

***Machine learning* et droit d'auteur : la quête d'un équilibre**

Reza Moradinejad*

RÉSUMÉ	1505
INTRODUCTION	1507
1. <i>MACHINE LEARNING</i> ET DROIT D'AUTEUR : CONFRONTATION INCONTOURNABLE.....	1511
1.1 Développement des systèmes d'intelligence artificielle : une brève description du processus technique.....	1511
1.2 « The right to read is [not] the right to mine »	1515
2. <i>MACHINE LEARNING</i> ET DROIT D'AUTEUR : COEXISTENCE SOUHAITÉE.....	1518
2.1 Les conditions de coexistence dans le cadre juridique actuel : incertitude et inadaptation.....	1519
2.1.1 L'utilisation équitable : un cadre limité et incertain.....	1520
2.1.2 Les reproductions temporaires pour des processus technologiques : un cadre inadapté..	1525
2.2 Vers la création d'une solution certaine et adaptée . . .	1527

© Reza Moradinejad, 2021.

* Avocat, professeur adjoint à la Faculté de droit de l'Université Laval. L'auteur tient à remercier son auxiliaire de recherche Luis Urréa, étudiant au baccalauréat en droit à l'Université Laval, de sa contribution à cet article.

[Note : cet article a été soumis à une évaluation à double anonymat.]

2.2.1	Diversité de solutions en droit comparé : de l'utilisation de la notion de <i>fair use</i> à la création d'une exception dédiée	1527
2.2.2	Pistes de réflexion pour le Canada.	1537
	CONCLUSION.	1541

RÉSUMÉ

Le présent article vise à explorer les avenues possibles pour trouver un équilibre entre la nécessité de protéger le droit d'auteur et celle d'encourager le développement des systèmes d'intelligence artificielle (système IA), lequel pourrait impliquer la reproduction de données protégées par le droit d'auteur au cours de processus de *machine learning* (ML). L'étude du droit positif démontre le caractère incertain et inadapté du cadre juridique actuellement en vigueur. En l'absence d'une disposition spécifique, l'exception générale de l'utilisation équitable devient pratiquement la seule possibilité pour reproduire des données protégées sans violer les droits des titulaires du droit d'auteur. Or, cette solution place le développement des systèmes IA à la merci des tribunaux et de leur appréciation du caractère équitable ou non de chaque cas d'espèce. D'où la nécessité de trouver une avenue garantissant une meilleure stabilité juridique. Le survol des solutions retenues en droit comparé nous amène à privilégier une solution fondée sur la création d'une exception spécifique applicable à la ML.

INTRODUCTION

Annoncée comme une nouvelle révolution, l'intelligence artificielle – propulsée par de nouvelles techniques de *machine learning* (ML) – monopolise désormais l'actualité scientifique. Toutes disciplines confondues, les chercheurs appliquent déjà des systèmes d'intelligence artificielle¹ (système IA) à leur domaine de recherche. Le recours aux systèmes IA ne se limite cependant pas à la recherche scientifique. Nouvelle ruée vers l'or, l'IA attire les entreprises dans différents secteurs de l'économie, et ce, dans le but de développer de nouvelles applications pratiques. De la médecine² jusqu'à la mode³ en passant par le commerce électronique⁴, les systèmes IA rendent possibles des avancées jusqu'ici difficilement imaginables. L'efficacité des solutions fondées sur les systèmes IA – ou peut-être l'aura magique qui les entoure – fait en sorte qu'une compétition internationale s'amorce autour du développement des systèmes IA. Faisant preuve d'imagination, les États tentent de trouver une façon d'attirer

1. Dans ce texte, pour définir le « système d'intelligence artificielle », on retient la définition proposée par l'Office québécois de la langue française : « Système conçu pour simuler le fonctionnement de l'intelligence humaine afin d'exécuter des fonctions relevant normalement de celle-ci. » (« système d'intelligence artificielle », en ligne : <http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=23798323> (consulté le 17 mai 2021).
2. Par exemple, citons cette recherche qui suggère l'utilisation de *deep learning* pour le dépistage de la Covid-19 : Tahereh JAVAHERI, Morteza HOMAYOUNFAR, Zohreh AMOOZGAR, Reza REIAZI, Fatemeh HOMAYOUNIEH, Engy ABBAS, Azadeh LAALI, Amir Reza RADMARD, Mohammad Hadi GHARIB, Seyed Ali Javad MOUSAVI, Omid GHAEMI, Rosa BABAEI, Hadi Karimi MOBIN, Mehdi HOSSEINZADEH, Rana JAHANBAN-ESFAHLAN, Khaled SEIDI, Mannudeep K. KALRA, Guanglan ZHANG, L. T. CHITKUSHEV, Benjamin HAIBE-KAINS, Reza MALEKZADEH et Reza RAWASSIZADEH, « CovidCTNet: An Open-Source Deep Learning Approach to Identify Covid-19 Using CT Image », *arXiv:2005.03059 [cs, eess]* 2020, en ligne : <<http://arxiv.org/abs/2005.03059>> (consulté le 10 mai 2021).
3. Yuli LIANG, Seung-Hee LEE et Jane E. WORKMAN, « Implementation of Artificial Intelligence in Fashion: Are Consumers Ready? », (2020) 38-1 *Clothing and Textiles Research Journal* 3.
4. Vishal Dineshkumar SONI, « Emerging Roles of Artificial Intelligence in E-commerce », (2020) 4-5 *International Journal of Trend in Scientific Research and Development* 223.

davantage de développeurs et de capitaux sur leur territoire afin de stimuler la recherche et le développement dans le domaine de l'IA⁵.

La réglementation juridique fait partie des éléments qui peuvent stimuler ou encore freiner le développement des systèmes IA⁶, d'où l'intérêt des juristes à étudier les implications juridiques de l'utilisation des systèmes IA⁷. L'un des domaines où la réglementation peut jouer un rôle clé est celui du droit de la propriété intellectuelle et plus particulièrement celui du droit d'auteur. Plusieurs points de croisement peuvent être identifiés entre le droit d'auteur et l'IA. Tout naturellement, nous pouvons penser à la question de la protection des algorithmes d'un système IA ou des codes sources⁸. Songeons également à la protection des œuvres générées par l'exploitation d'un système IA doté d'un algorithme de type *generative adversarial networks*⁹. Ces algorithmes sont capables de générer des œuvres originales. Dans l'hypothèse où l'œuvre est véritablement le fruit du travail de la machine et où aucun homme ne peut démontrer qu'il a personnellement fait preuve de jugement dans la création de l'œuvre, faut-il accorder un droit d'auteur sur l'œuvre ? Si oui, qui

-
5. Citons par exemple, la stratégie américaine renforcée par un décret présidentiel adopté en 2019 afin de promouvoir le développement des systèmes IA sous la présidence de Donald Trump. « Artificial Intelligence for the American People », en ligne : <<https://trumpwhitehouse.archives.gov/ai/>> (consulté le 4 mai 2021).
 6. Michael GEIST, « Why Copyright Law Poses a Barrier to Canada's Artificial Intelligence Ambitions – Michael Geist », *Michael Geist*, 18 mai 2017, en ligne : <<https://www.michaelgeist.ca/2017/05/copyright-law-poses-barrier-canadas-artificial-intelligence-ambitions/>> (consulté le 20 février 2021).
 7. Certains sujets abordés par les juristes retiennent davantage l'attention dans la sphère publique. Il en est ainsi des questions d'ordre éthique et déontologique qui se posent lors de l'utilisation d'un système IA (voir Keith FRANKISH et William M. RAMSEY, *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2014). Pensons également aux enjeux liés aux régimes de la responsabilité soulevés en lien avec l'utilisation des voitures autonomes (Jérôme PERRIN, « Éthique de responsabilité et de sollicitude dans la conception et l'usage des véhicules autonomes », (2020) 307-3 *Revue d'éthique et de théologie morale* 43). Plus près de l'objet de notre étude qui porte sur les données utilisées par les systèmes IA, nous constatons que beaucoup d'encre est coulée pour traiter des problématiques liées à la protection des renseignements personnels (voir, par exemple, Pierre-Luc DÉZIEL, « Les limites du droit à la vie privée à l'ère de l'intelligence artificielle : groupes algorithmiques, contrôle individuel et cycle de traitement de l'information », (2018) 30-3 *C.P.I.* 829).
 8. Les programmes d'ordinateur, définis à l'article 2 de la *Loi sur le droit d'auteur* (L.R.C. (1985), ch. C-42 [ci-après « LDA »]), peuvent faire l'objet d'une protection en vertu de l'article 3 LDA. Voir en ce sens, *Apple Computer, Inc. c. Mackintosh Computers Ltd.* ; *Apple Computer, Inc. c. 115778 Canada Inc.*, [1990] 2 R.C.S. 209.
 9. Ian J. GOODFELLOW, Jean POUGET-ABADIE, Mehdi MIRZA, Bing XU, David WARDE-FARLEY, Sherjil OZAIR, Aaron COURVILLE et Yoshua BENGIO, « Generative adversarial networks », (2020) 63-11 *Commun. ACM* 139.

serait le titulaire des droits ? Plusieurs solutions sont proposées par les auteurs¹⁰, mais le débat n'est pas encore clos. Cela dit, aucune de ces questions ne fait l'objet de notre étude.

Dans cet article, nous allons nous intéresser à la question de la conformité du processus du développement d'un système IA du point de vue des règles du droit d'auteur. À la base du fonctionnement des systèmes IA, lorsqu'il s'agit de mobiliser les algorithmes de ML, se trouve un processus impliquant l'utilisation des données. La question qui retiendra notre attention est celle de l'accès aux données lorsque celles-ci sont protégées par le droit d'auteur. La protection des données par le droit d'auteur peut être perçue comme un obstacle au développement des systèmes IA. Outre les bases de données déjà constituées qui peuvent, sous certaines conditions, bénéficier de la protection offerte par le droit d'auteur¹¹, les données comprises dans une base de données peuvent également être considérées comme une œuvre protégée. Dans l'hypothèse où de telles protections existent, la question générale se pose concernant le droit des développeurs d'utiliser ces données dans le cadre de l'élaboration d'un système IA.

Plus spécifiquement, nous nous interrogerons dans un premier temps sur les rapports entre le développement d'un système IA et le droit d'auteur. Autrement dit, les développeurs peuvent-ils procéder au développement d'un système IA sans se soucier des règles du droit d'auteur ? Ou bien ce processus doit-il se réaliser dans le cadre précis établi par la *Loi sur le droit d'auteur* (ci-après « LDA ») ? Du point de vue des développeurs, l'idéal serait de laisser le développement des systèmes IA échapper au giron du droit d'auteur. Or, la réalité est bien plus complexe : le droit d'auteur ne peut pas rester à l'écart dès lors qu'une activité implique l'utilisation d'œuvres protégées par le droit d'auteur.

