Université de Montréal

LES TENSIONS ENTRE LES PRINCIPES JURIDIQUES APPLICABLES AUX SYSTÈMES D'INTELLIGENCE ARTIFICIELLE EN DROIT QUÉBÉCOIS

(EXPLICABILITÉ, EXACTITUDE, SÉCURITÉ ET ÉQUITÉ)

Par

Nicolas Aubin

Faculté de droit

Mémoire présenté en vue de l'obtention du grade de Maîtrise en droit des technologies de l'information (LL.M.)

Août 2022

Université de Montréal

Faculté de droit

Ce mémoire intitulé

LES TENSIONS ENTRE LES PRINCIPES JURIDIQUES APPLICABLES AUX SYSTÈMES D'INTELLIGENCE ARTIFICIELLE EN DROIT QUÉBÉCOIS

(EXPLICABILITÉ, EXACTITUDE, SÉCURITÉ ET ÉQUITÉ)

Présenté par

Nicolas Aubin

A été évalué(e) par un jury composé des personnes suivantes

Nicolas Vermeys
Président-rapporteur

Vincent Gautrais
Directeur de recherche

Pierre Trudel
Membre du jury

Résumé

Le 21 septembre 2021, l'Assemblée nationale du Québec a adopté le projet de loi 64 afin de moderniser son régime de protection des renseignements personnels. S'inspirant du Règlement Général sur la Protection des Données européen, ce projet de loi renforce substantiellement les obligations des entreprises privées et des organismes publics à l'égard des renseignements personnels des Québécois. Ce projet de loi assure également le respect de certains principes juridiques applicables aux systèmes d'intelligence artificielle. Or, dans le cadre de ce mémoire, nous démontrons que des tensions existent entre quatre de ces principes. Ces principes sont : le principe d'explicabilité, le principe d'exactitude, le principe de sécurité ainsi que le principe d'équité et de non-discrimination. En effet, il est souvent difficile et parfois impossible d'assurer un respect conjoint de ces quatre principes.

La présente étude se divise en trois chapitres. Le premier explore les quatre principes pour ensuite identifier les obligations légales québécoises qui permettent d'en assurer le respect. Le second expose les tensions entre ces principes. Le dernier propose une solution permettant aux entreprises et aux organismes publics québécois de réaliser les arbitrages nécessaires entre ces principes tout en respectant la Loi.

Mots-clés : projet de loi 64 / intelligence artificielle / protection des renseignements personnels / explicabilité / exactitude / sécurité / discrimination / équité / vie privée / évaluation des facteurs relatifs à la vie privée.

Abstract

On September 21, 2021, the Quebec legislative passed Bill 64 to modernize its privacy regime. Inspired by the European General Data Protection Regulation, this bill strengthens the obligations of private companies and public bodies with respect to personal data. This bill also provides obligations protecting normative principles applicable to artificial intelligence systems. In this paper, we show that four of these principles exist in a state of tension. These principles are : explicability, accuracy, security and fairness and non-discrimination. Indeed, it is often difficult and sometimes impossible to ensure that these principles are respected together.

This study is divided into three parts. The first part defines the four principles to then identifies how these principles are translated into Quebec law. The second part sets out the tensions between these principles. The last part provides a solution that would allow Quebec businesses and public bodies to make the necessary trade-offs between these principles in a matter that complies with their legal obligations.

Keywords: bill 64 / artificial intelligence / data protection / accuracy / security / discrimination / fairness / privacy / impact assessment / privacy impact assessment.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abréviations

AA : Apprentissage automatique

AAAI : Association for the Advancement of Artificial Intelligence

ACM : Association for Computing Machinery Conference on Fairness, Accountability and Transparency

ACM FAccT : Association for Computing Machinery

ACM SIGKDD : Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining

ACM Comput. Surv. : Association for Computing Machinery Computing Surveys

Artif Intell Rev : Artificial Intelligence Review

API : Interface de programmation

Attributs protégés : Renseignements se rattachant à un motif de discrimination reconnu tels que la couleur de peau, l'origine ethnique, le genre, les opinions politiques…

BBC : British Broadcasting Corporation

BESC M : Programme de bourses d'études supérieures du Canada au niveau de la maitrise

CAI : Commission d'accès à l'information du Québec

CCC : Cybersecurity and Cyberforensics Conference

CDPDJ *:*Commission des droits de la personne et des droits de la jeunesse

IAD : Institute of Artificial Intelligence

CFPB : Consumer Financial Protection Bureau

Charte québécoise : Charte des droits et libertés de la personne

Charte canadienne : *Charte canadienne des droits et libertés*

CNIL : Commission nationale de l'informatique et des libertés

CNN : Convolutional neural network

COMPAS : Correctional Offender Management Profiling for Alternative Sanctions

CV : curriculum vitae

CVPC : Commissariat à la protection de la vie privée du Canada

ECAI : European Conference on Artificial Intelligence

EFVP : Évaluations des facteurs relatifs à la vie privée

Enisa : European Union Agency for Network and Information Security

FAT* : Conference on Fairness, Accountability and Transparency

FICD : Fonderie de l'innovation dans le commerce au détail

IA : Intelligence artificielle

IBM : International Business Machines Corporation

ICO : Information Commissioner's Office

ICS : Information & Computer Security

IEEE : Institute of Electrical and Electronics Engineers

IJCAI : International Joint Conference on Artificial Intelligence AI and autonomy track

Int J Data Sci Anal : International Journal of Data Science And Analytics

J Am Med Inform Assoc : Journal of the American Medical Informatics Association

JMIR : Journal of Medical Internet Research

JMIR Med Inform : Journal of Medical Internet Research Medical Informatics

Loi sur l'accès : Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels

Loi sur le privé : Loi sur la protection des renseignements personnels dans le secteur privé

LSTM : Long short-term memory (artificial neural network)

Nat Commun : Nature Communications

NeurIPS : Conference on Neural Information Processing Systems

OBVIA : Observatoire international sur les impacts sociétaux de l'IA et du numérique

OPC : Office of the Privacy Commissioner of Canada

Philos Trans A Math Phys Eng Sci : Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences

PIA : Privacy impact assessment

PL64 *:* Le projet de loi 64 ou *Loi modernisant des dispositions législatives en matière de protection des renseignements personnel*

Proxys : Diminutif de « *proxys variables* »

RGPD : *Règlement Général sur la Protection des Données*

RNN *:* Recurrent neural networl

SCC : Service correctionnel du Canada

SIA : Système d'intelligence artificielle

SPVM : Service de Police de la Ville de Montréal

SSCI : Symposium Series on Computational Intelligence

WHI : Workshop on Human Interpretability in Machine Learning

XAI : Explainable Artificial Intelligence

À ma conjointe Sara Eve Levac, pour son soutien indéfectible.

Remerciements

Je tiens à remercier le Conseil de recherches en sciences humaines du Canada et son Programme de bourses d'études supérieures du Canada au niveau de la maitrise (BESC M) pour son soutien financier tout le long de l'écriture de ce mémoire. Je tiens également à remercier mon directeur de recherche le professeur Vincent Gautrais pour son immense soutien et les nombreuses opportunités qu'il m'a offertes. C'est grâce à nos travaux sur les Évaluations des facteurs relatifs à la vie privée que le sujet de ce mémoire est né. Sans son soutien, ce travail de recherche n'aurait jamais vu le jour.

Introduction

Émergence normative. Lors des deux dernières décennies, nous avons assisté à un accroissement de l'intérêt porté à l'encadrement du développement et des impacts des outils d'intelligence artificielle (ci-après « IA »). Cet intérêt succède une série de scandales qui ont brusqué la confiance du public à l'égard des organisations qui développent ou utilisent ces outils. Nous n'avons qu'à penser :

> Au scandale des « *Google Cars* » où en 2009 et 2010, Google a été sanctionné pour avoir collecté illégalement des données privées parmi lesquelles des renseignements sensibles relatifs à l'état de santé ou à l'orientation sexuelle d'usagers[1].
>
> Au scandale de Clearview AI une entreprise spécialisée en reconnaissance faciale utilisée par plusieurs autorités policières dont la Gendarmerie royale canadienne. Selon plusieurs institutions publiques, l'entreprise collectait et traitait des renseignements personnels au mépris des règles et des droits des personnes concernées par les renseignements[2] et
>
> Au scandale lié au *Correctional Offender Management Profiling for Alternative Sanctions* (ci-après « COMPAS »). Cet outil assiste les tribunaux américains dans leur évaluation des risques qu'un accusé récidive s'il est mis en libéré dans l'attente de son procès. L'outil a fait l'objet d'une controverse lorsqu'une enquête journalistique a

[1] Emilie MOUCHARD, *L'accountability ou le principe de responsabilité en matière de protection des renseignements personnels*, Thèse de doctorat, Université de Montréal et Université Paris-Sud XI, 2018, p. 271.

[2] COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (CNIL), « Facial recognition: the CNIL orders CLEARVIEW AI to stop reusing photographs available on the Internet » (16 décembre 2021), en ligne : <https://www.cnil.fr/en/facial-recognition-cnil-orders-clearview-ai-stop-reusing-photographs-available-internet> (consulté le 7 août 2022); INFORMATION COMMISSIONER'S OFFICE (ICO), « ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted » (31 mai 2022), en ligne : <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/> (consulté le 7 août 2022); COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, COMMISSARIAT À L'INFORMATION ET À LA PROTECTION DE LA VIE PRIVÉE DE LA COLOMBIE-BRITANNIQUE, et COMMISSARIAT À L'INFORMATION ET À LA PROTECTION DE LA VIE PRIVÉE DE L'ALBERTA, *Conclusions en vertu de la LPRPDE no 2021-001 : Enquête conjointe sur Clearview AI, Inc. par le Commissariat à la protection de la vie privée du Canada, la Commission d'accès à l'information du Québec, le Commissariat à l'information et à la protection de la vie privée de la Colombie-Britannique et le Commissariat à l'information et à la protection de la vie privée de l'Alberta*, 2021, en ligne : <https://priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2021/lprpde-2021-001/> (consulté le 4 juillet 2022); Commissariat à la protection de la vie privée du CANADA, « Communiqué : L'utilisation par la GRC de la technologie de reconnaissance faciale de Clearview AI contrevenait à la Loi sur la protection des renseignements personnels, selon une enquête » (10 juin 2021), en ligne : <https://www.priv.gc.ca/fr/nouvelles-du-commissariat/nouvelles-et-annonces/2021/nr-c_210610/> (consulté le 5 juin 2022).

soutenu, entre autres, qu'il identifiait plus souvent erronément les personnes noires comme étant à risques de récidives que les personnes blanches[3].

Explosion normative. En conséquence, dans la dernière décennie, des dizaines de lignes directrices portant sur le développement et l'utilisation éthique des systèmes d'intelligence artificielle (ci-après « SIA ») ont été publiées par plusieurs institutions civiles et gouvernementales en Amérique, en Europe et en Asie[4]. Parallèlement, la refonte du régime de protection des renseignements personnels entamée par l'Union européenne par la mise en œuvre du *Règlement Général sur la Protection des Données (*ci-après « RGPD ») a poussé plusieurs gouvernements, dont celui du Québec, à renforcer les règles encadrant la circulation des renseignements personnels[5].

Une limite importante aux lignes directrices. C'est donc dans ce contexte de méfiance et de développement normatif que les outils d'IA sont désormais utilisés et développés. Pourtant, malgré leur multiplication, très peu de lignes directrices s'attardent à explorer comment implémenter les différents principes qu'elles désirent imposer ou protéger. Ainsi, sur les 21 lignes directrices à vocation universelle les plus souvent référencées dans la littérature académique

---

[3] Jeff Larson MATTU Julia Angwin,Lauren Kirchner,Surya, « How We Analyzed the COMPAS Recidivism Algorithm », *ProPublica* (2016), en ligne : <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=J5k3utYNqmWvBjTaTBs4TylpiUAiFx2o> (consulté le 13 avril 2022); Julia Angwin MATTU Jeff Larson,Lauren Kirchner,Surya, « Machine Bias », *ProPublica*, en ligne : <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (consulté le 13 avril 2022).

[4] Par exemple, les *Lignes directrices pour une IA digne de confiance*, les *Beijing AI Principles* de 2019, le *AI4People - an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, l*es *Asilomar AI Principles* du *Future of Life Institute* de 2017, le rapport *AINow* de 2017, les *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, le *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Artificial Intelligence and Autonomous Systems* du *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* de 2016 et le *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers* de 2018. Voir Thilo HAGENDORFF, « The Ethics of AI Ethics: An Evaluation of Guidelines », (2020) 30-1 *Minds & Machines* 99-120, 102.

[5] *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels (anciennement Projet de Loi 64)*, 64 (2021) [*PL64*]; BARREAU DU QUÉBEC, *Mémoire du Barreau du Québec - Projet de loi no 64 — Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, CI–031M C.P. – PL 64 Protection des renseignements personnels, Québec, Barreau du Québec, 2020, p. 18-19; Jean-François TRUDEL, Anastasia BERWALD, Marie CARPENTIER et Mathieu FORCIER, *Mémoire à la Commission des institutions de l'Assemblée nationale - Projet de loi n° 64, Loi modernisant les dispositions législatives en matière de protection des renseignements personnels*, Québec, Commission des droits de la personne et des droits de la jeunesse, 2020, p. 64-69; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Projet de loi no 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels - Mémoire de la Commission d'accès à l'information*, Québec, 2020, p. 3, 9, 12, 20 et 36.

portant sur l'IA[6] seules deux d'entre elles s'intéressent à l'implémentation technique des principes qu'elles formulent[7].

Impacts de cette limitation. Cette incapacité diminue fortement l'utilité de ces instruments, ce qui pourrait expliquer en partie pourquoi, malgré leur prolifération, l'impact véritable de ces lignes directrices sur les pratiques des développeurs et utilisateurs de SIA reste controversé[8]. Surtout, en ne se penchant pas sur les considérations pratiques, plusieurs de ces instruments normatifs contournent une problématique majeure à laquelle les développeurs et utilisateurs de SIA devront faire face. En effet, les mesures devant être mises en œuvre afin d'assurer le respect de certains principes réclament très souvent de sacrifier le respect d'un autre principe. C'est cette réalité que nous nous efforcerons de démontrer dans le présent mémoire.

Des principes en tension. Ainsi, nous observons que les principes de sécurité, d'explicabilité, d'exactitude et d'équité et de non-discrimination sont en tension puisque les mesures devant être mises en œuvre afin de respecter un de ces principes imposent souvent de sacrifier le respect d'un autre principe.

Justification du choix des quatre principes. À noter que ces principes ne sont pas les seuls principes applicables à l'IA. Néanmoins, nous avons choisi ces quatre principes pour deux principales raisons. D'une part, il s'agit de principes universels bénéficiant d'une excellente reconnaissance au Québec et à l'international[9]. D'autre part, il s'agit, selon nous, des principes

---

[6] Ces informations proviennent de T. Hagendorff. Ce dernier n'a pas considéré les « *documents that only refer to a national context – such as position papers of national interest groups (Smart Dubai 2018) or the report of the British House of Lords (Bakewell et al. 2018)* » ; T. HAGENDORFF, préc., note 4, 3.

[7] *Id.*, 2.

[8] En effet, une étude empirique publiée en 2018 n'a trouvé aucune preuve appuyant l'hypothèse selon laquelle les lignes directrices influencent les comportements et décisions des étudiants et des professionnels en informatique. Andrew MCNAMARA, Justin SMITH et Emerson MURPHY-HILL, *Does ACM's code of ethics change ethical decision making in software development ?*, *Conference: the 2018 26th ACM Joint Meeting*, New York, États-Unis, 26 octobre 2018, p. 729-733, p. 732, DOI : 10.1145/3236024.3264833.

[9] À noter que de tous ces principes, le principe d'exactitude est le moins bien reconnu dans les lignes directrices internationales. Il est, en revanche, un principe reconnu par le droit québécois. Jianlong ZHOU, Fang CHEN, Adam BERRY, Mike REED, Shujia ZHANG et Siobhan SAVAGE, *A Survey on Ethical Principles of AI and Implementations*, *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, Australia, 2020, p. 8, p. 3012-3013; T. HAGENDORFF, préc., note 4, 2; Vincent GAUTRAIS et Nicolas AUBIN, *Modèle d'évaluation des facteurs relatifs à la circulation des données*, Montréal, Chaire L.R. Wilson sur le droit des technologies de l'information et du commerce électronique de l'Université de Montréal et l'Observatoire international sur les impacts sociétaux de l'IA et du numérique, 2022, en ligne : <https://www.gautrais.com/conferences/modele-devaluation-des-facteurs-relatifs-a-

qui sont le plus souvent en tension. Cette situation s'explique en partie par leur caractère concret. En effet, le respect de ces principes réclame l'implémentation de mesures spécifiques. Or, c'est principalement par l'implémentation de ces mesures que les tensions émergent. À ce titre, nos quatre principes se distinguent donc de principes plus abstraits tels que la solidarité et la cohésion sociale eux-aussi reconnus par plusieurs lignes directrices internationales[10].

Développements législatifs récents au Québec. Nous explorons les tensions entre ces principes dans le contexte du régime québécois de protection des renseignements personnels. Ce régime a été le sujet de modifications majeures à la suite de l'adoption en septembre 2021 de la *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels* anciennement connue sous l'appellation « projet de loi 64 » (ci-après « PL64 »)[11]. L'un des objectifs de PL64 était de moderniser la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (ci-après « *Loi sur l'accès* ») et la *Loi sur la protection des renseignements personnels dans le secteur privé* (ci-après « *Loi sur le privé* »). Ces lois règlementent respectivement la protection des renseignements personnels auprès des organismes publics et des personnes qui exploitent une entreprise[12].

Paysage institutionnel québécois. L'implémentation de ces deux lois est assurée par la Commission d'accès à l'information du Québec (ci-après « CAI ») qui est dotée de pouvoirs d'enquête et de sanctions à l'égard des organisations publiques et privées[13]. En revanche, la CAI n'est pas l'unique institution susceptible d'assurer le respect des quatre principes évalués. En effet, plusieurs principes tels que la sécurité et l'équité sont intrinsèquement liés au respect des droits et libertés garanties par la *Charte des droits et libertés de la personne* (ci-après « *Charte québécoise* »). Or, celle-ci prévoit que la Commission des droits de la personne et des droits de

---

la-circulation-des-donnees/> (consulté le 10 juillet 2022); Pierre-Luc DÉZIEL, Karim BENYEKHLEF et Eve GAUMOND, *Repenser la protection des renseignements personnels à la lumière des défis soulevés par l'IA*, Montréal, Observatoire international sur les impacts sociétaux de l'IA et du numérique et le Laboratoire de cyberjustice, 2020, p. 21-28.

[10] T. HAGENDORFF, préc., note 4, 102.

[11] *PL64*, préc., note 5, p. 1.

[12] *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels*, chapitre A-2.1, art. 1 [*Loi sur l'accès*]; *Loi sur la protection des renseignements personnels dans le secteur privé*, chapitre P-39.1, art. 1 [*Loi sur le privé*].

[13] *Loi sur le privé*, préc., note 12, art. 81; *Loi sur l'accès*, préc., note 12, p. 123, 127, 158-164; *PL64*, préc., note 5, art. 159. Nouvel article 90.1. de la *Loi sur le privé.*

la jeunesse (ci-après « CDPDJ ») a comme mandat d'assurer le respect des droits qu'elle énonce[14]. À ce titre, un intérêt particulier sera dédié à la documentation fournie par la CAI et la CDPDJ .

Importance du droit mou. Notre mémoire ne se limite toutefois pas à l'analyse de la Loi. En fait, nous considérons que dans un domaine aussi technique et émergent que celui des IA, toute analyse qui se limite qu'au droit positif restera nécessairement incomplète[15]. En effet, chartes de principes, codes de conduites et lignes directrices font désormais partie du paysage normatif et peuvent être porteurs d'effets juridiques[16]. En fait, « dans les domaines caractérisés par un certain degré de complexité et voués à des mutations fréquentes »[17], comme l'IA, le droit est influencé par une forme de « droit mou » (en anglais « *soft law* ») qui « se manifeste [de plus en plus] par des textes énonçant des principes généraux »[18]. Ces textes vont parfois même prendre « part au processus de densification normative en acquérant une certaine force contraignante qui, bien que graduée, intègre la catégorie du droit dur. »[19] À ce titre, nous proposons qu'une évaluation par principes issus de lignes directrices permet d'offrir une vision plus complète du paysage normatif appliqué aux SIA au Québec.

Structure du mémoire. Le mémoire se divise en trois chapitres : (1) l'exploration des quatre principes susmentionnés, (2) l'analyse des tensions entre ces principes et (3) la présentation d'une piste de solution permettant aux organisations de mieux gérer les tensions identifiées.

Premier chapitre. D'abord, nous explorons les quatre principes susmentionnés. Cette exploration permet d'expliciter ce que sont ces principes et d'identifier les différentes mesures qui assurent leur respect. De plus, nous y identifions les obligations imposées par le régime québécois de protection des renseignements personnels qui se rattachent à chacun de ces principes.

Second chapitre. Ensuite, nous exposons les zones de tensions entre les principes. Nous constatons une tension entre les principes d'explicabilité et d'exactitude, entre l'exactitude et la

[14] *Charte des droits et libertés de la personne*, RLRQ c C-12, art. 57, en ligne : <https://canlii.ca/t/6c3nj>.
[15] Pierre TRUDEL, « Quel droit et quelle régulation dans le cyberespace ? », (2000) 32-2 *Sociologie et sociétés* 21, 201.
[16] E. MOUCHARD, préc., note 1, p. 50-51.
[17] P. TRUDEL, préc., note 15, 201.
[18] *Id.*
[19] E. MOUCHARD, préc., note 1, p. 371.

sécurité et entre l'exactitude et l'équité. Nous observons également des tensions opposant les principes de sécurité et d'équité ainsi que des tensions entre différentes conceptions de l'équité.

Troisième chapitre. Finalement, nous proposons aux organisations d'utiliser les Évaluations des facteurs relatifs à la vie privée (ci-après « EFVP ») afin de mieux gérer les tensions identifiées. Dans ce chapitre, nous exposons ce qu'est une EFVP pour ensuite évaluer ses différents avantages au regard des tensions. Enfin, nous formulons quelques recommandations qui permettent selon nous de renforcer l'utilité des EFVP dans ce contexte de tensions.

Limites à l'analyse. Avant de débuter notre analyse, il apparait opportun de définir ce que nous entendons par « IA» et d'expliciter le contexte législatif dans lequel s'insère notre analyse.

IA : une notion à préciser. Généralement, l'IA est définie par sa relation avec l'intelligence humaine. Ainsi, l'Office québécois de la langue française propose que les SIA sont des systèmes numériques qui tentent de « simuler le fonctionnement de l'intelligence humaine afin d'exécuter des fonctions relevant normalement de celle-ci »[20]. Similairement le Larousse définit l'IA comme « l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine »[21]. Les auteurs Nguyen et al. proposent également que l'IA réfère à « Toutes techniques visant à permettre aux ordinateurs d'imiter le comportement humain. (traduction libre) »[22]. Aucune de ces définitions ne nous est utile. En effet, elles renvoient toutes à un concept très mal défini soit la notion « d'intelligence humaine ». Si nous interprétons trop strictement cette notion alors l'IA ne relève que de la science-fiction puisqu'aucune machine créée, pour le moment, n'est sentiente. Si, au contraire, nous interprétons la notion trop librement alors presque tous les outils informatiques peuvent être qualifiés d'IA. En effet, il faut comprendre que :

[20] OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - intelligence artificielle » (2018), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26552508> (consulté le 30 janvier 2022).

[21] « Intelligence artificielle », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257> (consulté le 21 juillet 2022).

[22] Giang NGUYEN, Stefan DLUGOLINSKY, Martin BOBÁK, Viet TRAN, Álvaro LÓPEZ GARCÍA, Ignacio HEREDIA, Peter MALÍK et Ladislav HLUCHÝ, « Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey », (2019) 52-1 *Artif Intell Rev* 77-124, 78, DOI : 10.1007/s10462-018-09679-z.

> prior to the middle of the twentieth century the word « computer » was a job title referring to any person who was engaged in systematic calculation of values like those found in mathematical tables[23].

Ces « ordinateurs » ne détenaient aucune connaissance mathématique avancée. Ils ne faisaient que répéter mécaniquement des séquences d'additions et de soustractions[24]. Qu'à cela ne tienne, il aurait été loisible pour les contemporains des premiers ordinateurs modernes, à considérer que ceux-ci imitaient ou simulaient l'intelligence humaine. Après tous, ces derniers réalisaient des tâches qui ne pouvaient être réalisées que par des humains avant leur création. En revanche, qualifier les premiers ordinateurs modernes d'« IA » apparait absurde tant ceux-ci nous semblent à la fois primitifs et familiers. Bref, malgré la popularité de la définition de l'IA qui se fonde sur une comparaison à l'intelligence humaine, nous proposons que, dans les faits, cette définition ne reflète pas adéquatement comment l'expression est utilisée en pratique. De plus, elle est n'est pas très utile puisqu'elle ne permet pas de distinguer quels sont les systèmes numériques qui appartiennent à la famille des IA de ceux qui n'y appartiennent pas.

Ce que « IA » signifie pour nous. Plutôt que de tenter de définir l'IA, nous proposons de limiter la portée de la notion. Ainsi, lorsque nous traitons de SIA, dans le cadre de ce mémoire, nous référons aux systèmes : *« endowed with learning, reasoning and adaptation capabilities*. »[25] En effet, les tensions sous-mentionnées concernent principalement des systèmes qui sont dotés de capacités dites d'« apprentissage ». Parmi ceux-ci, nous retrouvons principalement les systèmes d'apprentissage automatique (en anglais « *machine learning* ») (ci-après « AA »). Selon Géron « le *machine learning* est la science (et l'art) de programmer les ordinateurs de sorte qu'ils puissent apprendre à partir de données »[26]. Similairement, Nguyen propose que l'AA est « *a subset of AI techniques that enables computer systems to learn from previous experience (i.e. data observations) and improve their behaviour for a given task.* »[27] Nous considérons donc

[23] John S. CONERY, *Explorations in Computing: An Introduction to Computer Science*, CRC Press, Boca Raton, FL, CRC Press, 2010, p. 2.
[24] *Id.*
[25] Alejandro BARREDO ARRIETA, Natalia DÍAZ-RODRÍGUEZ, Javier DEL SER, Adrien BENNETOT, Siham TABIK, Alberto BARBADO, Salvador GARCIA, Sergio GIL-LOPEZ, Daniel MOLINA, Richard BENJAMINS, Raja CHATILA et Francisco HERRERA, « Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI », (2020) 58 *Information Fusion* 82-115, 2, DOI : 10.1016/j.inffus.2019.12.012.
[26] Aurélien GÉRON, *Deep Learning Avec Keras et Tensorflow 2ed.*, Paris, Dunod, 2020, p. 6.
[27] G. NGUYEN et al., préc., note 22, 78.

comme des SIA les systèmes d'AA qui emploient différents types de modèles ou d'algorithmes tels que :

> les modèles de régression linéaire[28],
>
> les modèles de régression logistique[29],
>
> la méthode d'apprentissage ensembliste[30],
>
> les machines à vecteurs de support[31],
>
> la méthode des k plus proches voisins (ci-après « kNN »)[32],
>
> le modèle « *Naïve Bayes* »[33];
>
> les modèles utilisant des arbres de décision[34] et
>
> les modèles fondés sur des réseaux de neurones artificiels parmi lesquels les modèles d'apprentissage profond (en anglais « *deep learning* »)[35] ;

Descriptions terminologiques. Les données qui « alimentent » les SIA sont qualifiées de « données d'entrainement »[36]. Celles-ci sont regroupées dans un jeu de données (en anglais « *dataset* ») qualifié de « jeu d'entrainement » (en anglais « *training set* »)[37]. Chaque exemple utilisé par le SIA pour son entrainement est qualifié d'« instance »[38]. Chacune de ces instances est accompagnée de « variables » qui sont, en quelque sorte, des caractéristiques spécifiques qui se rapportent à chaque instance[39]. Par exemple, il est possible d'entrainer un modèle de SIA afin

---

[28] A. BARREDO ARRIETA et al., préc., note 25, 31; A. GÉRON, préc., note 26, p. 8-14, 38-47.
[29] *Id.*
[30] Hoa Khanh DAM, Truyen TRAN et Aditya GHOSE, « Explainable Software Analytics », *International Conference on Software Engineering'18 New Ideas and Emerging Results* 2018, 2, en ligne : <http://arxiv.org/abs/1802.00603> (consulté le 17 février 2022).
[31] *Id.*; G. NGUYEN et al., préc., note 22, 78.
[32] A. BARREDO ARRIETA et al., préc., note 25, 31.
[33] Jenna BURRELL, « How the machine 'thinks': Understanding opacity in machine learning algorithms », (2016) 3-1 *Big Data & Society* 2053951715622512, 5, DOI : 10.1177/2053951715622512.
[34] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2; A. BARREDO ARRIETA et al., préc., note 25, 31.
[35] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2; A. BARREDO ARRIETA et al., préc., note 25, 31; G. NGUYEN et al., préc., note 22, 78.
[36] A. GÉRON, préc., note 26, p. 6.
[37] Xue YING, « An Overview of Overfitting and its Solutions », (2018) 1168-022022 *Journal of Physics: Conference Series* 7, 1; A. GÉRON, préc., note 26, p. 6.
[38] A. GÉRON, préc., note 26, p. 6.
[39] *Id.*, p. 8. La Commission Nationale de l'Informatique et des Libertés (CNIL) identifie d'ailleurs les « variables » utilisées par un SIA de « caractéristiques ». COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS(CNIL), « Caractéristique », en ligne : <https://www.cnil.fr/fr/definition/caracteristique> (consulté le 17 août 2022).

de prédire le prix des maisons sur un jeu de données composé de plusieurs maisons, de leurs prix, de leur superficie et du revenu médian des habitants de leur quartier. Dans cet exemple, chacune des maisons constituerait une « instance » alors que le revenu médian et la superficie rattachés à chacune de ces maisons seraient des « variables »[40].

Les tâches des SIA : Classification. Les SIA sont appelés à réaliser plusieurs fonctions parmi lesquelles la classification et la prédiction. Classifier une donnée réclame le plus souvent au SIA de catégoriser les données évaluées par le système dans différents ensembles. Par exemple, les filtres anti-pourriels performent une fonction de classification. En effet, l'objectif de ces filtres est de différencier les courriels légitimes des pourriels frauduleux (en anglais « *spam* »).

Les tâches des SIA : Prédiction. Les SIA sont également souvent appelés à prédire une valeur ou un événement précis[41]. C'est ce que l'on qualifie de « régression »[42]. La valeur que le SIA tente de prédire est qualifiée de « valeur cible »[43]. Par exemple, la police américaine utilise des outils de « *predictive policing* » qui leur permettent, entre autres, de détecter la probabilité qu'un crime soit commis dans un espace précis[44].

Autres régimes législatifs : le régime fédéral. Nous n'explorons pas, en l'espèce, les régimes de protection des renseignements personnels canadien et européen. En effet, le régime canadien sera probablement modifié dans les prochains mois. Ainsi, le 16 juin 2022, le projet de loi C-27 a été présenté à la Chambre des communes du Canada. Ce projet de loi prévoit édicter, entre autres, une *Loi sur l'intelligence artificielle et les données* et une loi instituant un tribunal de la protection des renseignements personnels[45]. Or, ce projet de loi n'en est qu'à sa première lecture[46]. Même s'il sera adopté, ce qui n'est pas garanti, il sera fort probablement sujet à de multiples amendements. À titre illustratif, malgré n'avoir été opposé par aucun parti politique,

[40] A. GÉRON, préc., note 26, p. 6-8.
[41] *Id.*, p. 7.
[42] *Id.*
[43] *Id.*
[44] Andrew Guthrie FERGUSON, *The Rise of Big Data Policing*, New York, New York University Press, 2017, ch. 4.
[45] *Projet de Loi C-27 : Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*, Première session, quarante-quatrième législature, Première lecture le 16 juin 2022, p. 80 et 85.
[46] *Id.*, p. 1.

PL64 a été modifié par plus d'une centaine d'amendements qui ont, par moment, modifiés considérablement les obligations imposées aux organisations publiques et privées[47].

Autres régimes législatifs : le régime européen. Nous n'analysons pas non plus le régime européen dans le cadre de ce mémoire. Cependant, nous mobilisons parfois ce régime à des fins comparatives et interprétatives. En effet, l'influence du RGPD sur PL64 est indéniable. Ainsi, les mémoires déposés lors des travaux portant sur PL64 par le Barreau du Québec, la CAI et la CDPDJ dressent tous des parallèles entre le projet de loi et le RGPD[48]. Qui plus est, la documentation produite par les institutions européennes sur les SIA est particulièrement riche et permet de mieux explorer les tensions identifiées dans ce texte ainsi que les différentes mesures qui permettent d'implémenter les principes évalués.

[47] Par exemple, le législateur a diminué considérablement les circonstances dans lesquelles il est obligatoire de rédiger une Évaluation des facteurs relatifs à la vie privée (EFVP) et prévoit désormais qu'elles doivent être « proportionnées » à plusieurs facteurs. Les amendements ont également élargi les personnes pouvant exercer la fonction de responsable de protection des renseignements personnels, modifié les informations devant être communiquées à la personne visée par la collecte des renseignements personnels, modifié les options de paramétrage par défaut, clarifié la notion de renseignement sensible, limité les usages des renseignements anonymisés… COMMISSION DES INSTITUTIONS, *Amendements adoptés: Projet de loi no 64 : Loi modernisant les dispositions législatives en matière de protection des renseignements personnels*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci/mandats/Mandat-43711/index.html>; *PL64*, préc., note 5, arts. 13, 15, 28, 103,107, 108, 110 et 119. Nouveaux articles 59, 63.5 et 73 de la *Loi sur l'accès* et 3.1, 3.3, 8, 9.1, 23 de la *Loi sur le privé.*

[48] BARREAU DU QUÉBEC, préc., note 5, p. 18-19; J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 64-69; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 5, p. 3, 9, 12, 20 et 36.

Chapitre 1 - Le paysage normatif : quatre principes encadrant le développement des SIA

Structure du premier chapitre. Dans le premier chapitre de ce mémoire, nous explorons les principes (I) d'explicabilité, (II) d'exactitude, (III) de sécurité ainsi que (IV) d'équité et de non-discrimination. Dans chacune de ces sections, nous définissons chacun de ces principes pour ensuite exposer comment ceux-ci se traduisent en droit québécois.

Section I - Le principe d'explicabilité

Reconnaissance universelle du principe. Le principe d'explicabilité bénéficie d'une reconnaissance universelle. Plusieurs lignes directrices, compagnies, agences gouvernementales, éthiciens et juristes reconnaissent, en effet, que les SIA doivent être, en principe, « explicables »[49].

Définir le principe. Pourtant, si plusieurs s'entendent pour souligner l'importance du principe d'explicabilité, il n'existe actuellement aucun consensus sur ce qu'il signifie. De plus, peu s'entendent sur les mesures qu'un développeur ou utilisateur de SIA doit implémenter pour favoriser une plus grande explicabilité[50]. À sa plus simple expression, le principe réclame de s'assurer qu'un SIA ou que ses résultats restent compréhensibles aux êtres humains[51]. Or, ce qu'est une explication, un résultat « compréhensible », un SIA « explicable » ou « interprétable »,

[49] Selon T. HAGENDORFF, préc., note 4, 102. le principe est reconnu notamment par : les *Lignes directrices pour une IA digne de confiance*, les *Beijing AI Principles* de 2019, le *AI4People - an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,* les *Asilomar AI Principles* du *Future of Life Institute* de 2017, le rapport *AINow* de 2017, les *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, la *Déclaration de Montréal pour un développement responsable de Artificial l'intelligence artificielle*, le *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Intelligence and Autonomous Systems* du *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* de 2016 et le *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers* de 2018.
Voir aussi: Scott ROBBINS, « A Misdirected Principle with a Catch: Explicability for AI », (2019) 29-4 *Minds & Machines* 495-514, 1, DOI : 10.1007/s11023-019-09509-3.

[50] Brent MITTELSTADT, Chris RUSSELL et Sandra WACHTER, *Explaining Explanations in AI*, *CM FAT* Conference 2019*, Atlanta, 2019, p. 1, DOI : 10.1145/3287560.3287574.

[51] Christian HERZOG, « On the risk of confusing interpretability with explicability », *AI Ethics* 2021, 1, DOI : 10.1007/s43681-021-00121-9; B. MITTELSTADT, C. RUSSELL et S. WACHTER, préc., note 50, p. 2.

à qui s'adresse ces explications et à quelles fins sont toutes des questions qui restent vivement débattues dans la littérature. À ce titre :

> one of the most striking aspects of research into explainable AI (xAI) is how many different people, be they lawyers, regulators, machine learning specialists, philosophers, or futurologists, are all prepared to agree on the importance of explainable AI. However, very few stop to check what they are agreeing to, and to find out what explainable AI means to other people involved in the discussion[52].

Positionnement. Il n'appartient pas à ce texte de proposer une définition claire du principe[53]. Plutôt, nous nous contenterons d'explorer le principe afin de permettre au lecteur d'arriver à ses propres conclusions. Pour ce faire, nous identifions (A) les principales sources d'opacité des SIA, (B) les principaux objectifs poursuivis par le principe et (C) comment il se traduit en droit québécois.

A. Les sources d'opacité des SIA

Raisons d'être du principe. Afin de comprendre le principe d'explicabilité et ses contours, il convient de saisir les sources d'opacité des outils d'IA. En effet, le principe d'explicabilité est avant tout une réponse au manque de transparence entourant le développement, l'utilisation et le déploiement de ces outils. Cette opacité est principalement causée, selon l'auteure Jeanna Burrel, par trois phénomènes soit : (i) l'influence des intérêts mercantiles et institutionnels affectant le développement et l'utilisation des SIA, (ii) le manque de connaissances généralisé sur les technologies d'information et (iii) l'opacité inhérente au fonctionnement de certains SIA[54].

i. L'opacité intentionnelle

Maintenir une opacité afin de garder son avantage compétitif. L'opacité des SIA est en partie intentionnelle[55]. Les organisations qui développent et utilisent des SIA hésitent souvent à se

[52] B. MITTELSTADT, C. RUSSELL et S. WACHTER, préc., note 50, p. 1.

[53] À des fins de concisions, le présent texte traite de l' « interprétabilité » et de la « compréhensibilité » comme des composantes du principe d'*explicabilité*. À noter que dans la littérature scientifique, ces termes ("*interpretability*", *"comprehensibility*" et "*explicability*") sont parfois utilisés de façon interchangeable et parfois comme des termes distincts. Pour une proposition de distinctions entre le sens de ces termes consultez : A. BARREDO ARRIETA et al., préc., note 25, sect. 2.1.

[54] J. BURRELL, préc., note 33, 2.

[55] *Id.*

montrer transparentes quant à leur technologie, et ce, afin de maintenir leurs avantages compétitifs sur des concurrents potentiels[56]. Faire preuve de trop de transparence comporte le risque de voir ses compétiteurs recourir à des techniques de contre-ingénierie afin de répliquer leur technologie, une situation à la fois coûteuse et risquée une organisation[57]. Les organisations mobilisent donc plusieurs mécanismes légaux afin de maintenir cette opacité parmi lesquels le secret commercial[58], le droit criminel[59] et l'insertion dans les conditions d'acceptations de leurs produits de clauses prohibant ou pénalisant toutes tentatives de contre-ingénierie[60].

ii. L'opacité causée par un manque de connaissances généralisé

Des connaissances rares dans la population. L'opacité des SIA s'explique également par l'absence généralisée de connaissances informatiques dans la population[61]. En effet, la programmation informatique reste une compétence détenue par des spécialistes. Les personnes détenant des connaissances en IA sont encore plus rares en partie parce que pour véritablement comprendre le comportement de certains SIA, il faut disposer de certaines compétences dans plusieurs domaines, parmi lesquels l'informatique, les mathématiques et les statistiques[62].

Des connaissances rares pour les surveillants. Cette problématique est si sérieuse que ces connaissances s'avèrent rares même pour les organismes qui surveillent l'usage et le développement des SIA. Ainsi, dans un rapport rédigé en réponse à une consultation publique dans laquelle la CAI a invité plusieurs organisations à partager leurs préoccupations au sujet de l'IA, celle-ci a précisé que :

> Malgré la position très enviable du Québec dans le secteur de l'intelligence artificielle, les ressources qualifiées pour répondre aux enjeux de l'intelligence artificielle sont

[56] Nicholas DIAKOPOULOS, « Algorithmic Accountability Reporting: On the Investigation of Black Boxes », *Tow Center for Digital Journalism Publications* 2014, 12, DOI : 10.7916/D8ZK5TW2.
[57] Sandra WACHTER et Brent MITTELSTADT, « A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI », (2019) 2019-2 *Columbia Business Law Review* 494-620, 503-504, DOI : 10.7916/cblr.v2019i2.3424; N. DIAKOPOULOS, préc., note 56, 26-27.
[58] S. WACHTER et B. MITTELSTADT, préc., note 57, 504.
[59] N. DIAKOPOULOS, préc., note 56, 27.
[60] *Id.*, 26-27.
[61] J. BURRELL, préc., note 33, 4.
[62] Voir par exemple, A. I. PUBLISHING, *Deep Learning Crash Course for Beginners with Python: Theory and Applications of Artificial Neural Networks, CNN, RNN, LSTM and Autoencoders usingTensorFlow 2*, AI Publishing LLC, 2020; A. GÉRON, préc., note 26.

> rares et fortement disputées sur le marché du travail. Certains des répondants émettent des doutes quant à la disponibilité des ressources (humaines, financières et matérielles) nécessaires pour effectuer ce travail de vérification par les organismes de contrôle[63].

À cette crainte manifestée que la CAI ne détienne pas les ressources suffisantes pour répondre aux problèmes liés à l'intelligence artificielle, celle-ci s'est contentée de répondre qu'elle avait déjà recommandé au gouvernement d'envisager l'inclusion de « dispositions lui permettant de disposer d'un budget pérenne »[64] et qu'elle allait « poursuivre également ses démarches pour s'assurer que ses ressources disposent de l'expertise nécessaire »[65]. À noter que malgré les récents changements importants apportés au régime de protection des renseignements personnels au Québec et les demandes répétées de la CAI afin d'obtenir des ressources supplémentaires[66], le Gouvernement du Québec a choisi en 2022 de ne pas augmenter le budget de l'organisme[67].

Des connaissances rares pour les vulgarisateurs. Cette rareté des connaissances affecte également la profession journalistique[68]. La principale conséquence probable de cette situation est qu'elle alimente la mécompréhension des SIA puisque ce sont surtout les journalistes qui vulgarisent les enjeux et les controverses affectant ces outils. S'en suit une difficulté pour la population à comprendre les SIA et à appréhender leurs comportements, ce qui alimente, en retour, des espoirs irréalistes ou la méfiance à leur égard[69].

---

[63] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Intelligence artificielle et protection des renseignements personnels - Retour sur la consultation sur les principes en intelligence artificielle*, Québec, 2021, p. 12, en ligne : <https://www.cai.gouv.qc.ca/publications-et-documentation/documents-de-reflexion-et-danalyse/>.
[64] *Id.*, p. 12-13.
[65] *Id.*
[66] Voir e.g. *Id.*
[67] Charles LECAVALIER, « Données personnelles des Québécois: Le chien de garde réclame « des bras » et son indépendance », *La Presse*, sect. Politique (29 avril 2022), en ligne : <https://www.lapresse.ca/actualites/politique/2022-04-29/donnees-personnelles-des-quebecois/le-chien-de-garde-reclame-des-bras-et-son-independance.php> (consulté le 5 juin 2022).
[68] N. DIAKOPOULOS, préc., note 56, 26.
[69] J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3; Cynthia RUDIN et Joanna RADIN, « Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition », (2019) 1-2 *Harvard Data Science Review*, 5, DOI : 10.1162/99608f92.5a8a3a3d.

iii. L'opacité technique

Opacité inhérente à certains SIA. Surtout, il faut comprendre que l'opacité des outils d'AA existe également pour des raisons purement techniques. Même un auditeur ayant accès au code source de ces outils ne sera pas nécessairement capable d'expliquer leurs comportements et de prédire leurs résultats. En effet :

> the opacity of machine learning algorithms is challenging at a more fundamental level. When a computer learns and consequently builds its own representation of a classification decision, it does so without regard for human comprehension. Machine optimizations based on training data do not naturally accord with human semantic explanations.[70]

À ce titre, plusieurs SIA sont qualifiées de « boites noires »[71] afin de souligner la réalité que même les programmeurs et les développeurs responsables de leur création sont incapables de comprendre et d'expliquer comment ils parviennent à leurs conclusions[72].

Causes de la boite noire. Les causes de ce phénomène sont multiples. En revanche, afin de les exposer convenablement, nous devons expliquer comment un SIA d'AA approche une problématique. Afin de faciliter cette démonstration, nous reprenons l'exemple de filtre anti-pourriel utilisé par l'auteure Jeanna Burrel dans l'un des articles les plus influents sur ce sujet[73]. Ce filtre est un SIA d'AA qui différencie les courriers légitimes des pourriels frauduleux de type « *Nigerian 419* »[74]. *Nigerian 419* est une forme d'escroquerie dans laquelle le fraudeur réclame à sa victime de débourser un montant en lui promettant un gain monétaire futur[75].

---

[70] J. BURRELL, préc., note 33, 10.

[71] *Id.*; Liam G. MCCOY, Connor T. A. BRENNA, Stacy S. CHEN, Karina VOLD et Sunit DAS, « Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based », (2022) 142 *Journal of Clinical Epidemiology* 252-257, 252, DOI : 10.1016/j.jclinepi.2021.11.001; Amina ADADI et Mohammed BERRADA, « Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) », (2018) 6 *IEEE Access* 52138-52160, DOI : 10.1109/ACCESS.2018.2870052; C. RUDIN et J. RADIN, préc., note 69, 2.

[72] HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: Main capabilities and Disciplines*, Brussels, European Commission, 2019, p. 5; C. RUDIN et J. RADIN, préc., note 69, 2; J. BURRELL, préc., note 33, 2; L. G. MCCOY, C. T. A. BRENNA, S. S. CHEN, K. VOLD et S. DAS, préc., note 71, 252.

[73] J. BURRELL, préc., note 33.

[74] *Id.*, 7-8.

[75] Richard G. BRODY, Sara KERN et Kehinde OGUNADE, « An insider's look at the rise of Nigerian 419 scams », (2020) 29-1 *Journal of Financial Crime* 202-214, 203-204, DOI : 10.1108/JFC-12-2019-0162.

Une approche contre-intuitive. À partir de cet exemple, Burrell démontre que ce SIA d'AA approche la problématique de la reconnaissance de pourriels frauduleux d'une façon complètement différente qu'un humain. Ainsi, le SIA n'accorde aucune importance à l'ordre des mots d'un courriel ou à leur signification[76]. Chaque mot du courriel est traité comme une variable qui se voit assigner une valeur numérique qualifiée de « poids » ou de « coefficient de pondération »[77]. L'agrégat pondéré des poids assignés aux mots du courriel permet ensuite au SIA de reconnaitre le courriel comme étant frauduleux ou légitime[78]. Une telle approche à la classification des courriels apparait contre-intuitive pour un être humain, surtout si ce dernier n'est pas initié aux statistiques et aux mathématiques[79].

Déterminations contre-intuitives. En outre, le choix des mots identifiés par le SIA comme étant les plus « problématiques » apparait contre-intuitif. Ainsi, malgré être utilisé afin de détecter une fraude de type « *Nigeria 419* » le filtre considère que le mot « nigeria » est un terme essentiellement neutre[80]. De plus, des mots qui nous apparaissent intuitivement problématiques comme « *fraud* » ou « *trust worthy* (sic.) » ne sont pas considérés comme tel par le filtre[81]. En fait, le SIA de Burrel associe les pourriels frauduleux à des termes tels que « *our* », « *most* », « *will* » ou « *visit* ». Or, ces termes n'apparaissent pas suspects aux yeux d'un humain puisqu'ils peuvent très facilement se retrouver dans le contenu de courriels légitimes[82].

Complexité née de la quantité de variables évaluées. Or, il faut comprendre que la simple présence d'un ou de quelques termes jugés « problématiques » dans un courriel ne va pas nécessairement influencer la classification du SIA au point où le courriel sera reconnu comme un courriel frauduleux. C'est l'agrégation de tous les poids qui détermine la classification. À ce titre, simplement disposer d'une liste des variables qui détiennent les plus grands « poids » ne va pas nécessairement permettre de comprendre ou de prédire les résultats du SIA. En effet, leur

[76] J. BURRELL, préc., note 33, 9.
[77] A. GÉRON, préc., note 26, p. 8.
[78] J. BURRELL, préc., note 33, 8.
[79] Voir également C. RUDIN et J. RADIN, préc., note 69, 2.
[80] J. BURRELL, préc., note 33, 8.
[81] *Id.*
[82] *Id.*

influence relative ne sera probablement pas suffisante en soi pour prédire le résultat[83]. En revanche, connaitre toutes les variables utilisées par le SIA ne va pas nécessairement faciliter la compréhension de son choix puisque celles-ci peuvent être trop nombreuses pour être comprises par un humain. Dans l'exemple de Burrel, le SIA reconnaissait et assignait un poids spécifique à tous les mots les plus fréquemment utilisés d'un dictionnaire[84]. Toutefois, l'auteure note que des SIA peuvent utiliser des centaines voire des milliers de variables[85].

