réponse à cette demande

Vous devez répondre à cette demande par écrit, personnellement ou par avocat, au palais de justice de Montréal situé au 1, rue Notre-Dame Est, Montréal (Québec) H2Y 1B6, district de Montréal dans les 15 jours de la signification de la présente demande ou, si vous n'avez ni domicile, ni résidence, ni établissement au Québec, dans les 30 jours de celle-ci. Cette réponse doit être notifiée à l'avocat du demandeur.

Défaut de répondre

Si vous ne répondez pas dans le délai prévu, de 15 ou de 30 jours, selon le cas, un jugement par défaut pourra être rendu contre vous sans autre avis dès l'expiration de ce délai et vous pourriez, selon les circonstances, être tenu au paiement des frais de justice.

Contenu de la réponse

Dans votre réponse, vous devez indiquer votre intention, soit :

- de convenir du règlement de l'affaire;
- de proposer une médiation pour résoudre le différend;
- de contester cette demande et, dans les cas requis par le Code de procédure civile, d'établir à cette fin, en coopération avec le demandeur, le protocole qui régira le déroulement de l'instance. Ce protocole devra être déposé au greffe de la Cour du district mentionné plus haut dans les 45 jours de la signification du présent avis. Toutefois, ce délai est de 3 mois en matière familiale ou si vous n'avez ni domicile, ni résidence, ni établissement au Québec;
- de proposer la tenue d'une conférence de règlement à l'amiable.

Cette réponse doit mentionner vos coordonnées et, si vous êtes représenté par un avocat, le nom de celui-ci et ses coordonnées.

Lieu et dépôt de la demande en justice

Cette demande est, sauf exceptions, entendue dans le district judiciaire où est situé votre domicile ou, à défaut, votre résidence ou le domicile que vous avez élu ou convenu avec le demandeur. Si elle n'a pas été déposée dans le district où elle peut être entendue et que vous voulez qu'elle y soit transférée, vous pouvez présenter une demande au tribunal à cet effet.

Cependant, si cette demande porte sur un contrat de travail, de consommation ou d'assurance ou sur l'exercice d'un droit hypothécaire sur l'immeuble vous servant de résidence principale, elle est entendue dans le district où est situé le domicile ou la résidence du salarié, du consommateur ou de l'assuré, qu'il soit demandeur ou défendeur, dans le district où est situé cet immeuble ou dans le district où le sinistre a eu lieu s'il s'agit d'une assurance de biens. Si cette demande n'a pas été déposée dans le district où elle peut être entendue et que vous voulez qu'elle y soit transférée, vous pouvez, sans qu'une convention contraire puisse vous être opposée, présenter une demande à cet effet au greffier spécial de ce district.

Transfert de la demande à la Division des petites créances

Si vous avez la capacité d'agir comme demandeur suivant les règles relatives au recouvrement des petites créances, vous pouvez communiquer avec le greffier du tribunal pour que cette demande soit traitée selon ces règles. Si vous faites cette demande, les frais de justice du demandeur ne pourront alors excéder le montant des frais prévus pour le recouvrement des petites créances.

Convocation à une conférence de gestion

Dans les 20 jours suivant le dépôt du protocole mentionné plus haut, le tribunal pourra vous convoquer à une conférence de gestion en vue d'assurer le bon déroulement de l'instance. À défaut, ce protocole sera présumé accepté.

MONTREAL, le 21 mars 2025



Audren Rolland s.e.n.c.r.l.

Avocats de la demanderesse Anne Robillard

**ATTESTATION D'INSCRIPTION AU
RÉPERTOIRE NATIONAL DES ACTIONS COLLECTIVES**
(Art. 55 du *Règlement de la Cour supérieure du Québec en matière civile*)

Les avocats soussignés attestent que la Demande pour autorisation d'exercer une action collective et pour être représentante sera inscrite au Répertoire national des actions collectives.

MONTRÉAL, le 21 mars 2025



Audren Rolland s.e.n.c.r.l.
Avocats de la demanderesse Anne Robillard

(Chambre des actions collectives)
COUR SUPÉRIEURE
DISTRICT DE MONTRÉAL
N° 500-06- 001369-251

ANNE ROBILLARD

Demanderesse

c.

META PLATFORMS, INC.

Défenderesse

**DEMANDE POUR AUTORISATION
D'EXERCER UNE ACTION COLLECTIVE ET
POUR ÊTRE REPRÉSENTANTE**
(Art. 574 et suivants CPC)

ORIGINAL

AUDREN | ROLLAND

Audren Rolland s.e.n.c.r.l.
393, rue St-Jacques, bureau 248
Montréal (Québec), H2Y 1N9
Tél. 514.974.3145
Télééc. 514.284.7771

Me Marie Audren
maudren@audrenrolland.com

BA1391