Dans cette optique, il faut se pencher sur une autre question spécifique, soit celle de la réglementation du processus de ML par

10. Georges AZZARIA, « Intelligence artificielle et droit d'auteur : l'hypothèse d'un domaine public par défaut », (2018) 30-3 *C.P.I.* 927 ; Kalin HRISTOV, « Artificial Intelligence and the Copyright Dilemma », (2016) 57-3 *IDEA* 43 ; Giorgio FRANCESCHELLI et Mirco MUSOLESI, « Copyright in Generative Deep Learning », *arXiv:2105.09266 [cs]* 2021, en ligne : <<http://arxiv.org/abs/2105.09266>> (consulté le 27 mai 2021).

11. Une base de données peut être considérée comme une compilation. L'article 2 LDA définit la compilation comme suit : « Les œuvres résultant du choix ou de l'arrangement de tout ou partie d'œuvres littéraires, dramatiques, musicales ou artistiques ou de données. »

le droit d'auteur. Faudrait-il obtenir une autorisation chaque fois qu'une donnée protégée sera reproduite au cours de ce processus ? Ou bien une telle reproduction tomberait-elle dans l'une des exceptions prévues par le droit d'auteur relative à l'utilisation équitable d'une œuvre ?

L'enjeu est de taille, puisque la réponse déterminerait la rigidité ou la souplesse d'une partie du cadre légal applicable au processus de développement d'un système IA¹². Cet enjeu constitue l'une des composantes des coûts d'exploitation de ces systèmes. Plus le cadre réglementaire est souple, moindre sera le coût du développement des systèmes IA et vice versa. Par ailleurs, outre l'enjeu financier, l'assujettissement de ce processus aux règles du droit d'auteur pourrait rendre cette entreprise pratiquement impossible considérant le nombre colossal des données qui sous-entend un nombre important de licences à obtenir¹³. Même si l'enjeu est capital, le cadre juridique applicable manque de clarté ; d'où l'importance d'une clarification. Nous essaierons modestement de contribuer à préciser l'applicabilité des règles ainsi qu'à amorcer une réflexion sur les pistes de solution envisageable.

Dans la première partie de cet article, nous démontrerons que le processus du développement des systèmes IA est assujéti aux règles établies par le droit d'auteur. C'est par la compréhension du processus technique du développement des systèmes IA que nous pouvons faire cette affirmation. Dans un second temps, nous dévoilerons l'incertitude du cadre légal applicable. Ce constat nous amènera à proposer une réflexion sur les avenues possibles pour trouver un

12. Sur une note générale, nous pouvons citer ce passage de la décision *Théberge c. Galerie d'Art du Petit Champlain inc.*, 2002 CSC 34, par. 32, qui confirme l'importance de trouver un équilibre entre les différents intérêts antagonistes : « Un contrôle excessif de la part des titulaires du droit d'auteur et d'autres formes de propriété intellectuelle pourrait restreindre indûment la capacité du domaine public d'intégrer et d'embellir l'innovation créative dans l'intérêt à long terme de l'ensemble de la société, ou créer des obstacles d'ordre pratique à son utilisation légitime. Ce risque fait d'ailleurs l'objet d'une attention particulière par l'inclusion, aux art. 29 à 32.2, d'exceptions à la violation du droit d'auteur ».

13. Voir Niva ELKIN-KOREN, « Copyright in a Digital Ecosystem: A User-Rights Approach », dans Ruth L. OKEDIJI (dir.), *Copyright Law in an Age of Limitations and Exceptions*, Cambridge, Cambridge University Press, 2017, p. 132, à la p. 142, qui évoque le même problème dans le contexte de l'éducation. Selon l'auteur, « [f]or instance, a school teacher who seeks to show a few video clips in class for educational purposes may give up the use altogether if she is required to acquire a license for each and every use: identify multiple owners, negotiate a license, and pay the license fee ».

équilibre entre le développement des systèmes IA et le droit d'auteur, et ce, à l'aide du droit comparé.

1. MACHINE LEARNING ET DROIT D'AUTEUR : CONFRONTATION INCONTOURNABLE

La description technique du développement des systèmes IA nous permet de dissiper tout doute quant à son assujettissement aux règles de la LDA.

1.1 Développement des systèmes d'intelligence artificielle : une brève description du processus technique

Qu'est-ce que l'IA ? Tout le monde a une idée approximative de ce qu'est l'IA. C'est souvent à travers des applications plus ou moins ludiques que le grand public peut avoir une idée concrète du fonctionnement d'un système IA¹⁴. Mais, au-delà des exemples anecdotiques, conceptuellement, l'IA réfère à des tentatives d'imiter les facultés cognitives humaines¹⁵. Dans les années 1950, Alain Turing, le pionnier dans le domaine de la science informatique, proposa un test connu sous le nom d'« *Intimidation Game* » ou le test de Turing. En vertu de ce test, il faut demander à une personne (A) de correspondre avec deux ordinateurs dont un fonctionne de façon autonome et l'autre est opéré par une autre personne (B). Si dans au moins 30 % des cas, A ne réussit pas à distinguer l'homme de la machine, il est possible de dire que cette machine est dotée d'un système IA. Malgré la difficulté de satisfaire les exigences de ce test¹⁶, les systèmes qui tendent à simuler le raisonnement des humains peuvent être qualifiés d'IA.

14. Pensons, par exemple, à ces applications qui réaniment les photos d'une autre époque ; voir Julien BALDACCHINO, « "DeepNostalgia", l'outil pour faire revivre les vieilles photos, comment ça marche ? », *France inter*, 5 mars 2021, en ligne : <<https://www.franceinter.fr/emissions/net-plus-ultra/net-plus-ultra-05-mars-2021>> (consulté le 5 mai 2021).

15. Stuart J. RUSSELL et Peter NORVIG, *Artificial Intelligence: A Modern Approach*, 4^e éd., coll. « Pearson series in artificial intelligence », Hoboken, Pearson, 2021. Un auteur dans son livre introductif à l'IA (Wolfgang ERTEL, *Introduction to Artificial Intelligence*, 2^e éd., coll. « Undergraduate Topics in Computer Science », Cham, Springer International Publishing, 2017) propose quelques tentatives de définitions, dont celle suggérée par L'Encyclopédie Britannica qui décrit l'IA comme « the ability of digital computers or computer-controlled robots to solve problems that are normally associated with the higher intellectual processing capabilities of humans » ou une autre proposée par Elaine Rich (Elaine RICH, *Artificial Intelligence*, New York, McGraw-Hill, 1983) selon laquelle « [a]rtificial Intelligence is the study of how to make computers do things at which, at the moment, people are better ».

16. Avec la mise en œuvre d'un nouveau modèle, connu sous le nom de GPT-3, on peut dire que le test de Turing est satisfait. Voir Tom B. BROWN, Benjamin MANN,

Parmi les différentes façons de concevoir un système IA figure une technique, le *machine learning* (ML)¹⁷ notamment sous forme d'apprentissage supervisé, qui requiert parfois l'utilisation massive de données lors de la phase d'apprentissage de l'algorithme. Selon un auteur, « [m]achine learning (ML) is a field of computer science that studies algorithms and techniques for automating solutions to complex problems that are hard to program using conventional programming methods »¹⁸. Les algorithmes de ML ont cette particularité qu'ils n'ont pas besoin d'avoir un « explicit detailed design »¹⁹ comme c'est le cas pour les programmes informatiques traditionnels. Les algorithmes de ML de type supervisés « learn the detailed design from a set of labelled data »²⁰. Autrement dit, grâce à des techniques d'analyse de données, ces algorithmes réussissent à dégager une logique, une intelligence ou des rapports – souvent de type de corrélation – entre les données. Le bon fonctionnement des algorithmes de ML ainsi que la précision de leur résultat dépend intimement de la qualité et de la quantité des données utilisées ; d'où la quête pour colliger davantage de données.

L'aspect technique du développement de ML qui intéresse véritablement le sujet de notre recherche concerne l'acquisition et le traitement des données. Sous ces aspects, le ML se rapproche de la technique utilisée dans le *Text and Data Mining* (TDM). Il ne s'agit pas de trouver une information apparente comme le numéro de téléphone d'une personne dans un répertoire téléphonique. Le TDM nous permet de générer une information à partir des relations qui existent entre les données. À titre d'illustration, prenons un jeu de données qui contient les données relatives à la circulation d'automobiles dans une

Nick RYDER, Melanie SUBBIAH, Jared KAPLAN, Prafulla DHARIWAL, Arvind NEELAKANTAN, Pranav SHYAM, Girish SASTRY, Amanda ASKELL, Sandhini AGARWAL, Ariel HERBERT-VOSS, Gretchen KRUEGER, Tom HENIGHAN, Rewon CHILD, Aditya RAMESH, Daniel M. ZIEGLER, Jeffrey WU, Clemens WINTER, Christopher HESSE, Mark CHEN, Eric SIGLER, Mateusz LITWIN, Scott GRAY, Benjamin CHESS, Jack CLARK, Christopher BERNER, Sam MCCANDLISH, Alec RADFORD, Ilya SUTSKEVER et Dario AMODEI, « Language Models are Few-Shot Learners », *arXiv:2005.14165 [cs]* 2020, en ligne : <<http://arxiv.org/abs/2005.14165>> (consulté le 6 mai 2021).

17. Certains suggèrent « l'apprentissage de machine » comme équivalent pour *machine learning*. Dans cet article, nous préférons utiliser le terme anglais qui est plus couramment utilisé dans la littérature scientifique.

18. Gopinath REBALA, Ajay RAVI et Sanjay CHURIWALA, « Machine Learning Definition and Basics », dans Gopinath REBALA, Ajay RAVI et Sanjay CHURIWALA (dir.), *An Introduction to Machine Learning*, Cham, Springer International Publishing, 2019, p. 1, à la p. 1.

19. *Ibid.*, p. 2.

20. *Ibid.*

ville. Grâce au TDM, il serait possible de prévoir l'heure de pointe à un moment déterminé dans le futur. Cette information existe de façon latente dans les données. Elle n'est pas immédiatement accessible, mais elle le sera à l'aide des techniques d'exploration des données qui dévoile les relations existantes entre les données. Ces techniques peuvent être utilisées aussi bien dans une quête aussi triviale que celle de l'exemple précédent que dans un système IA capable de générer une œuvre artistique originale²¹.