Connaitre les variables utilisées n'équivaut pas à comprendre leur pondération. Ensuite, même si une personne détient toutes les informations relatives aux variables utilisées et à leur poids, ces renseignements ne vont pas véritablement l'éclairer quant aux raisons sous-tendant le choix des valeurs assignées comme poids. En effet, ces valeurs ne sont pas toutes déterminées par des humains. Elles sont établies lors la phase d'apprentissage du modèle où celui-ci identifie, par l'analyse des données, les valeurs qui lui permettront de minimiser les erreurs dans la réalisation de la tâche qui lui est confiée[86]. Une fois ses poids identifiés le modèle sera ensuite testé lors d'une phase d'inférence auprès d'un « jeu de validation »[87] composé de nouvelles données dans l'espoir que « si le modèle fonctionne bien sur les données d'entrainement, il fonctionnera bien sur de nouvelles observations »[88]. À ce titre, nous pourrions argumenter qu'une compréhension parfaite des SIA nécessite l'accès aux jeux d'entrainement, car ce sont ceux-ci, qui, ultimement, déterminent l'importance accordée par l'outil d'IA aux différentes variables. En revanche, les SIA peuvent être entrainés à partir de milliards de données, une situation qui complexifie substantiellement la rédaction d'explications compréhensibles pour un humain[89].

Un exemple simple. Surtout, il convient de comprendre que le SIA présenté par Burrel est, de son propre aveu, très simple[90]. L'objectif du SIA était simple : il classifiait des courriels dans, uniquement, deux catégories possibles. Les variables utilisées étaient simples puisqu'elles

[83] *Id.*
[84] *Id.*
[85] Julia DRESSEL et Hany FARID, « The accuracy, fairness, and limits of predicting recidivism », (2018) 4-1 *Science Advances*, 3, DOI : 10.1126/sciadv.aao5580; J. BURRELL, préc., note 33, 5.
[86] A. GÉRON, préc., note 26, p. 8.
[87] *Id.*, p. 14.
[88] *Id.*, p. 8.
[89] J. BURRELL, préc., note 33, 5.
[90] *Id.*, 8.

disposaient de propriétés homogènes. En effet, le SIA de Burrel basait son analyse sur les termes contenus dans un courriel et n'évaluait pas d'autres types de données comme le « *header* »[91] du courriel. Enfin, le modèle de SIA décrit par l'auteure n'est pas le modèle le plus difficile à comprendre en AA. Plusieurs auteurs proposent, en effet, que les modèles les plus opaques sont les modèles d'apprentissage profond[92] qui sont « *often not just difficult but impossible to explain.* »[93].

B. Les objectifs du principe d'explicabilité

Objectifs du principe. Il convient maintenant de comprendre les principaux objectifs du principe d'explicabilité. Pour nos fins, il est essentiel de comprendre que l'explicabilité vise, entre autres, à garantir (i) l'assignation d'une responsabilité et (ii) une plus grande autonomie des personnes assujetties aux décisions des SIA.

i. La responsabilisation des acteurs

Assurer la justesse des décisions. Le principe d'explicabilité vise à responsabiliser les développeurs et les utilisateurs de SIA et à assurer qu'ils puissent être tenus responsables des préjudices qu'ils causent[94]. L'explication permet, en effet, à une personne visée par une décision de décider s'il convient ou non de contester cette dernière et sur quels motifs[95]. Elle permet également d'identifier si les développeurs ou utilisateurs ont commis une faute en cas de préjudice causé par le SIA. Respecter le principe d'explicabilité permet, en effet, de comprendre « les raisons pour lesquelles une décision d'IA était erronée »[96] et donc de pouvoir contester une

---

[91] *Id.*, 5.

[92] C. RUDIN et J. RADIN, préc., note 85, 4; INFORMATION COMMISSIONER'S OFFICE et THE ALAN TURING INSTITUTE, *Explaining decisions made with Artificial Intelligence*, Royaume-Uni, Information Commissioner's Office, 2020, annexe 2, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/annexe-2-algorithmic-techniques/> (consulté le 12 août 2022).

[93] Michael LANG, « Reviewing Algorithmic Decision Making in Administrative Law », (2021) 26-2 *Lex Electronica* 195, par. 5.

[94] Voir S. ROBBINS, préc., note 49, 501. référant à Filippo SANTONI DE SIO et Jeroen VAN DEN HOVEN, « Meaningful Human Control over Autonomous Systems: A Philosophical Account », (2018) 5 *Frontiers in Robotics and AI*, en ligne : <https://www.frontiersin.org/article/10.3389/frobt.2018.00015> (consulté le 13 avril 2022). ; S. ROBBINS, préc., note 49, 502.

[95] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, Bruxelles, Commission européenne, 2019, par. 53.

[96] *Id.*, par. 76.

décision inexacte[97] ou une décision qui se baserait sur des considérations impertinentes, illégales ou inacceptables[98]. Bref, l'explicabilité permet d'évaluer la légitimité et la fiabilité des décisions prises par le SIA et d'établir l'existence ou l'absence de faute commise par ces développeurs ou ces utilisateurs[99].

ii. Le renforcement de l'autonomie

Sauvegarder l'autonomie. L'explicabilité permet également de renforcer le principe d'autonomie. En effet, elle permet, d'une part, de protéger l'autonomie de la personne concernée par la décision du SIA et, d'autre part, d'assurer un contrôle humain sur ce dernier.

Protéger l'autonomie des personnes concernées. La capacité de contestation conférée par l'explication ne s'inscrit donc pas uniquement dans un objectif de responsabilisation. Il s'intègre également dans un objectif de renforcer l'autonomie des personnes concernées puisqu'elle leur permet d'agir à l'égard d'une décision qui les affecte. Ainsi, « *knowing why a particular decision was reached by an algorithm allows us to accept, disregard, challenge, or overrule that decision* »[100]. Dans le cas de décisions qui affectent grandement les droits des personnes concernées, une absence d'explication menace également « l'exercice d'importants droits fondamentaux procéduraux, tels que le droit à un recours effectif et à accéder à un tribunal impartial, ainsi que les droits de la défense et la présomption d'innocence »[101].

Protéger l'autonomie humaine. Cependant, l'explicabilité permet également de sauvegarder l'autonomie dans un sens plus large qui s'inscrit dans une relation entre les SIA et l'humanité. Ainsi, Robbins propose que le « *principle of explicability is primarily for the maintaining of meaningful human control over algorithms*. »[102] À ce titre, le principe permet d'assurer que les

[97] *Id.*, par. 53.
[98] S. ROBBINS, préc., note 49, 502.
[99] UNIVERSITÉ DE MONTRÉAL, « La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », *declarationiaresp*, en ligne : <https://www.declarationmontreal-iaresponsable.com/la-declaration-2> (consulté le 10 février 2022). sous Responsabilité.
[100] S. ROBBINS, préc., note 49, 496.
[101] *Proposition de règlement du Parlement européen établissant des règles harmonisées concernant l'Intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union*, (2021), p. 32, en ligne : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52021PC0206> (consulté le 10 février 2022).
[102] S. ROBBINS, préc., note 49, 501.

SIA restent subordonnés aux humains et que cette relation ne puisse, en aucun cas, s'inverser ou « [mettre] en péril l'autonomie humaine »[103].

C. La traduction légale du principe en droit québécois

Traduction du principe dans le régime québécois de protection des renseignements personnels. Le régime québécois de protection des renseignements personnels ne prévoit que quelques obligations pouvant se rattacher au principe d'explicabilité. Ainsi, des modifications prévues par PL64 assurent le respect du principe d'explicabilité en obligeant les entreprises et les organisations publiques à dévoiler certaines informations à l'égard d'une personne faisant l'objet d'une décision exclusivement automatisée[104].

Informations à fournir dans la *Loi sur le privé* et la *Loi sur l'accès.* Ainsi, les nouveaux articles 12.1. de la *Loi sur le privé* et 65.1. de la *Loi sur l'accès* prévoient plusieurs informations devant être offertes à la personne concernée par une décision fondée exclusivement sur un traitement automatisé. PL64 prévoit, en effet, qu'une personne visée par une décision fondée sur un tel traitement doit en être informée[105]. De plus, si elle en fait la demande, cette personne doit être informée :

> 1° des renseignements personnels utilisés pour rendre la décision;
>
> 2° des raisons, ainsi que des principaux facteurs et paramètres, ayant mené à la décision;
>
> 3° de son droit de faire rectifier les renseignements personnels utilisés pour rendre la décision.[106]

Signification de « raisons » et « principaux facteurs et paramètres ». En revanche, PL64 ne définit pas ce qu'il considère être une « raison », un « facteur » ou un « paramètre ».

---

[103] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 19-20. Voir également Luciano FLORIDI, Josh COWLS, Monica BELTRAMETTI, Raja CHATILA, Patrice CHAZERAND, Virginia DIGNUM, Christoph LUETGE, Robert MADELIN, Ugo PAGALLO, Francesca ROSSI, Burkhard SCHAFER, Peggy VALCKE et Effy VAYENA, « AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations », (2018) 28-4 *Minds & Machines* 689-707, 693-694 et 698, DOI : 10.1007/s11023-018-9482-5; UNIVERSITÉ DE MONTRÉAL, préc., note 99. sous Autonomie.

[104] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2. de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*

[105] *Id.*

[106] *Id.*

Malheureusement, les dispositions susmentionnées n'ont pas encore été mises en œuvre[107]. À ce titre, aucune jurisprudence n'a pour le moment été produite quant à la signification de ces notions dans le cadre de l'application de la *Loi sur l'accès* ou de la *Loi sur le privé.* Il est donc particulièrement difficile de définir ces notions surtout dans le contexte d'un SIA. Toutefois, le sens commun accordé à ces termes et le sens qui leur est octroyé dans la littérature sur l'AA permettent selon nous d'inférer certaines interprétations possibles quant à la signification de ces termes dans un contexte d'IA.

Divulguer plus que les renseignements personnels utilisés par le traitement. D'abord, on peut inférer que ces notions ne réfèrent pas qu'aux « renseignements personnels utilisés pour rendre la décision »[108]. En effet, les futurs articles 12.1 de la *Loi sur l'accès* et 65.2 de la *Loi sur le privé* distinguent très clairement les renseignements personnels utilisés par un processus décisionnel des raisons, facteurs et paramètres justifiant la décision.

Dévoiler les « raisons ». Le terme « raisons » renvoie, bien sûr à la justification de la décision[109]. En IA, cette notion peut néanmoins référer à une multiplicité d'éléments. Elle peut réclamer d'expliciter le fonctionnement du modèle, de justifier le choix des variables, de justifier le choix du modèle… Bref, dévoiler les « raisons » derrière la décision semble difficile à interpréter pour le moment tant les sens pouvant être appliqués au terme sont variés.

Signification possible de paramètres et facteurs. En revanche, nous proposons que les termes « paramètres » et « facteurs » sont plus faciles à interpréter. En effet, PL64 peut être interprété comme prévoyant une obligation d'expliciter les principales variables utilisées par un SIA ainsi que leur « poids » relatif sur le processus décisionnel. En effet, le poids assigné à une variable[110] est qualifié de « paramètre » dans la littérature en AA[111]. Si le terme « paramètre » est ainsi interprété, alors l'explication doit, bien sûr, être accompagnée d'une divulgation des variables

[107] *Id.*, art. 175.
[108] *Id.*, arts. 21 et 110. Nouveaux articles 65.2. de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*
[109] « Définitions : raison », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/dictionnaires/francais/raison/66270> (consulté le 17 août 2022).
[110] Ou à une neurone dans le cas de réseaux neuronaux. COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Paramètre (IA) », en ligne : <https://www.cnil.fr/fr/definition/parametre-ia> (consulté le 17 août 2022).
[111] A. GÉRON, préc., note 26, p. 8; COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (CNIL), « Paramètre (IA) », en ligne : <https://www.cnil.fr/fr/definition/parametre-ia> (consulté le 17 août 2022).

associées à chacun de ces poids. En effet, dévoiler les poids sans dévoiler les variables auxquels ils sont rattachés retirerait tout sens à l'explication. De plus, le sens courant du terme « facteur » renvoie à une cause ou un élément qui influence un résultat, un rôle identique à celui d'une variable dans un SIA[112]. À ce titre, nous proposons qu'il est possible de considérer les variables d'un SIA comme des « facteurs » et leurs poids comme des « paramètres ».

Maintien d'un flou normatif. Or, même si nous acceptons que ces termes doivent être interprétés ainsi, l'obligation reste floue. En effet, elle n'indique pas combien de variables devront être révélées. L'obligation se contente à prévoir que les « principaux » facteurs et paramètres doivent être explicités. Qui plus est, le « poids » d'une variable peut être présenté d'une multitude de façons. PL64 impose-t-il de dévoiler la valeur statistique exacte du « poids » ou, par exemple, de se contenter d'ordonner les variables selon leur ordre d'influence ?

Rattachement à l'autonomie individuelle et le principe de responsabilité. En revanche, les obligations d'information prévues par PL64 visent très clairement à sauvegarder les principes d'autonomie et de responsabilité susmentionnés. En effet, l'obligation de rappeler à la personne visée l'existence de son droit de rectifier ses renseignements personnels et l'obligation de lui expliciter les raisons ayant mené à la décision affichent clairement une volonté de s'assurer qu'une personne visée par une décision dispose de suffisamment d'informations pour réclamer sa révision[113]. À ce titre, ces nouvelles obligations traduisent une volonté de renforcer l'autonomie des personnes concernées par la décision et de responsabiliser l'organisme qui utilise le traitement automatisé.

Rattachement à l'autonomie humaine. Ces obligations se rattachent également à une volonté de maintenir l'autonomie humaine sur les processus décisionnels. Cela s'observe principalement

---

[112] Le Larousse défini facteur comme : un « Agent, élément qui concourt à un résultat ; cause ». Similairement, l'Office québécois de la langue française le définit comme une « Circonstance qui affecte les résultats d'une observation ou d'une expérience. » ou « une des conditions qui causent un événement ». « Définitions : facteur », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/dictionnaires/francais/facteur/32600> (consulté le 17 août 2022); OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - facteur (statistique) » (1975), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8477014> (consulté le 17 août 2022); OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - facteur (philosophie) » (1979), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8462443> (consulté le 17 août 2022).

[113] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*

par le fait que ces obligations ne concernent que les décisions qui sont fondées exclusivement sur un traitement automatisé[114].

Un principe peu traduit. L'explicabilité des SIA est, tel qu'identifié, très difficile à garantir et, surtout, à définir[115]. Au Québec, ces difficultés devront également composer avec la concision de la Loi. En effet, il est clair que le régime québécois de protection des renseignements personnels prévoit peu d'obligations rattachées à l'explicabilité des outils d'IA. De plus, le sens devant être accordé aux obligations prévues reste très difficile à déterminer.

Section II - Le principe d'exactitude

Absence de définition universelle. À l'instar du principe d'explicabilité, plusieurs lignes directrices et rapports soulignent l'importance de produire des SIA exacts ou précis[116]. Or, le « principe d'exactitude »[117] ne bénéficie pas non plus d'une définition universelle. Ainsi, afin de comprendre le principe d'exactitude, il convient, d'une part, (A) de définir ses deux principaux aspects et, d'autre part, (B) d'identifier comment il se traduit en droit québécois.

---

[114] *Id.*

[115] *Supra,* Chapitre 1, Section 1, Sous-section A.

[116] Par exemple, le principe est prévu par : les *Lignes directrices pour une IA digne de confiance,* la CDPDJ, l'Observatoire international sur les impacts sociétaux de l'IA et du numérique et le Laboratoire de cyberjustice, le *AI Now 2019 Report,* les *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,* les *Microsoft responsible AI principles* et les principes liées à l'intelligence artificielle de Google. Voir GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 21. ; J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 68. ; P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 35. ; Kate CRAWFORD, Roel DOBBE, Theodora DRYER, Genevieve FRIED, Ben GREEN, Elizabeth KAZIUNAS, Amba KAK, Varoon MATHUR, Erin MCELROY, Andrea Nill SÁNCHEZ, Deborah RAJI, Joy Lisi RANKIN, Rashida RICHARDSON, Jason SCHULTZ, Sarah Myers WEST et Meredith WHITTAKER, *AI Now 2019 Report*, New York, AI Now Institute, New York University, 2019, p. 39-40, 51. ; NICHOLAS DIAKOPOULOS, SORELLE FRIEDLER, MARCELO ARENAS, SOLON BAROCAS, MICHAEL HAY, BILL HOWE, H. V. JAGADISH, KRIS UNSWORTH, ARNAUD SAHUGUET, SURESH VENKATASUBRAMANIAN, CHRISTO WILSON, CONG YU, et BENDERT ZEVENBERGEN, *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, Fairness, Accountability, and Transparency in Machine Learning, en ligne : <https://www.fatml.org/resources/principles-for-accountable-algorithms> (consulté le 17 juillet 2022). ; MICROSOFT, *Microsoft responsible AI principles*, en ligne : <https://www.microsoft.com/en-us/ai/our-approach> (consulté le 17 juillet 2022). et GOOGLE, « AI at Google: our principles », *Google* (7 juin 2018), en ligne : <https://blog.google/technology/ai/ai-principles/> (consulté le 12 avril 2022). Voir aussi GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 21. et P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 35.

[117] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 21.

A. Les principaux aspects du principe d'exactitude

Ce qu'est le principe d'exactitude. Nous reconnaissons deux différents aspects au principe d'exactitude soit : l'exactitude des renseignements personnels utilisés et la précision du SIA. Dans le premier cas, le principe d'exactitude réclame (i) de s'assurer qu'une personne concernée par une décision d'un SIA bénéficie d'un traitement qui se base sur des renseignements exacts[118]. Dans le second, (ii) le principe demande que les SIA produisent des résultats exacts.

i. L'obligation d'utiliser des renseignements exacts, complets et à jour

Exactitude des renseignements utilisés. Ainsi le premier aspect prévoit qu'un SIA qui prend une décision à l'égard d'une personne doit fonder cette décision sur des renseignements exacts, complets, à jour et non équivoques[119].

Mesures pouvant être mises en œuvre. Afin de favoriser l'usage de renseignements véridiques, il est indiqué, par exemple, de permettre aux personnes concernées de rectifier ou de supprimer les renseignements personnels utilisés lorsque ceux-ci s'avèrent inexacts, incomplets, équivoques ou périmés[120].

ii. L'obligation de développer des SIA suffisamment précis

Ce qu'est la précision d'un SIA. Selon les *Lignes directrices en matière d'éthique pour une IA digne de confiance*, les SIA doivent également présenter un niveau de précision acceptable. Selon elles, la « précision » se définit en tant que : « capacité d'un système d'IA à poser un jugement correct […] ou de sa capacité à réaliser des prévisions, des recommandations ou des décisions correctes »[121]. À ce titre, si on développe un SIA pour prédire un évènement, alors l'évaluation de sa précision est liée à sa capacité de prédire adéquatement l'évènement en question. Si on développe un SIA pour distinguer, par exemple, des courriels légitimes de pourriels frauduleux, alors la précision du SIA se mesure à sa capacité à ne pas reconnaitre les pourriels frauduleux comme des courriels légitimes et les courriels légitimes comme des pourriels frauduleux.

---

[118] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 27.
[119] *Id.*
[120] *Id.*
[121] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 24.

Finalement, si on développe un SIA qui recommande des candidatures potentielles à un emploi alors la précision du SIA est liée à sa capacité de recommander les candidatures qui répondent le mieux aux critères recherchés pour l'emploi. Différentes métriques permettent de mesurer la précision d'un SIA parmi lesquelles : les taux de faux positifs, les taux faux négatifs, la valeur prédictive positive du SIA, la valeur prédictive négative du SIA...

Mesures permettant d'acquérir une plus grande précision. Plusieurs mesures peuvent être adoptées afin d'acquérir une plus grande précision des SIA. En l'espèce, il convient de comprendre deux mesures qui permettent de créer des SIA plus précis soit : d'entrainer les SIA sur des jeux d'entrainement étendus et profonds et de favoriser certains modèles de SIA.

Entrainer les SIA sur une grande quantité de données. D'une part, il est généralement admis que la création de SIA précis réclame d'entrainer ces derniers sur une grande quantité de données de qualité[122]. En effet, « *the accuracy of [machine learning] algorithms is known to improve with greater quantities of data to train on* »[123].

Distinctions terminologiques. Afin de bien comprendre le rôle des données dans un outil d'AA, il convient de distinguer les concepts de profondeur et d'étendue des jeux de données. La profondeur réfère au nombre d'instances peuplant un jeu de données alors que son étendue représente le nombre de variables attachées à chacune de ces différentes instances[124]. À ce titre, l'Autorité norvégienne de protection des données nous invite à conceptualiser le jeu d'entrainement comme une feuille de calcul (en anglais, « *spreadsheet* »)[125] :

> Like a spreadsheet, a dataset for machine learning may consist of rows and columns. If one has person-related data, the columns may well denote a person's age, gender, address, marital status, height, weight, nationality, etc. The rows will represent individual persons.[126]

[122] DATATILSYNET, *Artificial intelligence and privacy*, Oslo, Norvège, The Norwegian Data protection Authority, 2018, p. 11, en ligne : <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>.
[123] J. BURRELL, préc., note 33, 5.
[124] DATATILSYNET, préc., note 122, p. 11.
[125] *Id.*
[126] *Id.*

Nous pouvons illustrer, en partie, cet exemple par le Tableau 1 suivant :

Étendue

Profondeur

Instances	Âge	Genre	Statut matrimonial	Nationalité
John Doe	25	Homme	Marié	Canadien
Jane Doe	27	Femme	Célibataire	Britannique
Ronald George	32	Homme	Célibataire	Américain
Jacques Tremblay	45	Homme	Marié	Canadien

Tableau 1. – Feuille de calcul hypothétique

L'importance de la profondeur des jeux de données. L'usage de jeu d'entrainement profond permet de produire des SIA plus précis. En fait, « *the more training data we can feed into the model, the better the result* »[127] est un « *typical mantra frequently heard in connection with machine learning* »[128]. En effet, l'usage de jeux d'entrainement trop petits peut provoquer un « surajustement » des données d'entrainement (en anglais « *overfitting* »)[129]. On dit qu'un modèle a surajusté les données d'entrainements lorsqu'il performe substantiellement mieux à l'égard des données d'entrainement qu'à l'égard des nouvelles observations[130].

L'importance de l'étendue. Utiliser des jeux d'entrainement étendus permet également d'acquérir une meilleure précision pour de multiples raisons. D'abord, cela facilite la détection de variables utiles dans les jeux de données[131]. En effet, un SIA d'AA qui ne s'entraine pas auprès de jeux de données suffisamment étendus risque d'omettre certaines variables susceptibles

[127] *Id.*
[128] *Id.*
[129] A. GÉRON, préc., note 26, p. 8; X. YING, préc., note 37, 1.
[130] X. YING, préc., note 37, 1.
[131] Déziel, Benyekhlef et Gaumond soulignent qu'il est difficile pour un développeur de savoir, à l'avance, lesquelles des données seront pertinentes au traitement du SIA puisque cette connaissance réclame un traitement de ces mêmes données. À ce titre, collecter et entrainer des SIA sur de potentielles variables permet d'identifier la pertinence de chacune d'elles. Voir P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16; Pierre-Luc DÉZIEL, « Est-ce bien nécessaire ? Le principe de limitation de la collecte face aux défis de l'intelligence artificielle et des données massives », (2019) 465 *Développements récents en droit à la vie privée* 22, 22-24.

d'influencer l'analyse, ce qui peut biaiser ses résultats[132]. De plus, pour résoudre certaines problématiques, il ne sera pas toujours possible de prédire précisément une valeur cible en utilisant un nombre minimal de variables. Rappelons-nous, en effet, que dans l'exemple du filtre anti-pourriel de Burrel, celui-ci considérait tous les mots les plus souvent utilisés d'un dictionnaire[133]. À ce titre, la création de SIA précis peut dépendre d'une prise en considération de plusieurs centaines voire milliers de variables lors de son entrainement et lors de son déploiement[134].

Favoriser certains modèles et types de SIA. Finalement, certains SIA s'avèrent généralement plus précis que d'autres. À titre d'exemple, les auteurs Dam, Tran et Ghose proposent qu'en règle générale, les algorithmes d'apprentissage profond recourant à des réseaux de neurones sont plus précis que d'autres types de SIA comme les machines à vecteurs de support, les réseaux bayésiens ou les arbres de décisions[135]. Cette observation est partagée par Adadi et Berrada qui identifient que les algorithmes d'apprentissage profond qui utilisent des réseaux de neurones sont les « modèle[s] contemporain[s] d'apprentissage automatique ayant connu le plus de succès (traduction libre) »[136] et présentent les modèles de régressions linéaires ou logistiques comme étant généralement moins précis[137]. À noter que les algorithmes d'apprentissage profond réclament généralement de s'entrainer auprès d'un « *vastly greater volume of training data* »[138] que plusieurs autres types de SIA.

B. La traduction légale du principe en droit québécois

Traduction du principe en droit québécois. Le régime québécois de protection des renseignements personnels impose le respect du principe d'exactitude grâce à trois principaux mécanismes soit :

---

[132] Ninareh MEHRABI, Fred MORSTATTER, Nripsuta SAXENA, Kristina LERMAN et Aram GALSTYAN, « A Survey on Bias and Fairness in Machine Learning », (2021) 54-6 *ACM Comput. Surv.* 115:1-115:35, 115:5, DOI : 10.1145/3457607.
[133] J. BURRELL, préc., note 33, 8.
[134] J. DRESSEL et H. FARID, préc., note 85, 3; J. BURRELL, préc., note 33, 5.
[135] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2.
[136] A. ADADI et M. BERRADA, préc., note 71, 52145.
[137] *Id.*
[138] DATATILSYNET, préc., note 122, p. 11.

> une obligation de s'assurer que les renseignements personnels utilisés afin de prendre une décision à l'égard d'une personne concernée soient à jour et exacts,
>
> une obligation d'accorder à la personne concernée un droit de rectification à l'égard de ses renseignements personnels et
>
> une obligation d'accorder à la personne concernée l'occasion de présenter ses observations dans le cadre de toutes décisions exclusivement automatisées prises à son égard et d'en obtenir la révision.

i. L'obligation générale d'exactitude des renseignements

Obligation générale d'exactitude. Avant de prendre une décision à l'égard d'une personne, le régime québécois de protection des renseignements personnels impose à toute personne qui exploite une entreprise de :

> veiller à ce que les renseignements personnels qu'elle détient sur autrui soient à jour et exacts au moment où elle les utilise pour prendre une décision relative à la personne concernée (nos soulignements)[139].

De plus, elle impose aux organismes publics de :

> veiller à ce que les renseignements personnels qu'il conserve soient à jour, exacts et complets pour servir aux fins pour lesquelles ils sont recueillis ou utilisés (nos soulignements)[140].

Obligations à l'égard des agents de renseignements personnels. De plus, la *Loi sur le privé* prévoit des dispositions spécifiques à l'égard des « agents de renseignements personnels » soit des personnes qui exploite une entreprise au Québec qui par :

> elle-même ou par l'intermédiaire d'un représentant, fait le commerce de constituer des dossiers sur autrui, de préparer et de communiquer à des tiers des rapports de crédit au sujet du caractère, de la réputation ou de la solvabilité des personnes concernées par ces dossiers[141].

La *Loi sur le privé* prévoit que ceux-ci doivent « établir et appliquer des modalités d'opérations propres à garantir que les renseignements qu'ils communiquent sont à jour et exacts »[142].

[139] *Loi sur le privé*, préc., note 12. art 11.
[140] *Loi sur l'accès*, préc., note 12. art 72.
[141] *Loi sur le privé*, préc., note 12. art 70 al 2.
[142] *Id.* 71 al 2.

Les résultats de SIA sont parfois visés par l'exactitude. Au regard de ce qui précède, il apparait, à première vue, que le droit québécois n'impose de respecter que le premier aspect du principe d'exactitude soit de s'assurer que les renseignements utilisés par un SIA soient exacts. Or, la situation n'est pas si simple. En effet, le droit réclame de s'assurer de l'exactitude des résultats des SIA dans deux circonstances soit : lorsque ce résultat peut être défini comme un renseignement personnel et lorsque ce résultat constitue une décision « fondée exclusivement sur un traitement automatisé »[143].

L'exactitude des résultats de SIA qui sont des renseignements personnels. D'une part, il faut comprendre que l'origine d'un renseignement personnel n'affecte pas sa qualification. Un renseignement créé ou inféré par un SIA reste un renseignement personnel dès lors qu'il se rattache à une personne physique et permet de l'identifier[144]. À ce titre, si le résultat du SIA répond à la définition d'un renseignement personnel alors ce dernier est soumis aux mêmes obligations d'exactitude que celles qui sont prévues pour tous les autres renseignements personnels. Subséquemment, si un SIA infère un renseignement personnel se rattachant à une personne physique et que cette inférence s'avère fausse, rien n'empêche, à première vue, cette personne de demander la rectification de son renseignement personnel[145]. De ce fait, l'usage de SIA précis reste pertinent puisqu'ils permettent de produire des renseignements personnels qui seront moins susceptibles être contestés.

L'influence du droit de révision. D'autre part, PL64 prévoit l'existence de certaines obligations à l'égard des décisions fondées exclusivement sur un traitement automatisé[146]. Tel qu'il sera explicité sous peu les modifications appliquées à la *Loi sur l'accès* et la *Loi sur le privé* prévoient qu'en cas de décision fondée exclusivement sur un traitement automatisé l'organisme public ou l'entreprise devra donner « à la personne concernée l'occasion de présenter ses observations à un membre du personnel […] en mesure de réviser la décision. »[147] Clairement, ces nouvelles

[143] *PL64*, préc., note 5, arts. 21 et 110.
[144] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 63, p. 5.
[145] *Code civil du Québec*, chapitre CCQ-1991, art. 40, en ligne : <https://www.legisquebec.gouv.qc.ca/fr/document/lc/ccq-1991> (consulté le 10 juillet 2022); *PL64*, préc., note 5, art. 121. Nouvel article 28 de la *Loi sur le privé ; Loi sur l'accès*, préc., note 12, arts. 89, 94, 98, 128 et 141.
[146] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2. de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*
[147] *Id.*

dispositions permettent de contester des résultats des SIA. Il n'est pas clair cependant que ce nouveau droit de révision permettra de contester une décision en soulignant purement le manque de précision de l'outil utilisé. Les modifications liées à ces dispositions ne sont toujours pas entrées en vigueur[148]. Il n'existe donc pas de jurisprudence capable de nous éclairer sur cette question.

Ewert c. Canada **- Aucune distinction entre prédictibilité et exactitude des renseignements utilisés.** Il faut également comprendre que l'exactitude des renseignements produits par des outils d'évaluation est parfois imposée par d'autres régimes juridiques que le régime de protection des renseignements personnels. Par exemple, la *Loi sur le système correctionnel et la mise en liberté sous condition* prévoit des obligations rattachées à l'exactitude. Ainsi, cette loi prévoit à son article 24(1) que :

> Le Service est tenu de veiller, dans la mesure du possible, à ce que les renseignements qu'il utilise concernant les délinquants soient à jour, exacts et complets[149].

Or, dans l'arrêt *Ewert c. Canada*, la Cour suprême a conclu que cet article imposait au Service correctionnel du Canada (ci-après « SCC ») de recourir à des outils « de prédiction des risques de récidive »[150] suffisamment précis. Dans cet arrêt, la Cour a refusé de distinguer les notions de « validité prédictive » (en anglais « *predictive validity* »[151]) et d'« exactitude ». En effet, dans *Ewert* la Couronne a tenté d'argumenter que l'obligation d'exactitude contenue à l'article 24(1) ne pouvait pas s'appliquer puisqu'en matière de science actuarielle, il ne convenait pas de parler d'exactitude, mais bien de « validité prédictive »[152]. Cet argument a clairement été rejeté par la Cour qui a précisé que[153] :

> Même si nous acceptons que la science actuarielle établit une distinction entre les notions d'« exactitude » et de « validité prédictive », il n'est pas inopportun d'assujettir les résultats de tests actuariels à l'obligation prévue au par. 24(1) : dans ce

[148] *Id.*, arts. 21, 110 et 175. Nouveaux articles 65.2. de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*
[149] *Loi sur le système correctionnel et la mise en liberté sous condition*, L.C. 1992, ch. 20. art 24(1).
[150] Pierre-Luc DÉZIEL, « L'utilisation de renseignements personnels dans le contexte de la justice prédictive : le cas des outils actuariels d'évaluation des risques de récidive », (2018) 60-1 *Archives de philosophie du droit* 253-269, 261, DOI : 10.3917/apd.601.0268.
[151] *Ewert c. Canada*, [2018] 2 RCS 165, par. 43 (Cour suprême du Canada), en ligne : <https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/17133/index.do> (consulté le 7 juillet 2022).
[152] *Id.*
[153] *Id.*

> contexte, l'obligation de veiller, « dans la mesure du possible », à ce que les renseignements soient « exacts » signifierait que le SCC doit veiller à ce que les résultats de tests sur lesquels il s'appuie aient une capacité forte plutôt que faible à prédire les risques. (nos soulignements)[154]

Lien entre « validité prédictive » et précision. Or, la « validité prédictive » permet d'évaluer la capacité d'un outil à prédire convenablement une variable ou une valeur comme le risque de récidive[155]. En d'autres termes, il s'agit d'une notion qui permet d'évaluer la précision d'un outil.

Conclusion *d'Ewert*. En revanche, parce que le SCC savait depuis longtemps qu'il était possible « que ces outils soient empreints d'un préjugé culturel »[156], la Cour suprême a conclu que l'organisme avait violé ses obligations en vertu de la *Loi sur le système correctionnel et la mise en liberté.* En effet, selon la Cour, le SCC aurait dû « mener des recherches pour savoir si, et le cas échéant dans quelle mesure, ces outils sont susceptibles de donner lieu à de la variance interculturelle lorsqu'on les utilise à l'égard de délinquants autochtones »[157].

Les implications d'Ewert c. Canada sur le régime de protection des renseignements personnels québécois. *Ewert* indique très clairement que dans d'autres régimes que le régime de protection des renseignements personnels, un organisme peut être obligé à n'utiliser que des SIA précis. En revanche, il convient de se questionner quant aux implications d'*Ewert* sur le régime de protection des renseignements personnels québécois en tant que tel.

Des ressemblances identifiées par la doctrine. En effet, le lien entre cet arrêt et le régime de protection des renseignements personnels a déjà été établi par Déziel. En fait, selon ce dernier, la Cour suprême a, dans *Ewert,* mis en œuvre une interprétation du principe d'exactitude qui tire son origine des régimes canadiens de protection des renseignements personnels[158].

---

[154] *Id.*

[155] Voir, par exemple : Mark Q. THOMPSON, Olga THEOU, Graeme R. TUCKER, Robert J. ADAMS et Renuka VISVANATHAN, « FRAIL scale: Predictive validity and diagnostic test accuracy », (2020) 39-4 *Australasian Journal on Ageing* e529-e536, e533-e535, DOI : 10.1111/ajag.12829. ; « Predictive validity », dans *American Psychological Association Dictionary of Psychology*, American Psychological Association, en ligne : <https://dictionary.apa.org/predictive-validity> (consulté le 26 août 2022).

[156] *Ewert c. Canada*, préc., note 151, par. 50.

[157] *Id.*, par. 67.

[158] P.-L. DÉZIEL, préc., note 150, 266.

Similitudes entre les dispositions. De plus, les similitudes entre les formulations des obligations imposées par la *Loi sur l'accès* et la *Loi sur le système correctionnel et la mise en liberté sous condition* sont manifestes. Ainsi, la *Loi sur l'accès* prévoit une obligation de « veiller à ce que les renseignements personnels […] soient à jour, exacts et complets …(nos soulignements) »[159], alors que la *Loi sur le système correctionnel et la mise en liberté sous condition* demande, elle, de « veiller, dans la mesure du possible, à ce que les renseignements qu'il utilise concernant les délinquants soient à jour, exacts et complets[160] (nos soulignements) ». À ce titre, tant l'article 24(1) de la *Loi sur le système correctionnel et la mise en liberté sous condition* que l'article 72 de *Loi sur l'accès* demandent de « veiller » à ce que les renseignements utilisés soient « à jour, exacts et complets »[161]. La formulation prévue par la *Loi sur le privé* est assez similaire et demande d'utiliser des renseignements personnels « à jour et exacts »[162] avant de prendre une décision à l'égard d'un individu. À ce titre, il est possible de proposer, à la vue de l'interprétation de la Cour suprême dans *Ewert,* que les obligations d'exactitudes formulées par la *Loi sur l'accès* et la *Loi sur le privé* imposent aux organisations de recourir à des SIA précis dans la mesure où les renseignements qu'ils produisent sont des renseignements personnels.

Nuance. Il convient cependant de nuancer cette proposition. L'obligation d'exactitude évaluée dans *Ewert* n'est pas issue d'une loi liée à un régime de protection de renseignements personnels. À notre connaissance, la décision *Ewert* n'a jamais été utilisée pour interpréter les obligations liées à l'exactitude dans la *Loi sur l'accès* et la *Loi sur le privé*[163]. Ensuite, dans *Ewert*, les outils

[159] *Loi sur l'accès*, préc., note 12. art 72.
[160] *Loi sur le système correctionnel et la mise en liberté sous condition*, préc., note 149, art. 24(1).
[161] *Loi sur l'accès*, préc., note 12. art 72. ; *Loi sur le système correctionnel et la mise en liberté sous condition*, préc., note 149. art 24(1).
[162] *Loi sur le privé*, préc., note 12. art 11.
[163] Par exemple, rechercher le terme *Ewert* en conjonction avec chacune de ces lois ne procure aucun résultat pertinent ni dans les publications de la section enquête de la CAI ni dans les bases de données de *CanLii* et *SOQUIJ.* COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC - SECTION SURVEILLANCE, « Recherche avancée : Ewert », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=ewert&ref=&d1=&d2=&p=&col=162&or=> (consulté le 27 août 2022); CANLII, « Ewert; citing Loi sur la protection des renseignements personnels dans le secteur privé, RLRQ c P-39.1 », en ligne : <https://www.canlii.org/fr/#search/text=Ewert&origin1=/fr/qc/legis/lois/rlrq-c-p-39.1/derniere/rlrq-c-p-39.1.html&section1=11&linkedNoteup=> (consulté le 27 août 2022); CANLII, « Ewert; citing Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels, RLRQ c A-2.1 »; SOQUIJ, « Recherche Juridique : Ewert c. Canada (C.S. Can., 2018-06-13), 2018 CSC 30, SOQUIJ AZ-51502366, 2018EXP-1629, [2018] 2 R.C.S. 165 et Accès aux documents des organismes publics et sur la protection des renseignements personnels (Loi sur l'), (RLRQ, c. A-2.1) », en ligne : <https://soquij.qc.ca/portail/recherchejuridique/Rechercher/5283005> (consulté le 27 août 2022); SOQUIJ,

évalués avaient la prétention de produire des résultats objectifs. Cette prétention d'objectivité a influencé la Cour à conclure que les résultats qu'ils produisaient devaient être exacts :

> Lors de sa plaidoirie, la Couronne a fait valoir que les tests actuariels sont un outil important parce que les renseignements qui en découlent sont objectifs et qu'ils viennent donc atténuer les distorsions propres aux évaluations cliniques subjectives. Autrement dit, les outils contestés sont jugés utiles *parce que* les renseignements qu'ils produisent peuvent être scientifiquement validés. À mon avis, il faut en conclure à plus forte raison que le par. 24(1) impose au SCC l'obligation de veiller, dans la mesure du possible, à ce que les renseignements soient exacts.[164]

Or, pas tous les SIA auront une telle prétention d'objectivité. Enfin, l'article 24(1) de la *Loi sur le système correctionnel et la mise en liberté sous condition* mentionne que les renseignements utilisés doivent être à jour, exacts et complets « dans la mesure du possible »[165]. Cette mention n'est pas présente dans la *Loi sur l'accès* et la *Loi sur le privé.* Elle a pourtant influencé l'évaluation de la cour qui a identifié que : « Le fait que le législateur ait jugé cette limite nécessaire tend à indiquer que le par. 24(1) exige davantage qu'une simple bonne tenue de dossiers. »[166]

Les « degrés d'exactitude » réclamés dépendent des impacts des décisions prises par l'outil. De plus, même si nous acceptons qu'*Ewert* puisse influencer l'interprétation du principe d'exactitude tel que prévu par la *Loi sur le privé* et par la *Loi sur l'accès*, cela ne signifie pas que l'obligation de produire des résultats précis s'appliquerait en toutes circonstances. Dans *Ewert,* les enjeux étaient élevés. Les résultats de l'outil pouvaient porter atteintes aux droits et libertés fondamentales d'individus. Or, c'est « en insistant sur l'impact dramatique que peuvent avoir les décisions prises par le SCC sur la vie des contrevenants que la Cour conclut que le SCC doit s'assurer d'utiliser des renseignements qui affichent un haut degré d'exactitude »[167]. À ce titre,

---

« Recherche Juridique : Ewert c. Canada (C.S. Can., 2018-06-13), 2018 CSC 30, SOQUIJ AZ-51502366, 2018EXP-1629, [2018] 2 R.C.S. 165 et Protection des renseignements personnels dans le secteur privé (Loi sur la), (RLRQ, c. P-39.1) », en ligne : <https://soquij.qc.ca/portail/recherchejuridique/Rechercher/5283005>.

[164] *Ewert c. Canada*, préc., note 151, par. 41.

[165] *Loi sur le système correctionnel et la mise en liberté sous condition*, préc., note 149. art 24(1).

[166] *Ewert c. Canada*, préc., note 151, par. 42.

[167] P.-L. DÉZIEL, préc., note 150, 266.

plus un outil a un impact important et potentiellement préjudiciable sur les personnes concernées par ses décisions, plus son « degré d'exactitude » devra être élevé[168].

Les « degrés d'exactitude » réclamés dépendent des droits en cause. De façon similaire, la CDPDJ propose que si une technologie porte atteinte à un droit garanti par la *Charte québécoise,* tel que le droit à la vie privée, celle-ci ne pourrait se justifier que si elle « permet d'atteindre l'objectif pour lequel elle a été conçue sans commettre un nombre inacceptable d'erreurs »[169]. Or, si nous reprenons la définition de la précision susmentionnée proposée par les *Lignes directrices en matière d'éthique pour une IA digne de confiance*[170], la CDPDJ propose, au final, de n'utiliser que des outils précis en cas d'atteintes importantes aux droits et libertés fondamentales.

ii. Le droit de rectification

Ce qu'est le droit de rectification. Le droit québécois confère aux personnes concernées par les renseignements personnels un droit de rectification qui leur permet de supprimer[171] ou de rectifier les renseignements personnels inexacts, incomplets, équivoques ou périmés qu'une personne ou un organisme détient sur elle[172]. De plus, ce droit permet à la personne concernée d'apposer des commentaires à ses propres renseignements[173]. Aussi, le droit de rectification permet à une personne visée par une décision prise au terme d'un traitement exclusivement automatisé de faire « rectifier les renseignements personnels utilisés pour rendre la décision »[174].

---

[168] *Id.*; Cour fédérale du Canada, 20 décembre 2010, T-246-10, *Nammo c. TransUnion of Canada Inc.*, *RCF* 3.600, 39-40, en ligne : <https://canlii.ca/t/2f3n8> (consulté le 27 août 2022).

[169] Jean-François TRUDEL, Ramon AVILA, Geneviève ST-LAURENT et Ramon AVILA, *Mémoire à la Commission d'accès à l'information sur le document de consultation « Intelligence artificielle »*, Cat. 2.412.133, coll. Document adopté à la 681e séance de la Commission, tenue le 15 mai 2020, par sa résolution COM-681-4.2.1, Québec, Commission des droits de la personne et des droits de la jeunesse, 2020, p. 4.

[170] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 24. ; *Supra,* Chapitre 1, Section II, Sous-section A.

[171] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 27.

[172] *Code civil du Québec*, préc., note 145, art. 40; *PL64*, préc., note 5, art. 121. Nouvel article 28 de la *Loi sur le privé. Loi sur l'accès*, préc., note 12, arts. 89, 94, 98, 128 et 141.

[173] *Code civil du Québec*, préc., note 145. art 40.

[174] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*

[174] *Code civil du Québec*, préc., note 145. art 40. ; *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*

Rattachement de la rectification à la circulation des renseignements personnels. Demander la rectification d'un renseignement personnel ne se limite d'ailleurs pas qu'à l'organisation directement visée par la demande. Ainsi, le Code civil prévoit que pareille demande doit être, même si contestée par l'organisation visée, « notifiée, sans délai, à toute personne qui a reçu les renseignements dans les six mois précédents »[175].

Rattachement à l'exactitude. Le droit de rectification se rattache évidemment au principe d'exactitude puisqu'il permet de supprimer des renseignements inexacts ou de les modifier afin qu'ils deviennent exacts. De plus, la nécessité de notifier à toute personne qui a reçu les renseignements les changements apportés par la personne concernée intègre le droit de rectification à l'écosystème socio-économique de partage des renseignements personnels et permet aux participants de ce système de bénéficier de l'usage de ce droit afin d'acquérir des renseignements exacts.

iii. Le droit de révision

Le droit de révision vise les décisions fondées sur un traitement exclusivement automatisé. PL64 introduit un droit de révision pour toute décision fondée exclusivement sur un traitement automatisé. Celui-ci permet à la personne concernée par la décision de « présenter ses observations à un membre du personnel […] en mesure de réviser la décision »[176]. À noter que les traitements automatisés et les SIA ne sont pas des catégories identiques. Tout traitement automatisé n'est pas nécessairement un SIA et un SIA n'est pas nécessairement utilisé pour prendre une décision à l'égard d'une personne physique. En revanche, dans la mesure où un SIA est utilisé afin de produire des décisions, celles-ci risquent fort probablement d'être soumises au droit de révision prévu par PL64.

Pauvreté des sources. Tel qu'identifié, les dispositions prévoyant le droit de révision n'ont toujours pas été mises en œuvre[177]. À ce titre, la portée exacte de ce droit reste, pour le moment, assez floue puisqu'aucun tribunal n'a été appelé à s'y prononcer. Nous devons donc nous

[175] *Code civil du Québec*, préc., note 145. art 40 al 2.
[176] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1 de la *Loi sur le privé.*
[177] *Id.*, art. 175. ; *Supra*, Chapitre 1, Section I, Sous-section C et Chapitre 1, Section II, Sous-section B(i).

contenter d'explorer l'interprétation de ce nouveau droit au regard de communiqués gouvernementaux et des débats parlementaires.

Circonstances où le droit de révision s'applique. Une limite importante au droit de révision prévu par PL64 est qu'il ne s'applique qu'aux décisions prises au terme d'un traitement exclusivement automatisé. Selon le Gouvernement du Québec[178], le droit de révision ne vise donc pas les outils d'aide à la décision et les traitements lors desquels une personne physique a une réelle influence sur la décision[179].

Révision automatique par un humain ? De plus, il est clair à la lecture des débats parlementaires entourant PL64, que l'intention du législateur n'était pas d'accorder un droit de faire automatiquement réviser la décision par un être humain « sans motif, du simple fait que c'est un système automatisé »[180]. En effet, le mémoire soumis par le Barreau du Québec proposait, lors des consultations publiques ayant précédé PL64, de prévoir un droit à une révision automatique par un être humain[181]. Cependant, un tel amendement a été explicitement refusé par les députés de l'Assemblée nationale au motif que celui-ci élimine toute pertinence à recourir à un outil décisionnel exclusivement automatisé[182].

Qu'est-ce qui peut être révisé : les renseignements utilisés. Toutefois, PL64 ne précise pas clairement les motifs pouvant justifier une révision d'une décision fondée exclusivement sur un traitement automatisé. Il appert cependant que le droit de révision permet de réviser une décision si celle-ci se fonde sur l'usage de renseignements personnels inexacts. En effet, la personne concernée par une décision fondée exclusivement sur un traitement automatisé doit, sur demande, être informée « de son droit de faire rectifier les renseignements personnels

---

[178] GOUVERNEMENT DU QUÉBEC, « Décision fondée exclusivement sur un traitement automatisé », en ligne : <https://www.quebec.ca/gouvernement/travailler-gouvernement/services-employes-etat/conformite/protection-des-renseignements-personnels/technologie-et-droit-a-la-protection-des-renseignements-personnels/decision-traitement-automatise> (consulté le 20 janvier 2022).

[179] *Id.*

[180] ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels», (17 mars 2021) 45-126 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210317.html> (consulté le 6 juin 2022).

[181] *Id.* ; BARREAU DU QUÉBEC, préc., note 5, p. 15.

[182] ASSEMBLÉE NATIONALE DU QUÉBEC, préc., note 180.

utilisés pour rendre la décision »[183]. Cette mention nous permet d'inférer qu'une décision peut être révisée si elle s'est basée sur des renseignements personnels inexacts[184].

Qu'est-ce qui peut être révisé : le traitement en soi. En revanche, nous proposons qu'il pourrait également être possible de réclamer la révision d'une décision en soulevant l'inexactitude du traitement automatisé en soi. En effet, PL64 prévoit que la personne concernée par une décision fondée exclusivement sur un traitement automatisé doit aussi être informée, sur demande, des principaux facteurs et paramètres ayant fondé la décision[185]. Tel qu'identifié, la formulation des dispositions suggère que les informations rattachées à ces obligations ne se limitent pas qu'aux renseignements personnels utilisés par le SIA et renvoie à d'autres types de renseignements comme les « poids » conférés aux différentes variables[186]. Or, les dispositions qui prévoient ces explications sont les mêmes que celles qui prévoient le droit de révision[187]. Cette présence suggère que la révision permet de contester davantage que l'exactitude des renseignements personnels utilisés par le SIA. Il apparait absurde, en effet, de prévoir le dévoilement des raisons, facteurs et paramètres de la décision dans les mêmes dispositions qui confèrent le droit à sa révision si ceux-ci ne pouvaient être contestés.

Les limites du régime de protection des renseignements personnels en matière d'exactitude. À ce titre, il est clair que le régime québécois de protection des renseignements personnels demande de respecter le principe d'exactitude. Or, ce régime reste, somme toute, assez succinct et vague sur l'exactitude réclamée des SIA. Ainsi, si l'exactitude des renseignements personnels est clairement réclamée, il n'est pas si clair dans quelles circonstances les résultats d'un outil d'IA seront visés par une telle obligation. De plus, s'il est apparent que la personne concernée possède un droit à la révision d'une décision exclusivement automatisée, comment ce droit pourra être exercé en pratique reste assez nébuleux.

[183] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*
[184] *Loi sur le privé*, préc., note 12, art. 11; *Loi sur l'accès*, préc., note 12, art. 89.
[185] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*
[186] *Supra,* Chapitre 1, Section I, Sous-section C.
[187] *PL64*, préc., note 5, arts. 21 et 110. Nouveaux articles 65.2 de la *Loi sur l'accès* et 12.1. de la *Loi sur le privé.*

Section III - Le principe de sécurité

Universalité du principe. Comme les principes précédents, le principe de sécurité jouit d'une reconnaissance universelle. De nombreuses lignes directrices et rapports soulignent, en effet, l'importance de sécuriser les renseignements personnels utilisés par le SIA[188].

Objectif du principe. Le principe de sécurité poursuit un objectif de protection de la vie privée des personnes concernées par des renseignements personnels collectés ou utilisés dans le cadre du développement ou de l'usage de SIA. Un « renseignement personnel » est un « renseignement qui concerne une personne physique et permet de l'identifier. »[189] Cette protection des renseignements personnels permet, en retour, de protéger certains aspects du droit à la vie privée des personnes concernées parmi lesquels :

> la « *privacy of personal information* » qui permet aux personnes concernées par des renseignements personnels d'exercer un certain degré de contrôle sur l'usage de ses renseignements[190],
>
> la « *privacy of personal behaviour* » qui protège les individus contre des systèmes de surveillance béhavioraux. Cet aspect de la vie privée confère une protection contre la divulgation de certains types de renseignements concernant, par exemple, les opinions religieuses et politiques ainsi que les habitudes et préférences sexuelles[191] et
>
> la « *privacy of personal communications* » qui protège les individus contre l'interception et l'enregistrement de leurs communications[192].

---

[188] Selon T. HAGENDORFF, préc., note 4, 102., les principes de sécurité ou de protection de la vie privée sont reconnus dans, entre autres : Les *Lignes directrices pour une IA digne de confiance*, le *Report on the Future of Artificial Intelligence,* les *Beijing AI Principles* de 2019, les recommandations du Conseil sur l'intelligence artificielle de l'OCDE, les *Usages malicieux de l'intelligence artificielle, l*e *AI4People - an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,* les *AsilomarAIPrinciples* du *Future of Life Institute* de 2017, le rapport de *AINow* de 2016, le rapport de *AINow* de 2017, le rapport de *AINow* de 2018, le rapport de *AINow* de 2019, les *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, *Ethically Aligned Design : A vision for Prioritizing Human Well being with Autonomous and intelligent systems (Version for Public Discussion), Ethically Aligned Design : A vision for Prioritizing Human Well being with Autonomous and intelligent systems (First Edition), l*es *AI Policy Principles* du *Information Technology Industry Council* de 2017, les *Microsoft responsible AI principles;,* les principes liés à l'intelligence artificielle de Google, le *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers* de 2018, *Partnership on AI.*

[189] *Loi sur le privé*, préc., note 12, art. 2.

[190] INFORMATION COMMISSIONER'S OFFICE, *Privacy Impact Assessment Handbook*, Royaume-Uni, Privacy Impact Assessment Handbook, 2009, p. 15.

[191] *Id.*

[192] *Id.*

Le principe de sécurité en droit québécois. Le régime québécois de protection des renseignements personnels impose de mettre en place plusieurs mesures de sécurité. Ainsi, le régime impose (A) d'adopter une « approche minimaliste »[193] à l'égard des renseignements personnels (B) de prévoir des règles de gouvernance des renseignements personnels et (C) de mettre en place des mesures de sécurité raisonnables « propres à assurer la protection des renseignements personnels »[194]. Après cette exploration des obligations prévues par le régime québécois de protection des renseignements personnels, il conviendra (D) d'exposer différentes mesures applicables afin d'assurer la sécurité des renseignements personnels et (E) les défis que présentent les SIA en matière de sécurité.

A. L'approche minimaliste

L'approche minimaliste. Le droit québécois réclame d'adopter une « approche minimaliste »[195] à l'égard des renseignements personnels. Selon l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (ci-après « OBVIA ») et le Laboratoire de cyberjustice, cette approche découle de trois principes : « le principe de détermination des fins, le principe de limitation de la collecte et le principe de limitation de l'utilisation, de la communication et de la conservation [des renseignements personnels]. » [196] Comme OBVIA et le Laboratoire de cyberjustice, nous proposons d'évaluer ces deux derniers principes en un seul, qualifié de « principe de limitation »[197], à des fins de clarté et de concision.

Le principe de détermination des fins. Le principe de détermination des fins réclame à une entreprise privée d'identifier les objectifs poursuivis par la collecte d'un renseignement personnel avant de procéder à celle-ci. En d'autres termes, avant de collecter un renseignement personnel, une entreprise privée doit déterminer pourquoi ce renseignement est collecté. Bien que l'existence de ce principe précède l'adoption de PL64[198], ce dernier a néanmoins précisé son existence en prévoyant que :

---

[193] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[194] *Loi sur l'accès*, préc., note 12, art. 63.1.; *Loi sur le privé*, préc., note 12, art. 10.
[195] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[196] *Id.*
[197] *Id.*
[198] *Id.*

> Toute personne qui exploite une entreprise et qui, en raison d'un intérêt sérieux et légitime, recueille des renseignements personnels sur autrui doit, avant la collecte, déterminer les fins de celle-ci[199].

Le principe de limitation. Ensuite, la *Loi sur l'accès* et la *Loi sur le privé* prévoient que le renseignement personnel ne peut être utilisé, sauf exception, « qu'aux fins pour lesquelles il a été recueilli »[200]. Ce principe de limitation s'applique à toutes les opérations concernant un renseignement personnel parmi lesquelles : la collecte, l'utilisation, la conservation et la communication[201]. Les modifications apportées à la *Loi sur l'accès* et la *Loi sur le privé* prévoient également qu'un organisme public ou une entreprise doit détruire le renseignement personnel ou l'anonymiser lorsque les fins pour lesquelles il a été collecté ou utilisé sont accomplies[202].

Le critère de nécessité. Surtout, le principe de limitation réclame de se limiter qu'aux renseignements personnels dits « nécessaires ». Ce critère de nécessité doit être respecté même si la personne concernée consent à la collecte et à l'usage du renseignement personnel[203]. Ainsi, le nouvel article 5 de la *Loi sur le privé* prévu par PL64 propose que :

> La personne qui recueille des renseignements personnels sur autrui ne doit recueillir que les renseignements nécessaires aux fins déterminées avant la collecte (nos soulignements)[204].

Le nouvel article 64 de la *Loi sur l'accès* propose quant à lui que :

> Nul ne peut, au nom d'un organisme public, recueillir un renseignement personnel si cela n'est pas nécessaire à l'exercice des attributions de cet organisme ou à la mise en œuvre d'un programme dont il a la gestion (nos soulignements)[205]

Pluralité d'interprétations du critère de nécessité. Malheureusement, il n'existe actuellement aucune définition claire et unique du critère. Déziel observe trois interprétations distinctes du critère dans la jurisprudence soit[206] : une interprétation « stricte et littérale »[207], une

[199] *PL64*, préc., note 5, art. 104. Nouvel article 4 de la *Loi sur le privé.*
[200] *Id.*, arts. 20 et 110. Nouveaux articles 12 de la *Loi sur le privé* et 65.1. de la *Loi sur l'accès.*
[201] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[202] *PL64*, préc., note 5, arts. 28 et 119. Nouveaux articles 73 de la *Loi sur l'accès* et 23 de la *Loi sur le privé*.
[203] Voir, par exemple, Commission d'accès à l'information, 30 septembre 2021, 1015556-S, *Enquête à l'égard de Bruneau Électrique inc.*, par. 18, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/514817/index.do>.
[204] *PL64*, préc., note 5, art. 105. Nouvel article 5 de la *Loi sur le privé.*
[205] *Loi sur l'accès*, préc., note 12, art. 64.
[206] P.-L. DÉZIEL, préc., note 131, 9.
[207] *Id.*, 9-10.

interprétation « contextuelle et relative »[208] et une interprétation « souple et dynamique »[209]. À noter que ces interprétations « cohabitent et, dans une certaine mesure, se chevauchent »[210] dans la jurisprudence. En effet, Déziel note qu'il :

> ne s'agit donc pas d'interprétations qui se sont succédé ou qui se sont remplacées; au contraire, nous verrons que la jurisprudence semble aujourd'hui encore hésitante à confirmer la prépondérance d'une d'entre elles sur les autres[211].

Interprétation stricte et littérale. L'interprétation stricte et littérale prévoit qu'un renseignement ne peut être collecté que si ce dernier est indispensable aux objectifs de l'entreprise ou de l'organisation publique. Ainsi, en 2010 dans l'arrêt *M.L. c Ville de Gatineau,* la CAI reprenait l'interprétation du critère par le Tribunal d'arbitrage qui a proposé en 1998 que :

> le mot « nécessaire » doit être interprété de façon restrictive. C'est donc dans son sens d'« indispensable », d'« essentiel » ou de « primordial » qu'il doit être retenu. N'est donc pas « nécessaire », au sens de l'article 5 [de la *Loi sur le privé*] (et par ricochet aux fins de l'article 37 C.c.Q.) ce qui est simplement commode, utile, avantageux ou expédient : est nécessaire ce qui est indispensable, essentiel et dont la présence « rend seule possible une fin ou un effet », pour reprendre les mots du Petit Robert[212].

Interprétation contextuelle et relative. Toujours en 2010, la Cour du Québec a favorisé, dans *PB c Lepage,* une interprétation « contextuelle et relative »[213] du critère de nécessité. Selon le tribunal, le critère réclame de prouver « que les renseignements recueillis […] sont nécessaires à l'objet du dossier en ce sens qu'ils ne sont pas superflus, sans objet ni pertinence. »[214]

Interprétation souple et dynamique. Enfin, une dernière interprétation a été développée en 2003 par la Cour du Québec dans l'arrêt *Société de Transport de la Ville de Laval c. X*.[215] Cette interprétation, que Trudel qualifie de « souple et dynamique »[216], prévoit qu' « un

---

[208] *Id.*, 10-11.
[209] *Id.*, 12-13.
[210] *Id.*, 9-10.
[211] *Id.*
[212] Commission d'accès à l'information du Québec, 16 mars 2010, 08 08 85, *M.L. c. Gatineau (Ville de)*, *68*, par. 79, en ligne : <https://canlii.ca/t/28r34> (consulté le 5 juillet 2022). citant Tribunal d'arbitrage, 31 octobre 1998, Grief n°: 97-02, *Syndicat des employées et employés professionnels et de bureau, section locale 57 et Caisse populaire St-Stanislas de Montréal*, en ligne : <https://canlii.ca/t/hnknp> (consulté le 5 juillet 2022).
[213] P.-L. DÉZIEL, préc., note 131, 10-11.
[214] Cour du Québec, 17 juin 2010, 450-80-000873-090, *P.B. c. Lepage*, par. 91, en ligne : <https://canlii.ca/t/2bljf> (consulté le 7 juillet 2022).
[215] P.-L. DÉZIEL, préc., note 131, 12.
[216] *Id.*; Cour du Québec, 21 février 2003, 500-02-094423-014, *Société de transport de la Ville de Laval c. X.*

renseignement sera donc nécessaire non pas lorsqu'il pourra être jugé absolument indispensable, ou au contraire simplement utile. »[217] Plutôt, selon cette interprétation, la collecte de renseignements personnels « s'inscrit dans une logique de pondération des intérêts de l'organisme public ou de l'entreprise et du droit à la vie privée de la personne »[218]. À ce titre, l'interprétation souple et dynamique impose une évaluation du critère de nécessité qui s'inspire du test développé par la Cour suprême dans l'arrêt *Oakes*[219]. Ainsi, l'analyse de la nécessité réclame, d'abord, d'évaluer les objectifs qui motivent les opérations prévues aux renseignements personnels pour ensuite évaluer si ces opérations constituent une solution proportionnée à la poursuite de ces objectifs.

Premier critère de l'interprétation souple et dynamique - Évaluation des objectifs. En premier lieu, l'organisation visée par une enquête doit prouver donc que les objectifs poursuivis par les renseignements personnels sont suffisamment « légitime[s], important[s], urgent[s] et réel[s] »[220]. Le caractère « réel » réclame de prouver que les objectifs sont plus que simplement « appréhendés »[221]. Par exemple, une entreprise qui collecte des renseignements personnels afin de contrer le vol de temps de ses employés doit faire la démonstration de l'existence d'une telle problématique au sein de son entreprise[222].

---

[217] Cour du Québec, 21 février 2003, *Société de transport de la Ville de Laval c. X.*, préc., note 216, par. 44.
[218] P.-L. DÉZIEL, préc., note 131, 13.
[219] Cour du Québec, 21 février 2003, 500-02-094423-014, *Société de transport de la Ville de Laval c. X.*, par. 44, en ligne : <http://citoyens.soquij.qc.ca/> (consulté le 5 juillet 2022); *R. v. Oakes*, [1986] 1 SCR 103 (Supreme Court of Canada), en ligne : <https://canlii.ca/t/1ftv6> (consulté le 30 janvier 2022).
[220] Commission d'accès à l'information du Québec, 9 février 2015, 111756, *Garderie Excelsiori Daycare inc.*, par. 21, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/351801/index.do?q=Garderie+Excelsiori+Daycare+inc> (consulté le 11 août 2022); Cour du Québec, 21 février 2003, *Société de transport de la Ville de Laval c. X.*, préc., note 216, par. 44; Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, *1014137-S*, 4-5, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520898/index.do>; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Fiche d'information sur les pièces d'identité : entreprises*, p. 1, en ligne : <https://www.cai.gouv.qc.ca/documents/CAI_FI_pieces_identite_entreprises.pdf> (consulté le 20 janvier 2022); Commission d'accès à l'information du Québec, 12 mai 2021, 1019622-S, *Enquête à l'égard de 9055-4635 Québec inc. (Alimentation Larouche)*, par. 16-17, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/500003/index.do> (consulté le 30 janvier 2022).
[221] Commission d'accès à l'information, 30 septembre 2021, *Enquête à l'égard de Bruneau Électrique inc.*, préc., note 203, par. 23.
[222] *Id.*, 6-7.

Deuxième critère de l'interprétation souple et dynamique - la proportionnalité. En second lieu, l'atteinte à la vie privée doit être proportionnelle aux fins poursuivies[223]. La proportionnalité réclame de prouver :

> que « la collecte de ces renseignements est un moyen efficace d'atteindre chaque objectif poursuivi »[224] (ci-après « critère d'efficacité »),
>
> que l'atteinte est minimisée[225] (ci-après « atteinte minimale ») et
>
> que « les avantages de mettre en place ce système surpassent l'atteinte aux droits et les conséquences préjudiciables susceptibles de résulter de cette collecte pour les personnes concernées »[226] (ci-après « proportionnalité de l'atteinte »).

Critère d'efficacité. Ainsi, la collecte des renseignements personnels doit constituer un moyen efficace permettant d'atteindre les objectifs poursuivis par un organisme public ou une entreprise[227]. Par exemple, si une entreprise collecte et conservent des pièces d'identité avec photo de ses clients afin de contrer la fraude, celle-ci devra prouver à la CAI qu'une telle mesure permet bel et bien de réduire la fraude dont elle se dit victime[228].

Atteinte minimale. Une organisation soumise à la *Loi sur le privé* et à la *Loi sur l'accès* doit également choisir le moyen le moins préjudiciable à la vie privée des personnes concernées dans

---

[223] Cour du Québec, 21 février 2003, *Société de transport de la Ville de Laval c. X.*, préc., note 216, par. 44; Commission d'accès à l'information du Québec, 12 mai 2021, *Enquête à l'égard de 9055-4635 Québec inc. (Alimentation Larouche)*, préc., note 220, par. 17.

[224] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 6; Commission d'accès à l'information, 5 avril 2022, 1011672-S, *Services financiers Globex 2000 inc.*, par. 58, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520899/index.do>.
La CAI a également formulé ce critère comme ceci : "la collecte est rationnellement liée aux fins visées"
Commission d'accès à l'information, 30 septembre 2021, *Enquête à l'égard de Bruneau Électrique inc.*, préc., note 203, par. 47; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 220, p. 1.

[225] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 6; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 58.

[226] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 6; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 58. La CAI formule également ce sous-critère de la façon suivante : « la divulgation du renseignement requis est-elle nettement plus utile à l'entreprise que préjudiciable à la personne concernée » ou « l'effet utile est plus grand que le préjudice susceptible d'être causé » et de la façon suivante « L'effet utile est plus grand que le préjudice susceptible d'être causé au client. » COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « La collecte de renseignements personnels », en ligne : <https://www.cai.gouv.qc.ca/la-collecte-de-renseignements-personnels/> (consulté le 10 mars 2022). COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 220, p. 1.

[227] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 6.

[228] Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 11, 34-40 et 60-61.

la poursuite de ses objectifs. En d'autres termes, l'organisation doit prouver qu'il « n'existe pas d'autre solution raisonnable portant moins atteinte à la vie privée »[229] que le moyen qu'elle a choisi. Cette obligation impose donc de favoriser les mesures qui limitent la quantité des renseignements personnels recueillis et qui évitent de collecter des renseignements sensibles[230].

Atteinte proportionnelle. Finalement, le critère de nécessité réclame de s'assurer que les avantages qui découlent de la collecte sont supérieurs à l'atteinte aux droits et aux préjudices qui pourraient en résulter[231]. Par exemple, dans son enquête sur l'*Auberge du lac Sacacomie inc.*, la CAI avait identifié que de recueillir des visages d'employés à des fins de gestion de paie ne constituait ni une atteinte minimale à la vie privée ni une pratique présentant des avantages capables de surpasser l'atteinte aux droits des employés[232]. En effet, l'entreprise avait déjà adopté d'autres méthodes de gestion de paie moins intrusives qui ne nécessitaient pas la collecte de visages[233].

Les modifications suites à PL64. L'approche minimaliste n'a pas été introduite par les modifications législatives prévues par PL64. En revanche, ce projet de loi renforce l'approche minimaliste en précisant que les organismes publics et les entreprises privées qui offrent au public un produit ou un service technologique qui prévoit des paramètres de confidentialité doivent dorénavant « s'assurer que, par défaut, ces paramètres assurent le plus haut niveau de confidentialité, sans aucune intervention de la personne concernée. »[234]

---

[229] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Fiche d'information sur les pièces d'identité : entreprises*, préc., note 220, p. 1.

[230] Commission d'accès à l'information, 30 septembre 2021, *Enquête à l'égard de Bruneau Électrique inc.*, préc., note 203, par. 43. ; Commission d'accès à l'information du Québec, 28 septembre 2016, 061063, *Banque Nationale du Canada*, par. 59-64, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/351226/index.do?q=%22atteinte+minimale%22> (consulté le 11 août 2022).

[231] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 6; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 58. La CAI formule également ce sous-critère de la façon suivante : « la divulgation du renseignement requis est-elle nettement plus utile à l'entreprise que préjudiciable à la personne concernée » ou « l'effet utile est plus grand que le préjudice susceptible d'être causé » et de la façon suivante « L'effet utile est plus grand que le préjudice susceptible d'être causé au client." COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 226. COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 220, p. 1.

[232] Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 7-8.

[233] *Id.*

[234] *PL64*, préc., note 5, arts. 15 et 108. Nouveaux articles 63.7. *Loi sur l'accès* et 9.1. *Loi sur le privé*.

B. L'adoption de règles de gouvernance

Prévoir des règles de gouvernance pour les organismes publics. PL64 impose de nouvelles obligations aux organisations publiques et aux entreprises relatives à l'établissement de règles de gouvernance à l'égard des renseignements personnels. Ainsi, les modifications prévues à la *Loi sur l'accès* imposent aux organismes publics de prévoir des « règles encadrant sa gouvernance à l'égard des renseignements personnels »[235]. Ces règles de gouvernance doivent être publiées sur le site internet de l'organisme public[236].

Prévoir des pratiques et politiques de gouvernance pour les entreprises. Les modifications prévues à la *Loi sur le privé* imposent quant à elles aux entreprises d'établir et de mettre en œuvre « des politiques et des pratiques encadrant [leur] gouvernance à l'égard des renseignements personnels et propres à assurer la protection de ces renseignements. »[237]. Les entreprises privées doivent aussi publier sur leur site web des « informations détaillées au sujet de ces politiques et de ces pratiques »[238] en termes simples et clairs. Si elles n'ont pas de site web, les entreprises doivent rendre ces informations « accessibles par tout autre moyen approprié »[239].

Contenu des pratiques, politiques ou règles de gouvernance. Politiques, pratiques et règles de gouvernance permettent de préciser les pouvoirs d'accès aux renseignements personnels des employés d'un organisme public ou d'une entreprise privée. En effet, les règles de gouvernance des organismes publics doivent prévoir « les rôles et les responsabilités des membres de son personnel »[240], une mesure qui risque de faciliter le respect de l'article 62 de la *Loi sur l'accès* qui « prévoit qu'un organisme doit prendre les mesures visant à restreindre l'accès aux seuls renseignements personnels nécessaires à l'exercice des fonctions des employés »[241]. Les politiques et pratiques des entreprises doivent similairement « prévoir les rôles et les

[235] *Id.*, art. 15. Nouvel article 63.3 de la *Loi sur l'accès.*
[236] *Id.*
[237] *Id.*, art. 103. Nouvel article 3.2. de la *Loi sur le privé.*
[238] *Id.*
[239] *Id.*
[240] *Id.*, art. 15. Nouvel article 63.3 de la *Loi sur l'accès.*
[241] Commission d'accès à l'information, 6 janvier 2022, *Enquête à l'égard du Centre intégré universitaire de santé et services sociaux de l'Estrie et du Centre hospitalier universitaire de Sherbrooke et Ministère de la Santé et des Services sociaux*, par. 23, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519701/index.do> (consulté le 7 juin 2022); *Loi sur l'accès*, préc., note 12, art. 62.

responsabilités des membres de son personnel tout au long du cycle de vie de ces renseignements »[242]. Les modifications à la *Loi sur le privé* imposent aussi à l'entreprise de prévoir l'« encadrement applicable à la conservation et à la destruction de ces renseignements »[243], un encadrement qui, bien sûr, devra respecter le principe de limitation susmentionné[244]. Enfin, la *Loi sur le privé* reconnait également que les politiques et pratiques encadrant la gouvernance à l'égard des renseignements personnels doivent être « propres à assurer [leur] protection» [245] et « proportionnées à la nature et à l'importance des activités de l'entreprise »[246]. Finalement, tant les règles des organismes publics que les politiques et pratiques de gouvernance des entreprises doivent prévoir un processus de traitement des plaintes relatives à la protection des renseignements personnels[247].

C. L'imposition de mesures de sécurité « raisonnables »

Obligations générales de protection des renseignements personnels. Tant la *Loi sur le privé* que la *Loi sur l'accès* imposent de « prendre les mesures de sécurité propres à assurer la protection des renseignements personnels collectés, utilisés, communiqués, conservés ou détruits... »[248]

Ce qu'est un incident de confidentialité. À ce titre, il convient de se prémunir contre ce que le régime québécois des renseignements personnels qualifie d'incidents de confidentialité. La *Loi sur l'accès* et la *Loi sur le privé* qualifient « d'incidents de confidentialité »[249] :

> l'accès non autorisé par la loi à un renseignement personnel,
>
> l'utilisation non autorisée par la loi d'un renseignement personnel,
>
> la communication non autorisée par la loi d'un renseignement personnel et
>
> la perte d'un renseignement personnel ou toute autre atteinte à la protection d'un tel renseignement[250].

---

[242] *PL64*, préc., note 5, art. 103. Nouvel article 3.2. de la *Loi sur le privé.*
[243] *Id.*, arts. 15 et 103. Nouvel article 3.2. de la *Loi sur le privé.*
[244] *Supra,* Chapitre 1, Section III, Sous-section A.
[245] *PL64*, préc., note 5, art. 103. Nouvel article 3.2. de la *Loi sur le privé*.
[246] *Id.*
[247] *Id.*, arts. 15 et 103. Nouveaux articles 63.3. de la *Loi sur l'accès* et 3.2. de la *Loi sur le privé*
[248] *Loi sur l'accès*, préc., note 12, art. 63.1.; *Loi sur le privé*, préc., note 12, art. 10.
[249] *PL64*, préc., note 5, art. 15, 103. Nouveaux articles 63.9 de la *Loi sur l'accès* et 3.6. de la *Loi sur le privé.*
[250] *Id.*

Menaces internes et externes. Ces incidents incluent donc des menaces internes, telles que la vente illicite de renseignements personnels par des employés de l'organisation. Ils incluent également les menaces externes comme les attaques par *ransomware* ou les cyberintrusions par lesquelles des acteurs malveillants accèdent ou altèrent illégalement des renseignements personnels.

Modulations des obligations. En revanche, il n'existe pas de « *one-size-fits-all approach* »[251] en matière de sécurité des renseignements. Au Québec, le cadre législatif prévoit que les mesures imposées pour assurer la sécurité des renseignements personnels doivent être « raisonnables »[252]. Le caractère « raisonnable » des mesures requises s'évalue principalement au regard de la « sensibilité [des renseignements personnels concernés], de la finalité de leur utilisation, de leur quantité, de leur répartition et de leur support »[253]. La notion de sensibilité des renseignements personnels mérite d'être détaillée. En effet, tel qu'il sera démontré, les difficultés liées à la collecte et l'usage de renseignements sensibles constituent une tension entre les principes de sécurité et d'équité[254].

Ce qu'est un renseignement sensible en droit québécois. Les modifications apportées à la *Loi sur le privé* et à la *Loi sur l'accès* définissent le renseignement sensible comme un renseignement qui « par sa nature notamment médicale, biométrique ou autrement intime, ou en raison du contexte de son utilisation ou de sa communication [...] suscite un haut degré d'attente raisonnable en matière de vie privée »[255]. À ce titre, le droit québécois n'identifie pas clairement quels sont les renseignements sensibles. Pour savoir quels renseignements sont considérés sensibles, il est nécessaire d'explorer la jurisprudence et les publications de la CAI, des tribunaux et du Commissariat à la protection de la vie privée du Canada (ci-après « CVPC »).

---

[251] INFORMATION COMMISSIONER'S OFFICE, *Guidance on AI and data protection*, Information Commissioner's Office, 2020, p. 52, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/>.
[252] *Loi sur le privé*, préc., note 12. art 10. ; *Loi sur l'accès*, préc., note 12. art 63.1.
[253] *Id.*
[254] *Infra,* Chapitre 2, Section IV.
[255] *PL64*, préc., note 5, arts. 13 et 110. Nouveaux articles 59 de la *Loi sur l'accès* et 12 de la *Loi sur le privé.*

Facteurs d'évaluation de la sensibilité : la nature. Au Québec comme au Canada, la sensibilité des renseignements s'évalue au regard de leur nature et du contexte[256]. Parmi les renseignements personnels qui ont été identifiés comme étant « naturellement » sensibles nous retrouvons, entre autres : les renseignements médicaux[257], les renseignements biométriques[258], les identifiants gouvernementaux uniques comme le numéro d'assurance sociale[259], les origines ethniques et raciales[260], les opinions politiques[261], la vie sexuelle ou l'orientation sexuelle[262] et les croyances religieuses ou philosophiques[263].

Éléments contextuels. Qui plus est, la jurisprudence reconnait plusieurs éléments contextuels pouvant influencer la sensibilité des renseignements personnels par exemple : les attentes raisonnables des usagers[264], le fait que la personne concernée appartient à un groupe

---

[256] *Id.* ; *Loi sur la protection des renseignements personnels et les documents électroniques*, LC 2000 ch 5 66, en ligne : <https://laws-lois.justice.gc.ca/PDF/P-8.6.pdf>. Annexe 4.3.4.

[257] *PL64*, préc., note 5, arts. 13 et 110.; *Loi sur la protection des renseignements personnels et les documents électroniques*, préc., note 256. Annexe 4.3.4.

[258] *PL64*, préc., note 5, arts. 13 et 110. Nouveaux articles 59 de la *Loi sur l'accès* et 12 de la *Loi sur le privé.*

[259] Cour supérieure, 19 septembre 2019, 500-06-000907-184, *Lévy c. Nissan Canada inc.*, par. 72, en ligne : <https://canlii.ca/t/j2klc> (consulté le 9 juillet 2022); Commissariat à la protection de la vie privée du Canada, 25 septembre 2007, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2007-389 : TJX Companies Inc./Winners Merchant International L.P.*, par. 41, en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2007/tjx_rep_070925/> (consulté le 9 juillet 2022).

[260] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Position de principe sur la publicité comportementale en ligne » (4 décembre 2015), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/technologie/protection-de-la-vie-privee-en-ligne-surveillance-et-temoins/pistage-et-publicite/bg_ba_1206/> (consulté le 4 novembre 2021); COMMISSION D'ACCÈS À L'INFORMATION, *Guide d'accompagnement : Réaliser une évaluation des facteurs relatifs à la vie privée*, Commission d'accès à l'information, 2021, p. 9.

[261] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 260.

[262] *Id.*; COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 9.

[263] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 263; COMMISSION D'ACCÈS À L'INFORMATION, préc., note 262, p. 9.

[264] Par exemple, dans son enquête sur Ashley Madison, le CVPC détermina que la sensibilité des adresses était renforcée par certaines déclarations de l'entreprise qui reconnaissaient que la "discrétion et la sécurité offertes à ses utilisateurs en tant qu'aspect clé de ses services." De plus, dans une enquête sur Facebook, le CVPC déclara que l'entreprise ne pouvait pas communiquer à des tiers les informations découlant des conversations privées de ses utilisateurs parce que des personnes ne s'attendraient pas raisonnablement à ce qu'une organisation communique des renseignements personnels, qu'ils ont considérés comme étant assez sensibles pour les limiter à des "amis uniquement".
Voir Commissariat à la protection de la vie privée du Canada, 22 août 2016, Rapport de conclusions d'enquête no 2016-005, *Enquête conjointe sur Ashley Madison menée par le commissaire à la protection de la vie privée du Canada et le commissaire à la protection de la vie privée/commissaire à l'information par intérim de l'Australie*, par. 50 [*Enquête conjointe sur Ashley Madison*], en ligne : <https://canlii.ca/t/h3p5k> (consulté le 4 juillet 2022). et Commissariat à la protection de la vie privée du Canada, 25 avril 2019, 2019-002, *Enquête conjointe du Commissariat*

vulnérable[265], l'impact de la compromission[266] et les risques de réidentification de la personne concernée[267].

Caractère vague et imprécis. La principale difficulté légale relative aux renseignements sensibles est leur imprécision. Le CVPC reconnait lui-même qu'il « n'y a pas de ligne de démarcation nette pour déterminer si un renseignement est sensible ou non. »[268] Plus encore, le concept de sensibilité reste contextuel et ne saurait être figé dans le temps. En effet, il se mesure au regard de l' « attente raisonnable en matière de vie privée »[269] et reste donc assujetti aux évolutions et aléas culturels.

La sensibilité influence la collecte et l'usage possibles des renseignements. Pour nos fins, il convient surtout de comprendre qu'il est plus complexe et plus dispendieux de collecter, d'utiliser, de communiquer ou de conserver des renseignements sensibles que des renseignements anodins en raison des obligations qui y sont associées. Tel qu'identifié, il est plus difficile de prouver la nécessité d'une opération affectant un renseignement sensible qu'un renseignement anodin[270]. Qui plus est, la présence de renseignements sensibles réclame de

---

à la protection de la vie privée du Canada et du Bureau du Commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique au sujet de Facebook, *35606*, 109, en ligne : <https://canlii.ca/t/hzzpl> (consulté le 4 juillet 2022).

[265] Commissariat à la protection de la vie privée du Canada, 8 janvier 2018, 2018-001, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-001 : Un manufacturier de jouets connectés améliore les mesures de sécurité pour protéger adéquatement les renseignements d'enfants*, par. 32 [*Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-001*], en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2018/lprpde-2018-001/> (consulté le 4 juillet 2022).

[266] À ce titre, les renseignements personnels d'une personnalité connue du public ont été identifiés comme étant plus sensibles en raison de l'incidence sur leur réputation, leur image, leur vie personnelle et sur leur vie professionnelle. Commissariat à la protection de la vie privée du Canada, 7 février 2018, 2018-006, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-006 : Intrusion dans la base de données de l'Agence mondiale antidopage*, par. 56 [*Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-006*], en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2018/lprpde-2018-006/> (consulté le 4 juillet 2022); Cour du Québec, 7 décembre 2009, 14676, *A c. B*, *200-22-044658-086*, par. 113, en ligne : <https://canlii.ca/t/27d0x> (consulté le 4 juillet 2022).

[267] Commissariat à la protection de la vie privée du Canada, 22 août 2016, *Enquête conjointe sur Ashley Madison*, préc., note 264, par. 47, 146-149.

[268] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Lignes directrices pour l'obtention d'un consentement valable » (24 mai 2018), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/collecte-de-renseignements-personnels/consentement/gl_omc_201805/> (consulté le 4 novembre 2021).

[269] *PL64*, préc., note 5, arts. 13 et 110. Nouveaux articles 59 de la *Loi sur l'accès* et 12 de la *Loi sur le privé.*

[270] *Supra,* Chapitre 1, Section III, Sous-sections A et C.

renforcer les mesures de sécurité devant être adoptées[271]. PL64 prévoit également que l'usage de renseignements sensibles réclame souvent d'obtenir un consentement exprès de la personne concernée[272], c'est-à-dire un consentement manifesté par un « geste positif clair »[273]. Surtout, la sensibilité des renseignements utilisée affecte les sanctions pouvant être appliquées. En effet, PL64 prévoit que la sensibilité des renseignements est l'un des éléments devant être considérés dans la détermination de la peine appropriée en cas d'infractions à la *Loi sur le privé*[274]. Cette peine peut s'élever jusqu'à « 25 000 000 $ ou du montant correspondant à 4 % du chiffre d'affaires mondial de l'exercice financier précédent si ce dernier montant est plus élevé »[275].

D. Les différentes mesures de sécurité pouvant être mises en œuvre

Multiplicité des mesures de sécurité possibles. Les mesures de sécurité qui permettent de se prémunir contre les incidents de confidentialité sont trop nombreuses pour être énumérées. Il convient néanmoins de comprendre que les mesures de sécurité regroupent à la fois des mesures organisationnelles et techniques.

Mesures organisationnelles. Nous avons déjà explicité certaines mesures organisationnelles lors de notre analyse des règles, politiques et pratiques de gouvernance susmentionnée[276]. Par exemple, sécuriser les renseignements personnels réclame de limiter les accès des employés aux renseignements personnels. Outre le contrôle des accès internes, PL64 propose à titre de mesures organisationnelles possibles :

> 1° la nomination d'une personne chargée de la mise en œuvre des mesures de protection des renseignements personnels;
>
> 2° des mesures de protection des renseignements personnels dans tout document relatif au projet, tel qu'un cahier des charges ou un contrat;

---

[271] *Loi sur l'accès*, préc., note 12. art 63.1 ; *Loi sur le privé*, préc., note 12. art 10.
[272] *Code civil du Québec*, préc., note 145; *PL64*, préc., note 5, art. 110. Nouvel article 12 de la *Loi sur le privé.*
[273] GOUVERNEMENT DU QUÉBEC, « Renseignements personnels sensibles », en ligne : <https://www.quebec.ca/gouvernement/travailler-gouvernement/services-employes-etat/conformite/protection-des-renseignements-personnels/consentement/renseignements-personnels-sensibles> (consulté le 4 novembre 2021).
[274] *PL64*, préc., note 5, art. 160. Nouvel article 91 de la *Loi sur le privé*.
[275] *Id.*
[276] *Supra*, Chapitre 1, Section III, Sous-section B.

> 3° une description des responsabilités des participants au projet en matière de protection des renseignements personnels;
>
> 4° la tenue d'activités de formation sur la protection des renseignements personnels pour les participants au projet.[277]

Mesures à implémenter pour se prémunir des menaces externes. Contrecarrer les menaces externes peut également demander d'implémenter différentes mesures de sécurité techniques. Il s'agit, par exemple, d'employer de faux serveurs qui permettent de fausser les adversaires (un « *honeypot* »), de renforcer la sécurité des comptes des utilisateurs par un système d'authentifications multiples, d'employer des SIA qui permettent de détecter les intrusions[278]… Il est également possible d'implémenter une architecture dite « *zero trust* »[279]. Plutôt que de se contenter de se prémunir contre des accès et des altérations non autorisées, cette approche à la sécurité :

> operates on the assumption that user identities or the network itself may already be compromised, and instead relies on AI and analytics to continuously validate connections between users, data and resources.[280]

Limiter l'identification. Il est également possible d'implémenter des mesures qui visent à limiter l'identification des personnes concernées par les renseignements personnels en cas d'incidents de confidentialités en employant, par exemple, des techniques cryptographiques qui permettent d'altérer la forme des renseignements personnels[281]. À ce titre, la *Loi sur l'accès* et la *Loi sur le privé* prévoient qu'il est possible d'anonymiser ou de dépersonnaliser les renseignements personnels afin de diminuer les préjudices pouvant résulter des incidents de confidentialités.

---

[277] *PL64*, préc., note 5, arts. 15 et 103. Nouveaux articles 63.6. de la *Loi sur l'accès* et 3.4. de la *Loi sur le privé*.

[278] Voir par exemple : ARTICLE 29 DATA PROTECTION WORKING PARTY et DIRECTORATE C (FUNDAMENTAL RIGHTS AND UNION CITIZENSHIP) OF THE EUROPEAN COMMISSION, *Opinion 05/2014 on Anonymisation Techniques*, WP216, 0829/14/EN, Brussel, 2014, p. 20; Chris MOORE, *Detecting Ransomware with Honeypot Techniques*, *2016 Cybersecurity and Cyberforensics Conference (CCC)*, Amman, Jordanie, 1 août 2016, p. 77-81, p. 4, DOI : 10.1109/CCC.2016.14; Alexandre DEY, Marc VELAY, Jean-Philippe FAUVELLE et Sylvain NAVERS, « Adversarial vs behavioural-based defensive AI with joint, continual and active learning: automated evaluation of robustness to deception, poisoning and concept drift », *European Cyber Week- C&ESAR/IAD Conference- Artificial Intelligence and Defence*, Rennes, France, hal-02432377, 2019, p. 2.

[279] IBM SECURITY, *Cost of a Data Breach : Report 2022*, IBM, 2022, p. 30, en ligne : <https://www.ibm.com/>.

[280] *Id.*

[281] ENISA, *Pseudonymisation techniques and best practices: recommendations on shaping technology according to data protection and privacy provisions.*, Union Européenne, European Union Agency for Network and Information Security - Publications Office, 2019, p. 21-26, en ligne : <https://data.europa.eu/doi/10.2824/247711> (consulté le 30 janvier 2022).

Anonymisation. Suite aux modifications prévues par PL64 la *Loi sur l'accès* et la *Loi sur le privé* prévoient toutes deux qu'un renseignement est anonymisé :

> lorsqu'il est, en tout temps, raisonnable de prévoir dans les circonstances qu'il ne permet plus, de façon irréversible, d'identifier directement ou indirectement cette personne.[282]

Ces mêmes lois prévoient également que l'anonymisation doit se faire selon « les meilleures pratiques généralement reconnues et selon les critères et modalités déterminés par règlement. »[283]

Dépersonnalisation. Un renseignement dépersonnalisé quant-à-lui est défini comme un renseignement qui « ne permet plus d'identifier directement la personne concernée. »[284] Le régime québécois impose aussi aux organismes publics et aux entreprises qui utilisent des renseignements dépersonnalisés de « prendre les mesures raisonnables afin de limiter les risques que quiconque procède à l'identification d'une personne physique à partir de renseignements dépersonnalisés »[285].

Sanctions. Finalement, PL64 prévoit des sanctions pour toute personne qui sans autorisation de l'organisme public ou de la personne qui détient des renseignements personnels « procède ou tente de procéder à l'identification d'une personne physique à partir de renseignements dépersonnalisés »[286]. Des sanctions sont également prévues en cas d'identification de personne physique à partir de renseignements anonymisés. Or, contrairement aux renseignements dépersonnalisés, il est interdit de « réidentifier » les renseignements anonymisés même en cas d'autorisation de l'organisme public ou de l'entreprise qui détient les renseignements[287].

E. Les défis et les mesures spécifiques aux SIA

Les défis liés à la réidentification. Les développements récents en matière d'IA complexifient la sécurité des renseignements personnels. En effet, les SIA posent certains défis au regard de

[282] *PL64*, préc., note 5, arts. 28, 119. Nouveaux articles 28 de la *Loi sur l'accès* et 23 de la *Loi sur le privé*.
[283] *Id.* Nouveaux articles 28 de la *Loi sur l'accès* et 23 de la *Loi sur le privé*.
[284] *Id.*, arts. 20 et 110. Nouveaux articles 65.1 de la *Loi sur l'accès* et 12 de la *Loi sur le privé.*
[285] *Id.*
[286] *Id.*, arts. 69 et 160. Nouveaux articles 159 *Loi sur l'accès* et 91 *Loi sur le privé.*
[287] *Id.*

l'anonymisation et de la dépersonnalisation. En fait, ce sont en partie ces avancés qui ont motivé le législateur québécois à amender PL64 qui prévoyait originellement qu'un renseignement ne pouvait être considéré anonymisé que s'il n'était plus possible, de façon irréversible, d'identifier la personne concernée par le renseignement[288]. Plusieurs intervenants avaient, en effet, énoncé lors des travaux ayant précédé PL64 que d'anonymiser au point où il n'était pas possible, de façon irréversible, de réidentifier la personne concernée était « pratiquement impossible »[289] en raison notamment des développements récents en intelligence artificielle[290].

Des risques propres aux SIA. Ensuite, en matière de SIA, les mesures de sécurité ne devraient pas se limiter à l'implémentation de barrières limitant l'accès direct aux jeux d'entrainement. En effet, des acteurs malveillants peuvent accéder aux renseignements personnels composant le jeu d'entrainement en interagissant avec un SIA déjà entrainé. En d'autres termes, les « *outputs* »[291] d'un SIA peuvent révéler des renseignements personnels utilisés lors de l'entrainement du système[292]. Ainsi, plusieurs auteurs ont révélé qu'il était possible de mener des « *membership inference attacks »*[293] qui permettent de « *deduce whether a given individual was present in the training data of a ML [machine learning] model*. »[294] De plus, il a été démontré que si des attaquants détiennent certaines informations sur des personnes incluses dans les données d'entrainement alors ceux-ci peuvent « *infer further personal information about those same*

[288] ASSEMBLÉE NATIONALE DU QUÉBEC, « Étude détaillée du projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », *Journal des débats de la Commission des institutions* 45-132 (31 mars 2021), en ligne : <http://assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210331.html> (consulté le 21 janvier 2022).

[289] Voir, par e.g. les interventions de Diane Poitras de la CAI, de Jocelyn Maclure d''Option consommateurs, de Marie-Pierre Grignon du Bureau d'assurance du Canada… ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels - 29 septembre 2020 », (2020) 45-96 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-200929.html> (consulté le 30 janvier 2022).

[290] *Id.* Voir, par e.g. l'intervention de Jocelyn Maclure d'Option consommateurs.

[291] INFORMATION COMMISSIONER'S OFFICE, préc., note 251, p. 56.

[292] *Id.*

[293] *Id.*, p. 57. Michael VEALE, Reuben BINNS et Lilian EDWARDS, « Algorithms that remember: model inversion attacks and data protection law », (2018) 376-2133 *Philos Trans A Math Phys Eng Sci* 20180083, 5, DOI : 10.1098/rsta.2018.0083; Hongsheng HU, Zoran SALCIC, Lichao SUN, Gillian DOBBIE, Philip YU et Xuyun ZHANG, « Membership Inference Attacks on Machine Learning: A Survey », *ACM Computing Surveys* 2022, 1, DOI : 10.1145/3523273.

[294] INFORMATION COMMISSIONER'S OFFICE, préc., note 251, p. 57.; M. VEALE, R. BINNS et L. EDWARDS, préc., note 293, 5; H. HU et al., préc., note 293, 1.

individuals by observing the inputs and outputs of the ML [machine learning] model »[295]. Ce type d'attaque est qualifiée quant à elle de « *model inversion attack* »[296]. Ces attaques réclament donc de mettre en place des mesures de sécurité même après l'entrainement du SIA.

Mesures à implémenter. Malheureusement, il n'est pas très clair quelles mesures peuvent être mises en œuvre afin de limiter les risques liés aux deux types attaques susmentionnées. Ainsi, l'Information Commissioner's Office (ci-après « ICO ») du Royaume-Uni soutient que :

> Security and ML researchers are still working to understand what factors make ML models more or less vulnerable to these kinds of attacks, and how to design effective protections and mitigation strategies.[297]

Cet organisme propose surtout d'implémenter des stratégies et mesures liées à l'interface de programmation (en anglais « API ») du SIA[298]. Il propose, entre autres, de limiter le nombre de requêtes à un SIA issues d'un même usager au cours d'une même période[299]. L'organisation propose également d'implémenter des mesures permettant de « *monitor queries from the API's users, in order to detect whether it is being used suspiciously* »[300]. Agir ainsi permettra de bloquer ou de suspendre tout utilisateur suspect ou malicieux.

Le principe le mieux traduit. À la vue de ce qui précède, il est clair que le régime québécois des renseignements personnels accorde une grande importance à la sécurité des renseignements personnels collectés et utilisés par les entreprises et les organismes publics. Alors qu'une poignée de dispositions de PL64 traduisent les principes d'explicabilité et d'exactitude, ce projet de loi, la *Loi sur le privé* et la *Loi sur l'accès* prévoient une multiplicité d'obligations se rattachant à la sécurité des renseignements personnels. Pourtant, malgré cette quantité de dispositions les obligations imposées aux organismes publics et entreprises restent parfois mal définies. Ainsi, le critère de nécessité, pourtant central à l'approche minimaliste, fait l'objet de trois interprétations concurrentes. Le droit québécois prévoit également une modulation des mesures devant être imposées pour assurer la sécurité des renseignements personnels. Le droit prévoit, en effet,

[295] INFORMATION COMMISSIONER'S OFFICE, préc., note 251, p. 56.
[296] *Id.*
[297] *Id.*, p. 58.
[298] *Id.*
[299] *Id.*, p. 59.
[300] *Id.*

l'implémentation de mesures « raisonnables ». Or, cette raisonnabilité doit être interprétée au regard de notions floues comme la sensibilité des renseignements personnels utilisés. Toutefois, comme nous le verrons, le caractère nébuleux des obligations rattachées à la sécurité est source de tensions à l'égard des autres principes[301].

Section IV - Le principe d'équité et de non-discrimination

Importance du principe. L'importance du principe d'équité et de non-discrimination (ci-après « principe d'équité ») est universellement reconnue. Il s'agit, en fait, de l'un des principes les plus souvent identifiés dans les lignes directrices destinées au développement ou à l'usage responsables de SIA[302].