De quoi parle-t-on lorsqu'on utilise le terme « données » ? S'agit-il nécessairement des suites de chiffres insérées dans un tableau ? Pas nécessairement. Une donnée peut référer aux « numbers, figures, images, sounds, computer programs (viewed as collections of data interpreted as instructions), etc. »²². Le concept de « donnée » est assez large et peut englober tout élément qui contient de l'information. Ce constat est encore plus important concernant le fonctionnement des systèmes IA dont une grande partie est consacrée à l'analyse des éléments d'information écrits, graphiques ou sonores. Du point de vue du droit d'auteur, certains proposent de faire la distinction entre deux types de données, les *high-level data* et les *low-level data* :

High-level data would consist of « rich » contents, such as newspaper or journal articles, books, music, photographs or cinematographic works. Low-level data would consist of more « raw » data, such as genetic information, measurement data in any field of science, name and address data, GPS coordinates, phone numbers or financial data.²³

Le type de données qui interpelle davantage le droit d'auteur est évidemment celui des *high-level data*. Cela dit, la reproduction de données de types *low-level* peut également susciter de l'intérêt du point de vue du droit d'auteur dès lors que celles-ci sont compilées dans une base de données dont la constitution témoigne de l'exercice du talent et du jugement d'une personne²⁴.

21. Voir I. GOODFELLOW *et al.*, préc., note 9.

22. Gorunescu FLORIN, *Data mining: concepts, models and techniques*, coll. « Intelligent systems reference library », n° 12, Berlin, Heidelberg, Springer-Verlag, 2011, p. 45.

23. Marco CASPERS, Lucie GUILBAULT, Kiera MCNEICE, Stelios PIPERIDIS, Kanella POULI, Maria ESKEVICH et Maria GAVRIILIDOU, « Baseline report of policies and barriers of TDM in Europe (extended version) », D3.3+, *FutureTDM*, 2017, p. 19, en ligne : <<https://www.futuretdm.eu/knowledge-library/>>.

24. *Robertson v. Thomson Corp.*, 2006 CSC 43, par. 37.

Que fait-on avec ces données dans un processus de ML ou de TDM ? Selon un auteur, on peut schématiser ce processus en trois phases :

1. Exploring data, consisting of data « cleansing », data transformation, dimensionality reduction, feature subset selection, etc. ;
2. Building the model and its validation, referring to the analysis of various models and choosing the one who has the best performance of forecast – competitive evaluation of models ;
3. Applying the model to new data to produce correct forecasts/ estimates for the problems investigated.²⁵

La phase qui nous concerne principalement est la première que l'on désigne communément comme l'étape de la préparation des données. Avant d'entamer cette phase, le développeur doit d'abord et avant tout avoir accès aux données. Il existe plusieurs façons d'obtenir les données. L'une d'entre elles serait d'acquérir une base de données déjà constituée, gratuitement ou moyennant une contrepartie. Il y a également la possibilité d'utiliser les données accessibles sur Internet. Les données ainsi colligées, si elles sont de type *high-level*, peuvent être protégées par le droit d'auteur. Le développeur doit-il alors obtenir une autorisation ou même payer des redevances aux ayants droit pour constituer une base de données ou pour utiliser de telles données ?

Une fois les données rassemblées, le développeur doit les préparer à être utilisées dans l'élaboration d'un modèle ou dans l'apprentissage de l'algorithme. Lors de cette phase, le développeur doit « nettoyer » les données. Autrement dit, il doit les rendre compatibles avec ses objectifs de recherche. Cela peut consister à supprimer ou ajouter des données ou encore à modifier certaines données. Au cours du processus, le développeur peut revenir en arrière pour corriger ses choix. Toutes ces manipulations pour préparer et « nettoyer » les données nécessitent la reproduction des données dans des fichiers différents. Ainsi, au cours du processus, une donnée peut être reproduite à de multiples reprises. Il y aura ensuite d'autres reproductions qui seront faites par l'ordinateur lors de l'étape d'apprentissage.

Faut-il payer des redevances chaque fois qu'une œuvre est reproduite pour les fins de la préparation des données ?

25. G. FLORIN, préc., note 22, p. 7.

1.2 « The right to read is [not] the right to mine »

Certains acteurs du milieu défendent l'idée selon laquelle « the right to read is the right to mine »²⁶. Cette idée part d'un constat assez simple. La simple lecture d'une œuvre écrite, l'observation d'un tableau, le visionnement d'un film ou l'écoute d'un morceau de musique ne constitue pas en soi une violation des droits d'auteur. On se demande alors pourquoi l'exploration de ces œuvres par une machine – qui en fin de compte ne fait que lire et observer les données afin d'y trouver des relations entre les divers éléments d'information – ne devrait pas être assimilée à la lecture d'une œuvre faite par une personne humaine. Prenons un exemple peu original : un développeur souhaite concevoir un système IA pour distinguer les chats des chiens. Lors de la phase d'apprentissage supervisé de l'algorithme, des milliers d'images de chats et de chiens seront présentées au modèle pour qu'il apprenne à trouver les traits communs des chats et des chiens. Conceptuellement, cet exercice ressemble à l'apprentissage d'un tout petit être humain qui, à force de regarder les différentes images des chats et des chiens, apprend à les distinguer. Cette ressemblance conceptuelle s'efface lorsqu'on fait un examen sérieux du cadre juridique applicable à l'analyse de données.

Partant de l'hypothèse que l'ensemble des données ou un certain nombre d'entre elles sont protégés en vertu de la LDA, la question se pose à savoir si les gestes posés en lien avec celles-ci pourraient être considérés comme une violation des droits d'auteur. La LDA reconnaît un nombre exhaustif de droits sur une œuvre originale. Parmi les droits d'auteur accordés, l'article 3 LDA prévoit notamment le droit exclusif de « reproduire la totalité ou une partie importante de l'œuvre, sous une forme matérielle quelconque ». Elle reconnaît également à l'artiste-interprète le droit d'auteur exclusif de reproduire sa prestation²⁷. La reproduction d'une œuvre sans l'autorisation du titulaire constitue une violation des droits d'auteur.

La LDA ne définit cependant pas la notion de reproduction. La Loi énonce simplement que le droit d'auteur comporte le droit exclusif

26. Ross MOUNCE, « The Right to Read is the Right to Mine: Text and Data Mining Copyright Exceptions Introduced in the UK », *Impact of Social Sciences Blog*, 4 juin 2014, en ligne : <<http://blogs.lse.ac.uk/impactofsocialsciences/>> (consulté le 14 mai 2021) ; M. GEIST, préc., note 6 ; Peter MURRAY-RUST, « The Right to Read Is the Right to Mine », *Open Knowledge Foundation blog*, 1^{er} juin 2012, en ligne : <<https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/>> (consulté le 20 février 2021).

27. Art. 15 LDA.

« de reproduire la totalité ou une partie importante de l'œuvre sous une forme matérielle quelconque ». La notion de reproduction a été précisée par la Cour suprême du Canada dans l'affaire *Théberge*. Le juge Binnie, au nom de la majorité, définit la reproduction « comme l'action de produire des copies supplémentaires ou nouvelles de l'œuvre sous une forme matérielle quelconque »²⁸. La *multiplication des copies* devient alors l'élément central pour déterminer si une œuvre est reproduite ou non. C'est le nombre de copies qui subsiste à la fin de l'opération qui détermine s'il y a eu un acte de reproduction ou non. Cette conception de reproduction ne fait pas pourtant l'unanimité. Dans la même affaire, les juges minoritaires, sous la plume du juge Gonthier, considèrent que « reproduire [une] œuvre, sous une forme matérielle quelconque » équivaut simplement à « rematérialiser » ce qui existait déjà sous une première forme matérielle »²⁹. C'est donc l'action de *rematérialisation de l'œuvre* qui constitue le critère déterminant pour conclure qu'il y a eu une reproduction ou non. Selon cette conception, peu importe le nombre de copies qui subsistent à la fin de l'opération, si au cours du processus, l'œuvre a été rematérialisée, on peut conclure qu'il y a eu reproduction³⁰.

Pour les fins de notre étude, que l'on retienne le critère de la multiplication de copies ou celui de la rematérialisation de l'œuvre, force est de constater qu'il y aura nécessairement une reproduction de l'œuvre protégée. Pensons d'abord à ce qui arrive au tout début du processus. On acquiert ou on constitue une base de données. Cette opération consiste à enregistrer une copie des données dans un ordinateur. Si cela concerne la reproduction de données dont le contenu constitue une œuvre protégée, la reproduction nécessite l'autorisation des ayants droit. Plus tard, au cours du processus et au moment de la préparation des données, la création d'autres copies des œuvres sera nécessaire. Ces reproductions constituent une violation de droit d'auteur à moins d'être autorisées ou de faire partie d'une exception légale.

En sus des reproductions faites par le développeur lors de l'étape de la constitution et de la préparation des données, le processus

28. *Théberge c. Galerie d'Art du Petit Champlain inc.*, préc., note 12, par. 42.

29. *Ibid.*, par. 146.

30. Cet avis est partagé par Normand TAMARO, *2020 Annotated Copyright Act*, Toronto, Thomson Reuters Canada, 2019. Selon M. Tamaro, « [w]ith all due respect, we must criticize the legal conclusion of Justice Binnie on the basis of the principle that mere fixation is not covered by copyright. Such a conclusion runs counter to the justification underlying copyright as established since its inception. And we cannot agree with his conclusion on the facts to the effect that, where there was a poster, there still is a poster ».

d'élaboration d'un système IA implique également la reproduction automatique des données par la machine au cours de l'apprentissage de l'algorithme. D'aucuns diraient que ces dernières opérations ne constituent pas une reproduction au sens de la LDA, car la copie de l'œuvre n'a jamais été visualisée par un être humain. Cet argument ne résiste pas à l'analyse, puisque la prise de connaissance de la copie par un être humain n'est pas un critère déterminant.

D'autres pourraient prétendre que la reproduction d'une œuvre protégée nécessite une autorisation dès lors que cette opération implique une perte économique pour le titulaire des droits d'auteur. Ou encore, qu'une autorisation est requise lorsque la personne qui reproduit l'œuvre en tire un avantage économique. Même si économiquement cet argument semble intéressant, il ne présente pas une grande utilité sur le plan juridique. Le monopole que le titulaire détient sur la reproduction de l'œuvre ne souffre d'exception que si cela est prévu par la LDA. En dehors des cas prévus expressément par la Loi – que nous allons passer en revue dans la seconde partie de notre article –, toute forme de reproduction de l'œuvre constitue une violation des droits du titulaire. Le titulaire du droit d'auteur n'a pas à justifier un quelconque intérêt, y compris économique, pour faire valoir ses droits. Selon la Cour suprême du Canada :

[...] un enregistrement [soit une forme de reproduction dans le contexte de la radiodiffusion d'une œuvre musicale] fait dans n'importe quel but, même non préjudiciable au titulaire du droit d'auteur, sans l'autorisation du titulaire du droit d'auteur constitue une violation de ses droits.³¹ (nos soulignements)

Autrement dit, ce n'est ni la valeur économique créée par l'exploitation de l'œuvre protégée ni le préjudice subi par le titulaire du droit d'auteur qui doit être prouvé pour conclure à la violation du droit d'auteur.