Considérations québécoises. Au Québec, plusieurs organisations publiques ou non gouvernementales sont également conscientes des risques d'iniquités et de discriminations associés à l'usage des SIA. En effet, lors des consultations publiques ayant précédé l'adoption de PL64 plusieurs participants, dont la CDPDJ, la CAI et la *Ligue des droits et libertés* ont déclaré craindre que les SIA et d'autres applications numériques puissent mener à des traitements discriminatoires et non équitables[303].

---

[301] *Infra,* Chapitre 2, Sections II et IV.

[302] Selon T. HAGENDORFF, préc., note 4, 102., ce principe est identifié, au moins, dans les lignes directrices et documents suivants : les *Lignes directrices pour une IA digne de confiance*, le *Report on the Future of Artificial Intelligence, l*es *Beijing AI Principles* de 2019, les recommandations du Conseil sur l'intelligence artificielle de l'OCDE, le *AI4People - an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, l*es *AsilomarAIPrinciples* du *Future of Life Institute* de 2017, le rapport de *AINow* de 2016, le rapport de *AINow* de 2017, le rapport de *AINow* de 2018, le rapport de *AINow* de 2019, les *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, *Ethically Aligned Design : A vision for Prioritizing Human Well being with Autonomous and intelligent systems (Version for Public Discussion), Ethically Aligned Design : A vision for Prioritizing Human Well being with Autonomous and intelligent systems (First Edition), l*es *Microsoft responsible AI principles; l*es principes liées à l'intelligence artificielle de Google, le *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers* de 2018 et *Partnership on AI*.

[303] ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels - 23 septembre 2020 », (2020) 45-96 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-200923.html> (consulté le 4 juin 2022); J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 63; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 5, p. 16.

Présentation du principe. Il n'existe pourtant actuellement aucun consensus permettant d'identifier avec précision quelle forme d'équité devrait être poursuivie par les SIA. En fait, il sera démontré que la difficulté à définir le principe est si prononcée qu'il est tout simplement impossible, dans plusieurs circonstances, d'assurer un traitement qui respecte toutes les interprétations de l'équité[304]. En d'autres termes, le respect de ce principe présente non seulement des tensions avec les autres principes, mais différentes conceptions du principe sont également en tension. Pour cette raison, il convient, pour nos fins, d'analyser (A) différentes conceptions du principe, (B) les principales raisons justifiant les difficultés des SIA à respecter ces conceptions du principe, (C) les mesures devant être mises en œuvre afin de privilégier une équité et (D) comment le principe d'équité se traduit en droit québécois.

A. Les différentes conceptions de l'équité

Comment définir le principe d'équité en IA. Bien que certains auteurs répertorient dans la littérature scientifique de l'AA jusqu'à dix conceptions du principe d'équité[305], il n'apparait pas utile, pour nos fins, d'expliciter chacune d'entre elles. Afin d'illustrer adéquatement les tensions qui concernent le respect de ce principe nous limiterons notre analyse à trois conceptions soit : (i) la parité de traitement, (ii) la parité de performance et la (iii) parité statistique.

i. La parité de traitement

Définir la parité de traitement. La « parité de traitement (traduction libre) »[306] propose de ne pas permettre aux SIA de considérer les renseignements qui se rattachent à des motifs de discrimination tels que la couleur de peau, le genre, l'orientation sexuelle (ci-après « attributs protégés »)... Selon cette conception, toutes les personnes évaluées par le SIA doivent bénéficier du même traitement sans considération aux groupes auxquels elles appartiennent.

---

[304] *Infra,* Chapitre 2, Sections IV et V.

[305] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:12.

[306] Michael Carl TSCHANTZ, *What is Proxy Discrimination ?*, *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*, Seoul, Republic of Korea, arXiv, 25 mai 2022, p. 1993-2003, p. 1993, DOI : 10.48550/arXiv.2205.05265.] citant Muhammad Bilal Zafar, Isabel Valera, Manuel Rodriguez, Krishna Gummadi, and Adrian Weller. 2017. From Parity to Preference-based Notions of Fairness in Classification. In Advances in Neural Information Processing System.

Motivation derrière cette vision: la crainte du profilage injustifié. L'espoir des partisans de la parité de traitement est qu'en empêchant le SIA de considérer les attributs protégés, on parviendra à prévenir des traitements discriminatoires[307]. En d'autres termes, les partisans d'une parité de traitement craignent que la connaissance de ces données permette un profilage injustifié de certains groupes[308].

Un problème d'actualité. Ces dernières années plusieurs entreprises ont fait l'objet de controverses pour leurs utilisations d'attributs protégés. Ainsi, *Facebook* infère certains renseignements tels que la race, les opinions politiques et les croyances religieuses à des fins commerciales, et ce, sans toujours obtenir un consentement valide des personnes concernées[309]. Il a également été révélé que l'entreprise *Target* détecte si ses usagères sont enceintes à des fins publicitaires[310]. *Amazon* a enregistré en 2018 un brevet sur une nouvelle version d'*Alexa*. Celle-ci permet à l'outil d'inférer l'état de santé d'un individu en analysant sa voix[311]. Bref, plusieurs entreprises tentent d'obtenir des attributs protégés ce qui alimente la crainte que ce type de renseignements puissent être utilisés malicieusement[312].

ii. La parité de performance

Définir la parité de performance. La parité de performance également qualifiée de « parité de classification (traduction libre) »[313] ou d'« *equalised odds »*[314] vise à ce « que les mesures

[307] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Pour un développement responsable de l'intelligence artificielle qui respecte le droit à la vie privée et responsabilise tous les acteurs impliqués - Présenté à la Déclaration de Montréal pour une intelligence artificielle responsable*, Montréal, Commission d'accès à l'information du Québec, 2018, p. 3, en ligne : <https://www.cai.gouv.qc.ca/publications-et-documentation/documents-de-reflexion-et-danalyse/>.

[308] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 5, p. 16-17. Voir aussi : J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 14.

[309] José González CABAÑAS, Ángel CUEVAS et Rubén CUEVAS, *Facebook Use of Sensitive Data for Advertising in Europe*, *27th USENIX Security Symposium (2018) 479-495,* Boston, États-Unis, 14 février 2018, p. 4-10, DOI : 10.48550/arXiv.1802.05030.

[310] S. WACHTER et B. MITTELSTADT, préc., note 57, 509.

[311] JAMES COOK, « Amazon patents new Alexa feature that knows when you're ill and offers you medicine », *The Telegraph* (2018), en ligne : <https://perma.cc/V346-HFWE> (consulté le 4 juin 2022). tel que cité par S. WACHTER et B. MITTELSTADT, préc., note 57, 506-510.

[312] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 76-77.

[313] Sam CORBETT-DAVIES et Sharad GOEL, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, arXiv, 14 août 2018, p. 2, en ligne : <http://arxiv.org/abs/1808.00023> (consulté le 22 mai 2022).

[314] Geoff PLEISS, Manish RAGHAVAN, Felix WU, Jon KLEINBERG et Kilian Q WEINBERGER, *On Fairness and Calibration*, *Advances in Neural Information Processing Systems*, Los Angeles, États-Unis, Curran Associates, Inc., 30, 2017, p. 1,

communes de la performance prédictive soient les mêmes pour tous les groupes définis par des attributs protégés. (traduction libre) »[315]. Cette perception de l'équité réclame, en effet, à ce que la « *measure of classification error is equal across groups defined by the protected attributes* »[316]. Des mesures souvent identifiées comme pertinentes dans la poursuite de cette forme d'équité sont :

> les taux de faux positifs entre différents groupes,
>
> les taux de faux négatifs entre différents groupes,
>
> les valeurs prédictives positives entre les différents groupes et
>
> les valeurs prédictives négatives entre les différents groupes[317].

À ce titre, la parité de performance réclame, par exemple, de garantir qu'un SIA soit aussi précis à l'égard d'une personne noire, qu'à l'égard d'une personne blanche.

Problème pratique. COMPAS, un SIA utilisé par des procureurs américains afin de calculer le risque de récidives de personnes détenues avant procès a fait controverse lorsque le journal ProPublica a dévoilé que les :

> noirs non récidivistes avaient deux fois plus de chances que les blancs d'être étiquetés comme présentant un risque supérieur et que le système faisait l'erreur inverse pour les blancs, chez qui les récidivistes avaient beaucoup plus de chances d'avoir été considérés comme présentant un risque inférieur[318].

De façon similaire, une étude du National Institute of Standards and Technology du US Department of Commerce identifiait en 2019 que les algorithmes de reconnaissance faciale identifient erronément 10 à 100 fois plus souvent les personnes amérindiennes, asiatiques ou afro-américaines que les personnes caucasiennes. Puisque les algorithmes de reconnaissance

---

en ligne : <https://papers.nips.cc/paper/2017/hash/b8b9c74ac526fffbeb2d39ab038d1cd7-Abstract.html> (consulté le 22 juin 2022).

[315] S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 1.

[316] *Id.*, p. 5.

[317] *Id.*, p. 5 et 11.

[318] Frederik Zuiderveen BORGESIUS, *Discrimination, intelligence artificielle et décisions algorithmiques*, coll. Publication de la Direction générale de la Démocratie, Strasbourg, Conseil de l'Europe, 2018, p. 25. citant J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3.

faciale sont parfois utilisés par les forces de l'ordre, cette iniquité risque de renforcer le profilage racial des populations plus vulnérables[319].

Importance de la conception: répartition équitable des risques et bénéfices. Cette conception de l'équité renvoie donc à une intention de ne pas désavantager un individu en raison de son appartenance à un groupe vulnérable. Les bénéfices et les risques créés par les SIA ne devraient pas favoriser certaines personnes simplement en raison de leur appartenance à certains groupes.

Importance de la conception: maintenir la pertinence des SIA. De plus, si un décideur, comme un juge ou un fonctionnaire, a recours à un SIA afin de l'assister pour prendre une décision, assurer une parité de performance permet à ce dernier de se fier aux résultats produits par le SIA sans avoir à considérer la couleur de peau, le genre ou l'orientation sexuelle de l'individu évalué[320]. Par exemple, en présence d'un SIA qui performe moins bien à l'égard des femmes, un décideur peut être motivé, voire contraint, à accorder moins d'importance aux décisions prises par l'outil à l'égard des femmes qu'à l'égard des hommes[321]. À ce titre, Pleiss et al. soulignent qu'il importe d'assurer une parité de performance chez les SIA qui évaluent les risques de récidives sans quoi une :

> probability estimate of p could carry different meaning for African-American and white defendants, and hence the tool would have the unintended and highly undesirable consequence of incentivizing judges to take race into account when interpreting its prediction[322]

iii. La parité statistique

Définir la parité statistique. La troisième perception de l'équité étudiée est celle de la parité statistique[323]. Selon cette perception, les résultats du SIA ne doivent pas surreprésenter ou sous-

[319] Voir par exemple, la controverse sur l'usage des services de *Clearview AI* par la GRC canadienne. COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 2.

[320] G. PLEISS, M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, préc., note 314, p. 1.

[321] Michael KEARNS et Aaron ROTH, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, New York, Oxford University Press, 2019, p. 77 (page pdf).

G. PLEISS, M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, préc., note 314, p. 1.[322] G. PLEISS, M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, préc., note 314, p. 1.

[323] Eleanor BIRD, Jasmin FOX-SKELLY, Nicola JENNER, Ruth LARBEY, Emma WEITKAMP, Alan WINFIELD, EUROPEAN PARLIAMENT, et DIRECTORATE-GENERAL FOR PARLIAMENTARY RESEARCH SERVICES, *The ethics of artificial intelligence: issues and initiatives.*, 2020, p. 31, en ligne : <https://op.europa.eu/publication/manifestation_identifier/PUB_QA0119779ENN> (consulté le 19 mai 2022); M. KEARNS et A. ROTH, préc., note 321, p. 65 (page pdf); Alexandra CHOULDECHOVA, « Fair Prediction

représenter les membres de groupes vulnérables. À ce titre, un SIA comme COMPAS qui évalue le risque de récidives des accusés devrait s'assurer qu'un « *equal proportion of defendants are detained in each race group* »[324]. À noter que ce que constitue un résultat « paritaire » peut varier selon les circonstances et l'auteur. Par exemple, il est possible d'assurer une parité « parfaite » soit de s'assurer que tous les groupes soient équitablement représentés dans les résultats du SIA ou d'assurer une parité « proportionnelle » soit « de veiller à ce que les groupes protégés soient sélectionnés proportionnellement à leur pourcentage dans la population. (traduction libre) »[325].

Exemple pratique. L'algorithme du système COMPAS susmentionné ne respectait probablement pas une parité statistique. En effet, ProPublica soutient qu'en utilisant COMPAS les Afro-américains ont 45 % plus de chances de recevoir un score de récidive plus élevé que les personnes blanches[326]. Un Afro-Américain a également 77,3 % plus de chances de recevoir un score plus élevé de récidives violentes qu'une personne blanche[327]. En conséquence, si un nombre égal d'Afro-Américains et de personnes blanches étaient arrêtés un nombre beaucoup plus élevé d'Afro-Américains seront incarcérés, une situation qui violerait la parité statistique.

B. Les causes des iniquités

Causes du problème. Les raisons sous-tendant les comportements inéquitables des SIA sont multiples et il n'est ni possible ni souhaitable de toutes les explorer afin d'illustrer les tensions entre ces différentes conceptions de l'équité. Pour nos fins, il convient d'identifier les trois principales causes d'iniquité soit (i) l'usage de proxys, (ii) la présence de biais dans les jeux d'entrainement et (iii) les biais créés par l'algorithme ou par l'interaction entre l'algorithme et l'usager.

---

with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments », (2017) 5-2 *Big Data* 153-163, 4, DOI : 10.1089/big.2016.0047.

[324] E. BIRD et al., préc., note 323, p. 31.

[325] Ainsi, le "*toolkit*" Aequitas permet aux développeurs de SIA en AA de réaliser des audits sur la base d'une parité statistique « parfaite » (qu'il qualifie d' « *equal parity* ») ou d'une parité statistique « proportionnelle » (qu'il qualifie de « *proportionnal parity* »). UNIVERSITY OF CHICAGO, CENTER FOR DATA SCIENCE AND PUBLIC POLICY, « Aequitas - Bias and Fairness Audit Toolkit » (2018), en ligne : <http://aequitas.dssg.io/audit/n5uppdoy/compas_for_aequitas/report-1.html> (consulté le 7 juin 2022).

[326] J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3.

[327] *Id.*

i. Les proxys

Le problème des proxys dans le contexte de l'IA. Simplement interdire à un SIA d'analyser des attributs protégés ne l'empêche pas de produire des résultats incapables de satisfaire une parité statistique ou une parité de performance. En effet, les SIA identifient et reproduisent les tendances observées dans les jeux d'entrainement[328]. Même si on empêche ces outils de considérer une variable spécifique, comme le genre ou l'origine ethnique, la valeur de cette variable reste néanmoins corrélée à d'autres données qui sont analysées par le SIA[329]. On qualifie ces données de « *proxys variables* » (ci-après « proxys »)[330]. À ce titre, si le jeu d'entrainement du SIA défavorise, par exemple, les femmes, la présence de ces proxys permet quand même au SIA de reproduire les iniquités à l'égard de celles-ci, et ce, même si on empêche l'outil de considérer directement le genre des personnes évaluées[331].

L'exemple d'Amazon. Par exemple, en 2014, la compagnie Amazon a entrainé un SIA d'AA sur un jeu d'entrainement composé de *curriculum vitae* (ci-après « CV ») accumulés sur une période de 10 ans. L'objectif était de développer un SIA capable d'évaluer les candidatures en analysant les mots utilisés dans leurs CV. Plus de 50 000 termes étaient considérés par le SIA. Or, l'outil s'est révélé inéquitable. Il défavorisait les CV qui contenaient le terme « femme » et notait défavorablement les CV qui mentionnaient des établissements scolaires exclusivement féminins. Il faut comprendre, en effet, que « l'immense majorité des ingénieurs engagés par Amazon les 10 années précédentes étaient des hommes (traduction libre) »,[332] une tendance que le SIA tentait d'émuler. L'entreprise a tenté de remédier au problème en prohibant au SIA d'évaluer les termes liés au genre. Malgré cette mesure Amazon n'a pas réussi à modifier le comportement discriminatoire de l'outil en raison du fait que le genre restait identifiable dans les termes utilisés

[328] M. KEARNS et A. ROTH, préc., note 321, p. 70 (page pdf).
[329] Richard BENJAMINS, Alberto BARBADO et Daniel SIERRA, *Responsible AI by Design in Practice*, *Proceedings of the Human-Centered AI: Trustworthiness of AI Models & Data (HAI) track at AAAI Fall Symposium*, Washington, États-Unis, 20 décembre 2019, p. 4, DOI : 10.48550/arXiv.1909.12838. ; M. KEARNS et A. ROTH, préc., note 321, p. 70 (page pdf).
[330] R. BENJAMINS, A. BARBADO et D. SIERRA, préc., note 329, p. 4.
[331] M. KEARNS et A. ROTH, préc., note 321, p. 62 (page pdf).
[332] Cet exemple est fourni par Kate CRAWFORD, *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, Yale University Press, 2021, p. 129-130.

dans les CV[333]. En effet, « *the model was biaised against women not just as a category but against commonly gendered forms of speech* » [334].

L'exemple de COMPAS. De façon similaire, certains auteurs considèrent que les résultats du système COMPAS susmentionné sont préjudiciés en raison du phénomène de profilage racial[335]. En effet, COMPAS considère les arrestations précédentes de l'accusé ainsi que les arrestations de ses amis et des membres de sa famille à titre d'indicateurs du niveau de risque de l'accusé[336]. Or, parce que les minorités visibles sont plus souvent contrôlées, surveillées et arrêtées par la police, ces indicateurs peuvent agir comme des « proxys » de leur origine ethnique[337]. Les résultats de COMPAS souffrent, donc, selon ces auteurs, d'un « *Measurement bias* » c'est-à-dire un biais qui « *arises from how we choose, utilize, and measure particular features* »[338]. Ce biais, en retour, provoque une surévaluation des risques associés aux personnes issues des communautés autochtones et afro-américaines[339].

Difficultés à identifier les proxys. La solution au problème peut apparaitre simple : il suffit de retirer tous les proxys de l'analyse des SIA. Or, la situation est bien plus complexe. En effet, prohiber l'usage de proxys réclame d'identifier précisément quels sont les renseignements qui peuvent agir comme proxys. Plusieurs d'entre eux sont déjà connus. Par exemple, il a été démontré que l'adresse permet d'inférer l'origine ethnique d'une personne[340] et que la longueur de cheveux peut révéler le genre[341]. En revanche, ces proxys sont compréhensibles sur une base

---

[333] *Id.*
[334] *Id.*
[335] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:4.
[336] *Id.*, 4.
[337] *Id.*
[338] *Id.*
[339] *Id.*; J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3.
[340] Francesco BONCHI, Sara HAJIAN, Bud MISHRA et Daniele RAMAZZOTTI, « Exposing the probabilistic causal structure of discrimination », (2017) 3-1 *Int J Data Sci Anal* 1-21, 7, DOI : 10.1007/s41060-016-0040-z; Kory JOHNSON, Dean FOSTER et Robert STINE, « Impartial Predictive Modeling: Ensuring Fairness in Arbitrary Models », *Statistical Science* 2016.1-29, 16.
[341] Zachary LIPTON, Alexandra CHOULDECHOVA et Julian MCAULEY, *Does mitigating ML's disparate impact require disparate treatment?*, *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada, 2018, p. 2; Sam CORBETT-DAVIES, Emma PIERSON, Avi FELLER, Sharad GOEL et Aziz HUQ, *Algorithmic Decision Making and the Cost of Fairness*, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, coll. KDD '17, New York, NY, USA, Association for Computing Machinery, 14 août 2018, p. 797-806, p. 8, DOI : 10.1145/3097983.3098095.

intuitive. Or, les SIA sont capables de reconnaitre des relations statistiques imperceptibles aux humains. Ainsi :

> intuition alone will often be inadequate to identify an AI's use of a proxy variable, event after the fact. No longer are the "traditional" proxies, like headgear, hairstyles, or height and weight, the only potential substitutes for our society's protected traits. Instead, AIs can generate proxies for directly predictive suspect traits based on all sorts of behavior, from what movies one streams online to the language one uses in social media posts[342].

Surtout, les proxys peuvent résulter de la combinaison de plusieurs variables[343]. À ce titre, *« proxies available to AIs may consist of numerous interacting pieces of data, whose significance as a proxy may be completely unintuitive*. »[344] Non seulement cela signifie que « *exempting any one variable does not result in a significant difference since associated variables still yield proxies that are nearly as strong* »[345], mais en plus, cela limite substantiellement l'espoir de pouvoir dresser une liste exhaustive de tous les proxys possibles. En effet, « *every covariate commonly used in predictive models is at least partially correlated with protected group status* »[346]. À ce titre :

> simply denying AIs access to the most intuitive proxies for directly predictive variables does little to thwart this process; instead it simply causes AIs to produce models that rely on less intuitive proxies[347].

Légitimité de certains renseignements. De plus, certains renseignements, que l'on sait être des proxys, sont malgré tout pertinents dans le cadre de nos processus décisionnels. Par exemple, le niveau d'éducation peut légitimement être réclamé dans le cadre d'une demande de candidature pour un emploi. Il s'agit cependant d'une variable corrélée au genre, au statut socio-économique et à l'origine ethnique d'un individu[348]. De plus, dans le cas de COMPAS susmentionné, les

---

[342] ANYA E. R. PRINCE et DANIEL SCHWARCZ, « Proxy Discrimination in the Age of Artificial Intelligence and Big Data », (2020) 105-1257 *Iowa Law Review*, 1304, en ligne : <https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data/> (consulté le 5 juin 2022).
[343] *Id.*
[344] *Id.*
[345] Samuel YEOM, Anupam DATTA et Matt FREDRIKSON, *Hunting for Discriminatory Proxies in Linear Regression Models*, *32nd Conference on Neural Information Processing Systems (NIPS 2018)*, Montréal, Canada., 2018, p. 11, p. 9.
[346] S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 9.
[347] *Id.*
[348] *Id.*

arrestations précédentes de la personne évaluée constituaient un proxy de l'origine ethnique[349]. En revanche, les antécédents criminels sont des renseignements considérés pertinents par la Cour suprême du Canada dans l'évaluation de la nécessité de détenir provisoirement un prévenu[350]. À ce titre, simplement bannir tous proxys est une démarche non seulement difficile, voire impossible, mais agir ainsi nous demande d'abandonner des variables potentiellement utiles.

ii. Les biais pouvant affecter les jeux d'entrainement

Causes du problème : les jeux de données biaisées. Les SIA d'AA tendent à produire des résultats discriminatoires en présence de biais latents dans les jeux d'entrainement. Ainsi, « *[I]n the cases where the underlying training data contains biases, the algorithms trained on them will learn these biases and reflect them into their predictions*. »[351]

Ce que l'on entend par « biais ». Il importe cependant de bien comprendre ce que le terme « biais » signifie en l'espèce. Lorsque les spécialistes en AA traitent de « biais », ceux-ci ne réfèrent pas à la conception du terme que l'on retrouve généralement dans le discours légal c'est-à-dire « a *preconceived notion or opinion, a judgement based on prejudices, as opposed to a decision come to from the impartial evaluation of the facts of a case* »[352]. Ils ne réfèrent pas non plus à l'usage du terme dans le discours populaire ou journalistique dans lequel le terme se limite, très souvent, à identifier les perceptions sexistes ou discriminatoires[353]. Un jeu de données « biaisé » n'est pas nécessairement un ensemble de données contaminé par les préjugés des personnes chargées de les collecter ou de les étiqueter. Lorsque les auteurs en AA utilisent le

[349] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:4.
[350] *R. c. St-Cloud*, [2015] 2 RCS 328, part. 71 (Cour suprême du Canada), en ligne : <https://canlii.ca/t/ghtdb> (consulté le 15 août 2022).
[351] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:3.
[352] K. CRAWFORD, préc., note 332, p. 135.
[353] Thomas HELLSTRÖM, Virginia DIGNUM et Suna BENSCH, *Bias in Machine Learning -- What is it Good for?*, *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), 2020. CEUR Workshop Proceedings 2659*, Santiago de Compostella, Espagne, arXiv, 20 septembre 2020, p. 1, DOI : 10.48550/arXiv.2004.00686.

terme « biais », ils réfèrent à une conception beaucoup plus large et englobante du terme qui inclut, entre autres, les biais statistiques[354], les biais cognitifs[355] et les biais algorithmiques[356].

Exemple de biais statistiques : le « *Representation bias* ». Un exemple de biais statistique est le « *Representation bias* »[357]. Ce biais émerge lorsqu'un jeu d'entrainement est trop homogène ou n'est pas représentatif de la population sur laquelle le SIA est utilisé. Tel serait le cas, par exemple, d'un jeu de données composé uniquement de Caucasiens destinés à l'entrainement d'un SIA diagnostiquant le cancer de la peau[358]. L'usage d'un jeu qui souffre de « *representation bias* » *« may result in a classifier that performs bad in general, or bad for certain demographic group* »[359].

Exemple de biais statistiques : l'« *Omitted variable bias* ». Également, le jeu de données peut omettre certaines variables qui seraient pertinentes à l'analyse. On parle alors d' « *Omitted variable bias* »[360]. Ainsi, un SIA qui prédit les ventes d'un service peut devenir imprécis en présence d'un nouvel événement qu'il ne considère pas à titre de variable, par exemple, l'émergence d'un nouveau compétiteur qui offre les mêmes services à plus bas prix[361].

Exemple de biais statistiques : l'« *Aggregation bias* ». Ensuite, présumer que les tendances observées à l'égard de la majorité de la population seraient applicables à tous individus peut mener à un « *Aggregation biais* » [362]. Mehrabi et al. donnent l'exemple des patients diabétiques. En effet, le taux de morbidité de patients diabétiques fluctue en fonction du sexe et de l'ethnicité.

---

[354] David DANKS et Alex John LONDON, *Algorithmic Bias in Autonomous Systems*, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence AI and autonomy track (IJCAI-17)*, Melbourne, Australia, 2017, p. 6, p. 4692, en ligne : <https://www.ijcai.org/proceedings/2017/654> (consulté le 14 juin 2022).

[355] M. KEARNS et A. ROTH, préc., note 321, p. 70 (page pdf).

[356] Pour nos fins, nous identifions à titre de "biais algorithmiques" les biais créés par l'algorithme (e.g. en raison de ses choix de conception) et qui n'existent pas dans les jeux d'entrainement. Voir par exemple, N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:7.

[357] T. HELLSTRÖM, V. DIGNUM et S. BENSCH, préc., note 353, p. 3; N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:5.

[358] Voir par exemple : David WEN, Saad M KHAN, Antonio JI XU, Hussein IBRAHIM, Luke SMITH, Jose CABALLERO, Luis ZEPEDA, Carlos DE BLAS PEREZ, Alastair K DENNISTON, Xiaoxuan LIU et Rubeta N MATIN, « Characteristics of publicly available skin cancer image datasets: a systematic review », (2022) 4-1 *The Lancet Digital Health* e64-e74, e71, DOI : 10.1016/S2589-7500(21)00252-1.

[359] T. HELLSTRÖM, V. DIGNUM et S. BENSCH, préc., note 353, p. 3.

[360] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:4-115:5.

[361] *Id.*

[362] *Id.*, 115:5.

Un SIA servant d'outil d'aide clinique doit donc être en mesure de considérer ces différences individuelles chez les patients afin de maintenir un bon niveau de précision[363].

Les biais cognitifs: Enfin, la présence de biais cognitifs dans les jeux de données peut également affecter les résultats d'un SIA. L'exemple de *Word2vec*, un modèle de SIA de Google qui permet de dresser des analogies entre certains mots, fournit un bon exemple de ce type de biais[364]. Ainsi, ce modèle de « *word-embedding* »[365] permet de dresser des analogies entre certains termes, par exemple, « un roi est à un homme » ce qu'une « reine est à une femme »[366]. Or, en 2016, un groupe de chercheurs a identifié plusieurs formes d'analogies regrettables commises par le modèle. Par exemple, à la question un « *Man is to Computer Programmer as woman is to Y* », le modèle modifiait la variable Y par « *Homemaker* »[367], une réponse pour le moins stéréotypé. Kearns et Roth proposent que ces résultats inéquitables ont été causés par les biais cognitifs des auteurs des textes qui ont permis d'alimenter les jeux d'entrainement du SIA. En effet, ce dernier n'avait que « *picked up on the ways in which human beings used language* »[368].

iii. Les biais créés par l'algorithme et par le processus de collecte des renseignements

Les biais algorithmiques. En revanche, les biais ne sont pas qu'introduits par des jeux d'entrainement des SIA. Ainsi, les « *algorithmic bias* » sont des biais créés ou amplifiés par les SIA en raison de choix conceptuels[369]. Selon Mehrabi et al. :

> The algorithmic design choices, such as use of certain optimization functions, regularizations, choices in applying regression models on the data as a whole or considering subgroups, and the general use of statistically biased estimators in algorithms, can all contribute to biased algorithmic decisions that can bias the outcome of the algorithms.[370]

[363] *Id.*, 115:5.
[364] M. KEARNS et A. ROTH, préc., note 321, p. 56-57 (page pdf).
[365] *Id.*, p. 54.
[366] *Id.*
[367] *Id.*, p. 53-56.
[368] *Id.*, p. 70.
[369] Martínez-Plumed FERNANDO, Ferri CÈSAR, Nieves DAVID et Hernández-Orallo JOSÉ, « Missing the missing values: The ugly duckling of fairness in machine learning », (2021) 36-7 *International Journal of Intelligent Systems* 3217-3258, 2, DOI : 10.1002/int.22415.
[370] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:7.

Exemple de biais algorithmique : le « *User interaction bias* ». De façon similaire, des biais peuvent être introduits en raison des procédures de collecte des renseignements qui permettent l'entrainement des modèles de SIA[371]. On parle alors de « *user interaction bias »* soit des biais qui peuvent-être « *triggered from two sources - user interface and through the user itself by imposing his/her self-selected biased behavior and interaction* »[372]. Par exemple, les procédures de collecte qui se fondent sur une interaction avec l'usager peuvent créer une forme de renforcement circulaire. C'est un biais assez courant chez les moteurs de recherche[373]. Ceux-ci priorisent les résultats de recherche que l'on risque de trouver pertinents ou intéressants en se basant sur nos recherches antérieures. Ce « *personal profile processing* »[374] permet au moteur de recherche de prioriser les résultats les plus pertinents pour l'utilisateur. C'est cette capacité qui permet, par exemple, au moteur de recherche de reconnaitre que son utilisateur juriste ne désire pas recevoir des résultats relatifs à la programmation lorsqu'il inscrit le terme « code » dans sa barre de recherche[375]. En revanche, parce que nos choix antérieurs influencent notre recherche actuelle il peut s'en suivre une forme de « *statistical ouroboros* »[376] qui peut nous empêcher d'accéder à certaines données importantes.

C. Les mesures favorisant une plus grande équité

Des mesures influencées par la conception d'équité désirée. Les mesures devant être mises en œuvre afin de produire des résultats plus équitables diffèrent selon la conception de l'équité recherchée. En effet, de façon générale, les mesures devant être mises en œuvre afin de favoriser (i) la parité de traitement sont distinctes (ii) de celles permettant d'acquérir une meilleure parité de performance ou une meilleure parité statistique.

---

[371] M.-P. FERNANDO, F. CÈSAR, N. DAVID et H.-O. JOSÉ, préc., note 369, 2.
[372] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:7.
[373] *Id.*
[374] Fons WIJNHOVEN et Jeanna VAN HAREN, « Search Engine Gender Bias », (2021) 4 *Frontiers in Big Data*, 1, en ligne : <https://www.frontiersin.org/articles/10.3389/fdata.2021.622106> (consulté le 14 août 2022).
[375] *Id.*
[376] K. CRAWFORD, préc., note 332, p. 131.

i. Les mesures permettant d'acquérir une parité de traitement

Limiter la collecte des attributs protégés. Sans surprise la poursuite d'une parité de traitement réclame de limiter sa collecte et son usage d' attributs protégés comme l'origine ethnique, la couleur de peau, le genre ou l'orientation sexuelle.

Ne pas utiliser et collecter les « proxy ». Or, les mesures proposées par les partisans de cette forme d'équité ne se limitent pas toujours aux renseignements qui concernent directement et explicitement un motif de discrimination prohibé. Plusieurs proposent, en effet, de limiter l'utilisation et la collecte de renseignements qui pourraient agir comme « proxys » à ce type de renseignements[377].

Compatibilité avec le principe de sécurité. À ce titre, cette perception de l'équité et les mesures qui s'y rattachent sont particulièrement compatibles avec les mesures permettant d'assurer le respect du principe de sécurité. Toutes deux demandent, en effet, de minimiser sa collecte et son utilisation de renseignements personnels surtout s'ils sont sensibles[378]. Or, il sera démontré que les attributs protégés sont généralement considérés comme des renseignements personnels sensibles[379].

ii. Les mesures permettant d'acquérir une parité de performance ou une parité statistique

Mesures à implémenter afin d'assurer une parité de performance ou une parité statistique. Plusieurs mesures ont été proposées afin de développer des SIA capables de satisfaire une parité statistique ou une parité de performance. Il a été proposé, entre autres :

> de détecter si les SIA présentent des performances distinctes ou des résultats inéquitables entre différents groupes[380],
>
> d'utiliser des jeux d'entrainement diversifiés[381],

---

[377] Voir notamment : F. BONCHI, S. HAJIAN, B. MISHRA et D. RAMAZZOTTI, préc., note 340, 7; Kory D. JOHNSON, Dean P. FOSTER et Robert A. STINE, *Impartial Predictive Modeling and the Use of Proxy Variables*, *17th International Conference, iConference 2022*, Évènement virtuel, 7 janvier 2022, p. 3, DOI : 10.48550/arXiv.1608.00528; S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 8.

[378] Voir notamment : COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 307, p. 3.

[379] *Infra,* Chapitre 1, Section IV, Sous-section D(ii).

[380] M. KEARNS et A. ROTH, préc., note 321, p. 76 (page pdf); J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3.

[381] M. KEARNS et A. ROTH, préc., note 321, p. 70 (page pdf).

de modifier les jeux d'entrainement afin qu'ils soient moins biaisés[382],

de prévoir différents seuils selon le groupe d'appartenance d'un individu[383] et

d'utiliser des « algorithmes de mitigation (traduction libre) »[384] permettant de modifier les résultats ou les entrées (en anglais « *inputs* ») des SIA[385].

Précisions. Trois de ces mesures méritent d'être précisées soit : l'utilisation de jeux de données diversifiés, prévoir différents seuils selon le groupe d'appartenance d'un individu et l'utilisation d'algorithmes de mitigation.

Précisions sur les jeux d'entrainement diversifiés. Afin de s'assurer qu'un SIA respecte une parité de performance, il ne suffit pas toujours d'employer des jeux d'entrainement composés de groupes statistiquement représentatifs de la population. Tel que noté par Kearns et Roth :

> lorsque l'on maximise la précision sur plusieurs populations distinctes, un algorithme optimisera naturellement mieux pour la population majoritaire au détriment de la population minoritaire puisque, par définition, il y a plus de personnes dans le groupe majoritaire. Celles-ci contribuent donc plus à la précision globale du modèle. (traduction libre)[386]

L'Executive Office of the President of The United States National Science and Technology Council soulève également cette problématique et souligne que les niveaux de précisions d'IA peuvent diverger entre les groupes majoritaires et groupes minoritaires « *not because the input data is unrepresentative of the overall population, but simply because the minority group is less numerous*. »[387]

Distinctions entre les jeux « représentatifs » et les jeux « diversifiés ». À ce titre, il convient de distinguer les jeux représentatifs des jeux diversifiés. En l'espèce, nous identifions à titre de jeux diversifiés les jeux d'entrainement qui représentent de façon paritaire chaque différent groupe.

---

[382] N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, préc., note 132, 115:13.

[383] S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 8-9; FAIRLEARN 0.7.0, « Mitigation », en ligne : <https://fairlearn.org/v0.7.0/user_guide/mitigation.html> (consulté le 27 juin 2022).; S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 6.

[384] FAIRLEARN 0.7.0, « Frequently asked questions », en ligne : <https://fairlearn.org/v0.7.0/faq.html> (consulté le 22 mai 2022).

[385] M. KEARNS et A. ROTH, préc., note 321, p. 67-68 (page pdf).

[386] *Id.*, p. 70 (page pdf).

[387] HOLDREN JOHN P., AFUA BRUCE, ED FELTEN, TERAH LYONS, et MICHAEL GARRIS, *Preparing for the Future of Artificial Intelligence*, Washington, D.C., Executive Office of the President of The United States National Science and Technology Council Committee on Technology, 2016, p. 31.

Par exemple, si 1000 Américains blancs peuplent le jeu d'entrainement, on doit aussi retrouver dans le jeu d'entrainement des données appartenant à 1000 Afro-Américains. Par jeux représentatifs, nous entendons des jeux dans lesquels les groupes sont distribués d'une façon correspondant à leur poids démographique. Si le SIA opère auprès d'une population composée à 30 % d'Afro-Américains, il faut retrouver 30 % d'Afro-Américains dans le jeu d'entrainement.

Définir « Modifier les seuils ». Ensuite certains auteurs soulignent qu'il est également possible de prévoir un seuil distinct pour les personnes appartenant à des groupes discriminés[388]. Reprenons l'exemple de COMPAS afin d'illustrer cette mesure. Ainsi, COMPAS assigne à une personne évaluée un score entre 1 et 10 qui illustre son risque de commettre un crime[389]. Un score élevé représente un risque élevé. Modifier les seuils signifie, par exemple, de prévoir la détention d'une personne blanche si elle présente un score de 4 ou plus, mais de prévoir la détention d'un Afro-Américain seulement s'il présente un score de 6 ou plus[390].

Définir « algorithmes de mitigation ». Enfin, un développeur désireux de respecter une parité statistique ou une parité de performance peut utiliser différents algorithmes de mitigation. Ces algorithmes modifient les résultats ou les entrées d'un SIA afin de s'assurer que ce dernier produise des résultats susceptibles de respecter une parité statistique ou une parité de performance[391]. Par exemple, Fairlearn, une initiative de Microsoft qui fournit une boite à outils aux développeurs de SIA, propose quatre types d'algorithmes de mitigation[392]. L'un de ces algorithmes s'intitule « *CorrelationRemover* »[393]. Ce dernier tente d'éliminer les corrélations entre les variables d'entrées et certains attributs protégés[394]. En d'autres termes, le

[388] M. KEARNS et A. ROTH, préc., note 321, p. 68-69 (page pdf); S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 1.
[389] S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 6.
[390] *Id.*, p. 1.
[391] M. KEARNS et A. ROTH, préc., note 321, p. 67-68 (page pdf).
[392] FAIRLEARN 0.7.0, préc., note 384.
[393] FAIRLEARN 0.8.0, « CorrelationRemover visualization », en ligne : <https://fairlearn.org/main/auto_examples/plot_correlationremover_before_after.html> (consulté le 8 juillet 2022); FAIRLEARN 0.7.0, « fairlearn.preprocessing package », en ligne : <https://fairlearn.org/v0.7.0/api_reference/fairlearn.preprocessing.html#fairlearn.preprocessing.CorrelationRemover> (consulté le 29 août 2022).
[394] FAIRLEARN 0.8.0, préc., note 393; FAIRLEARN 0.7.0, préc., note 393.

« *CorrelationRemover* » est « *a component that filters out sensitive correlations in a dataset*. »[395]. Le « *adversarial debiasing* »[396] est un autre exemple d'algorithme de mitigation. Celui-ci implique d'utiliser deux modèles de SIA en opposition soit un prédicteur et un adversaire[397]. Dans ce jeu à deux joueurs, le prédicteur tente d'optimiser la précision du modèle dans sa recherche de la valeur cible alors que l'adversaire tente de prédire à quel groupe le résultat appartient[398]. Ainsi, pour un SIA qui tente, par exemple, de prédire les « bons » candidats à un poste, l'adversaire tentera d'identifier si le candidat privilégié par l'algorithme est un homme ou une femme. L'objectif du prédicteur est de « tromper » l'adversaire tout en maintenant les meilleures prédictions possibles[399].

D. La traduction légale du principe en droit québécois

Traduction légale du principe. Des dispositions se rattachant au respect du principe d'équité sont présentes en droit québécois. En effet, d'une part, (i) des instruments de protection des droits et libertés fondamentales prohibent la discrimination sur divers motifs. D'autre part, (ii) une protection est conférée par notre régime de protection des renseignements personnels qui limite la collecte et l'utilisation d'attributs protégés.

i. Les protections conférées par les chartes

La Charte québécoise. Au Québec, le respect du principe d'équité est en partie assuré par des instruments qui n'adressent pas spécifiquement la protection des renseignements personnels tels que la *Charte québécoise*. Ainsi, l'article 10 de la *Charte québécoise* prévoit que[400] :

> Toute personne a droit à la reconnaissance et à l'exercice, en pleine égalité, des droits et libertés de la personne, sans distinction, exclusion ou préférence fondée sur la race, la couleur, le sexe, l'identité ou l'expression de genre, la grossesse, l'orientation sexuelle, l'état civil, l'âge sauf dans la mesure prévue par la loi, la religion, les convictions politiques, la langue, l'origine ethnique ou nationale, la condition sociale, le handicap ou l'utilisation d'un moyen pour pallier ce handicap.

---

[395] FAIRLEARN 0.7.0, préc., note 393.
[396] Brian Hu ZHANG, Blake LEMOINE et Margaret MITCHELL, *Mitigating Unwanted Biases with Adversarial Learning*, *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society.*, New York, 2018, p. 336.
[397] *Id.*
[398] *Id.*, p. 336 et 339.
[399] *Id.*, p. 339.
[400] *Charte des droits et libertés de la personne*, préc., note 14. art 10.

> Il y a discrimination lorsqu'une telle distinction, exclusion ou préférence a pour effet de détruire ou de compromettre ce droit. [401]

Interdiction de discriminer lors de l'embauche. La *Charte québécoise* prévoit donc une interdiction de discriminer sur la base de plusieurs attributs protégés. La charte interdit, par exemple, de discriminer lors d'un processus d'embauche puisqu'elle prévoit que :

> Nul ne peut exercer de discrimination dans l'embauche, l'apprentissage, la durée de la période de probation, la formation professionnelle, la promotion, la mutation, le déplacement, la mise à pied, la suspension, le renvoi ou les conditions de travail d'une personne ainsi que dans l'établissement de catégories ou de classifications d'emploi.[402]

Interdiction de discriminer lors de la formation de contrats. De façon similaire la charte prévoit que « nul ne peut par discrimination, refuser de conclure un acte juridique ayant pour objet un bien ou un service ordinairement offert au public »[403]. Pour la CDPDJ cet article signifie que « le recours à un SIA ne devrait donc pas entrainer un refus discriminatoire de contracter. »[404]

Interdiction de créer une distinction, une exclusion ou une préférence qui compromet l'exercice d'un droit. De plus, la CDPDJ identifie que :

> si un SIA entraine 1) une distinction, exclusion ou préférence 2) fondée sur un motif prohibé de discrimination 3) qui a pour effet de détruire ou compromettre l'exercice d'un droit par ailleurs protégé par la Charte, un constat de discrimination à première vue pourra être établi[405].

En pareilles circonstances, il reviendra à l'utilisateur ou au développeur du SIA de « démontrer que la distinction, exclusion ou préférence n'est pas contraire à l'égalité envisagée dans une perspective réelle ou encore qu'une disposition spécifique de la Charte l'autorise à appliquer ladite distinction »[406].

La Charte canadienne. Ensuite, la *Charte canadienne des droits et libertés* (ci-après « *Charte canadienne* »), qui s'applique aux entités publiques, prohibe la discrimination fondée sur une liste

---

[401] *Id.*, art. 10.
[402] *Id.*, art. 16.
[403] *Charte des droits et libertés de la personne*, préc., note 14. art 12. ; J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 15.
[404] J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 15.
[405] *Id.*, p. 10.
[406] *Id.*

non exhaustive de motifs parmi lesquels : la race, l'origine nationale ou ethnique, la couleur, la religion, le sexe, l'âge et les déficiences mentales ou physiques[407].

Ne pas se limiter à l'intention. Au Canada et au Québec, la discrimination s'évalue sans égard à l'intention[408]. En effet, lorsque l'on évalue la violation d'une disposition interdisant la discrimination :

> [l]a question n'est pas de savoir si la discrimination est intentionnelle ou si elle est simplement involontaire, c'est-à-dire découlant du système lui-même. Si des pratiques occasionnent des répercussions néfastes pour certains groupes, c'est une indication qu'elles sont peut-être discriminatoires[409].

Pour cette raison, la CDPDJ soutient qu'il importe « d'évaluer l'impact des SIA non pas uniquement en fonction de leurs objectifs ou des fins poursuivies, mais également en fonction des résultats effectifs qui découlent de leur utilisation »[410].

Imposition d'une parité statistique et d'une parité de performance. Les positions susmentionnées énoncées par la CDPDJ, la principale institution publique québécoise chargée de combattre la discrimination, impliquent que la poursuite d'une parité statistique ou d'une parité de performance pourra être imposée à des SIA dans certaines circonstances[411].

L'affaire *Ewert*. Cette analyse est cohérente avec les propos tenus par la Cour suprême dans l'arrêt *Ewert.* Cet arrêt, tel qu'identifié, a permis de contester l'application d'un outil d'évaluation des risques à l'égard d'Autochtones au motif qu'il ne produisait potentiellement pas des résultats « exacts » à leur égard[412]. Intuitivement, cet arrêt renvoie à une nécessité de promouvoir une forme de parité de performance. En revanche, il convient de noter que dans *Ewert* la Cour suprême n'a pas conclu à la présence d'une discrimination à l'égard des populations autochtones. En effet, selon la Cour, il n'était pas possible de conclure à cette discrimination en raison de l'absence de preuve que :

---

[407] *Loi constitutionnelle de 1982, Annexe B de la Loi de 1982 sur le Canada (R-U)*, c 11, art. 15(1) [*Charte canadienne*], en ligne : <https://canlii.ca/t/q3x8>.
[408] J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 6.
[409] *Id.*, p. 9. citant CN c. Canada (Commission canadienne des droits de la personne), [1987] 1 R.C.S. 1114, 1139.
[410] *Id.*, p. 8.
[411] *Charte des droits et libertés de la personne*, préc., note 14, arts. 57, 71 et 77-85.
[412] *Supra,* Chapitre 1, Section II, Sous-section B.

> les outils contestés surestiment effectivement le risque posé par les détenus autochtones ou mènent à des conditions d'incarcération plus sévères ou à la privation de possibilités de réadaptation en raison d'une telle surévaluation[413].

En l'espèce, il était uniquement possible d'établir qu'il existait un risque que les outils contestés produisent des « résultats moins exacts à l'égard des autochtones qu'à l'égard des autres détenus »[414]. Le SCC était conscient de ces risques, mais avait « continué de se fier à ces résultats pour prendre des décisions au sujet des délinquants sans examiner leur validité à l'égard des Autochtones »[415]. Cette détermination était suffisante pour conclure à une violation d'une obligation de « veiller » à l'exactitude des renseignements utilisés par le SCC, mais pas pour conclure à l'existence d'une discrimination[416]. Toutefois, malgré n'avoir pas identifié que les agissements du SCC étaient discriminatoires, les passages susmentionnés suggèrent que l'usage d'outils prédictifs aurait pu être déclaré discriminatoire si le demandeur avait fait la preuve qu'ils surestimaient effectivement les risques à l'égard des membres de groupes protégés. À ce titre, il est très clair qu'aux yeux de la Cour suprême, la *Charte canadienne* réclame d'utiliser des outils capables de respecter une forme de parité de performance sous certaines circonstances.

ii. Les protections conférées par le régime de protection des renseignements personnels

Les attributs protégés sont souvent des renseignements sensibles. Outre les chartes, le principe d'équité est également défendu par les protections accordées par le régime de protection des renseignements personnels. En effet, tel qu'identifié, le Québec confère une protection accrue aux renseignements sensibles[417]. Ce type de renseignements est plus difficile à utiliser puisque leur usage peut réclamer un consentement exprès de la personne concernée. De plus, il est plus difficile de prouver la nécessité de leur collecte et de leur utilisation[418]. La personne qui utilise ce type de renseignement s'expose également à des sanctions plus sévères en cas d'incidents de confidentialité[419]. Or, la CAI et le CVPC ont souvent considéré comme « sensibles » des

[413] *Ewert c. Canada*, préc., note 151, par. 79.
[414] *Id.*
[415] *Id.*, par. 80.
[416] *Id.*
[417] *Supra,* Chapitre 1, Section III, Sous-sections A et C.
[418] *PL64*, préc., note 5, arts. 13, 20 et 110. Nouveaux articles 59 et 65.1 de la *Loi sur l'accès* et 12 et 13 de la *Loi sur le privé*.
[419] *Id.*, arts. 69, 159 et 160. Nouveaux articles 160 de la *Loi sur l'accès* et 90.2 et 92.3 de la *Loi sur le privé.*

renseignements se rapportant à des motifs de discrimination identifiés par les chartes. Par exemple, ont été qualifiés de sensibles les renseignements rattachés à l'origine ethnique ou raciale[420], aux opinions politiques[421], à l'orientation sexuelle[422], aux croyances religieuses[423] et à la santé mentale[424].