Nous pouvons donc sans hésitation affirmer que le processus d'élaboration d'un système IA pourrait tomber sous le coup des règles de la LDA dès lors que les données exploitées ou la base de données utilisée sont protégées par le droit d'auteur.

L'assujettissement à la LDA ne signifie pas nécessairement que les développeurs sont systématiquement tenus d'obtenir une autorisation pour l'utilisation des données. La LDA prévoit un certain nombre

31. *Bishop c. Stevens*, [1990] 2 R.C.S. 467.

d'exceptions permettant aux utilisateurs de reproduire librement une œuvre. La question se pose alors de savoir si les reproductions faites dans le cadre de l'élaboration d'un système IA pourraient bénéficier de l'une des exceptions prévues par la LDA.

2. MACHINE LEARNING ET DROIT D'AUTEUR : COEXISTENCE SOUHAITÉE

La LDA est une loi qui s'inscrit principalement dans la tradition britannique de la protection des auteurs. Le *Copyright Act* de 1710³² (*Statute of Anne*) avait pour objectif d'accorder aux auteurs le droit exclusif de contrôler la publication et la reproduction de leurs œuvres³³. Les titulaires du droit d'auteur (principalement l'auteur et la maison d'édition) furent le personnage central de cet acte³⁴, l'objectif étant d'empêcher les autres de reproduire les œuvres concernées sans autorisation. Au Canada, dès 1921, la législation sur le droit d'auteur suivait cet objectif³⁵. Or, au fil du temps, cette conception fut modifiée. C'est dans la célèbre affaire *Théberge* que la Cour suprême du Canada récapitule cette évolution en procédant à un recentrage des objectifs de la LDA. Ainsi, les objectifs de la LDA étaient présentés par la Cour dans les termes suivants :

La Loi est généralement présentée comme établissant un équilibre entre, d'une part, la promotion, dans l'intérêt du public, de la création et de la diffusion des œuvres artistiques et intellectuelles et, d'autre part, l'obtention d'une juste récompense pour le créateur (ou, plus précisément, l'assurance que personne d'autre que le créateur ne pourra s'approprier les bénéfices qui pourraient être générés).³⁶

32. *Copyright Act 1710 (Statute of Anne)*, 1710 (UK), 8 Ann., c. 19.

33. « From its inception, copyright was a “professional right”: a right used by professionals against other professionals. » (Daniel J. GERVAIS, « The Purpose of Copyright Law in Canada », (2005) 2-2 *U. Ottawa L. & Tech. J.* 315, 326).

34. Voir Michael GEIST, « The Canadian Copyright Story : How Canada Improbably Became the World Leader on Users' Rights in Copyright Law », dans Ruth L. OKEDIJI (dir.), *Copyright Law in an Age of Limitations and Exceptions*, Cambridge, Cambridge University Press, 2017, p. 169, à la p. 172, qui met en lumière ce passage de la décision *Bishop c. Stevens*, préc., note 31, 478, où la CSC, citant la jurisprudence britannique (*Performing Right Society, Ltd. v. Hammond's Bradford Brewery Co.*, [1934] 1 Ch. 121, 127), énonce que « la *Copyright Act* de 1911 a un but unique et a été adoptée au seul profit des auteurs de toutes sortes, que leurs œuvres soient littéraires, dramatiques ou musicales ».

35. *Loi pour amender et refondre la Loi relative au droit d'auteur*, S.C. 1921, c. 42. Pour l'évolution législative du droit d'auteur, voir Érika BERGERON-DROLET, « Les exceptions de la *Loi sur le droit d'auteur* : rétrospective et état des lieux », (2016) 28-2 C.P.I. 301.

36. *Théberge c. Galerie d'Art du Petit Champlain inc.*, préc., note 12, par. 30.

Par la suite, la Cour suprême explique la façon dont cet équilibre peut être atteint :

On atteint le juste équilibre entre les objectifs de politique générale, dont ceux qui précèdent, non seulement en reconnaissant les droits du créateur, mais aussi en accordant l'importance qu'il convient à la nature limitée de ces droits.³⁷

Les multiples exceptions prévues aux articles 29 et suivants de la LDA s'inscrivent dans cette philosophie. Dans la même veine, la Cour suprême affirme dans l'affaire *CCH*³⁸ que ces exceptions constituent « un droit des utilisateurs » plutôt qu'un « moyen de défense ». Citant le professeur Vaver, la Cour rappelle :

Les droits des utilisateurs ne sont pas de simples échappatoires. Les droits du titulaire et ceux de l'utilisateur doivent donc recevoir l'interprétation juste et équilibrée que commande une mesure législative visant à remédier à un état de fait.³⁹

Dans l'optique où « [u]n acte visé par l'exception relative à l'utilisation équitable ne viole pas le droit d'auteur »⁴⁰, il serait intéressant d'examiner si les processus de ML et de TDM réalisés dans le cadre de l'élaboration d'un système IA peuvent être visés par l'une des exceptions actuellement prévues par la LDA. Un examen sommaire – et nécessairement *in abstracto*, en l'absence d'une jurisprudence en la matière – nous indique l'incertitude du droit positif applicable à l'élaboration des systèmes IA, d'où la quête pour de nouvelles solutions inspirées notamment du droit comparé.

2.1 Les conditions de coexistence dans le cadre juridique actuel : incertitude et inadaptation

Parmi de nombreuses exceptions prévues par la LDA permettant aux utilisateurs de reproduire une œuvre sans obtenir préalablement l'autorisation de l'ayant droit, deux pourraient théoriquement s'appliquer aux reproductions faites au cours d'un processus de l'élaboration d'un système IA. Nous verrons à tour de rôle l'applicabilité

37. *Ibid.*, par. 31.

38. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, [2004] 1 R.C.S. 339, 2004 CSC 13, par. 48.

39. David VAVER, *Copyright Law*, coll. « Essentials of Canadian Law », Toronto, Irwin Law, 2000, p. 171.

40. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 48.

de l'exception relative à l'utilisation équitable aux fins d'étude privée et de recherche (2.1.1) et de celle portant sur les reproductions temporaires pour processus technologiques (2.1.2).

2.1.1 L'utilisation équitable : un cadre limité et incertain

L'exception la plus importante – et historiquement l'une des quelques exceptions⁴¹ prévues dans la *Loi sur le droit d'auteur* de 1921⁴² – est celle qui concerne l'utilisation équitable (*fair dealing*) d'une œuvre protégée. Bien qu'elle soit présente dès la genèse de la législation nationale sur le droit auteur en 1921, l'utilisation équitable a connu, comme l'affirme le professeur Katz⁴³, une nouvelle vie avec la décision *CCH* de la Cour suprême du Canada. Avant d'aborder l'apport de cette décision dans le renouveau de l'article 29 LDA, il convient de reproduire le texte de cet article :

29. L'utilisation équitable d'une œuvre ou de tout autre objet du droit d'auteur aux fins d'étude privée, de recherche, d'éducation, de parodie ou de satire ne constitue pas une violation du droit d'auteur.

Dans l'affaire *CHH*, la Cour suprême établit un test en deux étapes pour déterminer si l'utilisation d'une œuvre protégée constitue une utilisation équitable de cette même œuvre. À cette fin, l'utilisateur doit prouver « (1) qu'il s'agit d'une utilisation aux fins d'étude privée ou de recherche et (2) qu'elle était équitable »⁴⁴.

Le premier volet du test vise à déterminer si l'utilisation correspond à l'une des fins prévues par la loi, soit l'étude privée ou la recherche. Selon la Cour, ce premier volet du test contient « un critère relativement peu strict »⁴⁵. Il n'est pas nécessaire que l'utilisation soit faite dans le cadre d'une recherche scientifique⁴⁶. Par ailleurs, selon la doctrine, le premier volet de ce test, soit l'identification des cas visés

41. Pour un historique des exceptions en droit d'auteur canadien, voir É. BERGERON-DROLET, préc., note 35.

42. S.C. 1921, c. 24, art. 16(1)(i).

43. Ariel KATZ, « Fair Use 2.0: The Rebirth of Fair Dealing in Canada », dans Michael GEIST (dir.), *The Copyright Pentology: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law*, Ottawa, University of Ottawa Press, 2013, p. 93, à la p. 94.

44. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 50.

45. *Société canadienne des auteurs, compositeurs et éditeurs de musique c. Bell Canada*, 2012 CSC 36, par. 27.

46. *1395804 Ontario Ltd. (Blacklock's Reporter) c. Canada (Procureure générale)*, 2016 CF 1255, par. 31 : « Il est donc possible de dégager les quatre principes

par l'exception de l'utilisation équitable, doit être interprété de façon large et libérale⁴⁷. À titre d'exemple, l'écoute préalable de l'extrait d'un morceau de musique avant que le consommateur décide d'acheter l'œuvre ou non est considérée comme un cas d'étude privée. Quant aux systèmes IA, il est certain que le premier volet du test est satisfait lorsque la reproduction est faite dans le cadre de l'exploitation des données s'inscrivant dans un projet de recherche scientifique. En revanche, lorsque cette exploitation vise une activité ayant des fins commerciales, la satisfaction de ce premier volet du test ne semble pas facilement acquise. Nous voyons donc déjà un premier obstacle à une application générale de cette exception au processus d'élaboration des systèmes IA.

Quant au second volet du test, soit l'évaluation du caractère équitable de l'utilisation, la Cour suggère l'analyse de six facteurs qui, par ailleurs, doivent être interprétés de façon large et libérale⁴⁸ : « (1) le but de l'utilisation ; (2) la nature de l'utilisation ; (3) l'ampleur de l'utilisation ; (4) les solutions de rechange à l'utilisation ; (5) la nature de l'œuvre ; (6) l'effet de l'utilisation sur l'œuvre »⁴⁹. Cette liste n'est pas exhaustive et le respect de l'ensemble des facteurs n'est pas indispensable pour conclure que l'utilisation de l'œuvre est équitable. Cela dit, les différentes juridictions ayant subséquentement tranché un litige portant sur l'article 29 LDA analysent systématiquement les six facteurs tels qu'expliqués par la Cour suprême⁵⁰ tout en reconnaissant par moment le caractère non exhaustif de ces facteurs.

supplémentaires suivants d'après la discussion du terme "recherche" que l'on trouve dans l'arrêt *SOCAN* :

1. Il n'est pas nécessaire que la recherche soit tributaire de la poursuite d'une fin créative ;
2. La recherche n'est pas limitée aux fins créatives, mais elle peut être "fragmentaire, informelle, exploratoire ou confirmative", et elle peut même être entreprise pour aucun autre motif que l'intérêt personnel ;
3. Le critère appliqué pour le premier volet de l'analyse relative à l'utilisation équitable est relativement souple et il ne nécessite pas la création d'une œuvre nouvelle ;
4. L'analyse devrait être entreprise du point de vue de l'utilisateur ou de la fin que poursuit le consommateur ».