Nuance. Cependant, les obligations se rattachant aux renseignements sensibles ne s'inscrivent pas strictement dans une logique d'équité. En effet, les renseignements personnels se rattachant à des motifs de discrimination ne sont pas les seuls renseignements considérés sensibles. Par exemple, les renseignements financiers[425], les portraits du visage[426] et certaines données de géolocalisation[427] sont également considérés sensibles. Cette réalité contraste largement avec le RGPD qui considère à titre de « catégories particulières de données à caractère personnel »[428],

---

[420] Reconnu comme un motif de discrimination par les deux chartes. *Charte des droits et libertés de la personne*, préc., note 14, art. 10; *Charte canadienne*, préc., note 407, art. 15(1); Commission d'accès à l'information du Québec, 14 mars 2018, 111494, *X c. Ministère de la Santé et des Services sociaux et Bibliothèque des archives nationales du Québec et Société de généalogie canadienne-française et Société de généalogie de Québec et Institut généalogique Drouin*, par. 141-142, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/350960/index.do> (consulté le 7 juin 2022); COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 260.

[421] Les convictions politiques sont reconnues comme un motif de discrimination par la *Charte québécoise*. *Charte des droits et libertés de la personne*, préc., note 14, art. 10; COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 260.

[422] Reconnu comme un motif de discrimination par les deux chartes (à la suite de l'arrêt *Egan* dans le cadre de la *Charte canadienne*). *Charte canadienne*, préc., note 407, art. 15; *Egan c. Canada*, [1995] 2 RCS 513, 514-515 (Cour suprême du Canada), en ligne : <https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/1265/index.do> (consulté le 29 août 2022); *Charte des droits et libertés de la personne*, préc., note 14, art. 10; COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 260.

[423] La religion est un motif de discrimination reconnue par les deux chartes. COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 260; Commission d'accès à l'information du Québec, 14 mars 2018, *X c. Ministère de la Santé et des Services sociaux et Bibliothèque des archives nationales du Québec et Société de généalogie canadienne-française et Société de généalogie de Québec et Institut généalogique Drouin*, préc., note 420, par. 141-142; *Charte des droits et libertés de la personne*, préc., note 14, art. 10; *Charte canadienne*, préc., note 407, art. 15(1).

[424] Reconnu comme un motif de discrimination par les deux chartes. *Charte des droits et libertés de la personne*, préc., note 14, art. 10; *Charte canadienne*, préc., note 407, art. 15(1); Commission d'accès à l'information, 6 janvier 2022, *Plainte à l'endroit de la Fondation Vipassana de l'Est du Canada*, par. 4, 20 et 21, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/500008/index.do> (consulté le 7 juin 2022).

[425] *Loi sur la protection des renseignements personnels et les documents électroniques*, préc., note 256. art 4.3.4. (Annexe 1).

[426] Commissariat à la protection de la vie privée du Canada, 22 août 2016, *Enquête conjointe sur Ashley Madison*, préc., note 264, par. 74; Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 3-4.

[427] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « Pointage électronique des employés par géolocalisation », en ligne : <https://www.cai.gouv.qc.ca/pointage-electronique-employes-par-geolocalisation-regles-a-respecter/> (consulté le 8 juin 2022).

[428] *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces*

aussi nommées « données sensibles »[429], des données qui sont presque exclusivement rattachées à des motifs de discrimination[430]. En fait, les débats parlementaires ayant mené à l'adoption de PL64 révèlent que l'intention du législateur n'était pas de rattacher la notion de sensibilité des renseignements aux motifs interdits de discrimination. Ainsi, les mémoires soumis par quelques organisations, dont le Barreau du Québec et la CDPDJ lors des travaux parlementaires portant sur PL64, invitèrent le gouvernement à préciser la notion de sensibilité en le rattachant aux motifs de discrimination[431]. Ces propositions n'ont toutefois pas été retenues par le législateur. De plus, l'un des co-auteurs du projet de loi avait même, lors des débats parlementaires, appelé à ne pas confondre la notion de « motifs interdits » à celle de « renseignements sensibles » en précisant que : « la sensibilité d'une information et le fait que […] ça pourrait potentiellement être un motif de discrimination sont deux choses différentes. »[432]

Une absence de clarté dans les obligations. Bref, le droit québécois prévoit plusieurs obligations se rattachant au principe d'équité et de non-discrimination. En revanche, ces obligations ne sont pas toujours imposées par le régime de protection des renseignements personnels en soi. C'est surtout par l'entreprise des chartes que le respect du principe pourra être protégé en droit québécois. Or, bien que les obligations prévues par ces instruments s'appliquent très clairement dans certains contextes, comme un processus d'embauche[433] ou l'évaluation de risques de récidives[434], il ne sera pas toujours évident d'identifier dans quelles circonstances précises les obligations prévues par les chartes s'appliqueront. Par exemple, la *Charte québécoise* prévoit qu'il peut y avoir une discrimination lorsqu'une distinction, une exclusion ou une préférence qui se fonde sur un motif de discrimination reconnu compromet l'exercice d'un droit protégé par

---

données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données), (2018) Journal officiel de l'Union européenne [*RGPD*]. art 9.

[429] *Id.*, considérant no 10.

[430] *Id.*, art 9.

[431] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 82; BARREAU DU QUÉBEC, préc., note 5, p. 20.

[432] ASSEMBLÉE NATIONALE DU QUÉBEC, « Étude détaillée du projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », (21 février 2021) 45-120 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210217.html> (consulté le 30 janvier 2022).

[433] Voir la discussion se rapportant à la Charte québécoise (les obligations rattachées à son article 16). *Id.*

[434] Voir la discussion se rapportant à la décision *Ewert* de la Cour suprême de la présente section. *Supra,* Chapitre 1, Section IV, Sous-section D(i).

celle-ci[435]. En revanche, certains droits protégés par cette charte sont parfois assez nébuleux. Par exemple, la Cour suprême a, dans l'affaire *Ward*, qualifié le droit à la sauvegarde de la dignité[436] comme « un droit dont la portée est particulièrement difficile à cerner »[437]. De façon similaire, il existe, tel qu'identifié, plusieurs conceptions de l'équité[438]. Or, celles-ci, nous le verrons, sont très souvent en tension et il n'est généralement pas possible d'assurer un respect conjoint de toutes les conceptions pour un même SIA[439]. À ce titre, il ne sera pas toujours clair dans quelles circonstances le principe s'appliquera pour un SIA et quelle conception de l'équité devra être privilégiée.

[435] J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 10.

[436] *Charte des droits et libertés de la personne*, préc., note 14, art. 4.

[437] Il peut paraitre étrange de proposer qu'un SIA puisse porter atteinte à la dignité d'une personne. En revanche, il faut comprendre que les SIA produisent des résultats sans compréhension des contextes sociohistoriques, des dynamiques de pouvoirs et des stéréotypes qui affectent certaines populations. Par exemple, en 2015 une application de Google qui utilisait un SIA, « *Google photos* », a fait controverse lorsqu'il a été révélé qu'il reconnaissait erronément les personnes noires comme des « gorilles ». À ce titre, il n'est pas impossible de voir apparaitre des résultats de SIA qui pourraient compromettre la « sauvegarde de la dignité » de membres de groupes marginalisés.
Ward c. Québec (Commission des droits de la personne et des droits de la jeunesse), [2021] 43 CSC, par. 48-49 (Cour suprême du Canada), en ligne : <https://canlii.ca/t/jk1tm> (consulté le 29 août 2022); « Google apologises for Photos app's racist blunder », *BBC News*, sect. Technology (1 juillet 2015), en ligne : <https://www.bbc.com/news/technology-33347866> (consulté le 29 août 2022).

[438] *Supra,* Chapitre 1, Section IV, Sous-section A.

[439] *Supra,* Chapitre 2, Sections IV et V.

Chapitre 2 - La problématique : Exploration de tensions entre les principes

Structure du second chapitre. Dans ce second chapitre, nous explorons différentes tensions entre les principes susmentionnés. Ainsi, alors que le premier chapitre explore le paysage normatif imposé aux SIA, le second identifie comment les différentes mesures devant être adoptées sont en tensions. En effet, les mesures devant être mises en œuvre afin d'assurer le respect de ces principes se chevauchent et contrarient très souvent le respect d'autres principes. Ainsi, nous observons des tensions opposant (I) le principe d'explicabilité au principe d'exactitude, (II) le principe d'exactitude au principe de sécurité et (III) le principe d'exactitude au principe d'équité. Nous observons également des tensions (IV) entre les mesures devant être mises en œuvre afin de satisfaire le principe de sécurité ainsi que la parité de traitement et les mesures devant être mises en œuvre afin d'assurer la parité statistique et la parité de performance. Finalement, nous exposons les tensions qui existent entre deux conceptions du principe d'équité soit (V) la parité de performance et la parité statistique.

Section I - Les tensions opposant l'explicabilité à l'exactitude

Les trois sources de tension entre l'explicabilité et l'exactitude. La relation entre les principes d'explicabilité et d'exactitude est tendue. En effet, le Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle de la Commission européenne reconnait que « des arbitrages peuvent s'avérer nécessaires entre le renforcement de l'explicabilité d'un système (qui pourrait réduire sa précision) et l'amélioration de sa précision (au détriment de l'explicabilité). »[440] En l'espèce, nous identifions trois sources de tensions soit : (A) le fait qu'il puisse être nécessaire de maintenir une certaine opacité afin de conserver l'exactitude des SIA, (B) la nécessité d'utiliser de grandes quantités de données afin d'obtenir des résultats exacts et (C) le fait que les modèles les plus difficiles à comprendre et à expliquer sont souvent les plus précis.

[440] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 22.

A. La nécessité de maintenir une opacité afin de conserver l'exactitude

L'opacité intentionnelle comme source de tensions. Tel que précisé, l'opacité des SIA est en partie volontaire et résulte, entre autres, d'un désir des entreprises de maintenir leurs avantages compétitifs sur leurs concurrents[441]. En revanche, il serait simpliste de proposer qu'il s'agisse de l'unique raison pour laquelle les entreprises agissent ainsi. L'opacité intentionnelle des SIA est parfois liée à un désir de promouvoir une plus grande exactitude.

Environnements dynamiques. En effet, pas tous les SIA évoluent dans des environnements dits statiques. Certains opèrent dans des environnements dynamiques où ils sont appelés à modifier leurs comportements en fonction des agissements des usagers ou d'acteurs malveillants. Tel est le cas, par exemple, des filtres anti-pourriel et des systèmes de détection d'intrusions. Ces SIA jouent un « *game of cat-and-mouse* »[442] avec les criminels puisque ces derniers développent continuellement de nouvelles méthodes qui visent à contourner les capacités de détection mises en place par les SIA[443]. De façon similaire, des plateformes comme YouTube ou Google luttent continuellement contre des utilisateurs désireux de tirer profit des comportements de leurs algorithmes[444]. Ces mauvais joueurs vont tenter de « jouer le système » (en anglais « *game the system* »[445]) pour tromper l'algorithme afin que ce dernier puisse les avantager au détriment de leurs compétiteurs[446].

Le prix de la transparence : la perte d'exactitude dans des environnements dynamiques. En pareilles circonstances, offrir trop d'informations et de détails sur les algorithmes utilisés facilite la tâche à ces criminels et mauvais joueurs. En effet, les modèles transparents sont « plus faciles

---

[441] *Supra*, Chapitre 1, Section I, Sous-Section A (i).
[442] Christian SANDVIG, Kevin HAMILTON, Karrie KARAHALIOS et Cedric LANGBORT, *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, *Data and Discrimination: Converting Critical Concerns into Productive Inquiry, a preconference at the 64th Annual Meeting of the International Communication Association*, Seattle, États-Unis, 2014, p. 23, p. 7, en ligne : <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>.
[443] *Id.*, p. 9. ; A. DEY, M. VELAY, J.-P. FAUVELLE et S. NAVERS, préc., note 278, p. 2.
[444] C. SANDVIG, K. HAMILTON, K. KARAHALIOS et C. LANGBORT, préc., note 442, p. 9.
[445] Jane BAMBAUER et Tal ZARSKY, « The Algorithm Game », (2018) 94-1 *Notre Dame Law Review* 1, 1.
[446] *Id.*; C. SANDVIG, K. HAMILTON, K. KARAHALIOS et C. LANGBORT, préc., note 442, p. 9.

à manipuler, ce qui réduit leur pouvoir prédictif (traduction libre) »[447]. Révéler, par exemple, l'entièreté des termes considérés « suspects » par un filtre anti-pourriel permet aux fraudeurs de modifier leurs pratiques et d'éviter d'utiliser de tels termes. À ce titre, Burrell note que « les applications d'apprentissage automatique […] qui traitent explicitement des pourriels, des escroqueries et de la fraude doivent être opaques (traduction libre) (nos soulignements) »[448]. Sans cette opacité, ces SIA ne pourront pas différencier avec exactitude, par exemple, les courriels légitimes des pourriels frauduleux[449].

Le prix de la transparence : la perte d'exactitude et la perte de sécurité. La même situation s'observe pour les SIA utilisés afin de détecter les intrusions et les anomalies. Ceux-ci constituent, tel qu'identifié, des mesures de sécurité permettant de protéger des renseignements personnels[450]. Or, les SIA utilisés afin de détecter les intrusions et les anomalies dans les réseaux informatiques sont tous confrontés à la possibilité qu'un adversaire tente de leurrer le système de défense en tentant d'imiter un comportement « normal » qui ne sera pas détecté par le SIA[451]. En revanche, leurrer un SIA est beaucoup plus simple lorsque l'adversaire connait le mécanisme de défense mis en place[452]. À ce titre, non seulement l'explicabilité peut menacer l'exactitude des SIA, mais en plus, lorsque ces SIA sont mobilisés afin d'assurer une protection des renseignements personnels, leur perte d'exactitude affecte aussi le respect du principe de sécurité.

B. Les tensions créées par l'immense quantité de variables utilisées

Opacité née de la quantité de variables. Ensuite, comme il a été proposé, développer des SIA exacts réclame d'utiliser des jeux d'entrainement suffisamment étendus[453]. Or, comme nous l'avons identifié, la présence d'un nombre élevé de variables contribue à l'opacité des SIA[454]. En

[447] Smitha MILLI, Ludwig SCHMIDT, Anca D. DRAGAN et Moritz HARDT, *Model Reconstruction from Model Explanations*, *Proceedings of the Conference on Fairness, Accountability, and Transparency*, coll. FAT* '19, New York, NY, USA, Association for Computing Machinery, 29 janvier 2019, p. 1-9, p. 1, DOI : 10.1145/3287560.3287562.
[448] J. BURRELL, préc., note 33, 4.
[449] *Id.*
[450] *Supra,* Chapitre 1, Section III, Sous-section D.
[451] A. DEY, M. VELAY, J.-P. FAUVELLE et S. NAVERS, préc., note 278, p. 2.
[452] *Id.*
[453] *Supra,* Chapitre 1, Section II, Sous-section A.
[454] *Supra,* Chapitre 1, Section I, Sous-section A (iii).

effet, « *the more data fed into the algorithm as input makes the output that much harder to explain* »[455] puisque chaque nouvelle variable va « subtilement et imperceptiblement (traduction libre) »[456] modifier les résultats du SIA. À ce titre, la seule masse de variables considérées par le SIA peut créer ce que Burrell qualifie de « *opacity as the complexity of scale* »[457] en ce que le « *number of possible features to include in a classifier rapidly grows way beyond what can be easily grasped by a reasoning human* »[458]. Il semble, en effet, particulièrement difficile d'expliquer de façon compréhensible une décision qui résulte de relations entre des centaines voire des milliers de variables qui sont, tel qu'identifié, souvent sélectionnées et pondérées de façon contre-intuitive par le SIA[459].

Variables et pondérations contre-intuitives. De surcroit, comme il a été mentionné[460], c'est très souvent le SIA et non ses développeurs qui identifie lui-même les variables pertinentes à son analyse et qui pondère leur influence relative sur le processus décisionnel[461]. En fait, l'auteur Robbin propose que c'est justement dans de telles circonstances que les outils d'AA trouvent leur pertinence puisque :

> If we already know which considerations are acceptable, then there is no reason to use ML in the first place. We could simply hard-code the considerations into an algorithm—giving us an automated decision using pre-approved, transparent, reasoning.[462]

L'approche québécoise : Limiter une explication aux principales variables. Or, le nouveau régime québécois des renseignements personnels limite substantiellement l'explication devant être offerte aux personnes visées par une décision fondée exclusivement sur un traitement automatisé. L'explication devant être accordée à la personne concernée par une telle décision ne demande de révéler que les renseignements personnels utilisés pour prendre la décision, les « raisons et principaux facteurs et paramètres ayant mené à la décision (nos soulignements) »[463]

[455] S. ROBBINS, préc., note 49, 507.
[456] J. BURRELL, préc., note 33, 9.
[457] *Id.*
[458] *Id.*
[459] *Supra,* Chapitre 1, Section I, Sous-section A(iii).
[460] *Id.*
[461] S. ROBBINS, préc., note 49, 509-510; P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[462] S. ROBBINS, préc., note 49, 509-510.
[463] *PL64*, préc., note 5, arts. 21 et 110.Les articles 65.2 de la *Loi sur l'accès* et 12.1 de la *Loi sur le privé.*

ainsi que « son droit de faire rectifier les renseignements personnels utilisés pour rendre la décision. »[464]. Tel qu'identifié, cette obligation peut être interprétée comme exigeant de révéler les renseignements personnels utilisés et les variables les plus influentes au processus décisionnel ainsi que leur pondération respective[465]. À ce titre, on pourrait penser que le régime contourne la problématique susmentionnée en limitant l'explication aux éléments les plus importants.

PL64 prévoit-il une explication utile? Le problème, c'est qu'en limitant l'explication à l'étude des principales variables, celle-ci devient potentiellement inutile. Dans l'exemple de Burrel susmentionné, une explication limitée aux principales variables utilisées ne permettait pas de comprendre les décisions prises par le SIA[466] puisque la présence ou l'absence d'une seule variable peut modifier complètement l'analyse[467]. Elle ne permettait pas non plus de comprendre pourquoi ces variables ont été choisies et les raisons qui sous-tendent leur pondération[468]. À ce titre, il est possible de se questionner quant à la capacité des nouvelles dispositions prévues par PL64 de fournir une explication véritablement utile pour contester la décision à moins, bien sûr, que le SIA visé n'utilise qu'un nombre très peu élevé de variables.

Tensions dans le choix de l'explication. À ce titre, une autre source de tension s'observe entre la nécessité de fournir une explication compréhensible qui permet à une personne de comprendre le processus décisionnel auquel elle a été soumise et celle de fournir une explication suffisamment rigoureuse qui lui permet de contester adéquatement la décision qui l'affecte. Une explication compréhensible, telle qu'une exposition se limitant aux principales variables utilisées, permet de garantir certains aspects du principe d'explicabilité. Ce type d'explication, qui semble être privilégié par le législateur québécois, répond, en partie au désir de renforcer l'autonomie des personnes concernées par les décisions automatisées puisqu'elle leur permet de

[464] *Id.* ; *Supra,* Chapitre 1, Section I, Sous-section C
[465] *Supra,* Chapitre 1, Section I, Sous-section C. Dans la mesure où ces variables ne sont pas des renseignements personnels bien sûr. Il est possible que toutes les variables qui sont des renseignements personnels devront être explicitées.
[466] *Supra,* Chapitre 1, Section I, Sous-section A(iii).
[467] J. BURRELL, préc., note 33, 8-9.
[468] *Supra,* Chapitre 1, Section I, Sous-section A(iii).

comprendre, dans une certaine mesure, les décisions auxquelles elles sont soumises[469]. En revanche, fournir une telle explication limite la capacité de contester la décision et d'identifier les motifs qui permettraient d'en réclamer la révision. Or, le droit à la révision constitue, tel qu'identifié, l'une des principales mesures permettant de renforcer l'exactitude en droit québécois[470].

Une solution possible, mais imparfaite. À noter qu'il est possible pour un développeur de réduire le nombre de variables utilisées par un SIA en faisant usage de méthodes de « *feature selection* »[471]. Cette pratique permet de reconnaitre et de limiter l'analyse du SIA aux variables les plus « influentes »[472]. En réduisant ainsi le nombre de variables utilisées, il devient possible de fournir des explications à la fois compréhensibles et rigoureuses. Le « *feature selection* »[473] permet même de produire, à l'occasion, des SIA plus exacts en ce qu'elle représente une solution possible au « fléau de la dimension » (en anglais « *Curse of Dimensionality* »), un problème courant dans le développement des SIA[474]. Succinctement, le fléau de la dimension se manifeste lorsque « *an excessive number of features will result in matches being lost amongst all the non-matching data. This will mean that enormous volumes of data will be needed by way of compensation* »[475]. En revanche, utiliser le « *feature selection* » peut, dans certaines circonstances, limiter l'exactitude future des SIA puisque:

> One disadvantage of reducing the scope of feature selection is that one may lose possible matches, or patterns, that were not previously known or which had not been thought of[476].

[469] S. ROBBINS, préc., note 49, 496.
[470] *Supra*, Chapitre 1, Section I, Sous-section A(iii).
[471] Samina KHALID, Tehmina KHALIL et Shamila NASREEN, *A survey of feature selection and feature extraction techniques in machine learning*, *2014 Science and Information Conference*, Londres, Royaume-Uni, août 2014, p. 372-378, p. 1, DOI : 10.1109/SAI.2014.6918213.
[472] J. BURRELL, préc., note 33, 9. ; S. KHALID, T. KHALIL et S. NASREEN, préc., note 471, p. 1 et 3.
[473] S. KHALID, T. KHALIL et S. NASREEN, préc., note 471, p. 1.
[474] DATATILSYNET, préc., note 122, p. 11.
[475] *Id.*
[476] *Id.*

À ce titre, l'usage du « *feature selection* » peut être utile pour la création de SIA à la fois exacts et explicables[477]. Cependant, il ne sera pas toujours souhaitable d'agir ainsi si l'on désire produire des SIA exacts.

C. La nécessité de favoriser certains types de SIA au détriment d'autres

Privilégier les SIA explicables. Enfin, pas tous les SIA sont des « boites noires » qui présentent les mêmes défis en termes d'explicabilité[478]. Certains auteurs, tels que Rudin et Radin proposent donc de privilégier l'usage et le développement de modèles plus transparents et plus simples[479].

Nécessité d'un « *trade-off* » entre explicabilité et exactitude. Malheureusement, plusieurs auteurs ont identifié que, souvent, les modèles de SIA les plus interprétables sont également les moins précis[480]. Ainsi, Barredo Arrieta et al.[481], Adadi et Barrada[482] et Dam et al.[483] notent tous la nécessité d'effectuer un « *trade-off* »[484] entre l'interprétabilité d'un modèle et sa précision dans le choix de modèle de SIA. En fait, Barredo Arrieta et al.[485] et Dam et al.[486] illustrent à l'aide de tableaux une corrélation entre le niveau d'interprétabilité des modèles et leur exactitude (voir Figure 1 et Figure 2) :

[477] J. BURRELL, préc., note 33, 9. ; S. KHALID, T. KHALIL et S. NASREEN, préc., note 471, p. 1 et 3.
[478] C. RUDIN et J. RADIN, préc., note 69, 6.
[479] *Id.*, 1-3.
[480] A. BARREDO ARRIETA et al., préc., note 25, 31. ; H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2.
[481] A. BARREDO ARRIETA et al., préc., note 25, 31.
[482] A. ADADI et M. BERRADA, préc., note 71. figure 4.
[483] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2.
[484] A. BARREDO ARRIETA et al., préc., note 25, 31.
[485] *Id.*
[486] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2.

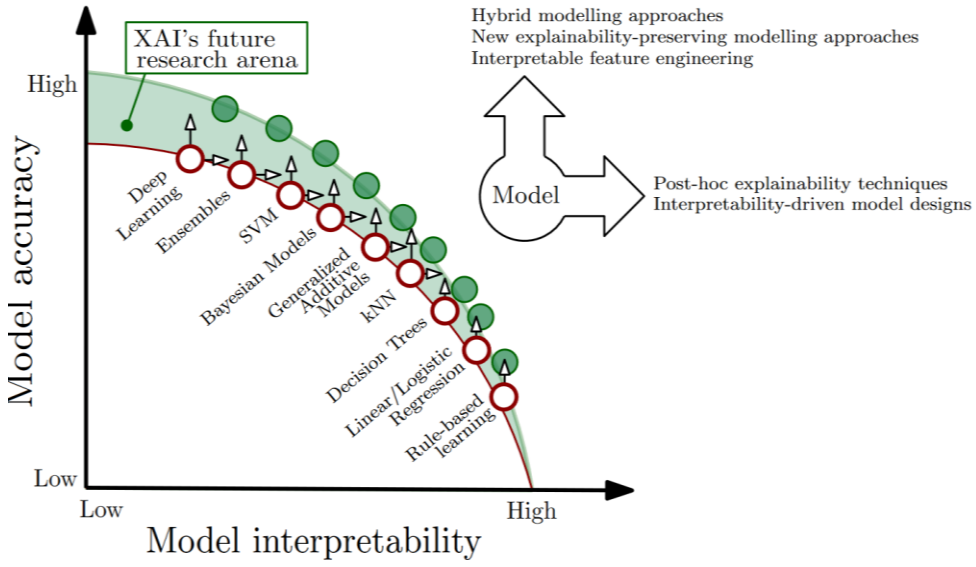

Figure 1. – Figure de Barredo Arrieta et al. (2020) intitulé : « *Trade-off between model interpretability and performance, and a representation of the area of improvement where the potential of XAI techniques and tools resides.* »[487]

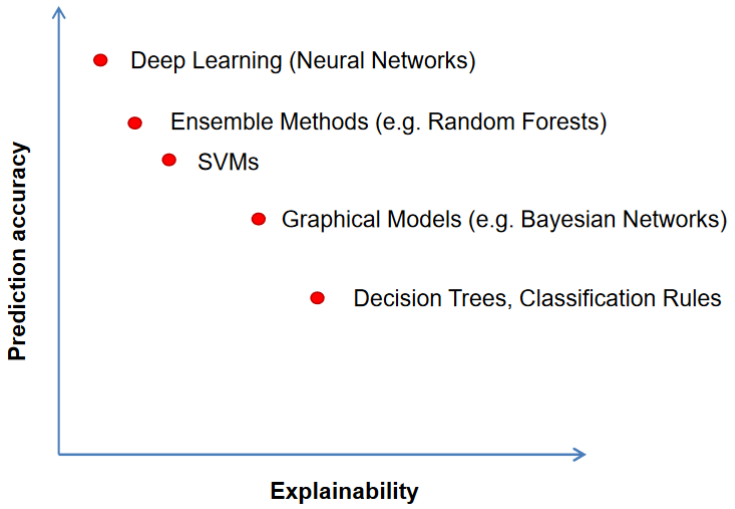

Figure 2. – Figure de Dam et al. (2018) intitulé : « *Prediction accuracy versus explainability* »[488]

Favoriser un modèle exact c'est sacrifier une alternative plus simple. Certains modèles de SIA sont donc plus précis que d'autres. Malheureusement, cette capacité à produire des résultats

[487] A. BARREDO ARRIETA et al., préc., note 25, 31.
[488] H. K. DAM, T. TRAN et A. GHOSE, préc., note 30, 2.

exacts s'accompagne généralement d'une complexité qui rend ces modèles difficiles à expliquer. À ce titre, une tension existe entre la volonté de favoriser des SIA explicables et celle de favoriser des SIA exacts.

Nuance : multiplicité de définitions. À noter toutefois que l'existence de cette corrélation est contestée dans la littérature. D'une part, tel qu'identifié, il n'existe pas de définition universellement acceptée sur ce qu'est un SIA explicable[489]. À ce titre, l'existence de cette corrélation dépend fortement du sens que l'on accorde à cette notion.

Nuance : des modèles plus simples peuvent parfois être aussi exacts. Des auteures telles que Rudin et Radin proposent également que même dans l'usage de réseaux neuronaux profonds qui sont, selon elles, « *the most difficult kind of black box model to explain* »[490], certaines contraintes peuvent être imposées afin d'assurer un processus plus transparent sans avoir à sacrifier la précision du modèle[491]. Surtout, ces auteures notent que, pour accomplir certaines tâches, des modèles plus transparents se sont révélés capables d'arriver à des résultats aussi précis ou plus précis que des « *black box model* »[492]. Ainsi, Rudin et Radin sont d'avis que « *[t]he belief that accuracy must be sacrificed for interpretability is inaccurate* »[493] et présentent le « *so-called accuracy–interpretability tradeoff* »[494] comme une fausse dichotomie[495]. De façon similaire, les auteurs Barredo et al., qui ont proposé la Figure 1 susmentionnée, ne proposent pas que le « *trade-off between performance and interpretability* » [496] soit toujours exact en toutes circonstances. Selon eux, le « *trade-off* » n'existe qu'en présence de certains facteurs très spécifiques. Par exemple, ils proposent que le « *trade-off* » puisse ne pas exister en présence de

[489] *Supra,* Chapitre 1, Section I, Sous-section A ; Voir aussi Zachary C. LIPTON, *The Mythos of Model Interpretability*, *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016),* New York, NY, USA, 6 mars 2017, p. 7, en ligne : <http://arxiv.org/abs/1606.03490> (consulté le 11 février 2022).
[490] C. RUDIN et J. RADIN, préc., note 69, 4.
[491] *Id.*, 4-5.
[492] *Id.*, 4-6. Voir aussi : J. DRESSEL et H. FARID, préc., note 85, 3.
[493] C. RUDIN et J. RADIN, préc., note 69, 3.
[494] *Id.*, 6-7.
[495] C. RUDIN et J. RADIN, préc., note 69.
[496] A. BARREDO ARRIETA et al., préc., note 25, 30.

circonstances dans lesquelles « *the data is well structured and features at our disposal are of great quality and value* »[497].

Réponse à la nuance. Il convient néanmoins de souligner qu'il n'est pas essentiel en l'espèce de prouver que le choix de modèle de SIA demande nécessairement de réaliser un arbitrage entre les principes d'exactitudes et d'explicabilité. Il suffit d'établir, pour nos fins, qu'une tension peut exister entre ces deux principes. Or, si Rudin et Radin soutiennent que les « modèles plus interprétables sont souvent plus (et non moins) précis (traduction libre) »[498] que les modèles plus complexes, elles n'argumentent pas que ceux-là sont toujours plus précis[499]. Bref, il convient de reconnaitre que le choix des modèles à utiliser doit bel et bien faire face à une certaine tension entre les principes d'exactitude et d'explicabilité, mais que cette tension ne se présente pas en toutes circonstances.

Nuance utile pour les développeurs, mais pas pour les utilisateurs. Qui plus est l'argumentaire de Rudin et Radin ne permet pas de solutionner la tension pour les utilisateurs de SIA. Comme elles le notent elles-mêmes, plusieurs développeurs ne tentent pas d'utiliser des modèles plus interprétables puisqu'ils présument qu'ils seront moins précis[500]. À ce titre, elles soutiennent que « *It is possible that an interpretable model can always be constructed—we just have not been trying* »[501]. Or, lorsqu'une organisation choisit d'utiliser un modèle, elle doit le faire au regard de ce qui est actuellement disponible et des ressources qu'elle possède. Un utilisateur de SIA ne dispose pas nécessairement des ressources nécessaires pour développer ses propres applications d'IA. En pareilles circonstances, ce dernier peut faire face à une situation où le modèle le plus précis disponible sur le marché n'est pas le modèle le plus facile à interpréter. Même dans une situation où il serait possible, dans l'absolu, de produire un modèle aussi performant tout en étant plus interprétable, l'utilisateur de SIA, lui, ne sait pas si cela est possible et doit choisir parmi les options qui lui sont disponibles présentement. À ce titre, même en présumant que la position

[497] *Id.*
[498] C. RUDIN et J. RADIN, préc., note 69, 7.
[499] *Id.*
[500] *Id.*, 6.
[501] *Id.*, 7.

de Rudin et Radin s'avère exacte, une organisation ou entreprise peut quand même être confrontée à cette tension entre les principes d'explicabilité et d'exactitude.

De multiples sources de tensions. Nous pouvons donc observer plusieurs sources de tensions entre les principes d'exactitude et d'explicabilité. La poursuite de l'exactitude réclame souvent, en effet, de déployer des outils disposant d'un certain niveau de complexité. Elle peut réclamer l'usage de modèles plus complexes, d'un grand nombre de variables et de maintenir intentionnellement une certaine opacité afin d'assurer la production de résultats exacts. L'explicabilité favorise, quant-à-elle, exactement l'inverse. Elle demande de privilégier des modèles simples qui utilisent peu de variables et de favoriser une approche plus transparente.

Quand favoriser l'un ou l'autre des principes ? Or, il convient de se questionner quant aux circonstances dans lesquelles un principe devra être privilégié au détriment de l'autre. Le problème, en l'espèce, c'est que les situations qui réclament de privilégier l'exactitude demandent généralement de privilégier, elles aussi, l'explicabilité. Ainsi, il est notable que le Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle de la Commission européenne, qui reconnait pourtant la nécessité d'effectuer des arbitrages entre ces deux principes[502], demande de respecter ceux-ci dans les mêmes circonstances : soit les situations dans lesquelles les SIA affectent de façon importante la vie des personnes concernées par ses décisions. Ainsi, le groupe d'experts souligne l'importance de pouvoir exiger « une explication appropriée du processus de décision du système d'IA »[503] dès lors qu'un SIA a « une incidence importance sur la vie des personnes »[504]. Toutefois, il reconnait également qu'un « niveau élevé de précision est particulièrement essentiel dans les situations où le système d'IA a une incidence directe sur des vies humaines »[505]. La source de ce paradoxe s'explique par le fait que ces deux principes visent à assurer la légitimité des décisions produites par les SIA. Tel qu'identifié, le principe d'explicabilité se rattache à un objectif de responsabilisation des acteurs qui déploient

[502] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 22.
[503] *Id.*, p. 22.
[504] *Id.*
[505] *Id.*, p. 21.

les outils d'IA afin d'assurer que ces derniers produisent des résultats légitimes[506]. Or, l'inexactitude d'une décision affecte sa légitimité. À ce titre, les tensions entre les principes d'explicabilité et d'exactitude sont profondément inconfortables puisque, quelle que soit l'avenue privilégiée, la légitimité des décisions se trouve affectée.

Section II - Les tensions opposant l'exactitude à la sécurité

Quelques tensions mineures associées aux coûts liés à la sécurité des données. Les sources de tension entre les principes de sécurité et d'exactitude sont multiples. En effet, la sécurité réclame de mettre en œuvre des mesures de sécurité « raisonnables » afin de protéger les renseignements personnels[507]. Bien sûr, des coûts sont rattachés à l'adoption de telles mesures ce qui limite les capacités des entreprises privées et des organismes publics de collecter des renseignements personnels pour garnir leurs jeux d'entrainement. Or, tel qu'identifié, les SIA doivent s'entrainer auprès de jeux d'entrainement étendus et profonds afin de produire des résultats exacts[508]. À ce titre, nous pouvons déjà observer une tension entre la nécessité de collecter de grandes quantités de données pour produire des SIA exacts et les coûts associés aux mesures de sécurité devant être mises en œuvre.

La principale tension entre principes de sécurité et d'exactitude. En revanche, nous proposons que cette tension n'est pas particulièrement sévère et est loin de constituer le principal obstacle à la création de jeux d'entrainement suffisamment étendus et profonds au Québec. En fait, nous proposons que la principale tension entre le principe de sécurité et l'exactitude nait de l'imposition de l'approche minimaliste. En effet, le développement de SIA précis repose en grande partie sur une « approche maximaliste de la collecte et de l'utilisation des données »[509]. En contrepartie, le respect de la sécurité réclame de minimiser sa collecte et son usage des renseignements personnels. En droit québécois, c'est surtout l'interprétation qui est faite du critère de nécessité par la CAI qui pose obstacle à la création de SIA exacts. En effet, il limite les capacités des SIA à considérer un nombre élevé de variables lors de leur utilisation et lors de leur

[506] *Supra,* Chapitre 1, Section I, Sous-section B (i).
[507] *Supra,* Chapitre 1, Section III, Sous-section C.
[508] *Supra,* Chapitre 1, Section I, Sous-section B (i).
[509] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.

entrainement. Selon nous, le critère de nécessité cause des tensions avec le principe d'exactitude en raison (A) de son imprécision, (B) des imperfections rattachées à chacune de ses interprétations et (C) des difficultés liées à l'acquisition de la connaissance nécessaire pour respecter le critère de nécessité. De plus, il convient, selon nous, de (D) présenter les tensions liées à une mesure de sécurité promeut par plusieurs organisations gouvernementales de protection des renseignements personnels : l'usage de données synthétiques à des fins d'entrainement[510]. En effet, cette mesure est parfois présentée comme un « *silver-bullet solution to privacy-preserving data* »[511]. Or, elle n'est pas sans présenter certains défauts.

A. L'imprécision du critère de nécessité

Le caractère contradictoire du critère. D'abord, il convient d'identifier que le critère de nécessité est vague. En effet, les interprétations du critère qui coexistent actuellement dans la jurisprudence ne sont pas simplement distinctes, elles sont contradictoires.

Démonstration du caractère contradictoire des définitions. Ainsi, l'interprétation stricte et littérale[512] exige de prouver que la collecte est indispensable. Cette vision est clairement incompatible avec l'interprétation contextuelle et relative[513] qui propose qu'un renseignement est nécessaire lorsqu'il n'est pas superflu ou sans pertinence[514]. L'interprétation souple et dynamique[515], elle, propose que le critère de nécessité réclame de prouver plus que la simple utilité du renseignement, mais moins que son caractère indispensable[516].

---

[510] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 307, p. 2. ; EUROPEAN DATA PROTECTION SUPERVISOR, « Synthetic Data », en ligne : <https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_fr> (consulté le 27 juin 2022). ; INFORMATION COMMISSIONER'S OFFICE, *How should we assess security and data minimisation in AI?*, ICO, 2021, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/how-should-we-assess-security-and-data-minimisation-in-ai/> (consulté le 22 août 2022). ; DATATILSYNET, préc., note 122, p. 26.

[511] Theresa STADLER, Bristena OPRISANU et Carmela TRONCOSO, « Synthetic Data -- Anonymisation Groundhog Day », *arXiv: Learning* 2022, DOI : 10.48550/arXiv.2011.07018.

[512] P.-L. DÉZIEL, préc., note 131, 9-10.

[513] *Id.*, 10-11.

[514] Cour du Québec, 17 juin 2010, *P.B. c. Lepage*, préc., note 214, par. 91.

[515] P.-L. DÉZIEL, préc., note 131, 10-11.

[516] Cour du Québec, 21 février 2003, *Société de transport de la Ville de Laval c. X.*, préc., note 216, par. 44.

Conséquences des contradictions. Ainsi, trois interprétations contradictoires du même principe coexistent[517]. La conséquence de cette situation est qu'il reste difficile pour les organisations d'analyser efficacement la légalité de leurs pratiques ou de leurs projets. Des organisations soucieuses de respecter la Loi peuvent donc se montrer réticentes à collecter et utiliser des données tant et aussi longtemps que leurs opérations ne répondent pas aux interprétations du critère de nécessité les plus sévères. N'oublions pas, en effet, que l'entreprise privée qui collecte des renseignements qui ne sont pas nécessaires s'expose à des sanctions qui peuvent s'élever jusqu'à 25 millions de dollars ou jusqu'à « 4 % du chiffre d'affaires mondial de l'exercice financier précédent si ce dernier montant est plus élevé »[518].

B. Les défauts rattachés à chacune des interprétations du critère de nécessité

Problèmes de l'interprétation stricte et littérale en matière d'IA. Or, l'interprétation la plus sévère du critère de nécessité, c'est-à-dire l'interprétation stricte et littérale, est très difficilement applicable dans un contexte de SIA. En effet, cette interprétation réclame nécessairement aux SIA de n'utiliser que les variables qui sont « indispensable[s] ou essentiel[les], et dont la présence "rende[nt] seule possible une fin ou un effet" »[519]. En d'autres termes, elle réclame de se limiter aux variables qui permettent « le mieux d'atteindre la fin visée »[520] par le SIA. Or, il convient de rappeler que les SIA peuvent utiliser des milliers de variables afin de produire des résultats exacts et que chacune d'elles peut n'avoir qu'un « poids » très peu élevé sur le processus décisionnel[521]. Réclamer de se limiter à celles qui sont indispensables au processus est une requête qui va nécessairement réclamer de sacrifier la précision de l'outil. Il est également assez difficile de conceptualiser comment une organisation peut approcher cette obligation dans certaines circonstances. En effet, face à une situation où le SIA utilise une centaine de variables, mais que celles-ci détiennent toutes, individuellement, une

[517] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 17.
[518] *PL64*, préc., note 5, art. 160. Nouvel article 91 de la *Loi sur le privé*.
[519] Commission d'accès à l'information du Québec, 16 mars 2010, *M.L. c. Gatineau (Ville de)*, préc., note 212, par. 79. citant Tribunal d'arbitrage, 31 octobre 1998, *Syndicat des employées et employés professionnels et de bureau, section locale 57 et Caisse populaire St-Stanislas de Montréal*, préc., note 212.
[520] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[521] *Supra*, Chapitre 1, Section I, Sous-section A (iii).

influence très peu élevée sur la précision de l'outil comment est-il possible de qualifier quelconque de ces variables d'« indispensables » à l'analyse? Chacune d'elles ne rend pas seule « possible une fin ou un effet »[522]. Ainsi, l'interprétation stricte et littérale du critère de nécessité détient le potentiel d'empêcher pratiquement tout déploiement utile de SIA dès lors que ces derniers ne pourraient pas fonder leur analyse sur un nombre limité de variables.

Problème de l'interprétation souple et dynamique : sa rigidité. L'interprétation souple et dynamique n'est pas particulièrement plus généreuse. En effet, elle réclame de s'assurer que les moyens privilégiés constituent une « atteinte minimale »[523] à la vie privée des personnes concernées. Cette interprétation réclame *a priori* elle aussi de limiter l'analyse du SIA aux variables les plus influentes puisque, tel qu'identifié, plus un nombre élevé de renseignements personnels sont utilisés plus il devient difficile de prouver que les moyens privilégiés constituent une « atteinte minimale »[524]. À ce titre, cette interprétation limite elle aussi la capacité des SIA à produire des résultats exacts.

Problème de l'interprétation souple et dynamique : son imprécision. Un autre problème de cette interprétation est qu'elle est particulièrement vague. Cette imprécision est également héritée de l'imposition du critère d' « atteinte minimale ». En effet, il convient de se rappeler que ce critère tire son origine du test créé par la Cour suprême dans l'arrêt *Oakes*[525]. Or, dans l'arrêt *RJR MacDonald Inc. c. Canada,* la Cour suprême a interprété que l'atteinte minimale réclame de s'assurer « que l'atteinte aux droits ne dépasse pas ce qui est nécessaire. (nos soulignements) »[526]. À ce titre, Déziel propose que l'interprétation souple et dynamique souffre d'une certaine

[522] Commission d'accès à l'information du Québec, 16 mars 2010, *M.L. c. Gatineau (Ville de)*, préc., note 212, par. 79. citant Tribunal d'arbitrage, 31 octobre 1998, *Syndicat des employées et employés professionnels et de bureau, section locale 57 et Caisse populaire St-Stanislas de Montréal*, préc., note 212.
[523] *Supra,* Chapitre 1, Section III, Sous-section A.
[524] *Id* ; Commission d'accès à l'information du Québec, 28 septembre 2016, *Banque Nationale du Canada*, préc., note 230, par. 59-64.
[525] *Supra,* Chapitre 1, Section III, Sous-section A.
[526] *RJR-MacDonald Inc. c. Canada (Procureur général)*, [1995] 3 RCS 199, par. 160 (Cour suprême), en ligne : <https://canlii.ca/t/1frh0> (consulté le 21 août 2022).

« circularité »[527] qui la rendrait incapable d'aboutir « à une définition satisfaisante et opérationnelle du concept de nécessité »[528].

Les défauts de l'interprétation contextuelle et relative. L'interprétation contextuelle et relative, quant à elle, permet plus aisément de produire des SIA précis. En revanche, il convient de se questionner quant à sa capacité à procurer une quelconque forme de sécurité dans le contexte des SIA. En effet, prouver qu'un renseignement personnel n'est pas superflu à l'évaluation du SIA n'apparait pas particulièrement complexe. Ainsi, l'OBVIA et le Laboratoire de cyberjustice proposent que cette interprétation est incapable d'assurer l'imposition efficace d'une approche minimaliste au Québec puisqu'au regard de celle-ci « une entreprise traitant des renseignements personnels aux fins d'offrir des contenus personnalisés pourrait justifier la nécessité de pratiquement n'importe quels renseignements. »[529]

Observations des tensions entre sécurité et exactitude. Bref, quelle que soit l'interprétation privilégiée, une tension s'observe entre la nécessité de produire des SIA précis grâce à l'usage d'un nombre important de variables et le critère de nécessité qui vise à limiter la collecte et l'usage de celles-ci. Alors que les interprétations les plus strictes renforcent très clairement la sécurité des renseignements personnels, mais risquent de contrevenir à leur niveau de précision, l'interprétation contextuelle et relative, elle, semble incapable de fournir une forme de sécurité suffisante dans le contexte de l'IA.

Nuance : développements récents. En revanche, nous notons que la CAI semble de plus en plus privilégier l'interprétation souple et dynamique. Par exemple, toutes les décisions d'enquête parues jusqu'à maintenant en 2022 utilisent cette interprétation[530]. À ce titre, il serait loisible de

---

[527] P.-L. DÉZIEL, préc., note 131, 15-16.

[528] *Id.*, 15.

[529] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 17.

[530] Commission d'accès à l'information du Québec, 13 mai 2022, 1014505-S et 1016098-S, *Plainte à l'égard de l'Association québécoise des transports et de la Société de l'assurance automobile du Québec*, par. 21, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520910/index.do> (consulté le 26 août 2022); Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, préc., note 220, 4; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 52; Commission d'accès à l'information du Québec, 14 janvier 2022, 1016217-S, *Enquête à l'égard de Compagnie Selenis Canada*, par. 6-8, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519705/index.do> (consulté le 10 août 2022); Commission d'accès à l'information du Québec, 12 janvier 2022, 1011867-S, *X. c. MRC des Collines-de-l'Outaouais - Commission d'accès à l'information du Québec*, par. 9, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519698/index.do>

proposer que l'approche stricte et littérale susmentionnée est désormais révolue, une situation qui faciliterait la production de SIA exacts. Nous proposons néanmoins qu'il est trop tôt pour arriver à une telle conclusion. En effet, Déziel contemplait l'amoncèlement des interprétations en 2019[531] alors que l'OBVIA et le Laboratoire de cyberjustice l'observaient en 2020[532]. Qui plus est, il convient de rappeler que cette interprétation reste vague et assez stricte[533]. Même si elle devenait l'unique interprétation du critère de nécessité il est possible que des entreprises limitent leur collecte à un point où celle-ci ne concernera que les renseignements « indispensables » en raison des sanctions élevées prévues par PL64.

C. Le paradoxe créé par le critère de nécessité

Une prémisse problématique. Cependant, le critère de nécessité pose un problème plus fondamental dans le développement de SIA précis en raison de l'une de ses prémisses. Le critère réclame de savoir si un renseignement est nécessaire avant de procéder à sa collecte. En effet, le critère de nécessité réclame plus que la simple détermination des fins du renseignement[534]. Il réclame de prouver que la collecte du renseignement personnel peut réaliser les fins qui lui sont assignées. Selon l'interprétation privilégiée, cela réclame de faire la preuve de la pertinence du renseignement personnel (interprétation contextuelle et relative[535]), de son efficacité (interprétation souple et dynamique[536]) ou de son caractère indispensable (interprétation stricte et littérale[537]).

Difficultés à identifier les variables qui permettent le mieux d'atteindre la fin visée. Or, si nous interprétons le critère de nécessité au regard de l'approche stricte et littérale ou de l'approche souple et dynamique celui-ci réclame de ne collecter que les renseignements qui permettent le

(consulté le 26 août 2022). À noter que l'affaire *Enquête à l'égard du Centre intégré universitaire de santé et services sociaux de l'Estrie et du Centre hospitalier universitaire de Sherbrooke et Ministère de la Santé et des Services sociaux* ne traite pas du critère de nécessité des opérations visant les renseignements, mais de la nécessité de l'accès aux renseignements.