47. M. GEIST, préc., note 34, p. 181 ; N. TAMARO, préc., note 30, p. 791.

48. M. GEIST, préc., note 34, p. 199.

49. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 53.

50. *Université York c. Canadian Copyright Licensing Agency*, 2020 CAF 77 ; *Société québécoise de gestion collective des droits de reproduction (Copibec) c. Université Laval*, 2016 QCCS 900 ; *Société québécoise de gestion collective des droits de reproduction (Copibec) c. Université Laval*, 2017 QCCA 199 ; *Cedrom-SNI inc. c. Dose Pro inc.*, 2017 QCCS 3383 ; *Canadian Copyright Licensing Agency c. Ministère de l'éducation de la Colombie-Britannique*, 2017 CAF 16.

Une analyse sommaire de ces facteurs démontre l'incertitude quant à la possibilité d'appliquer cette exception à la reproduction dans le contexte de l'analyse de données. S'agissant d'une question mixte de faits et de droit, la détermination du caractère équitable d'une utilisation nécessite une analyse contextuelle dépendant intimement des faits de chaque litige. Cela dit, en appliquant la grille d'analyse détaillée proposée par la Cour suprême du Canada, nous pouvons identifier *prima facie* les obstacles potentiels à l'applicabilité de cette exception à la reproduction des œuvres au cours du processus de ML et TDM.

Le premier facteur porte sur le but de l'utilisation. Selon la Cour suprême, « il s'agit alors de déterminer objectivement le "but ou le motif réel" de l'utilisation de l'œuvre protégée »⁵¹. Quel est le but de l'exploitation des données dans un processus de ML ? Dans bien des cas, l'objectif est de trouver des rapports entre les données afin de dégager un nouveau savoir ; cet objectif relève plutôt du monde des idées que celui des expressions. En revanche, si le ML est développé pour concevoir un modèle ayant pour but d'imiter l'expression des idées des œuvres analysées, il serait plus difficile d'admettre que le but de l'utilisation de l'œuvre est équitable. Dans un autre registre, l'élaboration d'un système IA dans un but commercial pourrait également être considérée comme un facteur de complication. Il faut pourtant rappeler que l'importance de ce dernier obstacle est bien tempérée par la Cour suprême. Dans l'affaire SOCAN, cette dernière rappelle que :

[...] même si la recherche effectuée pour un motif commercial peut être moins équitable que celle effectuée à une fin non commerciale [...], l'utilisation peut tout de même être équitable si l'on « garantit raisonnablement » que les œuvres sont effectivement utilisées aux fins de recherche.⁵²

Ainsi, sans écarter complètement la possibilité d'avoir une utilisation équitable à but commercial, on comprend que ce premier facteur ne milite pas en faveur de l'application de l'article 29 dans un contexte de développement d'un système IA à but commercial.

Concernant le deuxième facteur, la Cour suprême explique en ces termes les éléments à considérer :

51. *Société canadienne des auteurs, compositeurs et éditeurs de musique c. Bell Canada*, préc., note 45, par. 33.

52. *Ibid.*, par. 36.

Pour déterminer la nature d'une utilisation, le tribunal doit examiner la manière dont l'œuvre a été utilisée. Lorsque de multiples copies sont diffusées, l'utilisation tend à être inéquitable. Toutefois, lorsqu'une seule copie est utilisée à une fin légitime en particulier, on peut conclure plus aisément que l'utilisation était équitable. Si la copie de l'œuvre est détruite après avoir été utilisée comme prévu, cela porte également à croire qu'il s'agissait d'une utilisation équitable.⁵³

Dans le cadre de l'élaboration d'un système IA, même si la préparation des données nécessite la création de plusieurs copies, force est de constater que celles-ci ne font pas l'objet d'une large diffusion. En effet, les copies servent uniquement à alimenter l'algorithme pour des fins d'apprentissage. De plus, les copies pourraient être détruites à la fin du processus. Ce facteur pourrait alors militer pour le caractère équitable de l'utilisation.

Le troisième facteur fait référence à l'importance de la reproduction de l'œuvre. Est-ce la totalité de l'œuvre qui est reproduite ou bien la reproduction concerne-t-elle une partie infime de l'œuvre ? Appliqué à l'exploitation de données, ce facteur tend à écarter l'hypothèse d'une utilisation équitable. En effet, l'exploitation de données nécessite la reproduction de l'ensemble de la donnée protégée.

Le quatrième facteur concerne l'existence ou l'accessibilité d'une solution de rechange. D'après la Cour suprême :

[...] l'existence de solutions de rechange à l'utilisation d'une œuvre protégée par le droit d'auteur peut avoir une incidence sur le caractère équitable ou inéquitable de l'utilisation. Lorsqu'un équivalent non protégé aura pu être utilisé à la place de l'œuvre, le tribunal devra en tenir compte.⁵⁴

Pour l'élaboration d'un système IA, la solution de rechange consisterait à limiter la fouille des données aux seuls jeux de données non protégés ; ce qui limiterait considérablement les possibilités de développement des systèmes IA. Une autre solution de rechange serait d'obtenir les autorisations requises avant de procéder à la fouille de données. Cette solution, bien que possible, augmente considérablement le coût de l'exploitation des données ainsi que la complexité de la

53. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 55.

54. *Ibid.*, par. 57.

gestion des données. De plus, d'autres pourraient dire que le fait de se contenter d'un nombre limité de données pourrait biaiser le modèle⁵⁵.

Quant au cinquième facteur, la Cour nous rappelle que :

Le tribunal doit également tenir compte de la nature de l'œuvre pour décider du caractère équitable de son utilisation. Bien qu'il ne s'agisse certainement pas d'un facteur décisif, l'utilisation d'une œuvre non publiée sera plus susceptible d'être équitable du fait que sa reproduction accompagnée d'une indication de la source pourra mener à une diffusion plus large de l'œuvre en question, ce qui est l'un des objectifs du régime de droit d'auteur. Par contre, si l'œuvre en question était confidentielle, la balance pourrait pencher en faveur du caractère inéquitable de l'utilisation.⁵⁶

Ce facteur s'applique plus ou moins à la fouille de données. Les données reproduites ne sont pas destinées à être diffusées ; elles sont préparées pour être lues par une machine. Par conséquent, la reproduction faite dans ce cadre ne peut jouer un rôle ni en faveur ni en défaveur de l'objectif de la protection de l'œuvre.

Finalement, le sixième facteur examine l'effet de l'utilisation sur l'œuvre. Selon la Cour suprême,

La concurrence que la reproduction est susceptible d'exercer sur le marché de l'œuvre originale peut laisser croire que l'utilisation n'est pas équitable. Même si l'effet de l'utilisation sur le marché est un facteur important, ce n'est ni le seul ni le plus important.⁵⁷

À l'instar du paragraphe précédent, ce sixième facteur revêt une importance minimale dans le cas d'exploitation de données dans l'exercice d'apprentissage de machine. En effet, les données sont simplement présentées à l'ordinateur, sans qu'il soit question de les exploiter et de les diffuser. Cela étant dit, dans l'hypothèse où le système IA serait conçu pour produire une œuvre semblable

55. Rachel MEADE, « Bias in Machine Learning: How Facial Recognition Models Show Signs of Racism, Sexism and Ageism », *Towards Data Science*, 13 décembre 2019, en ligne : <<https://towardsdatascience.com/bias-in-machine-learning-how-facial-recognition-models-show-signs-of-racism-sexism-and-ageism-32549e2c972d>> (consulté le 10 mai 2021).

56. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 58.

57. *Ibid.*, par. 59.

aux œuvres qui ont fait l'objet d'exploration de données (*Generative adversarial networks*), l'introduction de ces nouvelles œuvres pourrait faire perdre la valeur économique des données utilisées. Ainsi, en fonction de l'objectif du système IA, ce dernier facteur peut jouer en faveur ou en défaveur du caractère équitable de l'utilisation.

L'analyse globale – et, il faut la rappeler, abstraite – des facteurs proposés par la Cour suprême indiquent que l'application de l'exception de l'utilisation équitable aux cas d'exploitation de données dans le domaine de l'IA demeure, à tout le moins, hypothétique et incertaine, et ce, même lorsque la fin poursuivie n'est pas de nature commerciale.

2.1.2 Les reproductions temporaires pour des processus technologiques : un cadre inadapté

En 2012, la *Loi sur la modernisation du droit d'auteur*⁵⁸, dont l'un des objectifs fut d'adapter la *Loi sur le droit d'auteur* à l'ère numérique, ajouta une nouvelle exception au droit d'auteur spécifiquement applicable à la reproduction d'une œuvre dans le cadre d'un processus technologique. Le nouvel article 30.71 de la *Loi sur le droit d'auteur* se lit comme suit :

30.71 Ne constitue pas une violation du droit d'auteur le fait de reproduire une œuvre ou tout autre objet du droit d'auteur si les conditions suivantes sont réunies :

- a) la reproduction est un élément essentiel d'un processus technologique ;
- b) elle a pour seul but de faciliter une utilisation qui ne constitue pas une violation du droit d'auteur ;
- c) elle n'existe que pour la durée du processus technologique.

Cette nouvelle règle permet la reproduction non autorisée d'une œuvre dès lors qu'elle est indispensable au fonctionnement d'un processus technologique. La reproduction doit être temporaire et ne doit pas faciliter une utilisation qui constitue une violation du droit d'auteur. La jurisprudence sur cet article est relativement maigre. Dans une décision rendue le 21 avril 2016, la Commission du droit

58. L.C. 2012, c. 20.

d'auteur examine l'application de cette nouvelle règle. Elle rappelle d'abord que cette exception ne fait pas référence « à un processus commercial ou à une activité économique »⁵⁹. Ainsi, d'après la Commission, il ne suffit pas qu'un processus technologique soit utilisé dans une opération pour que la reproduction non autorisée tombe sous cette exception. De l'avis de la Commission :

La disposition vise les copies qui se font automatiquement ou sans le contrôle direct de l'utilisateur, ce qui se rapporte à l'exigence voulant que la reproduction soit un élément essentiel du processus technologique. Un « processus technologique » au sens de l'article 37.01 de la *Loi* ne fonctionnera généralement pas correctement ou efficacement sans qu'une reproduction ne soit effectuée. Ainsi, il n'appartient pas à l'utilisateur de déterminer si la copie doit être réalisée ou non.⁶⁰

Par ailleurs, concernant le caractère temporaire de cette exception, la Commission estime que « la destruction d'une telle reproduction est vraisemblablement automatique dans le cadre de la technologie utilisée »⁶¹. Finalement, cette exception vise les reproductions qui « n'ont pas de buts indépendants ou d'utilisations autres que de permettre le fonctionnement du processus technologique »⁶².