[531] P.-L. DÉZIEL, préc., note 131, 3-6.

[532] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16-18.

[533] *Supra*, Chapitre 2, Section II, Sous-section A.

[534] *Supra*, Chapitre 1, Section III, Sous-section B.

[535] P.-L. DÉZIEL, préc., note 150, 11-12.

[536] P.-L. DÉZIEL, préc., note 131, 12-13.

[537] *Id.*, 9-10.

mieux d'atteindre les objectifs[538]. Toutefois, c'est souvent au terme du traitement de ces mêmes renseignements que cette connaissance s'acquiert. Ainsi, l'OBVIA et le Laboratoire de cyberjustice proposent que :

> lorsque des techniques d'intelligence (sic.) sont déployées, il peut en certains cas être difficile de prévoir à l'avance quelles variables permettront le mieux d'atteindre la fin visée. Ce n'est qu'à la fin du processus de traitement de l'information qu'il devient possible d'identifier avec précision les données qui auront été nécessaires à l'atteinte des fins visées.[539]

Le paradoxe. Le critère de nécessité place donc le développeur de SIA devant un paradoxe. Le droit lui réclame de ne collecter que les renseignements nécessaires à la réalisation des fins poursuivies[540]. Cependant, en limitant ainsi sa collecte, il ne peut pas évaluer efficacement l'utilité des différents renseignements qu'il pourrait collecter. Il doit se résigner à n'utiliser que les renseignements qu'il sait être susceptibles d'atteindre la fin visée par le SIA. En revanche, ces renseignements ne seront pas nécessairement ceux qui lui permettront « le mieux d'atteindre la fin visée »[541]. Ainsi, pour limiter sa collecte, le développeur doit acquérir une connaissance qui ne peut être acquise que si lui ou d'autres personnes collectent ces renseignements afin d'en explorer la pertinence. Bref, le critère de nécessité réclame une connaissance dont il limite lui-même l'acquisition.

Contexte particulier des SIA : composer avec l'inconnu. Cette situation paradoxale est encore plus problématique lorsque le SIA est utilisé dans un environnement dynamique[542]. Dans de telles circonstances, tel qu'identifié, le SIA devra, afin de produire des résultats exacts, modifier ses comportements[543]. En présence d'un tel contexte, il est assez difficile, voire impossible, d'identifier dès le début du projet quelles seraient les « meilleures » données pour contrer les problématiques futures puisque celles-ci n'ont pas encore fait leur apparition.

---

[538] *Supra*, Chapitre 2, Section II, Sous-sections A et B.
[539] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[540] *Supra*, Chapitre 2, Section II, Sous-sections A et B.
[541] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 16.
[542] DATATILSYNET, préc., note 122, p. 10. Voir aussi *Supra,* Chapitre 2, Section I, Sous-section A.
[543] *Supra,* Chapitre 2, Section I, Sous-section A.

Nuance : Les exceptions à la recherche et à la production de statistiques. En revanche, il apparait opportun d'identifier que PL64 prévoit déjà certaines exceptions liées à la recherche et à la production de statistiques. Celles-ci permettent d'utiliser les renseignements à d'autres fins que celles initialement envisagées[544]. Il convient toutefois de nuancer l'impact réel de ces dispositions dans la résolution de la difficulté susmentionnée. En effet, ces exceptions concernent l'usage et la communication des renseignements, mais pas leur collecte[545]. Puisque quelqu'un doit nécessairement collecter les renseignements personnels directement auprès de la personne concernée, la collecte de ces renseignements devra satisfaire le critère de nécessité.

Une approche incompatible avec la réalité des SIA. Bref, nous pouvons constater à quel point l'approche minimaliste québécoise est difficilement réconciliable avec la nécessité de collecter de multiples variables dans l'objectif de produire des SIA suffisamment précis. En fait, le critère de nécessité constitue, selon nous, la principale source de tension qui s'oppose à la production de SIA précis. En effet, ce critère « laisse les entreprises et les organisations qui traitent des renseignements personnels dans une situation d'incertitude qui nuit tant à la protection de la vie privée qu'aux besoins des organisations. »[546]

D. Les défauts des données synthétiques

Position des institutions chargées de protéger les renseignements personnels. Enfin, tel qu'identifié, les institutions publiques assurant la protection des renseignements personnels proposent très souvent une solution aux développeurs de SIA qui permettrait, selon elles, de créer des SIA à la fois précis et sécuritaires[547]. Ainsi, la CAI[548], le European Data Protection Supervisor[549], le ICO[550] et l'Autorité norvégienne de protection des données[551] proposent toutes d'entrainer les outils d'IA auprès de jeux de données synthétiques.

---

[544] *PL64*, préc., note 5, arts. 20, 23 et 110. Nouveaux articles 65.1, 67.2.1 et 67.2.2 de la *Loi sur l'accès* et 12, 21 et 21.0.1. de la *Loi sur le privé.*
[545] *Id.*
[546] P.-L. DÉZIEL, K. BENYEKHLEF et E. GAUMOND, préc., note 9, p. 17.
[547] *Supra,* Chapitre 2, Section II.
[548] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 307, p. 2.
[549] EUROPEAN DATA PROTECTION SUPERVISOR, préc., note 510.
[550] INFORMATION COMMISSIONER'S OFFICE, préc., note 510.
[551] DATATILSYNET, préc., note 122, p. 26.

Ce que sont les données synthétiques. Les données synthétiques sont ainsi définies par le European Data Protection Supervisor *:*

> The concept of synthetic data generation is to take an original data source (dataset) and create new, artificial data, with similar statistical properties from it. Keeping the statistical properties means that anyone analysing the synthetic data, a data analyst for example, should be able to draw the same statistical conclusions from the analysis of a given dataset of synthetic data as he/she would if given the real (original) data.[552]

Pertinence des données synthétiques en matière de sécurité. L'espoir est que puisqu'elles ne se rattachent pas directement à des personnes physiques, un incident de confidentialité qui concerne des données synthétiques sera moins préjudiciable pour les personnes concernées par les renseignements personnels. En effet, ces données permettraient un :

> training of artificial intelligence models, in a manner that is less privacy-intrusive for the individuals because the data used in the training process does not directly refers to an identified or identifiable person.[553]

L'usage de données synthétiques peut affecter négativement l'exactitude des SIA. Malheureusement, les recherches empiriques relatives aux impacts des données synthétiques sur l'exactitude des SIA sont rares[554]. En revanche, quelques études récentes réalisées auprès de générateurs de données synthétiques dans le domaine de la santé tendent à révéler que l'usage de ces données peut affecter négativement les niveaux de précision des SIA. Ainsi, en utilisant trois différents générateurs de données synthétiques, les auteurs Rankin et al. ont démontré que *« [a] total of 92% of models trained on synthetic data have lower accuracy than those trained on real data. »*[555] De façon similaire, Benaim et al. ont étudié la validité des données synthétiques générées par un système spécifique (le « *MDClone system* »)[556]. Ces auteurs ont identifié que les données produites par ce système étaient souvent de qualité inférieure que les données réelles

[552] EUROPEAN DATA PROTECTION SUPERVISOR, préc., note 510.
[553] *Id.*
[554] Debbie RANKIN, Michaela BLACK, Raymond BOND, Jonathan WALLACE, Maurice MULVENNA et Gorka EPELDE, « Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing », (2020) 8-7 *JMIR Med Inform* e18910, 2, DOI : 10.2196/18910.
[555] *Id.*, 1.
[556] Anat Reiner BENAIM, Ronit ALMOG, Yuri GORELIK, Irit HOCHBERG, Laila NASSAR, Tanya MASHIACH, Mogher KHAMAISI, Yael LURIE, Zaher S. AZZAM, Johad KHOURY, Daniel KURNIK et Rafael BEYAR, « Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies », (2020) 8-2 *JMIR Medical Informatics* e16492, 2, DOI : 10.2196/16492.

et provoquaient parfois un « *selection bias* » dans les estimations produites[557]. Walonoski et al. quant à eux évaluèrent les données créées par Synthea « *an open-source software package that simulates the lifespans of synthetic patients* »[558] et constatèrent eux aussi plusieurs inexactitudes dans les données synthétiques produites par le système[559].

Nuance. Bien sûr, il n'est pas possible de prétendre qu'en toutes circonstances un SIA entrainé auprès d'un jeu composé de données synthétiques sera nécessairement moins précis qu'un SIA entrainé auprès d'un jeu composé de véritables données. En effet, tel que susmentionné, les études empiriques sur le sujet sont rares[560]. Notre objectif est d'appeler à la prudence auprès de ce type de données qui sont très souvent présentées comme une forme de solution miracle qui permettrait de répondre aux tensions entre la nécessité de collecter de grandes quantités de données pour produire des SIA précis et le désir de certaines institutions et auteurs de protéger la sécurité des données.[561]

Comprendre comment ces données sont créées. De plus, il convient d'identifier que même dans un cas dans lequel les données synthétiques seraient aussi précises que les données réelles, des tensions entre les principes de sécurité et d'exactitude pourront toujours être observées. En effet, les données synthétiques ne sont pas créées de façon aléatoire. Si tel était le cas, celles-ci seraient complètement inutiles. Pour être utiles, les données synthétiques doivent reproduire des propriétés statistiques identifiées dans des jeux de données réelles[562]. À ce titre :

> you will generally need to process some real data in order to determine realistic parameters for the synthetic data. Where that real data can be related to identified or identifiable individuals, then the processing of such data must comply with data protection laws.[563]

[557] *Id.*, 10.
[558] Jason WALONOSKI, Mark KRAMER, Joseph NICHOLS, Andre QUINA, Chris MOESEL, Dylan HALL, Carlton DUFFETT, Kudakwashe DUBE, Thomas GALLAGHER et Scott MCLACHLAN, « Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record », (2018) 25-3 *Journal of the American Medical Informatics Association* 230-238, 232, DOI : 10.1093/jamia/ocx079.
[559] *Id.*, 235.
[560] D. RANKIN et al., préc., note 554, 2.
[561] Voir, entre autres, T. STADLER, B. OPRISANU et C. TRONCOSO, préc., note 511.
[562] *Id.*, 2.
[563] INFORMATION COMMISSIONER'S OFFICE, préc., note 510.

La production des données hérite des sources de tensions susmentionnées. Cette situation crée deux problématiques. La première est que la production des données synthétiques n'échappe pas aux tensions susmentionnées. Afin de créer des données synthétiques capables d'imiter des propriétés statistiques identifiées dans des jeux de données réelles, des renseignements personnels devront être collectés et cette collecte devra composer avec les mesures de protection permettant d'assurer la sécurité des données.

Une sécurité imparfaite. La seconde est liée au fait que les données synthétiques ne parviennent pas à conférer une sécurité parfaite aux renseignements personnels. En effet, puisqu'elles doivent être produites à partir de vrais renseignements, les données synthétiques sont susceptibles aux « *membership infererence attacks* »[564]. Des acteurs malveillants peuvent donc réidentifier les personnes concernées par les données réelles à partir desquelles les données synthétiques ont été produites[565].

Un équilibre entre la sécurité et l'exactitude. En conséquence, la production de données synthétiques va nécessairement résulter d'un compromis entre les principes de sécurité et d'exactitude. Plus les données synthétiques s'avèreront capables de reproduire fidèlement les propriétés statistiques des données réelles, plus les SIA qui se basent sur ces données seront précis[566]. En revanche, plus les données synthétiques reproduisent fidèlement les données réelles plus il devient facile de réidentifier les personnes concernées par celles-ci[567].

Un compromis nécessaire entre les principes d'exactitude et de sécurité. Bref, les sources de tensions entre les principes d'exactitude et de sécurité sont multiples. Plusieurs mesures de protection imposées par le régime québécois des renseignements personnels limitent la capacité des organisations à entrainer des SIA précis. L'approche minimaliste reste, de loin, la principale source de tension entre ces deux principes. Il reste très difficile de contempler comment le critère

[564] T. STADLER, B. OPRISANU et C. TRONCOSO, préc., note 511, 1, 4 et 5. ; EUROPEAN DATA PROTECTION SUPERVISOR, préc., note 510. et *Supra*, Chapitre 1, Section III, Sous-section E.
[565] T. STADLER, B. OPRISANU et C. TRONCOSO, préc., note 511, 1.
[566] EUROPEAN DATA PROTECTION SUPERVISOR, préc., note 510; T. STADLER, B. OPRISANU et C. TRONCOSO, préc., note 511, 1 et 11.
[567] T. STADLER, B. OPRISANU et C. TRONCOSO, préc., note 511, 1, 4 et 5. ; EUROPEAN DATA PROTECTION SUPERVISOR, préc., note 510.

de nécessité se traduira dans un contexte de développement et d'utilisation de SIA. En effet, toutes ses interprétations semblent incapables de répondre de façon satisfaisante aux impératifs prescrits par les principes d'exactitude et de sécurité.

Section III - Les tensions opposant l'exactitude à l'équité

Le premier prix de l'équité, c'est l'exactitude. Il est généralement admis dans la littérature portant sur l'AA que de promouvoir une plus grande équité réclame nécessairement de sacrifier leur capacité à produire des résultats exacts, et ce, que l'on désire promouvoir une parité de performance ou une parité statistique[568]. Dans cette section, nous proposons (A) d'explorer les raisons du compromis (en anglais « *trade-off* ») entre l'exactitude et l'équité pour ensuite (B) illustrer cette tension par des exemples concrets pour, enfin, (C) nuancer notre propos.

A. Les raisons du compromis entre l'exactitude et l'équité

« Navigating the trade-offs ». Plusieurs mesures devant être mises en œuvre afin de renforcer une parité statistique ou une parité de performance réclament de sacrifier la capacité du SIA à produire des résultats exacts[569]. En effet, « *It has been shown that, under most frameworks of fairness, there is a trade-off between algorithmic performance and non-discrimination* »[570]. Ainsi, tel qu'identifié, il est possible d'utiliser des algorithmes de mitigation ou d'imposer des seuils distincts à différents groupes afin de produire des résultats plus équitables[571]. Or, ces deux mesures réclament nécessairement de réaliser des compromis entre l'exactitude du SIA et sa capacité à respecter une parité statistique ou une parité de performance[572].

Raison d'être du « *trade-off* ». La raison d'être de ce compromis est manifeste. L'usage d'algorithmes de mitigation ou de seuils distincts réclame de modifier les résultats du SIA. Or, puisque le SIA est entrainé afin de minimiser le risque d'erreurs, une modification de ses résultats réclame habituellement d'abandonner la solution dite « optimale »[573]. Ainsi, Meredith Whittaker

[568] *Supra,* Chapitre 1, Section II, Sous-section A.
[569] M. KEARNS et A. ROTH, préc., note 321, p. 73 (page pdf).
[570] G. PLEISS, M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, préc., note 314, p. 2.
[571] *Supra*, Chapitre 1, Section IV, Sous-section C.
[572] FAIRLEARN 0.7.0, préc., note 384.
[573] M. KEARNS et A. ROTH, préc., note 321, p. 68 (page pdf).

et al. proposent, qu'en pratique, une stratégie qui permet d'assurer une meilleure parité de performance « *often results in decreasing the "accuracy" for certain populations in order to match that of others.* »[574] Plus encore, des auteurs tels que Kearns et Roth observent, qu'il n'est pas possible d'échapper à ce dilemme opposant précision et équité puisque tous les bons modèles de SIA devront, par définition, y faire face[575]. Si un modèle ne fait pas face à ce dilemme, alors il s'agit tout simplement d'un mauvais modèle puisque l'un des principes peut être renforcé sans sacrifier l'autre[576].

Un problème d'optimisation. À ce titre, plusieurs auteurs approchent cette tension sous la forme d'un problème d'optimisation soumis à des contraintes : l'objectif étant de trouver l'algorithme le plus précis malgré les contraintes d'équité qui lui sont imposées[577]. Ainsi, les algorithmes de mitigation qui cherchent à produire des résultats satisfaisant une parité de performance ou une parité statistique « *seek to improve the fairness metrics without strongly affecting the accuracy, or more generally to navigate the trade-offs between performance and fairness metrics*. »[578]

B. Des exemples pratiques de la tension entre l'exactitude et l'équité

Exemple de Linkedln : perte de précision au profit d'une parité statistique. L'exemple d'une mesure récemment mise en œuvre par l'entreprise Linkedln afin de promouvoir une meilleure parité statistique permet d'illustrer cette problématique. Ainsi, Linkedln emploie un algorithme qui permet aux employeurs de trouver des candidats à des emplois potentiels. Le candidat qui répond le mieux au profil recherché par l'employeur sera présenté en priorité. L'entreprise a, en 2018, développé un outil qui modifie les résultats de ce SIA afin d'assurer une plus grande équité entre les candidats masculins et féminins. L'outil s'assure que si, par exemple, un employeur recherche un comptable dans une ville spécifique, qu'il existe 100 000 comptables dans cette

[574] À noter que le rapport utilise le terme « *classification parity* » et non « parité de performance ». Toutefois, tel qu'identifié, ces deux terminologies réfèrent à la même notion. MEREDITH WHITTAKER, KATE CRAWFORD, ROEL DOBBE, GENEVIEVE FRIED, ELIZABETH KAZIUNAS, VAROON MATHUR, SARAH MYERS WEST, RASHIDA RICHARDSON, JASON SCHULTZ, et OSCAR SCHWARTZ, *AI Now 2018 Report*, New York, AI Now Institute, New York University, 2018, p. 26. *Supra*, Chapitre 1, Section IV, Sous-section A(ii).
[575] M. KEARNS et A. ROTH, préc., note 321, p. 73 (page pdf).
[576] *Id.*, p. 72-73.
[577] S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 1.
[578] FAIRLEARN 0.7.0, préc., note **384**.

ville et que 40 % d'entre eux sont des femmes et 60 % sont des hommes, alors les candidatures présentées en priorité à l'employeur devront compter 40 % de femmes et 60 % d'hommes[579]. En agissant ainsi, le SIA assure une forme de parité statistique. En revanche, cette parité se fait au prix de la capacité du SIA à produire des résultats exacts. En effet, dans cet exemple, le rôle du SIA était d'identifier les candidats les plus susceptibles de répondre aux critères de l'employeur. La précision de l'outil se mesure donc à sa capacité à arranger adéquatement l'ordre de présentation des candidatures : les candidats répondant le mieux aux critères doivent être présentés en priorité[580]. Or, lui réclamer d'alterner la présentation des candidatures entre hommes et femmes interfère nécessairement avec sa capacité d'ordonner les candidatures purement sur la base de leur capacité à répondre aux attentes de l'employeur. L'application de cette mesure d'équité réclame donc au SIA de ne pas poursuivre la solution « optimale ». Il doit favoriser certaines candidatures non pas parce qu'elles sont reconnues par l'outil comme les candidatures les plus susceptibles de satisfaire les critères, mais parce qu'elles appartiennent à un groupe spécifique.

Exemple des SIA visant à évaluer les risques des accusés. Un autre exemple de cette tension s'illustre dans le développement des SIA comme COMPAS qui assistent juges et procureurs dans l'évaluation des risques des accusés[581]. Les auteurs Corbett-Davies et al. ont réalisé une expérience en appliquant un algorithme auprès des accusés d'un comté américain[582]. Comme COMPAS, le SIA utilisé assignait un score aux accusés entre 1 et 10. Ce score illustrait le risque qu'un accusé commette un crime violent si relâché avant la tenue de son procès. Si la personne analysée présente un score plus élevé qu'une valeur prédéterminée, alors elle est incarcérée. Or, puisque l'application d'un seuil identique à tous les accusés mène nécessairement à des résultats inéquitables à l'égard des détenus noirs, les auteurs proposent de prévoir différents seuils en

[579] Rosalie CHAN, « LinkedIn is using AI to make recruiting diverse candidates a no-brainer », *Business Insider*, en ligne : <https://www.businessinsider.com/linkedin-new-ai-feature-increase-diversity-hiring-2018-10> (consulté le 22 juin 2022); Miranda BOGEN, Aaron RIEKE et Shazeda AHMED, *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination*, *FAT* '20: Conference on Fairness, Accountability, and Transparency*, Barcelone, Espagne, 2020, p. 7, DOI : 10.1145/3351095.3372877.
[580] R. CHAN, préc., note 579; M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 7.
[581] S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 6.
[582] *Id.*

fonction de la couleur de peau de l'accusé[583]. En modifiant les seuils appliqués, les auteurs démontrent que le SIA peut respecter différentes conceptions de l'équité[584]. Leurs résultats sont non équivoques. Imposer le respect d'une contrainte d'équité implique nécessairement une hausse des crimes violents commis par les prévenus. Modifier le seuil afin d'assurer une contrainte d'équité mène également à une hausse du taux de détention des personnes qui présentent peu de risques, et ce, que l'on privilégie une parité statistique ou une parité entre taux de faux positifs[585]. Cependant, ne pas modifier les seuils afin de respecter l'une ou l'autre de ces formes d'équité implique de détenir 40 % des accusés noirs, mais 18 % des accusés blancs et de détenir 14 % des accusés blancs qui n'auraient pas commis un crime violent contre 32 % des accusés noirs qui n'auraient pas commis un crime violent, une situation qui viole à la fois la parité statistique et la parité de performance[586].

Autre exemple d'évaluation des risques. Les auteurs Kleinberg, Lakkaraju, Leskovec, Ludwig et Mullainathan ont également développé un SIA évaluant les risques qu'une personne libérée en attente de son procès commette un « crime ». Les auteurs définissent le « crime » prédit par leur algorithme comme le refus de l'accusé de se présenter à sa date de procès[587]. En effet, leur étude a été réalisée auprès de données de la ville de New York dans laquelle les juges ne peuvent détenir un accusé que s'ils prédisent qu'il ne se présentera pas à sa date de procès.[588] Les auteurs proposent, entre autres, trois possibilités de calibration pour leur algorithme :

> il peut être calibré afin de minimiser le taux de criminalité tout en assurant un taux de détention similaire à celui des juges de l'État de New York[589];
>
> il peut être calibré afin d'assurer une plus grande parité statistique afin que les personnes détenues « *have no higher share of black or Hispanic than that of the*

[583] *Id.*
[584] *Id.*
[585] *Id.*, p. 2 et 6.
[586] *Id.*, p. 6.
[587] Jon KLEINBERG, Himabindu LAKKARAJU, Jure LESKOVEC, Jens LUDWIG et Sendhil MULLAINATHAN, « Human Decisions and Machine Predictions », (2018) 133-1 *The Quarterly Journal of Economics* 237-293, 252, DOI : 10.1093/qje/qjx032.
[588] *Id.*, 239 et 257.
[589] *Id.*, 241, 276 et 277.

> *general defendant pool* »[590] tout assurant un taux de détention similaire à celui obtenu actuellement par les juges de l'État de New York[591];
>
> il peut être calibré afin de maximiser les libérations provisoires des accusés tout en assurant un taux de criminalité similaire à celui obtenu par les juges de l'État de New York[592];

Le premier scénario permet de réduire les crimes commis par des personnes libérées provisoirement de plus de 25 % en comparaison aux juges. Il assure aussi un taux de détention des personnes issues des minorités noires et hispaniques similaire à celui obtenu par les juges[593]. Le second permet de garantir une parité statistique entre les taux de détentions entre différents groupes en assurant un taux de détention moyen d'environ 26 %[594]. Cependant, l'algorithme garantit cette parité au prix d'une augmentation des crimes commis quoique cette différence est très mineure[595]. C'est le troisième scénario (et non le second) qui propose les résultats les plus avantageux pour les personnes issues de groupes minoritaires. Dans ce scénario le SIA détenait 38,8 % moins d'Afro-Américains et 44,6 % moins d' « *Hispanics »* que les juges[596].

C. Des nuances devant être formulées

Nuance. Il convient cependant de nuancer la tension entre les principes d'exactitude et d'équité. Trois points méritent, en effet, d'être explorés soit : ce que signifie « précision » dans le contexte des SIA, l'impact juridique de la tension observée et l'importance de la valeur cible recherchée par l'algorithme.

Ce que signifie « précision » dans le contexte d'un SIA. D'une part, il importe de contextualiser ce que l'on entend réellement par « précision » des SIA. Il importe, en effet, de comprendre qu'il n'est pas question ici d'une précision dans un sens absolu. Comme les humains, les SIA n'opèrent pas dans des environnements totalement transparents qui leur procurent une vue d'ensemble capable de dresser un reflet parfait de la réalité. À ce titre, les SIA doivent composer avec les

[590] *Id.*, 277.
[591] *Id.*
[592] *Id.*
[593] *Id.*, 241 et 277.
[594] *Id.*, 277.
[595] *Id.*, 277-278.
[596] *Id.*

réalités sociales qui peuvent affecter la collecte des renseignements qui les nourrissent. Les SIA d'évaluation de risques des accusés offrent un bon exemple de ce phénomène. Ceux-ci doivent se baser sur les renseignements qui leur sont connus et dévoilés. En revanche, un jeu de données ne peut pas réellement identifier qu'une personne libérée n'a pas « récidivée ». Plutôt, ce qu'il peut identifier, c'est que cette personne n'a pas été arrêtée par la police dans le cadre d'une « récidive »[597]. Or, pas tous les crimes commis mènent à une arrestation. Selon les données de 2019 du Federal Bureau of Investigation seulement « *45.5 percent of violent crimes and 17.2 percent of property crimes were cleared by arrest or exceptional means*. »[598] Le revers de cette réalité est que si les forces policières sont aux prises avec une problématique de profilage racial les personnes issues de minorités visibles sont plus à risque d'être arrêtées que les personnes issues de la majorité. À ce titre, les membres des minorités visibles « récidiveraient » plus (selon les données) non pas parce que c'est réellement le cas, mais parce qu'elles font l'objet d'une plus grande surveillance, et donc, la probabilité que leurs crimes soient identifiés par la police (ou que la police les accuse de commettre de tels crimes) est plus élevée[599]. Vues sous cet angle, les mesures d'équité susmentionnées peuvent être présentées non pas comme des obstacles à l'obtention de résultats exacts, mais au contraire, comme une prise en considération, certes imparfaite, de phénomènes sociaux connus qui affectent l'exactitude des données offertes au SIA.

Impact de la perte de précision. Qui plus est, l'exemple de LinkedIn susmentionné permet d'illustrer une autre nuance importante soit que l'impact des mesures d'équité ne sera pas nécessairement problématique au regard de la Loi. En effet, dans cet exemple, la capacité de cet algorithme à produire des résultats « précis » et « optimaux » était altérée. En revanche, il est difficile de concevoir comment cette décision pourrait s'avérer problématique au regard du droit québécois. Une modification à l'ordre des candidatures n'affecte pas l'exactitude de

---

[597] Marilyn ZHANG, *Affirmative Algorithms: Relational Equality as Algorithmic Fairness*, *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, Seoul, South Korea, Association for Computing Machinery, juin 2022, p. 13, p. 4, DOI : https://doi.org/10.1145/3531146.3533115.

[598] FEDERAL BUREAU OF INVESTIGATION, « 2019 Crime in the United States - Clearances », *FBI* (2020), en ligne : <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/clearances> (consulté le 24 juin 2022).

[599] M. ZHANG, préc., note 597, p. 4.

renseignements personnels. Aucun renseignement personnel n'est créé ou inféré dans cet exemple. Il serait également difficile de proposer qu'il s'agît d'une décision fondée exclusivement sur un traitement automatisé puisque le SIA ne fait qu'ordonner des candidatures potentielles. Il ne prend pas de décision à l'égard des candidats. Il aide l'employeur à prendre des décisions. Ainsi, il est difficile de percevoir comment, au regard de la Loi, l'imposition d'une telle mesure pourrait constituer un problème au Québec.

Importance des objectifs de l'algorithme. Finalement, il convient de comprendre que si notre objectif est de favoriser les groupes les plus vulnérables, alors la mise en application de mesures permettant d'acquérir une meilleure parité entre les groupes n'est pas la seule question qui devrait être étudiée. En effet, l'algorithme développé par Kleinberg et al. présente une nouvelle dimension dans la tension existante entre l'exactitude et l'équité. Dans son exemple, le scénario le plus favorable aux accusés issus de minorités visibles, c'est-à-dire celui qui garantissait le plus haut taux de libérations pour ceux-ci, n'était pas le scénario le plus paritaire. Le troisième scénario susmentionné ne poursuivait pas, en effet, une parité de performance ou une parité statistique. Il parvenait, en revanche, à réduire substantiellement les taux d'incarcération des Américains issus de minorités visibles. L'algorithme parvenait à ce résultat en modifiant sa valeur cible. Plutôt que d'optimiser la sécurité publique, l'algorithme cherchait à obtenir le plus haut taux de libération possible sans toutefois obtenir un taux de récidives plus élevé que le niveau actuellement atteint par les juges. Cette situation, sans surprise, provoque des résultats particulièrement favorables pour les personnes racisées ceux-ci étant surreprésentés dans le système carcéral américain. À ce titre, l'exemple de Kleinberg et al. illustre toute l'importance de la valeur cible du SIA. En favorisant les bonnes valeurs cibles, nous pouvons acquérir des résultats plus favorables aux membres des minorités visibles sans néanmoins sacrifier la capacité d'un SIA à produire des résultats exacts[600].

[600] Bien sûr, il est également possible de poursuivre de meilleures valeurs cibles et de satisfaire un traitement plus paritaire. Ce qu'il faut comprendre en l'espèce c'est qu'il importe de s'assurer que nos questionnements relatifs à l'équité n'obstruent une question parfois beaucoup plus importante : soit de déterminer si l'algorithme poursuit les cibles les plus désirables pour les personnes vulnérables dans les circonstances.

Tensions entre exactitude et équité. Bref, il est clair que la poursuite de certaines conceptions de l'équité réclame de sacrifier la capacité d'un SIA à produire des résultats exacts. En revanche, nous avons fortement nuancé cette tension. En effet, il importe de ne pas être aveuglé par l'apparence d'objectivité des outils d'IA. Tel qu'identifié, les résultats des SIA dépendent substantiellement des cibles poursuivies par ces systèmes et des données utilisées pour leur entrainement[601]. Ces données ne sont pas produites dans un contexte libéré de toutes contraintes sociales et économiques et il importe de considérer ce contexte lors du développement et de l'usage des outils d'IA[602].

Section IV - Les tensions opposant la parité statistique et la parité de performance à la sécurité et à la parité de traitement

Tensions entre équité et sécurité. Ensuite, nous observons des tensions qui opposent la parité statistique et la parité de performance à la sécurité et à la parité de traitement. En effet, les mesures devant être mises en œuvre afin de respecter une parité de traitement sont toutes des mesures qui permettent de garantir une meilleure sécurité des données[603]. La parité de traitement réclame de s'abstenir de collecter des attributs protégés ou des renseignements pouvant agir comme proxys[604]. Ces mesures sont donc compatibles avec l'approche minimaliste imposée par le régime québécois de protection des renseignements personnels. En revanche, les difficultés rattachées à la collecte d'attributs protégés limitent substantiellement les possibilités de garantir une meilleure parité statistique et une meilleure parité de performance. À ce titre, afin de bien comprendre les sources de tensions entre les principes d'équité et de sécurité il convient d'une part (A) d'expliciter pourquoi il est nécessaire de collecter des attributs protégés pour respecter certaines conceptions de l'équité, (B) d'exposer les obstacles imposés par le droit québécois et (C) d'illustrer les impacts pratiques des difficultés imposées à la collecte de ce type de renseignements.

[601] *Supra*, Chapitre 1, Section IV, Sous-section B.
[602] *Id.*
[603] *Supra*, Chapitre 1, Section IV, Sous-sections B et C.
[604] *Id.*

A. La nécessité de collecter et d'utiliser les attributs protégés

Pourquoi collecter et utiliser ces renseignements rattachés à des motifs de discrimination ? D'abord, il faut comprendre que la collecte (ou l'inférence) et l'utilisation des attributs protégés sont nécessaires afin de satisfaire une parité statistique ou une parité de performance. En effet, ces attributs permettent d'atteindre deux objectifs : la détection des iniquités et leur élimination.

Détecter les iniquités. Ainsi, la collecte et l'usage des attributs protégés permettent de détecter l'existence de divergences entre les performances des SIA ou de détecter la présence de résultats statistiquement non paritaires[605]. À ce titre, la collecte et l'usage de ce type de renseignements ne sont pas simplement utiles, mais très souvent nécessaires[606]. Sans la collecte d'attributs protégés, les développeurs et utilisateurs de SIA peuvent difficilement détecter les iniquités ou, même, identifier les renseignements qui agissent comme des proxys[607]. À ce titre, Fairlearn est très clair sur la nécessité de collecter ce qu'elle qualifie de « *sensitive features* »[608] telles que la « race », puisque :

> without having the race available in the dataset, it is hard to assess the model's impact across different groups defined by race or by race intersected with other demographic features[609].

Nécessité des attributs protégés pour corriger les iniquités. De façon similaire, la grande majorité des mesures proposées par les spécialistes de l'AA qui permettent d'acquérir une meilleure parité de performance ou une meilleure parité statistique réclament de collecter et d'utiliser les renseignements se rapportant à des motifs de discrimination[610].

Nécessaire pour diversifier les jeux d'entrainement. Ainsi, la diversification des jeux d'entrainement, une mesure qui permet, tel qu'identifié, d'atteindre une meilleure parité de performance, réclame de collecter et d'utiliser des attributs protégés[611]. En effet, afin de

---

[605] FAIRLEARN 0.7.0, préc., note 384.
[606] M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 7.
[607] Ainsi, la CDPDJ reconnait que : « les règlements sur les renseignements sensibles peuvent faire en sorte que des organisations ne collectent pas de telles données, ce qui complique l'identification de pratiques de discrimination par proxy » J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 63.
[608] FAIRLEARN 0.7.0, préc., note 384.
[609] *Id.*
[610] M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 1.
[611] *Supra*, Chapitre 1, Section IV, Sous-section C (ii).

s'assurer qu'un jeu de données soit diversifié, la personne chargée de collecter les renseignements peuplant le jeu doit nécessairement savoir à quels groupes les renseignements appartiennent. Le développeur qui utilise le jeu d'entrainement afin d'alimenter son SIA doit également avoir accès aux attributs protégés afin de pouvoir identifier quels sont les groupes qui peuplent le jeu d'entrainement et en quelles proportions[612].

Nécessaire aux algorithmes de mitigation. Il est également possible de recourir à des algorithmes de mitigation afin de garantir une plus grande parité statistique ou une plus grande parité de performance. Or, ces algorithmes réclament habituellement d'utiliser (et donc de collecter) des attributs protégés. Les auteurs Bofen, Rieke et Ahmed soulignent, en effet, que :

> les boites à outils en l'apprentissage automatique qui visent à garantir des résultats plus équitables tiennent souvent pour acquis, lors de la définition des problèmes et de leurs solutions, que des données d'attributs protégés sont disponibles. (traduction libre)[613].

Tel est le cas, par exemple, de la boite à outils «AI Fairness 360» fournie par IBM, de la librairie python « Pymetrics », du « What-If Tool » fournie par Google et de Fairlearn de Microsoft[614]. À ce titre, ne pas permettre à un SIA d'utiliser les attributs protégés l'empêche d'utiliser de nombreux algorithmes de mitigation.

Imposer des seuils distincts. Ensuite, il est nécessaire d'obtenir les renseignements qui se rattachent à des motifs de discrimination si l'on désire prévoir des seuils distincts pour différents groupes. Cette information est nécessaire non seulement afin de reconnaitre quel seuil devrait être imposé à la personne évaluée, mais également afin d'identifier quel sera le seuil « convenable » à imposer afin qu'il puisse garantir un résultat paritaire tout en maintenant un bon niveau d'exactitude. Si le développeur n'est pas en mesure de savoir avec précision à quel point un groupe est défavorisé, il risque soit de leur imposer un seuil trop sévère, qui s'avèrera

---

[612] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 63.

[613] M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 2.

[614] *Id.* citant GOOGLE AI BLOG, « The What-If Tool: Code-free probing of machine learning models », en ligne : <https://pair-code.github.io/what-if-tool/> (consulté le 21 août 2019); IBM, « AI Fairness 360 toolkit », en ligne : <https://github. com/IBM/AIF360> (consulté le 21 août 2019); Daniel DIAMOND, Lewis BAKER, Mark WARD, David WEINBERGER, Vishal MURALI et Joe NASO, *Pymetrics. 2019. audit-AI (Python library).*, Python, 2020, en ligne : <https://github.com/pymetrics/audit-ai> (consulté le 21 août 2019).; FAIRLEARN 0.7.0, préc., note 420; FAIRLEARN 0.7.0, préc., note 419;

incapable de corriger les iniquités, soit un seuil trop généreux qui sacrifierait inutilement la précision du SIA[615].

Nuance. À noter que certains auteurs tels que Lipton, Chouldechova et McAuley ont créé des applications d'AA qui parviennent à des résultats équitables sans avoir à utiliser ou collecter directement des attributs protégés[616]. Ils parviennent à ces résultats par l'utilisation de proxys. En effet, selon ces auteurs il est possible de garantir des résultats plus équitables par la collecte de renseignements qu'ils considèrent « triviaux ». En conséquence, les auteurs proposent qu'en présence d'un SIA qui évalue, par exemple, des candidats à un poste, celui-ci peut collecter la longueur des cheveux des candidats afin d'inférer lesquels des candidats sont des hommes et lesquelles sont des femmes. Grâce à ces informations, le SIA peut modifier son comportement afin de favoriser ces dernières[617].

Réponse à la nuance. Cette approche est néanmoins très problématique pour de nombreuses raisons. D'une part, le proxy utilisé ne constituera pas toujours un reflet parfait de l'attribut recherché. Si nous reprenons l'exemple proposé par ces auteurs nous pouvons, en effet, constater que la longueur des cheveux ne permet pas toujours d'inférer parfaitement le genre d'un individu. Pas toutes les femmes ont des cheveux longs et pas tous les hommes ont des cheveux courts. Pas toutes les femmes pourront donc bénéficier des mesures appliquées et certains hommes bénéficieront d'un avantage auquel ils ne devraient pas avoir droit[618]. D'autre part, la mesure peut renforcer des discriminations sur d'autres motifs prohibés. Nous n'avons qu'à prédire l'impact de considérer la « longueur des cheveux » à titre de proxy sur une personne souffrante d'alopécie féminine (parfois appelée calvitie féminine) pour concevoir à quel point cette mesure risque d'être inéquitable. Surtout, la décision devient difficile à justifier. Il est difficile de concevoir qu'une personne serait satisfaite de comprendre que la principale raison pour laquelle, par exemple, sa candidature à un emploi a été refusée est le fait que ses cheveux étaient trop courts. En contrepartie, la justification que, pour des raisons d'équité, les

[615] Voir M. KEARNS et A. ROTH, préc., note 321, p. 67-70 (page pdf).
[616] Z. LIPTON, A. CHOULDECHOVA et J. MCAULEY, préc., note 341, p. 2.
[617] *Id.*, p. 2-9.
[618] *Id.*, p. 9.

candidatures féminines ont été favorisées est beaucoup plus raisonnable. Finalement, ces méthodes sont loin d'être parfaites au regard de la parité de traitement et du principe de sécurité. En effet, bien que cette méthode permette de ne pas considérer les attributs protégés après le déploiement du SIA elle réclame d'utiliser des attributs protégés pour son entrainement[619]. À ce titre, le développeur du SIA devra toujours faire face aux limites imposées à la collecte de ce type de renseignements et aux obligations qui encadrent leurs usages.

B. Les obstacles à la collecte et à l'usage d'attributs protégés

Obstacles à la collecte et l'usage des attributs protégés. En l'absence d'attributs protégés, il est très difficile, voire impossible, de détecter les iniquités et d'y remédier. Cette nécessité d'utiliser ce type de renseignements[620] implique que la poursuite d'une parité statistique ou d'une parité de performance est en tension avec le principe de sécurité. La poursuite de ces deux formes de parité est également incompatible avec le respect d'une parité de traitement. En effet, cette dernière ne peut être respectée que si les attributs protégés ne sont pas explicitement considérés par le SIA[621]. Ces tensions se reflètent dans les difficultés créées par les obligations prévues par le droit québécois. À ce titre, il convient, d'une part, (i) d'observer les multiples difficultés qu'auront les organisations à collecter ce type de renseignements au Québec. D'autre part, il convient (ii) d'illustrer par des exemples pratiques les problèmes causés par les difficultés rattachées à la collecte et à l'usage de ce type de renseignements au niveau international.

i. Les multiples obligations imposées par le droit québécois qui limitent la collecte d'attributs protégés

Des coûts associés à la collecte et à l'usage d'attributs protégés en droit québécois. Au Québec, plusieurs obligations limitent la capacité des organisations à collecter et utiliser aisément des attributs protégés. Ainsi, tel qu'identifié, les attributs protégés sont très souvent des renseignements personnels sensibles[622]. Or, il est plus coûteux de collecter ce type de

[619] *Id.*, p. 1-2.
[620] À noter qu'il est également possible d'inférer les attributs protégés. En revanche, au Québec qu'un renseignement soit inféré ou collecté directement auprès de la personne concernée ne modifie pas les obligations légales qui y sont rattachées. Voir *Supra*, Chapitre 1, Section II, Sous-section B.
[621] *Supra*, Chapitre 1, Section IV, Sous-sections A à C.
[622] *Supra*, Chapitre 1, Section IV, Sous-section D.

renseignements. En effet, la *Loi sur le privé* et la *Loi sur l'accès* prévoient que la « raisonnabilité » des mesures de sécurité doit être évaluée au regard, entre autres, de la sensibilité des renseignements personnels collectés et utilisés[623].

Un usage ou une collecte qui ne peut procéder qu'au terme d'un acte positif de la part de la personne concernée. De façon similaire, PL64 prévoit que les fonctions d'une application technologique qui permettent à une entreprise ou à un organisme public d'effectuer un profilage devront être désactivées par défaut[624]. Le projet de loi définit le profilage comme :

> la collecte et […] l'utilisation de renseignements personnels afin d'évaluer certaines caractéristiques d'une personne physique, notamment […] [sa] situation économique, [sa] santé, [ses] préférences personnelles, [ses] intérêts ou [son] comportement...[625]

Or, cette définition incorpore clairement plusieurs attributs protégés parmi lesquels les données se rattachant aux handicaps et à la condition sociale[626]. PL64 prévoit également que l'usage de renseignements sensibles devra souvent faire l'objet d'un consentement manifesté par un « geste positif clair »[627]. Finalement, la *Charte québécoise* empêche de refuser de conclure un contrat avec un individu sur la base d'un motif de discrimination[628]. Bref, la concomitance de ces obligations indique que les attributs protégés ne pourront être collectés ou utilisés dans plusieurs circonstances qu'au terme d'un geste positif de la part de la personne concernée ce qui, bien sûr, limite les capacités des organisations à collecter ou utiliser ce type de renseignements personnels.

---

[623] *Supra*, Chapitre 1, Section III, Sous-section C.

[624] *PL64*, préc., note 5, arts. 15, 19, 107 et 108. Nouveaux articles 63.7. et 65.0.1. de la *Loi sur l'accès* et 8.1. et 9.1. de la *Loi sur le privé.* Voir aussi la discussion à cette effet dans ASSEMBLÉE NATIONALE DU QUÉBEC, « Étude détaillée du projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels (17 mars 2021) », (2021) 45-126 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210317.html> (consulté le 31 mai 2022).

[625] *PL64*, préc., note 5, arts. 19 et 107. Nouveaux art. 65.0.1. de la *Loi sur l'accès* et 8.1. de la *Loi sur le privé.*

[626] *Charte des droits et libertés de la personne*, préc., note 14, art. 10.

[627] *Supra,* Chapitre 1, Section III, Sous-section C.; GOUVERNEMENT DU QUÉBEC, préc., note 273. ; *Code civil du Québec*, préc., note 145, art. 37; *PL64*, préc., note 5, arts. 13, 20, 110. Nouveaux articles 59 et 65.1 de la *Loi sur l'accès* et 12 et 13 de la *Loi sur le privé*

[628] *Charte des droits et libertés de la personne*, préc., note 14, art. 12. ; J.-F. TRUDEL, R. AVILA, G. ST-LAURENT et R. AVILA, préc., note 169, p. 15.

Difficultés majeures liées à la collecte et à l'usage d'attributs protégés. En revanche**,** les obligations susmentionnées ne sont pas si problématiques. En effet, il reste possible de collecter et d'utiliser des attributs protégés malgré leur existence. Qui plus, plusieurs exceptions sont prévues par PL64 qui facilitent l'inférence, la collecte ou l'usage de renseignements sensibles. Par exemple, l'usage par une entreprise de renseignements sensibles à d'autres fins que celles pour lesquelles le renseignement a été collecté n'a pas besoin d'obtenir le consentement exprès des personnes concernées s'il est « nécessaire à des fins d'étude, de recherche ou de production de statistiques et qu'il est dépersonnalisé »[629]. Or, cette exception semble *a priori* permettre l'usage d'attributs protégés à des fins de détection de comportements discriminatoires lors de la phase d'entrainement ou d'inférence d'un SIA sans avoir à obtenir le consentement exprès des personnes concernées[630]. En effet, un tel usage peut aisément être interprété comme une fin d'étude, de recherche ou de production de statistiques. Selon nous, les réelles difficultés liées au régime québécois des renseignements personnels sont créées, en fait, par les sanctions prévues par PL64 en cas d'incidents de confidentialités et par le critère de nécessité[631].

Les sanctions prévues. En effet, les modifications législatives apportées à la *Loi sur le privé* prévoient que l'usage et la collecte de renseignements sensibles exposent les entreprises à des sanctions plus sévères en cas d'incidents de confidentialité[632]. Or, en raison de la sévérité des sanctions possibles, qui peuvent s'élever jusqu'à 4 % du chiffre d'affaires mondial d'une entreprise, il est peu probable que les entreprises se montrent désireuses de prendre le risque de collecter les attributs protégés afin de remédier aux comportements inéquitables de leurs SIA à moins que ce ne soit clairement requis par la Loi[633]. En effet, il faut comprendre que le seul fait de recueillir, d'utiliser ou de conserver des renseignements en contravention à la *Loi sur le privé* expose l'entreprise à de telles sanctions. Or, la Loi québécoise n'est pas très claire quant aux circonstances où il est légal pour une entreprise de collecter ce type de renseignement, et ce, en raison de l'interprétation par la CAI du critère de nécessité.

---

[629] *PL64*, préc., note 5, art. 110. Nouvel article 12(5) de la *Loi sur le privé.*
[630] *Id.* Nouvel article 12(5) de la *Loi sur le privé.*
[631] *Supra,* Chapitre 1, Section III, Sous-sections A et C.
[632] *PL64*, préc., note 5, arts. 159 et 160. Nouveaux articles 90.2, 90.12 et 91 de la *Loi sur le privé.*
[633] *Id.*, art. 160. Nouvel article 91 de la *Loi sur le privé.*

Les difficultés liées à la preuve du caractère important et réel des objectifs des renseignements personnels. Tel qu'identifié, au Québec, la collecte et l'usage de renseignements personnels doivent être justifiés au regard du critère de nécessité, et ce, même si la personne concernée consent de façon éclairée aux différentes opérations concernant ses renseignements personnels[634]. Or, non seulement le critère est appliqué de façons contradictoires par la CAI[635], mais en plus, son interprétation complexifie la collecte et l'usage de renseignements à des fins strictement préventives. En effet, il faut comprendre qu'il ne suffit pas de prouver qu'une collecte et l'usage d'un renseignement sont légitimes. L'interprétation souple et dynamique du critère impose également de prouver le caractère important et réel des objectifs des renseignements personnels collectés. Or, au regard de certaines décisions de la CAI, il semble difficile de prouver le caractère important et réel de la fin d'un renseignement si ce dernier répond à une problématique appréhendée.

Les fins des renseignements ne peuvent pas être qu'appréhendées. En effet, la CAI réclame de prouver que les objectifs poursuivis par un renseignement personnel doivent « être réels et non seulement appréhendés »[636]. Ainsi, dans son enquête sur *Bruneau Électrique inc.,* la CAI a précisé que le vol de temps ne pouvait constituer une fin réelle justifiant la collecte de renseignements puisqu'« aucune démonstration relative au vol de temps avant et après la mise en place de la collecte […] n'a été faite [par l'entreprise] » [637]. La CAI n'était, en effet, pas satisfaite du simple argument voulant « que le vol de temps est un problème généralisé dans toutes les entreprises »[638]. Ce passage implique que la commission réclamait la preuve que l'entreprise était spécifiquement aux prises avec une problématique liée au vol de temps. De plus, dans une enquête récente réalisée en avril 2022, la CAI a proposé que la collecte de données afin de détecter la fraude dans des transactions de moins de 3 000 $ ne poursuivait pas un objectif réel ou important en l'absence de documentation permettant, par exemple, de démontrer

[634] *Supra,* Chapitre 1, Section III, Sous-section A.
[635] *Supra,* Chapitre 2, Section II, Sous-section A.
[636] Commission d'accès à l'information, 30 septembre 2021, *Enquête à l'égard de Bruneau Électrique inc.*, préc., note 203, par. 23.
[637] *Id.*, par. 34.
[638] *Id.*

« l'ampleur des cas de fraude dont [l'entreprise] a été victime. »[639]. En 2015, la CAI a également conclu qu'une garderie n'avait pas fait la preuve du caractère « réel et urgent »[640] des objectifs poursuivis par une collecte de renseignements qui visait à assurer la sécurité d'enfants. La commission a spécifié que la garderie aurait pu démontrer que sa collecte de renseignements poursuivait des objectifs réels par la démonstration qu'elle était « motivée par un événement particulier ou une situation problématique relative à la sécurité des biens ou des personnes qui s'y trouvent (ex. : plainte d'un parent au sujet des agissements d'une éducatrice, vols, vandalisme) »[641].