La reproduction des données à des fins d'apprentissage de machine ne semble pas tomber exactement dans le champ d'application de cette exception. Celle-ci vise principalement des « copies » enregistrées dans le mémoire tampon.

Dans l'élaboration d'un système IA, il va y avoir de multiples reproductions des données par le développeur lors de la phase de la préparation des données. Ces copies du jeu de données ne sont pas indispensables au fonctionnement d'un système IA. Celles-ci visent à améliorer la qualité des données et à augmenter l'efficacité du système IA. De plus, c'est le développeur en personne qui décide quand il faut créer une nouvelle copie et à quel moment cette copie sera supprimée. Donc, l'aspect automatique de la destruction des données qui garantit son caractère temporaire ne semble pas exister dans le cas des reproductions faites lors de la phase de la préparation. Cette

59. *Tarif des redevances à percevoir par la SOCAN, Ré*, 2016 CarswellNat 1285, par. 178.

60. *Ibid.*, par. 180.

61. *Ibid.*, par. 181.

62. *Ibid.*, par. 188.

première sorte de reproductions ne peut donc pas être visée par cette exception.

2.2 Vers la création d'une solution certaine et adaptée

L'étude du droit positif nous montre qu'actuellement les activités relatives au développement des systèmes IA impliquant la reproduction des œuvres protégées tombent sous le couperet du droit d'auteur et qu'aucune exception ne leur permet de s'y soustraire sans équivoque. Ce constat – partagé par ailleurs dans beaucoup de systèmes juridiques – a amené plusieurs États à revoir la réglementation du droit d'auteur afin de faciliter le développement des systèmes IA. Le Canada est encore en phase de réflexion⁶³. L'idée a été évoquée par plusieurs parties prenantes devant un comité parlementaire de la Chambre des Communes, mais à l'heure actuelle aucune perspective de réforme n'est à l'horizon.

L'étude du droit comparé nous permettra d'amorcer une réflexion et une discussion critique sur le choix qui se présente au Canada afin de trouver un juste équilibre entre la protection des auteurs et le développement des systèmes IA.

2.2.1 *Diversité de solutions en droit comparé : de l'utilisation de la notion de fair use à la création d'une exception dédiée*

Malgré la diversité des solutions, le survol du droit comparé nous indique qu'il existe *grosso modo* deux catégories de solutions. La première solution qui se dégage est celle retenue en droit américain, où une interprétation large et libérale de l'exception générale de *fair use* permet de capter une grande partie des activités liées au développement d'un système IA. La seconde solution qui se dégage est celle consistant à prévoir un régime d'exception spécifiquement applicable au développement des systèmes IA.

Approche américaine – Interprétation large de l'exception de l'utilisation équitable : En droit américain, la section 107 de la

63. La question a été évoquée dans le rapport du Comité permanent de l'industrie, des sciences et de la technologie de la Chambre des communes (COMITÉ PERMANENT DE L'INDUSTRIE, DES SCIENCES ET DE LA TECHNOLOGIE, *Examen prévu par la loi de la Loi sur le droit d'auteur*, 16-INDU (42-1), Chambre des communes, 2019, en ligne : <<https://www.noscommunes.ca/DocumentViewer/fr/42-1/INDU/rapport-16>> (consulté le 19 février 2021)) qui a formulé une recommandation qui invite le législateur à agir dans ce domaine.

Copyright Law of the United States est consacrée à la doctrine de *fair use*. L'article se lit comme suit :

§ 107 Limitations on exclusive rights: Fair use

Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include –

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes ;

(2) the nature of the copyrighted work ;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole ; and

(4) the effect of the use upon the potential market for or value of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

La doctrine de *fair use*⁶⁴ est englobante et flexible. Contrairement à son équivalent canadien (utilisation équitable ou *fair dealing*), la section 107 ne procède pas à une énumération exhaustive des cas pouvant faire l'objet d'une utilisation équitable. La mention « such as » indique que les exemples cités au premier alinéa de cette section sont simplement à titre illustratif. Comme l'a rappelé un auteur, « [c]uriously, the first two Supreme Court decisions that construed section 107 do not appear to comport with the statutory

64. 17 U.S.C. § 107 (2012).

illustrations »⁶⁵. S'agissant d'une question mixte de fait et de droit⁶⁶, la Cour suprême des États-Unis laisse aux tribunaux le soin de décider le caractère équitable de l'utilisation selon les facteurs du deuxième alinéa de cette section⁶⁷.

Parmi les facteurs mentionnés dans la loi, c'est le premier facteur qui semble être plus controversé. Au fil du temps, la jurisprudence américaine est venue atténuer l'importance du premier facteur, soit le caractère commercial ou non commercial de l'utilisation⁶⁸. Ainsi, le simple fait que la reproduction d'une œuvre soit faite dans un but commercial n'exclut pas d'emblée sa qualification comme cas de *fair use*. L'approche américaine est axée sur la notion d'utilité et transformation⁶⁹ et le régime est interprété largement⁷⁰. Cette approche se distingue par son potentiel adaptatif et par le fait qu'elle répond aux nouvelles réalités à l'aide des principes fondamentaux qui guident le droit d'auteur américain⁷¹. En dépit de cet avantage indéniable, cette approche entraîne beaucoup d'incertitude dans le droit américain et est critiquée par la doctrine du fait qu'elle ne guide pas suffisamment les magistrats⁷². Un important rôle étant accordé aux juges, une disparité dans l'interprétation même des critères ainsi que dans la prépondérance accordée à chacun d'entre eux est observable d'état en état⁷³. Certains auteurs observent également une disparité au sein des cours de justice d'un même État ; les cours inférieures ayant tendance à favoriser le défendeur⁷⁴. D'autres prédisent que l'approche des

65. Robert GORMAN, *Copyright Law*, 2^e éd., Federal Judicial Center, 2006, p. 142. Les deux décisions mentionnées sont *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984) où le litige portait sur la légalité des copies maisons d'un film et *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985) où malgré le fait que l'utilisation correspondait à un cas énuméré (*news reporting*), la Cour suprême des États-Unis a refusé la qualification de *fair use*.

66. *Ibid.*

67. Matthew SAG, « Copyright Law's Impact on Machine Intelligence in the United States and the European Union Symposium: Intelligent Entertainment: Shaping Policies on the Algorithmic Generation and Regulation of Creative Works », (2020) 14-2 *FIU L. Rev.* 293, 297.

68. R. GORMAN, préc., note 65, p. 146 : « In examining the first factor, the Court downgraded the importance of the defendants' "commercial use", noting that essentially all fair use claims (and the uses enumerated in the first sentence) are made in the for-profit context of publishing and broadcasting ».

69. M. SAG, préc., note 67, 294.

70. *Ibid.*, 294 ; Mark A. LEMLEY et Bryan CASEY, *Fair Learning*, 2020, p. 134, en ligne : <<http://dx.doi.org/10.2139/ssrn.3528447>>.

71. M. SAG, préc., note 67, 297.

72. Stephen MCJOHN et Ian MCJOHN, « Fair Use and Machine Learning », (2020) 12-1 *N.E. U. L.R.* 99, 117.

73. *Ibid.*, 111 et s.

74. *Ibid.*, 119.

tribunaux dans les années à venir sera défavorable aux processus de ML du fait que cette utilisation ne constitue pas une transformation de l'œuvre en plus d'être un « free riding »⁷⁵, phénomène par lequel de grandes entreprises profitent des œuvres de « petits plaignants »⁷⁶.

Prévision d'une exception spécifique : Certaines autres juridictions comme le Royaume-Uni, l'Union européenne et le Japon ont déjà décidé de prévoir une règle claire applicable aux activités impliquant l'analyse informatique. Nous verrons, dans l'ordre, la règle prévue dans la *Copyright Law* du Royaume-Uni, celle suggérée dans la *Directive sur le droit d'auteur et les droits voisins*⁷⁷ et finalement les dispositions pertinentes de la *Copyright Law* du Japon.

En 2014⁷⁸, le Royaume-Uni a introduit un nouvel article 29A à la *Copyright Law* ayant pour objet de réglementer l'analyse des données. L'article se lit comme suit :

29A Copies for text and data analysis for non-commercial research

(1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that –

(a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and

(b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).

(2) Where a copy of a work has been made under this section, copyright in the work is infringed if –

75. M. A. LEMLEY et B. CASEY, préc., note 70, p. 107.

76. *Ibid.*

77. *Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE*, [2019] J.O. L 130/92.

78. L'article 29A est ajouté le 1^{er} juin 2014 par les *Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014*, S.I. 2014/1372, art. 3(2).

- (a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or
 - (b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.
- (3) If a copy made under this section is subsequently dealt with –
- (a) it is to be treated as an infringing copy for the purposes of that dealing, and
 - (b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.
- (4) In subsection (3) “dealt with” means sold or let for hire, or offered or exposed for sale or hire.
- (5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.

La disposition britannique présente quelques particularités. Tout d’abord, elle subordonne l’application du régime à l’obtention licite des données. Par exemple, si les données proviennent de la numérisation des pages d’un livre, l’acquisition de ce livre ne doit pas constituer, elle-même, une violation de droit d’auteur ; un exemplaire de ce livre doit être acheté ou emprunté dans une bibliothèque. Cette disposition accorde par la suite le droit au développeur de reproduire autant qu’il le souhaite – sous réserve des autres conditions de cet article – l’œuvre protégée. Qu’en est-il des données qui sont dès le départ de nature numérique ? Cette disposition permet-elle aux développeurs de constituer une base de données par le téléchargement non autorisé des œuvres qui se trouvent sur Internet ou sur un autre support ? *A priori*, la réponse est négative. L’article 29A ne constitue pas un passe-droit pour télécharger toute œuvre qui se trouve sur Internet. Citons à ce propos, l’*Intellectual Property Office* du Royaume-Uni qui affirme que les « [r]esearchers will still have to buy subscriptions to access material ; this could be from many sources including academic publishers »⁷⁹. En plus, cette première condition

79. INTELLECTUAL PROPERTY OFFICE, « Exceptions to copyright », *Guidance: Exceptions to copyright*, 12 juin 2014, en ligne : <<https://www.gov.uk/guidance/>

interdit également tout bris d'une mesure technique de protection pour accéder à l'œuvre.

La deuxième particularité du régime britannique applicable à l'analyse informatique concerne sa portée. Elle ne protège que l'analyse de données à but non commercial. C'est la recherche scientifique – ou à tout le moins la fouille de données à but non lucratif – qui fait l'objet de cette exception. En revanche, contrairement à l'approche américaine, l'approche britannique ne souffre pas d'incertitude. La création d'une exception spécifique a le mérite d'affirmer clairement le droit d'exploiter les données. Cela dit, cette approche semble plus restrictive que celle adoptée par le droit américain en ce qu'elle élimine toute possibilité d'exploitation des données, sans autorisation, dans le cadre d'un projet à but commercial.

La troisième particularité du régime britannique concerne son caractère impératif. L'article 29A (5) précise qu'aucun arrangement contractuel ne peut restreindre ce droit.

Emboîtant le pas au Royaume-Uni, l'Union européenne⁸⁰, dans sa directive adoptée en 2019 (DSM), va plus loin. Les articles 3 et 4 de la directive sont consacrés à l'exception relative à la fouille de données. L'article 3 de cette directive établit le régime applicable à la fouille de textes et de données à des fins de recherche scientifique :

Article 3

Fouille de textes et de données à des fins de recherche scientifique

1. Les États membres prévoient une exception aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.

exceptions-to-copyright> (consulté le 10 mai 2021).

80. Sur l'approche européenne, voir M. SAG, préc., note 67 ; Maurizio BORGHI, « Exceptions as users' rights? », dans Eleonora ROSATI (dir.), *The Routledge Handbook of EU Copyright Law*, Routledge, 2021, p. 263.

2. Les copies des œuvres ou autres objets protégés effectuées dans le respect du paragraphe 1 sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche.

3. Les titulaires de droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés. Ces mesures n'excèdent pas ce qui est nécessaire pour atteindre cet objectif.

4. Les États membres encouragent les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord des bonnes pratiques concernant l'application de l'obligation et des mesures visées aux paragraphes 2 et 3, respectivement.

S'agissant d'une disposition d'ordre public⁸¹, cet article précise bien que l'exception concerne la fouille de données effectuée dans le cadre de la recherche scientifique. Ainsi, contrairement à la disposition britannique, le domaine visé par cette règle ne se définit pas par rapport à son objectif commercial ou non commercial. C'est spécifiquement la recherche scientifique qui est ciblée. Par conséquent, la fouille de donnée réalisée dans le cadre d'une activité récréative ne saurait bénéficier du régime assez favorable de cet article. Par ailleurs, à l'instar de l'article 29A de la *Copyright Law* du Royaume-Uni, le droit des chercheurs à reproduire les données est conditionnel à la licéité de la méthode d'obtention des données. Finalement, le troisième paragraphe de cet article reconnaît explicitement le droit des ayants droit d'utiliser des mesures techniques de protection.

Parallèlement au régime plutôt libéral concernant la recherche scientifique, l'article 4 de la DSM prévoit un régime plus contraignant applicable aux autres activités impliquant la fouille de texte et de données :

Article 4

Exception ou limitation pour la fouille de textes et de données

81. *Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE*, préc., note 77, art. 7(1).

1. Les États membres prévoient une exception ou une limitation aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE, à l'article 4, paragraphe 1, points a) et b), de la directive 2009/24/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de manière licite aux fins de la fouille de textes et de données.
2. Les reproductions et extractions effectuées en vertu du paragraphe 1 peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données.
3. L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.
4. Le présent article n'affecte pas l'application de l'article 3 de la présente directive.

Ce régime, que l'on peut considérer comme le régime général applicable à la fouille de texte et de données, se distingue du régime applicable à la recherche scientifique par la possibilité d'*opt-out* prévue pour les ayants droit. Les titulaires de droits peuvent, en vertu du troisième paragraphe de l'article 4, limiter expressément l'utilisation de leurs œuvres de sorte que la fouille de données devienne impossible ou restreinte.

Il existe une autre distinction majeure entre les deux articles. Les scientifiques peuvent garder les données aussi longtemps qu'elles sont nécessaires à des fins scientifiques. Ceci dépasse passablement la période consacrée à la fouille de données. Le droit de conserver les données semble être bien plus limité dans le cadre des fouilles réalisées à des fins autres que la recherche scientifique. Le paragraphe 2 de l'article 4 précise que les données « peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données ».

Finalement, il existe un troisième modèle de l'exception dédiée à l'analyse informatique en droit japonais. Du fait de ses particulari-

tés, le choix du législateur japonais mérite d'être exposé. Précurseur en matière de législation sur le TDM, le Japon avait adopté, en 2009, une exception propre à l'analyse informatique. Le septième paragraphe de l'article 47 de la *Copyright Law* se lisait comme suit :

For the purpose of information analysis [...] by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis.⁸²

Conçue de façon assez large, cette exception ne faisait aucune distinction fondée sur le but commercial ou la recherche scientifique. Cette exception protégeait toute activité impliquant la fouille de données. En 2018, une modification de la *Copyright Law* japonaise est venue proposer de nouvelles solutions. Deux principaux articles sont désormais consacrés à l'analyse de données. Lisons d'abord la règle de l'article 30-4 :

Article 30-4 It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:

[...]

(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data ; the same applies in Article 47-5, paragraph (1), item (ii));

[...]

Visant les reproductions faites au cours d'un processus de TDM, le critère principal est la perception de l'œuvre par le développeur.

82. Cet extrait est cité par M. CASPERS *et al.*, préc., note 23, p. 75.

Cette exception concerne des cas où l'être humain n'utilise pas personnellement les œuvres. L'exception concerne des cas d'analyse informatique où la perception des œuvres par l'homme devient secondaire et accessoire par rapport à l'objectif de la reproduction. La copie est destinée à être utilisée par la machine ; sa consultation éventuelle par l'homme est incidente. Un autre critère est mentionné dans cet article : la reproduction de l'œuvre ne doit pas être préjudiciable au titulaire des droits sur l'œuvre.

La seconde règle applicable à TDM est prévue à l'article 47-4 où certaines reproductions sont faites de façon accessoire et dans le but de faciliter le fonctionnement de la machine. Il est encore précisé que l'utilisation visée par cette exception ne devrait pas être préjudiciable aux titulaires des droits d'auteur des œuvres qui font objet de fouille de données.

Malgré la modification législative, la solution japonaise demeure assez généreuse pour le développement des systèmes IA. Outre l'indifférence de la solution japonaise à l'égard de l'objectif de la fouille de données (commercial ou scientifique), le critère distinctif suggéré par le Japon concerne le rôle réservé au degré et à l'importance de la perception de l'œuvre par l'être humain qui se trouve derrière la machine. Dès lors que la consultation de l'œuvre par le développeur pendant le processus de TDM demeure accessoire à l'objectif de l'opération, les reproductions ne sont pas considérées comme une violation du droit d'auteur. De plus, la solution japonaise se distingue des autres en omettant de préciser une condition relative à la licéité de l'obtention des données. Ainsi, il serait possible pour les développeurs d'utiliser les œuvres accessibles sur Internet pour constituer une base de données. Finalement, la solution japonaise prend en considération les intérêts raisonnables des ayants droit. La reproduction des œuvres à cette fin ne devrait pas causer un préjudice aux intérêts légitimes des titulaires des droits d'auteur.

Pour conclure sur ce point, on peut constater que pour créer une exception propre au TDM, le droit comparé nous offre trois modèles qui se distinguent par la sévérité des conditions pour bénéficier de l'exception. Le modèle japonais se présente comme le modèle le plus permissif alors que le modèle britannique qui exclut toute activité ayant un but commercial est le modèle le plus contraignant ; le droit européen se présente comme une solution mitoyenne en scindant le régime selon l'objectif poursuivi.

2.2.2 Pistes de réflexion pour le Canada

Déjà en retard par rapport aux autres juridictions qui se sont dotées de mécanismes propres assurant le développement des systèmes IA, le Canada doit faire un choix. Comme l'illustre ce survol du droit comparé, le Canada se trouve devant l'éternel dilemme entre la « *rule* » et le « *standard* »⁸³ : adopter une interprétation large de l'exception existante relative à l'utilisation équitable, créer une nouvelle exception propre à la fouille de données ou adopter une approche mitoyenne.

La première avenue ne nécessite pas une modification législative. L'article 29 LDA prévoit déjà la possibilité d'utiliser une œuvre à condition que les critères d'une utilisation équitable soient satisfaits. Comme nous en avons discuté ci-dessus, l'application de cet article nécessite la satisfaction de deux conditions. D'abord, l'œuvre doit être utilisée pour l'un des objectifs mentionnés à l'article 29. Ensuite, l'utilisation doit être jugée équitable.

L'obstacle majeur concerne l'existence même de cette première condition. L'énumération – soi-disant – exhaustive des objectifs pour lesquels une œuvre peut être reproduite sans autorisation a pour effet que toute autre utilisation serait exclue de cette exception. Cet obstacle est-il insurmontable ? Cela dépend de l'interprétation que l'on fait de cet article. Une interprétation littérale de l'article mène à la conclusion qu'il s'agit d'une énumération exhaustive des cas admissibles à l'utilisation équitable. La jurisprudence de la Cour suprême du Canada tend également vers cette lecture littérale de l'article 29 LDA. Dans l'affaire *CCH*, tout en suggérant une interprétation libérale des termes « recherche » et « étude privée », la Cour exige que pour vérifier l'applicabilité de l'article 29, le juge doive s'assurer « qu'il s'agit d'une utilisation aux fins d'étude privée ou de recherche »⁸⁴. On pourrait prétendre que le choix terminologique de la version anglaise de la LDA confirme cette interprétation restrictive. Dans cette version, l'exception de l'article 29 LDA est appelée « *fair dealing* ». La terminologie employée ne correspond pas à celle utilisée en droit américain, où l'exception est connue sous le nom de « *fair use* ». Comme l'affirme un auteur, la distinction entre les deux doctrines américaine et canadienne est une « *conventional wisdom* »⁸⁵.

83. Justin HUGHES, « Fair Use and Its Politics – at Home and Abroad », dans Ruth L. OKEDIJI (dir.), *Copyright Law in an Age of Limitations and Exceptions*, Cambridge, Cambridge University Press, 2017, p. 234, à la p. 237.

84. *CCH Canadienne Ltée c. Barreau du Haut-Canada*, préc., note 38, par. 50.