Appréhension de la discrimination. La difficulté, en l'espèce, est que la collecte et l'usage d'attributs protégés à des fins de détection de comportements inéquitables peuvent être considérés comme relevant d'objectifs appréhendés plutôt que réels. À bien des égards, c'est parce que l'on appréhende qu'un SIA puisse discriminer que l'on collecte des attributs protégés. C'est grâce à ces renseignements que l'on peut confirmer ou infirmer cette crainte. Bref, la démonstration du caractère « réel » d'une collecte d'attributs protégés en l'absence de preuves que le SIA va bel et bien discriminer semble difficile à prouver. En effet, les passages susmentionnés semblent indiquer que la CAI pourrait ne pas être satisfaite d'une simple preuve par les entreprises que les difficultés des SIA de produire des performances ou des résultats paritaires sont des phénomènes connus et documentés. La démonstration du caractère important et réel semble réclamer une preuve qu'il existe de bonnes raisons de croire que le SIA fait spécifiquement face à une telle problématique. De plus, il convient de rappeler qu'il est plus difficile prouver la nécessité d'une collecte et d'une utilisation de renseignements personnels s'ils sont sensibles[642]. Quoiqu'il en soit, le droit n'est pas clair et cette absence de clarté conjuguée à

[639] Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 55.
[640] Commission d'accès à l'information du Québec, 9 février 2015, *Garderie Excelsiori Daycare inc.*, préc., note 220, par. 28.
[641] Il est à noter que dans les trois exemples mentionnés dans ce paragraphe il était manifeste que les entreprises employaient des méthodes disproportionnées ou inefficaces pour accomplir leurs objectifs. En revanche, l'étude du caractère « réel » ne demande d'évaluer que le caractère des objectifs poursuivis. Le caractère inefficace ou disproportionné des moyens employés dans la poursuite de ces objectifs est étudié postérieurement. Voir *Supra,* Chapitre 1, Section III, Sous-section A ; *Id.*, par. 30 et 46; Commission d'accès à l'information, 30 septembre 2021, *Enquête à l'égard de Bruneau Électrique inc.*, préc., note 203, par. 41; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 60.
[642] *Supra,* Chapitre 1, Section III, Sous-section A et Chapitre 1, Section IV, Sous-section D (ii).

l'application de sanctions sévères risque de motiver les entreprises à ne pas collecter ces attributs de crainte de contrevenir à la *Loi sur le privé* et de s'exposer aux sanctions qui y sont prévues.

ii. Le phénomène de rareté des attributs protégés

Des données plus rares que prévu. Enfin, il faut comprendre que la rareté de renseignements se rapportant à des motifs de discrimination n'est pas un problème hypothétique. Bien qu'il soit assez commun de penser qu'à notre époque de « *Big Data* » il reste relativement simple pour les organisations d'acquérir des renseignements personnels même sensibles, cette perception n'est pas toujours conforme à la réalité. En effet, pour bien comprendre la problématique de rareté qui affecte les attributs protégés, il convient d'observer la quantité d'attributs protégés devant être considérés par le SIA.

Impact de la multiplicité des groupes. Ainsi, il convient de comprendre qu'un algorithme ne garantit des résultats équitables à l'égard de groupes spécifiques que si nous lui réclamons explicitement de considérer ces groupes. Programmer un SIA afin qu'il produise des résultats équitables à l'égard, par exemple, des Afro-Américains ne va pas lui permettre de produire des résultats équitables à l'égard, par exemple, des personnes transgenres. Pour qu'un SIA soit équitable à l'égard de ces deux groupes, il doit pouvoir collecter et utiliser des attributs protégés se rapportant à ces groupes spécifiques[643]. En effet :

> algorithms generally, and especially machine learning algorithms, are good at optimizing what you ask them to optimize, but they cannot be counted on to do things you'd like them to do but didn't ask for, nor to avoid doing things you don't want but didn't tell them not to do. Thus if we ask for accuracy but don't mention fairness, we won't get fairness. If we ask for one kind of fairness, we'll get that kind but not others. […] This same theme holds true for the choice of which groups we protect[644].

Or, la *Charte québécoise* prévoit plus d'une dizaine de motifs de discrimination et la *Charte canadienne* ne prévoit pas de liste exhaustive de motifs. Chacun de ces motifs demande d'assurer une équité entre une multiplicité de groupes. Par exemple, un SIA désirant assurer une parité statistique sur la base de l'origine ethnique pourra être appelé à considérer les impacts de ces résultats à l'égard des personnes blanches, à l'égard des personnes noires, à l'égard des

[643] M. KEARNS et A. ROTH, préc., note 321, p. 78 (page pdf).
[644] *Id.*

personnes asiatiques, à l'égard des personnes autochtones… Bref, assurer une parité statistique ou une parité de performance peut demander de collecter une multiplicité d'attributs protégés.

Nuance. Il convient de nuancer la difficulté identifiée. En effet, il ne sera pas toujours nécessaire de considérer l'entièreté des motifs de discrimination en toutes circonstances. Contrecarrer la discrimination basée sur l'état civil, par exemple, un motif de discrimination prévu par la *Charte québécoise*[645], ne risque pas d'être pertinent pour un SIA évaluant les risques de récidives en matière pénale. En revanche, nous proposons qu'un tel SIA réclame quand même de considérer une multiplicité de motifs et de groupes. En effet, il conviendra sûrement d'évaluer, pour un tel SIA, les groupes qui font souvent l'objet d'un profilage social ou racial tels que : les personnes noires, les personnes autochtones, les personnes hispanophones, les personnes de descendances issues du Moyen-Orient, les personnes transgenres, les personnes aux prises avec des handicaps mentaux, les musulmans[646]…

Réponse à la nuance : Considérer les sous-groupes visés par une combinaison d'attributs protégés. Le phénomène de « *fairness gerrymandering* »[647] semble également complexifier la problématique. En effet, le « *fairness gerrymandering* » est un phénomène par lequel l'application de mesures permettant d'assurer une meilleure équité crée un SIA très inéquitable à l'égard des personnes qui disposent de caractéristiques se rapportant à plusieurs groupes vulnérables tel que les « femmes noires »[648]. Plusieurs auteurs en AA proposent donc que si nous désirons programmer un SIA équitable à l'égard, par exemple, du genre et de l'origine ethnique,

---

[645] *Charte des droits et libertés de la personne*, préc., note 14, art. 10.

[646] Voir, par exemple, Paul EID, Johanne MAGLOIRE et Michèle TURENNE, *Profilage racial et discrimination systémique des jeunes racisés: rapport de la consultation sur le profilage racial et ses conséquences*, Québec, Commission des droits de la personne et des droits de la jeunesse, 2011, p. 1 et 25-26; VICTOR ARMONY, MARIAM HASSAOUI, et MASSIMILIANO MULONE, *Les interpellations policières à la lumière des identités racisées des personnes interpellées : Analyse des données du Service de Police de la Ville de Montréal (SPVM) et élaboration d'indicateurs de suivi en matière de profilage racial (Rapport final remis au SPVM)*, Montréal, Canada, Équipe Armony-Hassaoui-Mulone (Chercheurs indépendants), p. 84 à 106, en ligne : <https://spvm.qc.ca/upload/Rapport_Armony-Hassaoui-Mulone.pdf>.

[647] Michael KEARNS, Seth NEEL, Aaron ROTH et Zhiwei Steven WU, *Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness*, *Proceedings of the 35th International Conference on Machine Learning*, PMLR, International Conference on Machine Learning, 80, 3 juillet 2018, p. 2564-2572, l'Abstract en ligne, en ligne : <https://proceedings.mlr.press/v80/kearns18a.html> (consulté le 24 août 2022); M. KEARNS et A. ROTH, préc., note 357, p. 79.

[648] M. KEARNS, S. NEEL, A. ROTH et Z. S. WU, préc., note 645, p. 2564; M. KEARNS et A. ROTH, préc., note 321, p. 79 (page pdf).

il n'est pas suffisant de demander au SIA de ne considérer que les « gros » groupes comme les « femmes », les « hommes », les « Caucasiens » et les « Autochtones ». Il faut également demander explicitement au SIA de protéger une équité entre sous-groupes tels que les « femmes autochtones », les « hommes autochtones », les « femmes caucasiennes » et les « hommes caucasiens »[649]. Si nous ne lui demandons pas d'agir ainsi, ces auteurs proposent que le SIA sera, certes, capable d'assurer, par exemple, une parité statistique entre « hommes et femmes » et « Autochtones et Caucasiens », mais que l'acquisition de cette parité se fera potentiellement « *at the expense of discrimination against some intersection of [overlapping groups]*. »[650] Ce phénomène réclame de collecter des attributs protégés même si, *a priori,* ces attributs ne sont pas normalement identifiés comme des motifs de discriminations courant dans le domaine dans lequel opère le SIA. Une personne pourrait donc argumenter que de collecter le genre, par exemple, serait peu pertinent pour un SIA qui évalue les risques de récidives puisque les statistiques ne révèlent pas, à Montréal du moins, que ce groupe fait l'objet d'un profilage aussi sévère que les personnes noires ou que les personnes autochtones[651]. Or, il pourrait s'avérer problématique d'agir ainsi puisqu'en raison du « *fairness gerrymandering* », un SIA pourrait produire des résultats équitables à l'égard, par exemple, des autochtones, tout en créant des résultats manifestement défavorables aux femmes autochtones[652].

Quantité importante de renseignements se rapportant à chacun de ces groupes. Après, non seulement est-il nécessaire de recueillir plusieurs attributs protégés se rapportant à une multiplicité de groupes et de motifs de discrimination, mais en plus, certaines mesures permettant d'acquérir une plus grande équité réclament de collecter un grand nombre de renseignements se rapportant à chacun de ces groupes. Tel est le cas, par exemple, de l'usage de jeux d'entrainement diversifiés. Tel qu'identifié, il ne suffit pas, pour assurer l'efficacité de cette mesure, d'utiliser des jeux d'entrainement représentatifs, c'est-à-dire des jeux dans lesquels les

[649] M. KEARNS et A. ROTH, préc., note 321, p. 78-79 (page pdf).
[650] *Id.*, p. 78.
[651] VICTOR ARMONY, MARIAM HASSAOUI, et MASSIMILIANO MULONE, préc., note 646, p. 66 et 88.
[652] À noter que même sans la problématique du « *fairness gerrymendering* » la collecte du genre devrait être réalisée ne serait-ce qu'en raison de la discrimination intersectionnelle vécue par les femmes autochtones. En effet, à Montréal des statistiques révèlent que les femmes autochtones ont 11 fois plus de chances d'être interpeler que les femmes blanches. Les hommes autochtones ont 3 fois plus de chance que les hommes blancs. *Id.*, p. 88.

groupes sont distribués de façon similaire à leur poids démographique[653]. Il faut utiliser des jeux dans lesquels les groupes sont également distribués[654]. Or, puisque les SIA réclament de collecter de grandes quantités de renseignements afin d'acquérir de bons niveaux de précisions, tenter d'entrainer des SIA sur des jeux diversifiés afin que ceux-ci puissent acquérir de bons niveaux de précision à l'égard de tous les groupes vulnérables semble être une démarche particulièrement complexe.

Volume de renseignements à collecter. Bref, lorsque nous considérons le volume de renseignements devant potentiellement être collectés et utilisés il devient aisé de concevoir pourquoi les limites légales liées à la collecte, à la communication et à l'utilisation d'attributs protégés sont en tension avec l'application de mesures permettant de poursuivre une plus grande parité statistique ou plus grande parité de performance. En fait, plusieurs exemples pratiques permettent d'illustrer l'existence d'un phénomène de rareté rattachée aux attributs protégés à l'international dans certains domaines.

Difficultés rencontrées par le gouvernement américain. En effet, le Québec n'est pas la seule juridiction qui prévoit des règles plus sévères à l'égard des attributs protégés. D'autres juridictions comme les États-Unis ou l'Europe prévoient également des contraintes sévères à l'égard de ce type de renseignements[655]. À ce titre, les auteurs Bogen, Rieke et Ahmed proposent que les spécialistes en AA tendent à sous-estimer les difficultés pratiques et légales limitant la collecte et la diffusion des attributs protégés[656]. En effet, ces auteurs soulignent que même des institutions gouvernementales américaines qui disposent de ressources importantes comme le Federal Reserve Board ont exprimé des difficultés à réaliser des études portant sur la discrimination en raison des difficultés à acquérir des attributs protégés[657]. Ces auteurs notent également qu'en 2013 le Consumer Financial Protection Bureau (ci-après « CFPB »), une institution gouvernementale américaine, avait imposé à une entreprise de rembourser plusieurs

[653] *Supra*, Chapitre 1, Section IV, Sous-section C.
[654] *Id.*
[655] À noter qu'aux États-Unis, ces limitations visent surtout certains domaines, comme le domaine bancaire. M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 2 et 3; *RGPD*, préc., note 428, art. 9.
[656] M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 3.
[657] *Id.*

millions de dollars à ses consommateurs appartenant à des minorités visibles. Cependant, le CFPB ne disposait pas de statistiques permettant d'identifier l'origine ethnique des consommateurs de l'entreprise. L'institution s'est résignée à devoir inférer l'origine ethnique des personnes qui devaient bénéficier du remboursement. Malgré avoir utilisé l'une des méthodes les plus précises pour inférer les attributs protégés, plusieurs chèques ont été erronément envoyés à des Américains blancs[658].

L'impact sur les plus petites entreprises. Or, pas toutes les organisations disposent de ressources et de moyens aussi importants que le gouvernement américain. Les activités de structures plus modestes, comme des « *startup* » ou des équipes de recherche peuvent dépendre de renseignements faciles d'accès. Or, si le gouvernement américain fait face à de telles difficultés pour collecter et utiliser des attributs protégés, il est assez difficile de concevoir comment une entreprise qui dispose de peu de ressources pourrait acquérir tous les attributs protégés nécessaires au regard de la *Charte québécoise* lors de l'entrainement de ses modèles.

Illustration récente du problème : cancer de la peau. Il faut également comprendre que les équipes de chercheurs et les entreprises plus modestes peuvent dépendre des jeux de données disponibles au public pour entrainer leurs modèles[659]. Or, il a été récemment révélé que les attributs protégés sont parfois absents de ces jeux de données même dans des circonstances où leur pertinence est pourtant évidente[660]. Cette étude portait sur des jeux de données composés d'illustrations de lésions de la peau disponibles publiquement aux chercheurs en dermatologie. Ces jeux de données permettent d'entrainer des SIA destinés à diagnostiquer le cancer de la peau[661]. Or, ces chercheurs ont démontré que sur 21 différents jeux de données étudiés seuls deux d'entre eux spécifiaient l'origine ethnique de ses participants[662]. Conséquemment, sur les 106 950 images offertes par ces jeux de données, les informations portant sur l'ethnicité du patient n'existaient que pour 1,3 % des images disponibles et la couleur de peau n'était indiquée

[658] *Id.*, p. 4.
[659] D. WEN et al., préc., note 358, e64.
[660] *Id.*, e64.
[661] *Id.*
[662] *Id.*, e71.

que dans 2,1 % des cas[663]. Des deux seuls jeux de données qui décrivaient l'ethnicité du patient, aucune image n'indiquait appartenir à un patient de descendance africaine, afrocaraïbéenne ou sud asiatique[664].

Utilité de la couleur de peau en dermatologie. Il est pourtant évident que la couleur de peau et l'origine ethnique sont des données pertinentes en dermatologie. En effet, il a été démontrer que demander à un SIA de détecter une lésion sur une peau de couleur distincte des observations sur lesquels il s'est entrainé ne produit pas des résultats précis[665]. Par exemple, en 2018, un SIA diagnostiquant le cancer de la peau avait été entrainé auprès d'un jeu de données composées de personnes asiatiques. Conséquemment, l'outil détenait un taux de précision de 81 % à l'égard de jeux de données composés de personnes asiatiques, mais de 56 % à l'égard de jeux composés de Caucasiens[666].

Une situation difficile. Bref, bien que nous vivions à une époque de « Big data » cela ne signifie pas que toutes les données sont disponibles en toutes circonstances pour toutes les organisations. Même des organisations disposant d'immenses ressources comme le gouvernement américain ont parfois de la difficulté à collecter certains types de renseignements. De plus, les bases de données disponibles au public ne révèlent pas toujours les attributs protégés pertinents à l'analyse. Il est donc clair qu'une entreprise n'aura pas nécessairement accès à ce type de renseignements au moment de l'entrainement de son modèle de SIA. Dans de pareilles circonstances, l'entreprise devra, au mieux, collecter ces attributs protégés après le déploiement de son outil d'IA. En revanche, tel qu'identifié, le droit québécois impose de nombreuses barrières à la collecte de ce type de renseignements afin, entre autres, d'en assurer la sécurité[667]. Au regard de ces difficultés, il devient apparent qu'il existe une tension entre les restrictions

---

[663] *Id.* La couleur de peau était mesurée au regard du "*Fitzpatrick skin type*".
[664] *Id.*
[665] *Id.*, e64.
[666] Manu GOYAL, Thomas KNACKSTEDT, Shaofeng YAN et Saeed HASSANPOUR, « Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities », (2020) 127 *Computers in Biology and Medicine* 104065, 5, DOI : 10.1016/j.compbiomed.2020.104065. citant Seung Seog HAN, Myoung Shin KIM, Woohyung LIM, Gyeong Hun PARK, Ilwoo PARK et Sung Eun CHANG, « Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm », (2018) 138-7 *Journal of Investigative Dermatology* 1529-1538, DOI : 10.1016/j.jid.2018.01.028.
[667] *Supra,* Chapitre 1, Section III, Sous-sections A et C et Chapitre 2, Section IV.

imposées afin d'assurer la sécurité des données (et une parité de traitement) et les mesures devant être mises en œuvre afin de poursuivre une parité statistique ou une parité de performance.

Section V - Les tensions opposant la parité de performance à la parité statistique

L'impossibilité de respecter parfaitement à la fois une parité de performance et une parité statistique. Finalement, il convient de comprendre que l'incompatibilité susmentionnée entre la parité de traitement et les autres conceptions de l'équité ne constitue pas l'unique source de tensions internes au principe d'équité[668]. En effet, la parité statistique et la parité de performance sont également en tension. Pour démontrer les sources de tensions qui affectent ces deux conceptions de l'équité, il convient à la fois d'exposer (A) les difficultés liées à l'évaluation de la parité de performance et (B) pourquoi il est impossible, pour des raisons mathématiques, de créer un SIA qui serait en tout temps capable de respecter parfaitement la parité statistique et la parité de performance.

A. Les difficultés liées à l'évaluation de la parité de performance

Comment mesurer la parité de performance. Tel qu'identifié, assurer une parité de performance réclame d'assurer la parité de la « performance prédictive »[669] d'un SIA entre membres de différents groupes[670]. Or, il n'est pas si simple de mesurer cette performance du SIA comme le démontre le débat entourant COMPAS.

Illustration de la problématique par le débat sur COMPAS. Comme nous l'avons mentionné, le journal ProPublica a accusé COMPAS d'être un outil discriminatoire[671]. Le journal est arrivé à cette conclusion en démontrant que l'outil produit plus de faux positifs et de faux négatifs à l'égard des personnes noires qu'à l'égard des personnes blanches[672]. Cette révélation a soulevé

[668] *Supra,* Chapitre 2, Section IV.
[669] S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 1.
[670] *Supra*, Chapitre 1, Section IV, Sous-section A(ii).
[671] *Id.*
[672] J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3. ; *Supra,* Chapitre 1, Section IV, Sous-section A.

un tollé et a permis à plusieurs de proposer que l'outil COMPAS était discriminatoire et biaisé[673]. Or, l'entreprise qui a créé COMPAS, Northpoint Inc., a répliqué à cette dénonciation en formulant que les taux de faux positifs et de faux négatifs n'étaient pas, en fait, des mesures appropriées pour évaluer la capacité d'un tel outil à respecter une parité de performance[674]. Selon l'entreprise c'est la valeur prédictive de l'outil qui devrait plutôt être considérée. Or, Northpoint Inc. propose que COMPAS possède une valeur prédictive similaire entre personnes blanches et personnes noires[675].

Distinctions entre les notions de taux de faux positifs/négatifs et valeurs prédictives. Pour comprendre le nœud du problème, il convient d'identifier brièvement ce qu'est une valeur prédictive ainsi que les notions de faux positifs et de faux négatifs.

Ce que signifie valeur prédictive. Prenons, à titre d'exemple, un SIA utilisé pour diagnostiquer un cancer. En pareilles circonstances, la valeur prédictive positive du SIA reflèterait la probabilité qu'une personne ayant testé positif soit véritablement atteinte d'un cancer. Ainsi, une valeur prédictive positive de 66 % indiquerait que si une personne a obtenu un résultat positif, il existe 66 % de chances que ce résultat s'avère exact. La valeur prédictive négative reflète, quant à elle, la probabilité qu'une personne ayant testé négatif ne soit pas atteinte du cancer.

Sa distinction avec les notions de faux négatifs et de faux négatifs. Si ce même SIA possède un taux de faux positifs de 33 %, cela signifierait que si 100 personnes sont analysées par le SIA et que ces dernières ne sont pas atteintes d'un cancer, alors, en moyenne, 33 d'entre elles seront identifiées à tort comme atteintes d'un cancer. Un taux de faux négatifs de 33 % identifie, quant à lui que si 100 personnes sont analysées alors qu'elles sont atteintes d'un cancer, alors le SIA identifiera en moyenne, erronément 33 de ces personnes comme n'étant pas atteintes du cancer.

Dans le contexte de COMPAS. Dans le contexte d'un outil d'évaluation des risques de récidives comme COMPAS, la valeur prédictive positive illustre donc la probabilité qu'une personne qui a

---

[673] Voir, par exemple : PIERRE-LUC DÉZIEL, préc., note 163, 258; K. CRAWFORD, préc., note 364, p. 128.

[674] WILLIAM DIETERICH, CHRISTINA MENDOZA, et TIM BRENNAN, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity : Performance of the COMPAS Risk Scales in Broward County*, Northpointe Inc., Northpointe Inc. Research Department, 2016, p. 2-3. Techniquement, les auteurs parlent de « *predictive parity* », mais il s'agit essentiellement du même concept.

[675] *Id.*

testé « positif » récidivera alors que la valeur prédictive négative illustre la probabilité qu'une personne qui a testé « négatif » ne récidivera pas[676]. Le taux de faux positifs quant à lui illustre le pourcentage d'individus qui testeront « positifs » alors qu'ils ne récidiveront pas et le taux de faux négatifs, le pourcentage d'individus qui testeront « négatifs » alors qu'ils récidiveront.

Personne n'a raison, personne n'a tort. À noter que la prétention de Northpoint Inc. voulant que COMPAS produit des résultats qui présentent des valeurs prédictives similaires entre personnes noires et personnes blanches est généralement acceptée dans la littérature scientifique[677]. À ce titre, accuser COMPAS d'être incapable de respecter une parité de performance réclame de statuer sur la question suivante : Quelle est la mesure appropriée pour évaluer cette parité : les taux de faux positifs/faux négatifs ou les valeurs prédictives? Malheureusement, tant les taux de faux positifs/faux négatifs, que les valeurs prédictives sont des métriques utiles et reconnues pour mesurer la performance d'un outil[678]. Ainsi, des auteurs tels que Corbett-Davies et Goel proposent que si la parité de performance (qu'ils nomment « *classification parity* »[679]) « *requires that measure be equal across group* »[680] alors les arguments en faveur de la valeur prédictive ou des taux de faux positifs/négatifs « *are all arguments for some type of classification parity*. »[681]. À ce titre, il n'est pas possible d'identifier qui a tort entre Northpoint Inc. et ProPublica. Ils ne font qu'utiliser des métriques distinctes pour mesurer la parité de performance.

Pourquoi ne pas maximiser toutes ces mesures ? Pourquoi alors ne pas réclamer au SIA de satisfaire à la fois une parité entre : taux de faux positifs, taux de faux négatifs, valeurs prédictives positives et valeurs prédictives négatives ? Agir ainsi n'est néanmoins pas toujours possible. En effet, l'auteur Chouldechova identifie que :

> if an instrument satisfies predictive parity—that is, if the PPV [positive predictive value] is the same across groups—but the prevalence differs between groups, the

[676] *Id.*, p. 34.
[677] Jon KLEINBERG, *Inherent Trade-Offs in Algorithmic Fairness*, *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, coll. SIGMETRICS '18, New York, NY, USA, Association for Computing Machinery, 12 juin 2018, p. 40, p. 1-2, DOI : 10.1145/3219617.3219634; S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 11.
[678] S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 11.
[679] *Id.*
[680] *Id.*
[681] *Id.*

> instrument cannot achieve equal false positive and false negative rates across those groups.[682]

À ce titre, l'auteur démontre que, dans le cas de COMPAS, les importantes iniquités entre les taux de faux positifs et de faux négatifs produits à l'égard des personnes blanches et des personnes noires étaient, en fait, la conséquence directe du désir de Northpoint Inc. de satisfaire une parité entre valeurs prédictives[683]. En effet, des théorèmes mathématiques prouvent une impossibilité de satisfaire une parité entre la valeur prédictive et les taux de faux positifs / faux négatifs dans la majorité des circonstances[684].

B. L'impossibilité mathématique de satisfaire la parité statistique et la parité de performance

Une impossibilité mathématique démontrée. Ainsi, la problématique de l'équité en matière de SIA est affectée par un enjeu fondamental. Un grand nombre d'auteurs spécialisés en AA expliquent que d'assurer un parfait respect conjoint de la parité statistique, d'une parité entre valeurs prédictives, d'une parité entre taux de faux positifs et d'une parité entre taux de faux négatifs est tout simplement mathématiquement impossible dans la majorité des circonstances[685]. En effet :

> It turns out there are certain combinations of fairness criteria that—although they are each individually reasonable—simply cannot be achieved simultaneously […]. There are mathematical theorems demonstrating this impossibility.[686]

L'impossibilité d'une parité de performance parfaite. Ainsi, les auteurs Kleinberg, Mullainathan et Raghavan auraient démontré l'impossibilité mathématique de satisfaire en même temps une

---

[682] A. CHOULDECHOVA, préc., note 323, 7.
[683] *Id.*, 2.
[684] M. KEARNS et A. ROTH, préc., note 321, p. 77-78 (page pdf).
[685] Voir par exemple : *Id.*, p. 76-78; Solon BAROCAS, Moritz HARDT et A. NARAYANAN, *Fairness and Machine Learning : Limitation and opportunities*, Fairmlbook.org, New York, 2019, p. 54, en ligne : <https://fairmlbook.org/> (consulté le 24 juin 2022); Kailash Karthik SARAVANAKUMAR, « The Impossibility Theorem of Machine Fairness -- A Causal Perspective », (2021) arXiv, 2021 *arXiv:2007.06024*, 3-5, DOI : 10.48550/arXiv.2007.06024; J. DRESSEL et H. FARID, préc., note 85, 1. référant à l'éditorial de SAM CORBETT-DAVIES, EMMA PIERSON, AVI FELLER, et SHARAD GOEL, « A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. », *Washington Post* (17 octobre 2016), en ligne : <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> (consulté le 24 juin 2022).
[686] M. KEARNS et A. ROTH, préc., note 321, p. 76 (page pdf).

parité entre taux de faux négatifs, une parité entre taux de faux positifs et une parité entre valeurs prédictives[687]. Ce « *Fairness Impossibility Theorem* »[688], implique que « *except in highly constrained special cases, it is not possible to satisfy these three constraints simultaneously* »[689]. Ainsi, dans l'immense majorité des cas, satisfaire une parité de performance réclame nécessairement de choisir l'un des trois scénarios suivants :

> (i) Allow unequal false negative rates to retain equal PPV's [positive predictive value] and achieve equal false positive rates
>
> (ii) Allow unequal false positive rates to retain equal PPV's [positive predictive value] and achieve equal false negative rates
>
> (iii) Allow unequal PPV's [positive predictive value] to achieve equal false positive and false negative rates[690].

L'impossibilité de satisfaire une parité de performance et une parité statistique. Finalement, plusieurs spécialistes en intelligence artificielle soulèvent également qu'il est impossible de satisfaire une parité entre valeurs prédictives et une parité statistique dans les circonstances où, en présence de deux groupes, un groupe bénéficie d'un taux plus élevé de résultats positifs que l'autre[691].

Conséquences de cette impossibilité. En d'autres termes, s'il est possible de concevoir des algorithmes capables de respecter certaines formes d'équité, il faut accepter qu'il sera souvent « impossible d'obtenir un algorithme capable de respecter chacune d'elles (traduction libre) »[692]. Cette impossibilité n'est pas liée à un manque de ressources ou même à l'usage d'algorithmes

---

[687] J. KLEINBERG, préc., note 677, p. 2, 4, 11-12; K. K. SARAVANAKUMAR, préc., note 685, 3; Mario BRCIC et Roman V. YAMPOLSKIY, « Impossibility Results in AI: A Survey », 2022, 5, DOI : 10.48550/arXiv.2109.00484. Les *balance for the negative class* et *balance for the positive class* identifié par le texte de Kleinberg et al. réfèrent respectivement à la parité entre taux de faux négatifs et à la parité entre taux de faux positifs. En effet, à la p. 2 les auteurs proposent que "*These balance conditions can be viewed as generalizations of the notions that both groups should have equal false negative and false positive rates*." Similairement, leur catégorie "*calibration within groups*" réfère à la parité entre valeurs prédictives (voir p. 1 et 4).

[688] M. BRCIC et R. V. YAMPOLSKIY, préc., note 687, 5.

[689] J. KLEINBERG, préc., note 677, p. 17.

[690] A. CHOULDECHOVA, préc., note 323, 12.

[691] S. BAROCAS, M. HARDT et A. NARAYANAN, préc., note 685, p. 51, 54; K. K. SARAVANAKUMAR, préc., note 685, 3-7; Ziyuan ZHONG, « A Tutorial on Fairness in Machine Learning », *Toward data science* (19 juin 2020), en ligne : <https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb> (consulté le 24 juin 2022).

[692] E. BIRD et al., préc., note 323, p. 32.

plutôt que d'humains. Il n'est tout simplement pas possible, pour des raisons mathématiques, d'assurer conjointement une parité statistique, une parité entre taux de faux négatifs, une parité entre taux de faux positifs et une parité entre valeurs prédictives parfaites dans la majorité des circonstances[693]. À ce titre, nous devons accepter que le respect du principe d'équité ne puisse jamais être atteint à la perfection. Il sera presque toujours nécessaire de faire des compromis non seulement au regard des autres principes, mais à l'égard du principe lui-même.

[693] M. KEARNS et A. ROTH, préc., note 321, p. 78 (page pdf).

Chapitre 3 - La solution : la rédaction d'analyses d'impacts

La nécessité de moduler nos attentes. Aux vues des tensions susmentionnées, il est clair que nous ne pouvons pas exiger des entreprises et des organisations publiques de produire et d'utiliser des SIA qui seraient à la fois parfaitement explicables, parfaitement exacts, parfaitement sécuritaires et parfaitement équitables[694]. L'existence des tensions entre ces principes nous impose de moduler nos attentes à l'égard des outils d'IA et d'adopter un regard réaliste quant à leurs limites.

Nécessité d'une solution. En même temps, cette opposition entre les principes place les développeurs et utilisateurs de SIA devant des choix difficiles. À bien des égards, ceux-ci peuvent, de façon compréhensive, avoir l'impression d'être placés devant un « *catch-22* »[695]. Puisque les mesures devant être mises en œuvre afin de respecter un principe réclament parfois d'en sacrifier un autre et que, de surcroit, le régime législatif applicable est vague et sujet à des interprétations contradictoires, même une organisation soucieuse de respecter la Loi s'expose à un risque d'être assujettie à des sanctions très sévères. Or, après avoir identifié pourquoi le respect des principes est si complexe, il convient maintenant de proposer comment les organisations peuvent développer des SIA en limitant le plus possible les risques de contrevenir à leurs obligations légales.

Solution proposée : les analyses d'impacts. Nous proposons d'évaluer les arbitrages nécessaires entre différents principes en incorporant cette analyse aux Évaluations des facteurs relatifs à la vie privée (en anglais « *Privacy Impact assessment* ») (ci-après « EFVP » et « PIA » respectivement)[696]. L'EFVP est une forme d'analyse d'impacts couramment utilisée en matière

[694] *Supra*, Chapitre 2.

[695] S. ROBBINS, préc., note 49, 507.

[696] En effet, le Guide du Commissariat à la vie privée du Canada au sujet du processus d'Évaluation des facteurs relatifs à la vie privée traduit les *Privacy Impact Assessement* à *Évaluation des facteurs à la vie privée* en français. Voir par e.g. COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Nos attentes : Guide du Commissariat au sujet du processus d'évaluation des facteurs relatifs à la vie privée » (13 octobre 2011), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/evaluations-des-facteurs-relatifs-a-la-vie-privee/gd_exp_202003/> (consulté le 28 juillet 2022); COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Expectations: OPC's Guide to the Privacy Impact Assessment Process » (13 octobre 2011), en ligne : <https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/gd_exp_202003/> (consulté le 28 juillet

de protection des renseignements personnels[697]. Nous proposons donc aux organisations de mobiliser ce processus d'analyse afin d'identifier les arbitrages nécessaires entre différents principes. Pour ce faire, nous présentons d'abord (I) ce qu'est une EFVP pour ensuite (II) explorer pourquoi cet outil permet de gérer efficacement les tensions identifiées. Finalement, nous proposons (III) quelques éléments qui devraient selon nous être incorporés à l'EFVP afin de faciliter les arbitrages entre les différents principes.

Section I - Ce qu'est une EFVP et sa place en droit québécois

Ce qu'est une EFVP. Une EFVP est un outil de gestion des risques[698] préventif et évolutif[699] qui « aide les institutions à s'assurer qu'elles respectent les exigences de la loi et à déterminer l'incidence éventuelle de leurs programmes et de leurs activités sur la vie privée d'individus »[700]. Ce processus réclame d'identifier et de mesurer les risques que présentent un projet, un programme, une politique, une technologie, un système ou une activité sur la vie privée des individus (ci-après « le projet »)[701]. L'EFVP permet aussi de recommander les mesures qui permettront de réduire ou d'éliminer les risques identifiés[702]. Dans cette section, nous explorons

2022); COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Privacy Impact Assessments (PIAs) » (16 septembre 2019), en ligne : <https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/> (consulté le 28 juillet 2022).

Le RGPD prévoit l'utilisation d'*Analyses d'impact à la protection des données* (AIPD). Les distinctions entre PIA et AIPD sont généralement

[697] En effet, le Guide du Commissariat à la vie privée du Canada au sujet du processus d'Évaluation des facteurs relatifs à la vie privée traduit les *Privacy Impact Assessement* à *Évaluation des facteurs à la vie privée* en français. Voir par e.g. COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696; COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696; COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696.

Le RGPD prévoit l'utilisation d'*Analyses d'impact à la protection des données* (AIPD). Les distinctions entre PIA et AIPD sont généralement

[698] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696, sect. 4.

[699] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 7.

[700] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696, sect. 4.

[701] David WRIGHT, « The state of the art in privacy impact assessment », (2012) 28 *Computer Law and Security Review: The International Journal of Technology and Practice* 54-61, 55; Konstantina VEMOU et Maria KARYDA, « Evaluating privacy impact assessment methods: guidelines and best practice », (2019) 28-1 *ICS* 35-53, 35, DOI : 10.1108/ICS-04-2019-0047.

[702] D. WRIGHT, préc., note 701, 55; K. VEMOU et M. KARYDA, préc., note 701, 35; COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 21; COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, préc., note 696, sect. 4.

(A) les nouvelles obligations relatives à la rédaction d'EFVP dans le régime québécois des renseignements personnels et (B) les différents éléments d'une EFVP[703].

A. Les nouvelles obligations relatives à la rédaction d'EFVP

L'importance des EFVP en droit québécois. Avec l'adoption de PL64, l'EFVP devient l'un des principaux piliers du régime de protection des renseignements personnels québécois. En effet, les organisations soumises à la *Loi sur l'accès* et à la *Loi sur le privé* devront rédiger des EFVP dans plusieurs circonstances. Ainsi, à compter du 22 septembre 2022 une EFVP devra obligatoirement être rédigée avant de communiquer des renseignements personnels « sans le consentement des personnes concernées à une personne ou à un organisme qui souhaite utiliser ces renseignements à des fins d'étude, de recherche ou de production de statistiques »[704]. De plus, à partir du 22 septembre 2023, il deviendra nécessaire pour ces organisations de procéder à une EFVP :

> avant « tout projet d'acquisition, de développement et de refonte de système d'information ou de prestation électronique de services impliquant la collecte, l'utilisation, la communication, la conservation ou la destruction de renseignements personnels »[705],

[703] Un autre outil d'analyse d'impact en matière de protection des renseignements personnels est l'évaluation de l'incidence algorithmique ou « analyse d'impact algorithmique » (ci-après « AIA »). À première vue, nous pouvons qualifier l'EFVP que nous proposons d'AIA. Plusieurs auteurs proposent l'AIA justement afin de tenir compte des difficultés spécifiques que présentent les SIA ce qui est exactement ce que nous proposons de faire. Toutefois, qualifier notre approche d'AIA ne serait pas utile. En effet, les AIA sont très mal définies dans la littérature et il n'existe actuellement aucun consensus sur les éléments constitutifs d'une AIA. Andrew D SELBST, « An Institutional View of Algorithmic Impact Assessments », (2021) 35-1 *Harvard Journal of Law and Technology* 75, 122 et 139; Emanuel MOSS, Elizabeth Anne WATKINS, Ranjit SINGH, Madeleine Clare ELISH et Jacob METCALF, *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest*, 3877437, Rochester, NY, Data & Society, 2021, p. 29, en ligne : <https://papers.ssrn.com/abstract=3877437> (consulté le 29 juillet 2022); Dillon REISMAN, Jason SCHULTZ, Kate CRAWFORD et Meredith WHITTAKER, *Algorithmic impact assessment*, AiNow, 2018, p. 3; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 5, p. 22; ÉCOLE DE LA FONCTION PUBLIQUE DU CANADA, « L'intelligence artificielle est à nos portes : Savoir quand et comment utiliser l'IA au gouvernement (vidéo) » (24 mai 2022), en ligne : <https://www.csps-efpc.gc.ca/video/artificial-intelligence-here-series/deciding-when-fra.aspx> (consulté le 29 juillet 2022).

[704] *PL64*, préc., note 5, arts. 23, 118 et 175(2); COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « Évaluation des facteurs relatifs à la vie privée », en ligne : <https://www.cai.gouv.qc.ca/espace-evolutif-modernisation-lois/thematiques/evaluation-facteurs-relatifs-vie-privee/> (consulté le 28 juillet 2022). ; Nouveaux articles 21 de la *Loi sur le privé* et 67.2.1. de la *Loi sur l'accès*.

[705] *PL64*, préc., note 5, arts. 15, 103 et 175. Nouveaux articles 3.3 de la *Loi sur le privé* et 63.5. de la *Loi sur l'accès*.

avant toute communication de renseignements personnels à l'extérieur du Québec[706] et

avant de confier « à une personne ou à un organisme à l'extérieur du Québec la tâche de recueillir, d'utiliser, de communiquer ou de conserver pour son compte »[707] un renseignement personnel.

Guides et modèles d'EFVP dans le contexte québécois. À noter que le législateur québécois ne propose aucune définition de l'EFVP dans PL64. Plus encore, nous constatons une absence de documentation gouvernementale destinée à assister les organisations publiques et privées dans leur rédaction d'EFVP. La CAI a publié en 2021 un Guide d'accompagnement à la réalisation d'EFVP. Or, ce très bref guide[708] précède l'adoption de PL64 et met en garde qu'il « sera revu à la lumière de l'adoption du projet de loi no 64 »[709]. Il n'a donc pas été rédigé au regard des modifications législatives apportées par le projet de loi. Un modèle rédigé par le professeur Vincent Gautrais et nous-mêmes est paru en mars 2022. Il s'agit, à notre connaissance, du seul modèle d'EFVP disponible publiquement qui incorpore les spécificités législatives propres au régime législatif québécois. Il s'agit cependant d'un modèle préliminaire qui est appelé à être modifié[710].

Portée étendue de l'analyse envisagée. Malgré tout, il est clair que l'analyse d'impacts que nous proposons à une portée beaucoup plus étendue que l'EFVP envisagée par le législateur québécois. En effet, l'EFVP envisagée par le législateur se résume à évaluer les risques associés à la vie privée. Les mémoires soumis par la CAI et la CDPDJ lors des travaux portant sur PL64 dénonçaient cette limite et recommandaient d'élargir l'évaluation à d'autres aspects dont les

[706] *Id.*, arts. 27, 111 et 175. Nouveaux articles 17 de la *Loi sur le privé* et 70.1. de la *Loi sur l'accès*.
[707] *Id.*
[708] Le guide ne fait que 34 pages. À titre de comparaison, le guide fourni par la Commission nationale informatique et libertés se divise en trois documents pour un total de 149 pages. COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260; COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - La méthode*, Commission nationale informatique et libertés, 2018; COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - Les modèles*, Commission nationale informatique et libertés, 2018; COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - Les bases de connaissances*, Commission nationale informatique et libertés, 2018.
[709] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 1.
[710] VINCENT GAUTRAIS et NICOLAS AUBIN, préc., note 9, voir le site web.

droits et libertés garanties par la *Charte québécoise*[711]. Or, le législateur n'a pas donné suite à ces propositions. En revanche, afin de mobiliser les EFVP afin d'arbitrer les tensions entre différents principes, il est impératif d'élargir, comme le proposent la CAI et la CDPDJ, l'évaluation aux autres principes susmentionnés.

B. Les différents éléments d'une EFVP

Les éléments d'une EFVP. Le contenu des EFVP varie selon les différents guides et modèles disponibles au public[712]. Cependant, après avoir recensé plusieurs documents portant sur la rédaction d'EFVP, les auteurs Vemou et Karyda proposent que les EFVP sont normalement composés des six éléments illustrés dans la Figure 3[713].

Figure 3. – Les éléments d'une EFVP

Il convient d'explorer succinctement chacun de ces éléments afin de mieux cerner ce qu'est une EFVP.

[711] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 89; COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 5, p. 22.
[712] K. VEMOU et M. KARYDA, préc., note 701, 36-41.
[713] *Id.*, 41, 45-50.

i. La préparation à l'EFVP

Identifier la nécessité de l'EFVP. D'abord, la préparation à l'EFVP réclame, entre autres, d'évaluer si l'organisation doit réaliser une EFVP. Ainsi, tel que susmentionné, PL64 ne réclame pas de créer des EFVP en toutes circonstances[714]. En revanche, il faut comprendre qu'il peut être opportun de produire une EFVP même dans les circonstances où elles ne sont pas expressément imposées par la Loi. En effet, réaliser une EFVP est très avantageux pour des raisons qui seront explicitées sous peu[715].

Délimiter son processus. Il convient également de délimiter le processus d'évaluation. En effet, la préparation à l'EFVP réclame d'évaluer à quel point celle-ci devra être détaillée. Ainsi, PL64 prévoit que l'EFVP doit être « proportionnée à la sensibilité des renseignements concernés, à la finalité de leur utilisation, à leur quantité, à leur répartition et à leur support. »[716] Il n'est donc pas nécessaire de déployer les mêmes efforts pour réaliser une EFVP à l'égard d'un petit projet qui n'utilise qu'une très faible quantité de renseignements personnels qu'à l'égard d'un immense projet qui utilise des renseignements personnels sensibles.

Identifier les participants. Il convient également d'identifier les personnes qui devront participer à l'EFVP ainsi que leurs responsabilités[717]. Le *Guide d'accompagnement de la CAI à la réalisation d'une EFVP* propose d'incorporer, entre autres : les personnes responsables du projet, les personnes responsables des affaires juridiques, les autorités compétentes de l'organisation et, dans certaines circonstances, les collègues de travail, les clients, les partenaires corporatifs et les sous-traitants[718].

Les personnes devant impérativement participer à l'EFVP. De plus, les modifications apportées par PL64 imposent à certaines entités de participer à la réalisation de l'EFVP. Celles-ci sont : le responsable de la protection des renseignements personnels et le comité sur l'accès à l'information et la protection des renseignements personnels.

---

[714] *Supra,* Chapitre 3, Section I, Sous-section A.
[715] *Infra,* Chapitre 3, Section II.
[716] *PL64*, préc., note 5, arts. 15 et 103. Nouveaux articles 3.3 de la *Loi sur le privé* et 63.5. de la *Loi sur l'accès*.
[717] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 1-5.
[718] *Id.*, p. 3-5.

Le responsable de la protection des renseignements personnels. Ainsi, les modifications apportées à la *Loi sur l'accès* et à la *Loi sur le privé* prévoient que la « personne ayant la plus haute autorité »[719] au sein de l'organisme public ou l'entreprise exerce la fonction de responsable de la protection des renseignements personnels[720]. Une entreprise privée peut déléguer cette fonction par écrit à n'importe qui alors qu'un organisme public ne peut le déléguer qu'à l'un de ses membres, à un membre de son conseil d'administration ou à un membre du personnel de direction[721].

Les fonctions du responsable dans une entreprise privée. Les modifications imposées à la *Loi sur l'accès* et à la *Loi sur le privé* prévoient que le responsable de la protection des renseignements personnels doit être consulté lors de la rédaction de l'EFVP, et ce, dès le début du projet[722]. Il doit également participer à l'évaluation des risques liés aux incidents de confidentialité[723] et peut suggérer des mesures de protection[724]. Or, il s'agit là de deux éléments centraux des EFVP qui seront analysées sous peu[725]. En vertu des modifications apportées à la *Loi sur le privé*, la personne qui occupe cette fonction détient plusieurs responsabilités dont : approuver les politiques et pratiques encadrant la gouvernance à l'égard des renseignements personnels[726], enregistrer certaines communications avec la CAI[727], répondre aux demandes d'accès ou de rectification[728] et motiver les refus d'acquiescer à ces demandes d'accès[729].

Les fonctions du responsable dans un organisme public. Les nouveaux articles de la *Loi sur l'accès* prévoient aussi que le responsable a la responsabilité d'enregistrer certaines

[719] *PL64*, préc., note 5, arts. 1 et 103. Nouveaux articles 8 de la *Loi sur l'accès* et 3.1. de la *Loi sur le privé.*
[720] *Id.*
[721] *Id.*
[722] *Id.*, art. 103. Nouvel article 3.7. de la *Loi sur le privé.* La *Loi sur l'accès* prévoit que c'est le comité sur l'accès à l'information et la protection des renseignements personnels qui doit participer au processus d'EFVP. En revanche, le responsable de la protection des renseignements personnels est un membre de ce comité. Voir *Id.*, arts. 1 et 15. Nouveaux articles 8.1 et 63.6. de la *Loi sur l'accès.*
[723] *PL64*, préc., note 5, arts. 15 et 103. Nouveaux articles 63.10 de la *Loi sur l'accès* et article 3.7. de la *Loi sur le privé.*
[724] *Id.*, art. 103. Nouvel article 3.4. de la *Loi sur le privé*.
[725] *Infra,* Chapitre 3, Section I, Sous-section B(iii).
[726] *PL64*, préc., note 5, art. 103. Nouvel article 3.2 de la *Loi sur le privé.*
[727] *Id.* Nouvel article 3.5. de la *Loi sur le privé.*
[728] *Id.*, art. 124. Nouvel article 32 de la *Loi sur le privé.*
[729] *Id.*, art. 126. Nouvel article 34 de la *Loi sur le privé.*

communications avec la CAI[730]. En revanche, c'est surtout à titre de membre du comité sur l'accès à l'information et la protection des renseignements personnels que le responsable de la protection des renseignements personnels d'une organisation publique exerce ses fonctions.