85. A. KATZ, préc., note 43, p. 95.

Dans un article consacré à l'étude des origines historiques de l'exception de l'utilisation équitable⁸⁶, le professeur Katz propose une nouvelle lecture de cet article. La thèse de l'auteur consiste à démontrer qu'en dépit de la phraséologie de l'article 29 – elle-même inspirée du *Statute of Anne* –, la bonne interprétation de cet article devrait mener à considérer l'énumération des cas comme une énumération non exhaustive. Il est d'avis que le caractère exhaustif de l'énumération n'a jamais été sérieusement affirmé par la jurisprudence britannique. Autrement dit, il soutient qu'il n'y a pas de distinction fondamentale entre la doctrine américaine de *fair use* et celle de *fair dealing* qui s'applique plutôt dans les pays du *Commonwealth*. Cette idée de rapprochement entre les deux doctrines est également soutenue par un autre auteur. Le professeur Michael Geist soutient qu'avec la pentalogie du droit d'auteur⁸⁷, la Cour suprême s'engage dans un chemin menant vers une approche similaire à celle de *fair use*. Selon lui :

While Canadian copyright law still involves the two-stage analysis, the first stage has become so easy to meet that Canada appears to be inching closer to fair use. Indeed, the breadth of the fair dealing purposes is now so wide – eight purposes covering most imaginable uses – that future Canadian fair dealing analyses are likely to involve only a perfunctory assessment of the first-stage purposes test together with a far more rigorous analysis (what the Court in *SOCAN v. Bell Canada* [Bell] described as “heavy-hitting”) in the second-stage, six factor assessment.⁸⁸

Il existe également un autre argument à l'appui de la thèse selon laquelle les cas mentionnés à l'article 29 LDA sont à titre illustratif. La Cour suprême reconnaît dans l'affaire *CCH* que les exceptions au droit d'auteur font partie des droits des utilisateurs.

86. *Ibid.*

87. *Alberta (Education) c. Canadian Copyright Licensing Agency (Access Copyright)*, 2012 CSC 37 ; *Entertainment Software Association c. Société canadienne des auteurs, compositeurs et éditeurs de musique*, 2012 CSC 34 ; *Rogers Communications Inc. c. Société canadienne des auteurs, compositeurs et éditeurs de musique*, 2012 CSC 35 ; *Société canadienne des auteurs, compositeurs et éditeurs de musique c. Bell Canada*, 2012 CSC 36 ; *Ré:Sonnet c. Fédération des associations de propriétaires de cinémas du Canada*, 2012 CSC 38.

88. Michael GEIST, « Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use », dans Michael GEIST (dir.), *The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law*, Ottawa, University of Ottawa Press, 2013, p. 93, à la p. 159, en ligne : <<https://muse.jhu.edu/book/22904>> (consulté le 12 avril 2021).

Dans cette perspective, ces exceptions doivent recevoir une interprétation large et libérale. Mentionnant l'adhésion des autres auteurs, le professeur Geist soutient que « a flexible provision in which the list of enumerated purposes would be illustrative rather than exhaustive would be more consistent with the Court's vision of fair dealing as a user's right »⁸⁹.

L'idée d'élargir la portée des exceptions est une avenue intéressante à portée de la main. Elle permet de trouver une solution sans requérir de modification législative. Elle aura pourtant l'inconvénient majeur de la solution américaine, à savoir l'incertitude juridique. La légalité de l'opération dépendrait alors de l'appréciation du juge qui devait apprécier les six facteurs déterminants pour établir le caractère équitable de l'utilisation. Cette avenue saura en revanche jouer un rôle important dans la phase transitoire, soit avant l'adoption d'une exception propre à l'analyse de données.

La seconde avenue consiste à modifier la LDA afin d'y introduire une nouvelle exception applicable aux activités impliquant la fouille de données. Faute d'initiative législative à l'heure actuelle, nous nous intéresserons au rapport du Comité permanent de l'industrie, des sciences et de la technologie de la Chambre des communes. S'intéressant, entre autres, aux enjeux liés à l'IA, le comité consacre quelques paragraphes à l'analyse informationnelle. S'inspirant de la définition suggérée par l'un des intervenants, *Element AI*, le comité définit l'analyse informationnelle « comme “la dérivation d'informations à partir de données”, notamment par l'analyse de textes et de données, “et non l'utilisation et la commercialisation réelles de ces données” »⁹⁰. Le terme « analyse informationnelle » semble être une expression assez large qui englobe une variété d'opérations réalisées sur les données, y compris le TDM et le ML. Faisant la synthèse des interventions, le comité se rallie à ceux qui militent en faveur de la modification de la LDA afin d'encadrer l'analyse informationnelle. Le comité rapporte que deux façons de faire sont proposées : ajouter une disposition indépendante, à l'instar de l'article 29A de la *Copyright Law* britannique, ou ajouter un sixième cas à l'article 29 LDA afin de reconnaître l'analyse de données comme un des cas possibles de l'utilisation équitable. Cela dit, au moment de formuler sa recommandation, le comité demeure assez prudent et ne favorise ni l'une ni l'autre des deux solutions. La 23^e recommandation se lit comme suit :

89. *Ibid.*, p. 158.

90. *Ibid.*

Que le gouvernement du Canada envisage de modifier la *Loi sur le droit d'auteur* afin de faciliter l'utilisation d'une œuvre ou d'un autre objet protégé à des fins d'analyse informationnelle.⁹¹

Cette recommandation démontre une intention générale de favoriser le développement des systèmes IA sans pour autant préciser comment trouver le bon équilibre entre les droits des titulaires des œuvres et la promotion du savoir.

Parmi les deux avenues évoquées, la modification de l'article 29 LDA afin de préciser l'analyse informationnelle comme un des cas visés par l'utilisation équitable nous semble bien critiquable. D'abord, cette solution renforcerait l'interprétation restrictive de cet article. Ainsi, l'implantation de cette solution transmettrait le message suivant : pour ajouter un cas à l'utilisation équitable visée à l'article 29 LDA, la seule possibilité est de le préciser expressément. Ensuite, cette solution est critiquable en ce qu'elle est incapable de dissiper l'incertitude inhérente à l'exception de l'article 29 LDA. Le développement des systèmes IA serait à la merci de l'appréciation par les tribunaux des six facteurs identifiés par la Cour suprême afin de déterminer si l'activité peut être qualifiée d'équitable ou non.

Quant à la deuxième avenue, le comité cite l'exception spécifique du droit britannique. Or, l'étude du droit comparé nous montre qu'il n'y a pas de solution idéale. Chacun des modèles comporte ses propres limites et forces. C'est en s'inspirant de l'ensemble de ces modèles qu'une solution optimale peut se dégager.

Établissant deux régimes distincts pour l'analyse de données selon les fins poursuivies par le développeur, la structure du régime européen apparaît bien équilibrée. L'élément distinctif le plus important entre les deux régimes est la possibilité accordée aux titulaires de droit de se soustraire au régime applicable à la fouille de données à des fins non scientifiques. Le régime européen a pourtant un inconvénient majeur. Pour trouver le point d'équilibre entre les intérêts divergents des parties prenantes, il ne suffit pas de s'intéresser uniquement au caractère lucratif ou non lucratif de l'utilisation.

Certains ajustements peuvent permettre au modèle européen d'atteindre un meilleur équilibre entre les droits des titulaires et ceux des développeurs. D'abord, concernant les conditions d'application, on peut également considérer l'impact de l'utilisation sur l'œuvre.

91. *Ibid.*

Ce critère est déjà présent en droit canadien, et ce, concernant les conditions d'application de l'article 29 LDA. Dans l'hypothèse où le développement d'un système IA viserait la création d'une œuvre similaire ou fortement inspirée des données utilisées⁹², la circulation des œuvres ainsi créées pourrait avoir un impact négatif sur la valeur des œuvres utilisées. Cette idée se retrouve également dans la solution japonaise en des termes plutôt généraux où le législateur exige que l'analyse des données ne cause pas un préjudice aux titulaires de droit. Ensuite, il conviendrait d'emprunter un autre élément du droit japonais, soit le rôle de l'être humain dans le processus. L'exception doit préciser que la collecte et le traitement des données ne visent pas l'utilisation et le plaisir personnel de l'humain derrière la machine. Ainsi, la consultation des œuvres par ce dernier serait accessoire et secondaire dans tout le processus. Ce deuxième ajout peut être complété par une mention expresse concernant l'interdiction de l'utilisation des données à des fins autres que l'analyse informationnelle ainsi qu'une interdiction formelle de transmettre les données aux tiers.

Cette deuxième avenue se trouve également être conforme aux exigences de l'article 9(2) de la Convention de Berne qui semble dire que les exceptions relatives à la reproduction d'une œuvre doivent être limitées à « certains cas spéciaux, pourvu qu'une telle reproduction ne porte pas atteinte à l'exploitation normale de l'œuvre ni ne cause un préjudice injustifié aux intérêts légitimes de l'auteur »⁹³. Ainsi, telle que nous la proposons, la deuxième avenue met en place une exception applicable à un cas spécial (TDM et ML) qui ne porte pas atteinte à l'exploitation normale de l'œuvre et qui prend en considération les intérêts légitimes de l'auteur en suggérant que l'exploitation des données ne devrait pas causer un préjudice à l'auteur.

CONCLUSION

Dans cet article, nous avons démontré que l'absence d'une réglementation précise applicable à l'analyse informationnelle impliquant la reproduction non autorisée des données protégées n'est pas nécessairement fatale pour les développeurs. L'exception relative à l'utilisation équitable, si elle est interprétée de façon large et libérale,

92. Paige LESKIN, « This AI is creating some surprisingly good bops based on music by Katy Perry and Kanye West – listen to some of the best », *Insider*, 4 mai 2020, en ligne : <<https://www.businessinsider.com/jukebox-ai-music-generator-realistic-songs-machine-learning-algorithm-deepfakes-2020-5>> (consulté le 10 mai 2021).

93. OMPI, *Convention de Berne pour la protection des œuvres littéraires et artistiques*, 9 septembre 1886, (1972) 828 R.T.N.U. 221, en ligne : <<https://wipo.int/fr/text/283699>>.

pourrait dans certaines circonstances – et notamment lorsque l’activité se déroule dans le cadre d’une recherche scientifique – protéger les développeurs. Cela dit, il faut se rappeler que le développement des systèmes IA gagne du terrain très rapidement et que ces systèmes s’introduisent de plus en plus dans tous les secteurs de l’économie. L’analyse informationnelle n’est plus cantonnée au sein des laboratoires de recherche ; l’utilisation commerciale de cette technique devient tranquillement la norme. Toute solution devrait également prévoir un régime applicable aux activités exercées à des fins commerciales. Une solution fondée sur l’application de l’article 29 LDA – à cause de la généralité de la règle – ne peut qu’être transitoire du fait de l’incertitude inhérente à son application. L’industrie a besoin d’un cadre clair et précis. Il incombe au législateur de prévoir un tel cadre. Dans cet article, nous avons démontré que le droit comparé peut grandement guider le législateur afin qu’il trouve le bon équilibre entre le droit d’auteur et le développement des systèmes IA.