Les participants à l'EFVP en vertu de la Loi sur l'accès. Pour les organismes publics, il sera également nécessaire d'incorporer le comité sur l'accès à l'information et la protection des renseignements personnels au processus d'EFVP. En effet, les modifications apportées à la *Loi sur l'accès* prévoient qu'un organisme public doit nommer un comité sur l'accès à l'information et la protection des renseignements personnels afin de l'assister dans « l'exercice de ses responsabilités et dans l'exécution de ses obligations »[731]. PL64 prévoit que ce comité doit être consulté lors la rédaction de l'EFVP, et ce, dès le début d'un projet[732]. Le comité peut également suggérer des mesures de protection des renseignements personnels identifiées lors de l'EFVP[733]. Le comité doit être composé du responsable de la protection des renseignements personnels, de la personne responsable de l'accès aux documents et de toute autre personne « dont l'expertise est requise incluant, le cas échéant, le responsable de la sécurité de l'information et le responsable de la gestion documentaire »[734].

ii. L'analyse du projet

Décrire le projet. Ensuite, l'EFVP réclame de décrire le projet ou le système qu'elle évalue. Cela demande, entre autres, d'identifier les renseignements personnels concernés par le projet, d'illustrer leur cycle de vie et d'expliciter chacune des différentes opérations prévues à leur égard[735]. Nous proposons également d'identifier très clairement le niveau de sensibilité des renseignements et des fins qui leur sont envisagées[736]. Il convient également d'identifier les

[730] *Id.*, art. 15. Nouvel article 63.8. de la *Loi sur l'accès.*
[731] *Id.*, art. 1. Nouvel article 8.1 de la *Loi sur l'accès.*
[732] *Id.*, art. 15. Nouvel article 63.5. de la *Loi sur l'accès.*
[733] *Id.* Nouvel 63.6. de la *Loi sur l'accès.*
[734] *Id.*, art. 1. Nouvel article 8.1 de la *Loi sur l'accès.*
[735] K. VEMOU et M. KARYDA, préc., note 701, 45; COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 6 et 8; Dariusz KLOZA, Alessandra CALVI, Simone CASIRAGHI, Sergi MAYMIR et Nikolaos IOANNIDIS, *Analyse d'impact relative à la protection des données dans l'Union européenne : élaboration d'un modèle de rapport du processus d'analyse*, d.pia.lab Note de Politique N° 1/2020, Bruxelles, Belgique, Laboratoire bruxellois de l'analyse d'impact relative à la protection des données et de la vie privée (d.pia.lab), 2020, p. 8, en ligne : <https://hal.archives-ouvertes.fr/hal-03332455/>.
[736] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 8, 13-15.

objectifs du projet, leur importance et leur légitimité. Enfin, il convient d'évaluer si le projet constitue une solution proportionnelle aux objectifs identifiés[737].

Identifier les cibles en matière de la vie privée. Des auteurs proposent également d'identifier les « cibles » du projet en matière de protection de la vie privée[738]. Ces cibles représentent des obligations issues de diverses sources telles que la Loi, les lignes directrices destinées au développement et à l'usage responsable de SIA et les politiques internes de l'entreprise[739]. Nous proposons de considérer la législation « *as a central starting point to define privacy targets* »[740]. Agir ainsi « *saves companies cost and time* »[741] et permet de s'assurer que « *If a company commits to invest in a (potentially cost intensive) PIA, the minimum outcome it expects is the legal compliance of its operations.* »[742]

iii. L'analyse des risques

Ce qu'est un risque. L'un des objectifs de l'EFVP est d'identifier les différents risques et menaces à la vie privée qui pourraient être créés par le projet. La Commission Nationale de l'Informatique et des Libertés (ci-après « CNIL ») définie le risque comme « un scénario hypothétique qui décrit un événement redouté et toutes les menaces qui permettraient qu'il survienne »[743]. L'EFVP impose donc d'analyser différents scénarios et sources potentielles de risques[744]. Le risque doit être évalué au regard de sa probabilité et de ses impacts appréhendés[745]. La probabilité du risque reflète la vraisemblance que le risque se manifeste. Un événement peu probable est un événement qui a peu de chances de se produire. Mesurer l'impact réclame d'analyser la gravité des conséquences de cette manifestation[746]. Au Québec, mesurer les impacts potentiels d'un risque réclame d'évaluer, notamment :

---

[737] *Id.*, p. 4.
[738] David WRIGHT et Paul DE HERT (dir.), *Privacy impact assessment*, coll. Law, governance and technology series, volume 6, Dordrecht ; New York, Springer, 2012, p. 338.
[739] K. VEMOU et M. KARYDA, préc., note 701, 46.
[740] D. WRIGHT et P. DE HERT (dir.), préc., note 738, p. 338.
[741] *Id.*, p. 338.
[742] *Id.*, p. 337.
[743] COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, préc., note 708, p. 6.
[744] K. VEMOU et M. KARYDA, préc., note 701, 46.
[745] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 19-20.
[746] COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, préc., note 708, p. 6.

« la sensibilité des renseignements concernés »[747],

« la probabilité que les renseignements concernés soient utilisés à des fins préjudiciables »[748] et

« les conséquences appréhendées qui pourraient résulter de leur utilisation »[749].

Les contraventions à la Loi sont-elles des risques ? Le guide fourni par la CAI se distingue de la documentation internationale relative aux EFVP sur un point important : elle traite le non-respect de la Loi comme un risque. En effet, la CAI expose comme « exemples de risques sur la vie privée »[750] des actes prohibés par la Loi parmi lesquelles : la « Divulgation non autorisée de renseignements personnels » [751], le fait que l'« Objectif du projet [n'est] pas suffisamment important ou non légitime »[752] et le fait que l'« Intrusion dans la vie privée [est] disproportionnée par rapport à l'objectif visé par le projet »[753]. D'autres modèles d'EFVP comme celui de la CNIL propose au contraire d'évaluer le non-respect de la Loi dans une optique distincte de la « gestion des risques »[754]. Selon elle, l'évaluation de la conformité à la Loi ne réclame pas de considérer, comme l'évaluation fondée sur les risques, « la nature, la gravité et la vraisemblance des risques encourus »[755] puisque les principes et droits fondamentaux fixés par la Loi représentent des éléments « non négociables »[756].

Risqué pour qui ? Une autre controverse concerne les personnes visées par les risques. Les risques évalués doivent-ils être limités à ceux qui visent les personnes concernées par les renseignements personnels ou doivent-ils évaluer un plus grand nombre d'acteurs ? Certains auteurs proposent que l'évaluation ne devrait pas être limitée aux personnes concernées et estiment que les « *impacts on groups of people and the society as a whole should be*

[747] *PL64*, préc., note 5, arts. 16 et 103. Nouveaux articles 3.7 de la *Loi sur le privé* et 63.10 de la *Loi sur le public*.
[748] *Id.*
[749] *Id.*
[750] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 16.
[751] *Id.*
[752] *Id.*
[753] *Id.*
[754] COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, préc., note 708, p. 3.
[755] *Id.*
[756] *Id.*

considered »[757]. Selon Vemou et Karyda, une minorité d'auteurs et d'organisations proposent également d'évaluer les risques guettant l'organisation responsable du projet[758].

iv. Les mesures mises en place pour mitiger ces risques

Identifier les mesures permettant de mitiger les risques. Les EFVP doivent également formuler des recommandations proposant différentes mesures ou stratégies[759] pouvant être mises en œuvre afin de mitiger les risques identifiés[760]. Celles-ci peuvent tenter de limiter la probabilité de manifestation des risques ou d'en réduire les impacts[761]. La CNIL propose d'évaluer trois « natures » de mesures soit :

> 1. mesures portant spécifiquement sur les données du traitement : chiffrement, anonymisation, cloisonnement, contrôle d'accès, traçabilité, etc. ;
>
> 2. mesures générales de sécurité du système dans lequel le traitement est mis en œuvre : sécurité de l'exploitation, sauvegardes, sécurité des matériels, etc. ;
>
> 3. mesures organisationnelles (gouvernance) : politique, gestion des projets, gestion des personnels, gestion des incidents et violations, relations avec les tiers, etc.[762]

Mesures imposées. Le régime québécois impose l'implémentation de certaines mesures organisationnelles. Celles-ci regroupent notamment les politiques, pratiques et règles de gouvernance que nous avons identifiées dans le premier chapitre[763]. Tel qu'identifié, les mesures plus techniques, quant à elles, ne sont généralement pas explicitées par la *Loi sur l'accès* et la *Loi sur le privé* qui se contentent de réclamer des mesures de sécurité « raisonnables »[764]. L'importance des mesures devant être mises en œuvre doit être évaluée, notamment, au regard

[757] K. VEMOU et M. KARYDA, préc., note 701, 47.
[758] *Id.*, 39 et 47.
[759] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 21.
[760] David WRIGHT, « Making Privacy Impact Assessment More Effective », (2013) 29-5 *The Information Society* 307-315, 312, DOI : 10.1080/01972243.2013.825687; COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, préc., note 708, p. 7; K. VEMOU et M. KARYDA, préc., note 701, 48-49.
[761] K. VEMOU et M. KARYDA, préc., note 701, 49.
[762] COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, préc., note 708, p. 7.
[763] *Supra,* Chapitre 1, Section III, Sous-section B.
[764] *Supra,* Chapitre 1, Section III, Sous-section C.

de la probabilité des risques et de leurs impacts[765]. L'EFVP doit également évaluer les risques dits résiduels soit les risques qui subsistent malgré l'implémentation des mesures[766].

v. Le rapport final

Documenter le processus. Enfin, il convient de documenter l'évaluation et de produire le rapport final de l'EFVP[767]. Ce rapport doit contenir l'information relative aux éléments susmentionnés.

Rendre l'EFVP disponible au public. Il convient également d'évaluer si l'EFVP doit être disponible au public. En effet, il se peut que certaines sections de l'EFVP soient trop sensibles pour être publiées[768]. Si tel est le cas, il est possible de simplement retirer ces sections ou les insérer dans une annexe confidentielle[769]. Il peut également être opportun de présenter une version du rapport avant sa complétion à différentes parties prenantes au projet. Agir ainsi permettra à celles-ci d'insérer leurs recommandations avant la rédaction finale du rapport[770].

Approbation. Il convient aussi de faire approuver le rapport final par l'organisation responsable du projet[771].

vi. Le suivi

Obtenir les autorisations nécessaires. Au Québec, l'EFVP n'a pas à être acceptée par les autorités de contrôle comme c'est le cas en Europe[772]. À ce titre, l'organisation n'a pas à attendre d'autorisation avant de pouvoir commencer son projet[773].

Nommer un responsable de l'implémentation des recommandations. En revanche, une personne de l'organisation responsable du projet (préférablement le responsable de la

[765] K. VEMOU et M. KARYDA, préc., note 701, 49.
[766] COMMISSION D'ACCÈS À L'INFORMATION, préc., note 260, p. 22; K. VEMOU et M. KARYDA, préc., note 701, 49.
[767] K. VEMOU et M. KARYDA, préc., note 701, 49.
[768] D. WRIGHT, préc., note 760, 312; K. VEMOU et M. KARYDA, préc., note 701, 49.
[769] D. WRIGHT, préc., note 760, 312; K. VEMOU et M. KARYDA, préc., note 701, 49.
[770] D. WRIGHT, préc., note 760, 312.
[771] K. VEMOU et M. KARYDA, préc., note 701, 50.
[772] *RGPD*, préc., note 428, art. 36.
[773] K. VEMOU et M. KARYDA, préc., note 701, 50.

protection des renseignements personnels) devrait être déclarée responsable de l'implémentation des recommandations formulées dans l'EFVP[774].

Changements importants. Finalement, l'EFVP devrait être révisée en présence de tout changement important affectant le projet[775].

Section II - Les avantages des EFVP dans le contexte des tensions.

Avantages des EFVP. Nous proposons que les EFVP peuvent assister les organisations désirant développer ou utiliser des SIA au regard des tensions susmentionnées. En effet, les EFVP permettent, selon nous, de mieux gérer ces tensions (A) parce qu'elles imposent une réflexion avant la réalisation du projet, (B) parce qu'elles sont suffisamment flexibles pour incorporer les réalités issues des tensions et (C) parce qu'elles permettent de protéger l'organisation utilisant ou développant le SIA.

A. Les avantages conférés par l'approche préventive

Approche *ex ante.* Une EFVP est un processus préventif. Elle doit être réalisée avant la conception et la mise en œuvre du projet étudié[776]. Son rôle est d'identifier les risques qui guettent un projet avant qu'ils ne se manifestent[777]. À ce titre, l'EFVP est un processus en amont qui repose sur une approche proactive plutôt que réactive face aux risques[778].

Moins coûteux. Or, l'adoption d'une approche en amont est particulièrement utile en raison des tensions susmentionnées. En effet, il est substantiellement moins coûteux et plus facile pour une entreprise ou un organisme public d'identifier les arbitrages nécessaires entre les tensions dès le début d'un projet.

Le choix du modèle, des jeux d'entrainement et de la valeur cible. Par exemple, la tension entre l'explicabilité et l'exactitude réclame d'évaluer quel serait le modèle de SIA qu'il conviendrait

[774] *Id.*
[775] *Id.*
[776] D. WRIGHT, préc., note 701, 55.
[777] *Supra,* Chapitre 3, Section I, Sous-section B(iii) et (iv).
[778] E. MOUCHARD, préc., note 1, p. 282.

d'utiliser[779]. Doit-on favoriser un modèle plus transparent ou favoriser un modèle plus précis[780]. De façon similaire, nous avons identifié l'importance du choix de la valeur cible. Tel que démontré, un choix judicieux de la valeur cible peut faire la différence entre un SIA qui renforce des iniquités à l'égard de certains groupes vulnérables et un SIA qui favorise un traitement qui leur est plus favorable sans néanmoins avoir à sacrifier sa précision[781]. De plus, nous avons identifié l'importance d'entrainer les SIA sur des jeux d'entrainement diversifiés afin de s'assurer que ceux-ci puissent garantir une certaine parité de performance à l'égard des membres de différents groupes[782]. En revanche, tous ces exemples de mesures devraient être mis en œuvre lors de la phase de conception ou d'entrainement des modèles de SIA.

Exercice moins coûteux. En effet, la décision de modifier son modèle de SIA, sa valeur cible ou ses jeux d'entrainement est une décision qui peut s'avérer très coûteuse une fois le modèle entrainé et déployé. Dans le cas des SIA, modifier un projet en cours de route implique d'avoir investi des ressources, de l'expertise et du temps dans le développement et l'entrainement d'un modèle qui ne sera plus utilisé. Cela peut également réclamer d'entrainer un nouveau modèle, de collecter de nouveaux renseignements et de devoir tester son nouvel outil. Pour ces mêmes raisons, le choix de modifier ou d'entrainer un nouveau modèle peut également retarder le déploiement de l'outil puisqu'il ne sera pas nécessairement possible de déployer le nouveau modèle immédiatement. À ce titre, l'un des avantages des EFVP est qu'elles permettent de réaliser des économies puisqu'il est « plus efficace et moins coûteux de cerner et de maitriser les risques d'atteinte à la vie privée au cours de la conception d'un programme que d'avoir à le modifier après sa mise en œuvre »[783]. En effet, les coûts liés à la modification du projet à l'étape

[779] *Supra,* Chapitre 2, Section I, Sous-section C.
[780] A. BARREDO ARRIETA et al., préc., note 25, 13.
[781] *Supra*, Chapitre 2, Section III, Sous-section C.
[782] *Supra,* Chapitre 1, Section IV, Sous-section C et Chapitre 2, Section IV, Sous-section A.
[783] COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE AU CANADA, *Rapport annuel au Parlement 2015-2016 concernant la Loi sur la protection des renseignements personnels et les documents électroniques et la Loi sur la protection des renseignements personnels: Le temps est venu de moderniser les outils du 20e siècle*, Gatineau (Québec), Commissariat à la protection de la vie privée au Canada, 2016, p. 15. Cette intuition est partagée par de nombreux praticiens en rédaction de PIA tels que Wright et De Hert. Voir D. WRIGHT et P. DE HERT (dir.), préc., note 738, p. 476.

de sa conception « ne représenteront qu'une fraction de ceux encourus plus tard (traduction libre) »[784].

Repérer l'impossibilité des projets avant leur déploiement. Pareillement, une approche en amont permet d'identifier si le déploiement du SIA est réalisable. En effet, il ne sera pas nécessairement possible de développer ou d'utiliser des SIA en toutes circonstances et pour toutes tâches. En raison des tensions susmentionnées, il pourrait exister des circonstances dans lesquelles aucun arbitrage entre principes ne sera acceptable au regard de la Loi[785]. Alternativement, il est possible que les risques légaux associés à un projet soient trop élevés. L'EFVP permet également d'évaluer les mesures de sécurité devant être mises en œuvre afin d'assurer la légalité des opérations qui concernent les renseignements personnels. Or, une entreprise peut décider que le SIA envisagé impose des coûts trop importants. Bref, l'EFVP permet à l'organisation d'explorer la faisabilité de son projet avant de débourser des montants substantiels sur un projet qui devra être abandonné.

Certaines mesures sont plus efficaces si elles sont mises en œuvre rapidement. Finalement, certaines mesures sont plus efficaces si elles sont mises en œuvre dès le début d'un projet. Par exemple, les mesures permettant d'acquérir une meilleure parité statistique ou une meilleure parité de performance réclament de collecter et d'utiliser des renseignements se rattachant à des motifs de discrimination reconnus[786]. Or, tel qu'expliqué, il ne sera pas nécessairement possible d'acquérir ces attributs protégés au stade de l'entrainement du SIA[787]. De plus, certaines mesures, comme l'entrainement des SIA sur des jeux d'entrainement diversifiés, réclament de collecter une certaine quantité de données appartenant à une multiplicité de groupes. Il convient donc de prévoir la collecte de ce type de renseignements dès le déploiement de l'outil afin de détecter dès que possible la présence d'iniquités pour les corriger et afin de garnir adéquatement

[784] D. WRIGHT, préc., note 701, 55.
"Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle de la GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, par. 54.
[786] *Supra,* Chapitre 1, Section IV, Sous-section C (ii) et Chapitre 1, Section IV, Sous-section C.
[787] *Supra*, Chapitre 2, Sections IV, Sous-section B.

de nouveaux jeux d'entrainement qui permettront d'entrainer de nouveaux modèles plus susceptibles de respecter une parité statistique ou une parité de performance.

B. Les avantages conférés par la flexibilité des EFVP

Les avantages découlant de la flexibilité des EFVP. Les EFVP sont particulièrement utiles dans un contexte de tensions en raison de leur flexibilité. Cette flexibilité existe parce qu'elles opèrent dans une optique de gestion des risques[788]. Ainsi, si ces évaluations doivent considérer la Loi et s'assurer que le projet la respecte, elles ne se limitent pas à un simple « contrôle de conformité »[789]. Elles ne sont pas de simples « *checklist[s]* »[790] qui se résument à vérifier que le projet respecte toutes les obligations légales qui lui sont imposées[791]. Bref, les EFVP ne reposent pas sur une forme de raisonnement dichotomique et invitent à réfléchir aux différents intérêts en jeu ce qui renforce leur capacité à arbitrer les tensions susmentionnées[792].

Les EFVP permettent de considérer plus que la Loi. Ainsi, en réclamant d'évaluer les risques, les EFVP permettent aux développeurs et utilisateurs de SIA d'évaluer les impacts de leurs projets au-delà des considérations imposées par la Loi. En effet, pas tous les principes jouissent de la même reconnaissance en droit québécois. Tel qu'identifié, le principe de sécurité est beaucoup plus développé et prévoit beaucoup plus d'obligations que les autres[793]. Or, le fait qu'une mesure ne soit spécifiquement imposée par la Loi ne signifie pas qu'elle devrait être ignorée. Le fait d'exposer les personnes concernées à certains risques peut préjudicier une organisation même si elle n'engage pas sa responsabilité légale. En effet, mettre sans justification des usagers à risque n'est ni une pratique éthique ni une pratique particulièrement avantageuse d'un point de vue économique. Il ne suffit pas de respecter la Loi pour ne pas être visé par une controverse et le dévoilement, par exemple, qu'une entreprise a recours à un SIA « raciste » peut entacher sa réputation de façon importante sans nécessairement l'exposer à des poursuites judiciaires. Les consommateurs réclament désormais aux entités privées de faire preuve de comportements

[788] D. KLOZA, A. CALVI, S. CASIRAGHI, S. MAYMIR et N. IOANNIDIS, préc., note 735, p. 2.
[789] D. WRIGHT et P. DE HERT (dir.), préc., note 738, p. 6.
[790] *Id.*, p. 22.
[791] *Id.*; D. KLOZA, A. CALVI, S. CASIRAGHI, S. MAYMIR et N. IOANNIDIS, préc., note 735, p. 2.
[792] D. KLOZA, A. CALVI, S. CASIRAGHI, S. MAYMIR et N. IOANNIDIS, préc., note 735, p. 2.
[793] *Supra*, Chapitre 1, Section III.

éthiques et peuvent boycotter ou défavoriser les sociétés qui s'avèrent incapables de se comporter ainsi[794].

Flexibilité nécessaire pour réaliser des arbitrages. Surtout, une approche flexible basée sur une évaluation des risques semble être plus adaptée à la nécessité d'arbitrer les différents principes qu'une « *checklist* ». En effet, puisqu'une mesure adoptée afin de respecter un principe demande potentiellement d'en sacrifier un autre, il n'est pas possible de développer une approche « *one size fits all* » à tous les SIA. Chaque mesure adoptée devra être évaluée aux regards des risques et des bénéfices qui y sont associés. Une organisation devrait, comme le demandent les EFVP[795], évaluer non seulement la probabilité des risques associés à ces mesures, mais également leur impact potentiel. À quoi bon, par exemple, collecter des attributs protégés dans un contexte où une évaluation « inéquitable » du SIA n'aurait aucun impact véritablement préjudiciable à l'égard de ses usagers. Agir ainsi peut porter inutilement atteinte à leur vie privée puisque cette mesure réclame de collecter des renseignements personnels sensibles.

La nécessité de développer une approche circonstancielle. Se sont donc les circonstances qui dicteront l'importance de favoriser un principe plutôt qu'un autre. Ainsi, nous offrons quelques exemples de contextes dans lesquels les SIA peuvent être déployés. Ces contextes réclament tous de favoriser le respect de certains principes en débit d'autres. Ces exemples permettent d'illustrer la nécessité de développer une approche flexible capable de considérer les circonstances dans lesquelles les SIA sont déployées.

Exemple de *Cortis* : favoriser l'exactitude. *Cortis*, un algorithme qui permet de détecter les arrêts cardio-vasculaires en analysant la voix est un exemple de SIA dans lequel l'explicabilité pourrait être raisonnablement sacrifiée au profit de l'exactitude. Cet algorithme est utilisé par des services d'urgence. Ces services l'utilisent afin de détecter si l'appelant souffre d'une attaque cardiaque. Cet outil « *has been shown to detect heart attacks on average 30 s faster than human operators*

[794] Voir par exemple : P. AUGER, T. M. DEVINNEY et J. J. LOUVIERE, « Global segments of socially conscious consumers: Do they exist? », dans *Global Challenges in Responsible Business*, Royaume-Uni, Cambridge University Press, 2010 aux pages 137-149, en ligne : <https://opus.lib.uts.edu.au/handle/10453/14347> (consulté le 15 août 2022).
[795] *Supra,* Chapitre 3, Section I, Sous-section B (iii) et (iv).

with an accuracy of 93% (human operators have a 73% accuracy rate). »[796] La détection des attaques permet ensuite aux services d'urgence de déployer les mesures appropriées pour venir en aide aux appelants. Bref, ce SIA sauve des vies. À ce titre, Robbins propose que cet algorithme doit être davantage jugé sur sa capacité à produire des résultats exacts que sur son explicabilité puisqu'il ne peut pas causer, selon lui, de préjudices. En effet, il propose que « *The worst-case scenario is that the algorithm does not identify someone as having a cardiac arrest who is indeed experiencing cardiac arrest.* »[797] une situation qui se serait de toute façon produite en cas de non-déploiement de l'outil[798]. À noter que nous ne sommes pas aussi catégoriques que cet auteur sur l'absence de préjudice possible. En effet, nous notons qu'il puisse exister des risques que les employés s'en remettent trop au jugement de l'outil au point qu'ils négligeraient d'évaluer eux-mêmes la voix des individus[799]. En revanche, nous partageons l'intuition de l'auteur que cet outil devrait être déployé en l'absence d'outils plus transparents qui produiraient des résultats aussi exacts.

Exemple de l'évaluation des risques de récidives : favoriser la parité de performance. De plus, il n'est pas nécessairement déraisonnable pour une entreprise chargée de développer des outils d'évaluation de risques de récidives de favoriser une parité de performance plutôt qu'une parité de traitement ou une parité statistique. Agir autrement, en effet, pousserait les juges à considérer, par exemple, la couleur de peau des accusés[800]. Qui plus est, dans *Ewert,* la Cour suprême a très clairement reconnu la possibilité de conclure à une discrimination contraire à la *Charte canadienne* en cas d'usage d'outils d'évaluation des risques moins précis à l'égard de certains groupes vulnérables[801].

En emploi : favoriser la parité statistique. En revanche, dans un contexte d'emploi il conviendrait sûrement de favoriser une parité statistique. En effet, une entreprise qui utilise un SIA biaisé en faveur, par exemple, des hommes blancs dans son processus de recrutement risque de

[796] S. ROBBINS, préc., note 49, 509.
[797] *Id.*, 509.
[798] *Id.*
[799] L'auteur reconnait cependant lui-même cette possibilité en note de bas de page. *Id.*
[800] G. PLEISS, M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, préc., note 314, p. 1.
[801] *Ewert c. Canada*, préc., note 151, par. 79.

contrevenir à la *Charte québécoise*[802]. Ici, le déploiement de SIA équitables pourrait s'avérer plus pertinent que le déploiement de SIA qui produisent des résultats dits « exacts » au regard des jeux de données qu'il dispose.

La nécessité d'une approche flexible pour évaluer la pertinence des différents arbitrages. Bref, la multiplicité des circonstances dans lesquelles les SIA peuvent se déployer ne permet pas d'identifier avec précision qu'un principe devrait, en toutes circonstances, être privilégié plutôt qu'un autre. Des arbitrages entre différents principes seront nécessaires et ceux-ci devront considérer le contexte et les circonstances qui entourent le projet.

La flexibilité permet d'évaluer la nécessité des arbitrages. Finalement, une flexibilité est requise afin d'identifier si un arbitrage entre principes est nécessaire. En effet, plusieurs nuances ont été formulées dans ce texte afin d'identifier que certaines tensions peuvent ne pas se matérialiser dans certains contextes. La décision d'utiliser un SIA moins complexe, mais plus transparent, par exemple, ne demande pas toujours de sacrifier l'exactitude[803]. De façon similaire, l'application de mesures plus équitables même au prix d'une exactitude ne va pas nécessairement violer les obligations légales d'une entreprise en matière d'exactitude[804]. Bref, la flexibilité conférée par l'approche de gestion des risques permet de faciliter l'identification d'une nécessité d'arbitrer et d'identifier quels principes devraient être favorisés lorsqu'un tel arbitrage s'avère nécessaire.

C. La possibilité de mobiliser les EFVP à des fins de protection

Mobiliser les EFVP pour se protéger. Enfin, les EFVP permettent de faciliter la défense de l'organisation responsable du projet au regard de la Loi et du public. En effet, la réalisation d'EFVP (i) impose une réflexion aux organisations qui leur permet de mieux maitriser les exigences légales qui leur sont imposées. De plus, (ii) elle produit une documentation qui permet à une entreprise ou une organisation publique de démontrer sa diligence à l'égard du public et des autorités.

[802] *Charte des droits et libertés de la personne*, préc., note 14, art. 16.
[803] A. BARREDO ARRIETA et al., préc., note 25, 30.
[804] *Supra*, Chapitre 2, Section III, Sous-section C.

i. L'imposition d'une réflexion

L'imposition d'une réflexion qui permet de faciliter le respect de la Loi. En premier lieu, les EFVP imposent une réflexion qui permet aux organisations de s'assurer que leurs activités respectent la Loi. En effet, l'EFVP impose de réfléchir à la fois à la conformité légale d'un projet et à ses impacts. À ce titre, cette analyse d'impacts incite l'organisation qui développe le SIA de « *build the capacity to understand and reflect upon what these systems actually do and whose lives are affected* »[805]. En effet, Bouchard propose que :

> les PIA constituent le premier point de contact entre le traitement envisagé par l'entreprise et la mise en œuvre de la protection adéquate de renseignements personnels traités. Les PIA s'ils s'ancrent dans un processus global, apparaissent alors comme l'élément clé d'une protection des renseignements personnels efficiente et efficace, donnant la possibilité d'un suivi durant les différentes étapes du traitement.[806]

Cette démarche de réflexion imposée par le processus d'EFVP est d'autant plus importante que le développement et l'usage de SIA opèrent dans un domaine qui connait un rythme d'innovation effréné. En effet, nous ne faisons que commencer à comprendre les impacts des SIA. Si nous pouvons reconnaitre qu'il existe des problèmes, il reste assez complexe de prédire quand ils feront surface[807]. Or, c'est précisément dans de tels contextes que les analyses d'impacts sont les plus utiles. En effet :

> Impact assessments are most useful when projects have unknown and hard-to-measure impacts on society, when the people creating the project are the ones with the knowledge and expertise to estimate its impacts but have inadequate incentives to generate the needed information[808].

Cette réflexion permet donc à l'organisation d'acquérir une vision plus précise des différentes actions qu'elle peut et ne peut pas poser et des risques légaux associés à chacune d'entre elles.

Identifier quand les mesures sont nécessaires. L'EFVP permet également à l'organisation d'évaluer adéquatement la nécessité des mesures devant être mises en œuvre. En effet, le droit

[805] E. MOSS, E. A. WATKINS, R. SINGH, M. C. ELISH et J. METCALF, préc., note 703, p. 29-30. référant à A. D. SELBST, préc., note 703.
[806] E. MOUCHARD, préc., note 1, p. 285.
[807] A. D. SELBST, préc., note 703, 121.
[808] *Id.*, 123-124.

québécois n'impose pas une obligation de sécurité absolue. Celle-ci doit être modulée et proportionnelle au regard de plusieurs facteurs[809]. À ce titre, la réflexion engagée par l'analyse d'impacts permet à une organisation non seulement d'identifier les actions qu'elle doit accomplir, mais également celles qui ne sont pas nécessaires, ce qui lui permet d'économiser des ressources.

Cette réflexion est imposée par le régime québécois de protection des renseignements personnels. Surtout, il faut comprendre que cette réflexion n'est pas simplement utile en droit québécois. Elle est imposée par la Loi, et ce, irrespectueusement des nouvelles obligations rattachées à la rédaction d'EFVP. En effet, l'approche de minimisation imposée par le droit québécois réclame une réflexion quant à la planification du cycle de vie des renseignements personnels, et ce, avant de procéder à leur collecte[810]. Sans cette réflexion il n'est pas possible pour une organisation de respecter, par exemple, le principe de détermination des fins[811]. Il serait également impossible de satisfaire l'obligation de déclarer les fins des renseignements personnels à la personne concernée avant de procéder à leur collecte[812].

ii. La production d'une documentation

Se protéger grâce à la documentation. Ensuite, les EFVP sont utiles en raison de la documentation qu'elles produisent[813]. Ainsi, Mouchard propose que le rapport produit par l'évaluation permet d'assurer « une vision d'ensemble des enjeux reliés à la protection des renseignements personnels par l'entreprise et permet un suivi tout au long du processus. »[814] Le rapport constitue donc un document de communication interne qui facilite une gestion de l'entreprise conforme aux impératifs juridiques.

Gagner la confiance du public. Surtout, une EFVP bien réalisée permet de montrer la diligence d'une organisation auprès du public. En effet, l'EFVP « *affirms that an organisation has addressed*

[809] *Supra,* Chapitre 1, Section III, Sous-section C.
[810] *Supra,* Chapitre 1, Section III, Sous-sections A et C. ; Commission d'accès à l'information, 5 avril 2022, *Services financiers Globex 2000 inc.*, préc., note 224, par. 51-52.
[811] *Supra,* Chapitre 1, Section III, Sous-section A.
[812] *PL64*, préc., note 5, arts. 104 et 107. Nouveaux articles 4 et 8 de la *Loi sur le privé.*
[813] E. MOUCHARD, préc., note 1, p. 282, 375 et 376.
[814] *Id.*, p. 282.

privacy issues and has taken reasonable steps to provide an adequate level of privacy protection. »[815]. Pour cette raison plusieurs auteurs et certaines institutions comme l'ICO recommandent aux entreprises de publier leur EFVP[816]. Agir ainsi permet à l'organisation de « diffuser ses bonnes pratiques auprès du public »[817] et de démontrer sa bonne foi[818]. L'EFVP permet également au public de mieux comprendre comment leurs renseignements personnels sont traités par l'organisation[819]. L'analyse d'impacts est donc un instrument permettant au public de mesurer la responsabilité d'une organisation et d'apprécier à quel point ses démarches sont sérieuses, ce qui permet à l'organisation, en retour, d'acquérir la confiance de la population à l'égard de ses activités[820].

Faciliter la mise en œuvre de certaines obligations. Or, acquérir la confiance du public est non seulement utile pour une organisation à des fins de « *marketing* », elle est également essentielle pour mettre en œuvre efficacement certaines mesures capables de favoriser les principes susmentionnés. Par exemple, il importe, afin d'implémenter plusieurs mesures d'équité, de collecter une quantité importante d'attributs protégés[821]. Or, tel qu'identifié, le droit québécois présente plusieurs limites à la collecte de ce type d'information. Si ces renseignements peuvent être utilisés, leur utilisation réclame très souvent un consentement exprès de la part de la personne concernée[822]. Toutefois, même si une organisation déclare n'utiliser ce type de renseignement qu'afin de favoriser le traitement de la personne concernée, la personne qui divulgue l'information, elle, ne peut pas savoir si l'entreprise se limite réellement à l'usage déclaré. Pour qu'elle divulgue l'information, celle-ci doit faire confiance à l'organisation. Or, l'acquisition d'une telle confiance ne va pas de soi. En effet, le domaine numérique et le domaine

[815] D. WRIGHT, préc., note 701, 55.
[816] Par exemple, D. WRIGHT, préc., note 760, 310 et 312. ; K. VEMOU et M. KARYDA, préc., note 701, 50. et INFORMATION COMMISSIONER'S OFFICE, préc., note 190, p. 39.
[817] E. MOUCHARD, préc., note 1, p. 385.
[818] *Id.*
[819] K. VEMOU et M. KARYDA, préc., note 701, 50.
[820] E. MOUCHARD, préc., note 1, p. 385.
[821] *Supra,* Chapitre 1, Section IV, Sous-section C et Chapitre 2, Section III.
[822] *Supra,* Chapitre 2, Section IV, Sous-section B.

des SIA suscitent la méfiance en raison de leur opacité, de l'absence de connaissances généralisées dans le public et des multiples controverses qui l'ont affecté[823].

Démontrer sa diligence aux autorités. Enfin, la rédaction d'une EFVP permet à une organisation de prouver sa diligence et sa capacité de rendre compte auprès d'organismes de contrôle comme la CAI. Comme l'indique Mouchard :

> Le principe de responsabilité en matière de protection des renseignements personnels renvoie à la personne raisonnable, prudente et diligente. Il se rattache à l'idée de la diligence raisonnable et de la preuve de celle-ci[824].

À ce titre, lorsqu'elles font l'objet d'une enquête, il convient pour les entreprises de mettre de l'avant les mesures proactives et préventives qu'elles ont adoptées afin de démontrer leur diligence raisonnable[825]. En retour, la démonstration d'une telle diligence :

> devrait aider les entreprises à éviter de faire les frais d'actions en justice en démontrant qu'elles ont pris toutes les mesures raisonnables pour ne pas prendre part à une atteinte présumée aux droits de l'homme[826].

Se protéger par la documentation. Ainsi, la documentation produite au terme de l'EFVP peut être offerte aux autorités de contrôle lorsque celles-ci enquêtent sur une organisation. Celle-ci peut être soumise à titre de preuve permettant à l'organisation de démontrer le sérieux de ses démarches et qu'elle a agi diligemment afin de limiter l'occurrence d'un incident. À ce titre, la documentation interne produite par l'EFVP peut « protéger l'entreprise en offrant une vision précise du contrôle assuré par cette dernière à chaque étape du traitement de l'information. »[827] Cette documentation « *can help to reduce or even eliminate any liability* »[828] en facilitant, notamment, la collaboration entre les autorités et l'organisation[829].

---

[823] *Supra*, Introduction et Chapitre 1, Section I, Sous-section A.
[824] E. MOUCHARD, préc., note 1, p. 398.
[825] *Id.*, p. 163.
[826] John RUGGIE, *Rapport du Représentant spécial du Secrétaire général chargé de la question des droits de l'homme et des sociétés transnationales et autres entreprises - Principes directeurs relatifs aux entreprises et aux droits de l'homme: mise en œuvre du cadre de référence «protéger, respecter et réparer»*, A/HR/17/31, Nations Unies, Conseil des droits de l'homme, 2011, p. 20.
[827] E. MOUCHARD, préc., note 1, p. 386.
[828] D. WRIGHT, préc., note 701, 55.
[829] E. MOUCHARD, préc., note 1, p. 385.

Importance particulière au Québec. La documentation se rapportant à de telles démarches est particulièrement importante au Québec en raison des multiples obligations qui réclament d'adopter une approche en amont[830]. Un autre élément qui renforce la pertinence de la démarche au Québec est le caractère vague de la Loi[831]. En effet :

> la documentation va venir jouer un rôle central de protection par l'incertitude. En laissant les entreprises dans le flou des objectifs, le législateur va exiger d'elles une meilleure diligence. Dans le doute, l'entreprise va ainsi prendre des mesures de protection qui vont au-delà d'une simple application stricte de la loi.[832]

Démonstrations pratiques. Cette capacité des EFVP à être mobilisées à titre d'éléments de preuve permettant de démontrer la diligence de l'entreprise s'observe à travers certains rapports d'enquêtes produits par des commissaires canadiens à la protection des données.

Les EFVP ont rarement été étudiées par la CAI. La CAI a rarement été appelée à évaluer des EFVP. En effet, dans le répertoire de la section surveillance de la CAI disponible au public l'expression « Évaluation des facteurs relatifs à la vie privée » n'apparait que dans le cadre d'une seule enquête et d'une seule ordonnance[833]. De plus, les termes « analyse d'impacts » ne renvoient qu'à deux avis[834]. De cette documentation, des analyses d'impacts ont été mobilisées par des organisations afin de démontrer leur diligence que dans deux circonstances soit : dans le cadre d'une enquête sur la Fonderie de l'innovation dans le commerce au détail (ci-après « FICD ») et dans le cadre d'un avis concernant un projet de la Régie des rentes du Québec[835]. Malheureusement, dans la cause sur FICD le projet visé par l'enquête a été abandonné par

---

[830] *Supra,* Chapitre 3, Section II, Sous-section C.

[831] *Supra*, Chapitre 1, Section I, Sous-section C et Chapitre 1, Section II, Sous-section B, Chapitre 1, Section III, Sous-section A et Chapitre 2, Section I.

[832] E. MOUCHARD, préc., note 1, p. 375.

[833] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC - SECTION SURVEILLANCE, « Recherche avancée - "Évaluation des facteurs relatifs à la vie privée" », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=%22%C3%89valuation+des+facteurs+relatifs+%C3%A0+la+vie+priv%C3%A9e%22&ref=&d1=&d2=&p=&col=162&or=> (consulté le 2 août 2022).

[834] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC - SECTION SURVEILLANCE, « Recherche avancée - "Analyse d'impacts" », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=%22Analyse+d%27impacts%22&ref=&d1=&d2=&p=&col=162&or=> (consulté le 2 août 2022).

[835] Commission d'accès à l'information du Québec, 15 mai 2020, 1019951-S, *Enquête sur l'utilisation de la technologie d'analyse de vidéo anonyme par Ivanhoé Cambridge Inc. et Innovations Galilei 2*, 2-3 ; 6-7, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/484279/index.do?q=%C3%A9valuation+des+facteurs+%C3%A0+la+vie+priv%C3%A9e> (consulté le 1 août 2022).

l'entreprise avant que la CAI ait pu évaluer sa conformité à la Loi québécoise[836]. Un rapport d'enquête a néanmoins été produit par la commission[837].

La documentation produite par les EFVP a permis aux organisations de démontrer leur diligence à la CAI. Or, malgré la rareté des décisions de la CAI se rapportant à des EFVP, il est très clair que dans les circonstances où elle a été mobilisée, la documentation produite par les EFVP a influencé l'attitude la CAI à l'égard des organisations qui les avaient rédigées. Ainsi, dans l'avis sur la Régie des rentes, la CAI a tenu à souligner « la qualité de la présentation des documents soumis, notamment l'analyse d'impacts au regard de la protection des renseignements personnels… »[838]. Dans son enquête sur la FICD, la CAI a été appelée à enquêter sur un projet très controversé. Ce dernier possédait, de l'avis de la CAI, une « faible acceptabilité sociale »[839] et souffrait d'une mauvaise couverture médiatique[840]. En revanche, la CAI a salué la rédaction de l'EFVP. En effet, elle a identifié que celle-ci était :

> une bonne pratique encouragée par la Commission avant de déployer un outil technologique ou une mesure qui implique la collecte, l'utilisation ou la communication de renseignements personnels. Elle permet d'identifier dès le début d'un projet les enjeux en ces matières et d'ajuster la solution de manière à respecter la loi et à minimiser les impacts sur la vie privée[841].

De plus, malgré l'aspect controversé du projet évalué, la CAI a néanmoins tenu à saluer les efforts des concepteurs ainsi qu'à « souligner la collaboration de la FICD tout au long de cette enquête »[842]. Ainsi, bien que la CAI ne se soit pas prononcée directement sur le projet en raison de son abandon, il est clair que la confection de l'EFVP aurait été bénéfique dans la défense de l'entreprise.

---

[836] *Id.*, 1.
[837] *Id.*
[838] Commission d'accès à l'information du Québec, 25 septembre 2002, *Avis de la Commission d'accès à l'information concernant le projet de confirmation d'identité de la clientèle lors de la prestation de services de la Régie des rentes du Québec*, 13, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/480877/index.do?q=%22Analyse+d%27impact%22> (consulté le 2 août 2022).
[839] Commission d'accès à l'information du Québec, 15 mai 2020, *Enquête sur l'utilisation de la technologie d'analyse de vidéo anonyme par Ivanhoé Cambridge Inc. et Innovations Galilei 2*, préc., note 835, 7.
[840] *Id.*
[841] *Id.*, 2-3.
[842] *Id.*, 7.

Répartition des risques entre partenaires. Le CVPC a également été appelé à enquêter sur des entreprises qui ont mobilisé, lors de leur défense, leur propre EFVP ou celle de leur partenaire. Par exemple, dans une enquête sur TransUnion, l'entreprise a présenté une EFVP réalisée par son partenaire Statistique Canada afin de démontrer la légalité des informations transmises à cette institution publique. Cette EFVP indiquait que Statistique Canada utilisait les renseignements communiqués par TransUnion « à des fins statistiques seulement, dans le but de réaliser son mandat »[843]. Cette preuve a, entre autres, permis au CVPC de conclure que la plainte qui visait ces renseignements était non fondée[844]. À noter qu'après cette enquête, le CVPC a également enquêté sur les pratiques de Statistique Canada. À l'issue de son enquête, le CVPC a déterminé que l'organisme public avait agi en toute légalité au cours de la communication susmentionnée[845].

L'importance de la documentation au regard des modifications législatives. Les modifications apportées par PL64 ne feront que renforcer la pertinence de la documentation relative aux EFVP. Ainsi, le simple fait que l'outil sera désormais obligatoire dans de nombreuses circonstances[846] risque de populariser et de normaliser son usage même dans les circonstances où il ne sera pas formellement imposé par la Loi. Surtout, les sanctions très élevées associées au non-respect de la *Loi sur le privé*[847] risquent d'encourager les entreprises à rédiger des EFVP afin de se doter des moyens nécessaires pour respecter celle-ci. En fait, dans les cas où l'EFVP sera imposée et, si l'usage de l'outil venait à être normalisé, il serait davantage approprié d'affirmer que c'est l'absence d'EFVP plutôt que sa présence qui sera remarquée par les autorités. D'outils facultatifs qui permettent actuellement aux organisations de prouver leur diligence, les EFVP risquent d'être mobilisées dans le futur non plus pour démontrer une diligence, mais plutôt pour démontrer une absence de négligence.

---

[843] Commissariat à la protection de la vie privée du Canada, 9 décembre 2019, 2019-007, *Une agence de crédit est autorisée à se prévaloir de l'exemption au consentement pour communiquer des renseignements sur le crédit à Statistique Canada*, par. 14, en ligne : <https://canlii.ca/t/j3wmq> (consulté le 2 août 2022).
[844] *Id.*, 14, 15 et 32.
[845] *Id.*, par. 20.
[846] *Supra,* Chapitre 3, Section I, Sous-section A.
[847] Voir par exemple : *PL64*, préc., note 5, art. 160. Nouvel article 91 de la *Loi sur le privé.*

L'importance de la documentation au regard des tensions. Nous proposons également que la documentation produite par l'analyse d'impacts soit particulièrement importante dans le domaine des SIA en raison des choix qui accompagnent les tensions entre principes. Dans un contexte où la protection d'un principe impose de contrevenir à un autre il apparait impératif de documenter pourquoi un arbitrage plutôt qu'un autre a été réalisé. La documentation produite par le processus d'EFVP permet à l'entreprise de démontrer que les arbitrages sont abordés « selon une réflexion raisonnée et fondée sur des éléments probants, plutôt que sur la base de l'intuition ou d'un jugement aléatoire. »[848] Ainsi, les *Lignes directrices en matière d'éthique pour une IA digne de confiance* proposent qu' « en cas de conflit, les arbitrages entre [principes] devraient être explicitement reconnus et évalués du point de vue du risque qu'ils posent pour les principes éthiques, y compris les droits fondamentaux »[849]. Or, cette approche fondée sur la gestion de risque est exactement ce qui est prévu par les EFVP.

Une protection imparfaite. Bien sûr, en agissant ainsi, une organisation n'éliminera pas nécessairement tous les risques légaux qui pourraient affecter son projet. Il existera toujours la possibilité que l'autorité évalue que l'arbitrage privilégié n'était pas celui qui aurait dû être favorisé. En revanche, en pareilles circonstances, la documentation peut être mobilisée afin de démontrer que l'organisation était positionnée dans une situation difficile et complexe qui ne présentait aucune réponse évidente. Minimalement, la documentation peut souligner l'absence d'inconscience ou de négligence dans les actions de l'organisation. Cette démonstration peut amoindrir la peine infligée. En effet, la démonstration de la diligence de l'entreprise par la documentation s'inscrit directement dans plusieurs facteurs identifiés par PL64 qui doivent être considérés lors de la détermination de la sévérité de la peine parmi lesquels :

> «le fait que le contrevenant ait agi intentionnellement ou ait fait preuve de négligence ou d'insouciance »[850],

[848] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 16.
[849] *Id.*, p. 25.
[850] *PL64*, préc., note 5, arts. 69 et 160.

« les tentatives du contrevenant de dissimuler l'infraction ou son défaut de tenter d'en atténuer les conséquences »[851],

« le fait que le contrevenant ait omis de prendre des mesures raisonnables pour empêcher la perpétration de l'infraction »[852] et

« le fait que le contrevenant, en commettant l'infraction ou en omettant de prendre des mesures pour empêcher sa perpétration, ait accru ses revenus ou ait réduit ses dépenses ou avait l'intention de le faire »[853].

Section III - Les ajouts réclamés par les tensions

Quelques ajouts au processus d'EFVP. Tel qu'identifié, les EFVP sont des processus particulièrement utiles pour les organisations désireuses de développer ou d'utiliser des SIA. En revanche, nous conseillons d'incorporer quelques éléments au processus d'analyse afin de pouvoir véritablement tirer profit des EFVP dans le contexte des tensions évaluées. Puisque ces éléments doivent être incorporés à différentes étapes du processus de l'analyse d'impacts nous accompagnons chacun de nos conseils d'un schéma afin de mieux situer à quelles étapes il conviendra de les considérer. Ainsi, nous conseillons d'incorporer divers éléments lors de la (A) préparation à l'évaluation, (B) de l'analyse des risques et (C) de l'identification des mesures devant être adoptées afin de mitiger ces risques.

A. L'élément à considérer lors de la préparation à l'EFVP

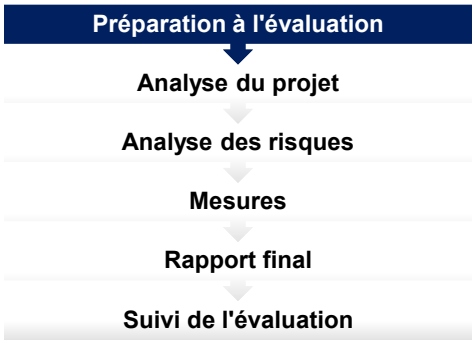

Figure 4. – Positionnement : Préparation à l'évaluation

[851] *Id.*
[852] *Id.*
[853] *Id.*

Consulter la société civile et les particuliers affectés. D'abord, nous désirons souligner l'importance de consulter à la fois des organisations en provenance de la société civile et des particuliers qui seront directement concernés par le traitement du SIA. Plusieurs auteurs recommandent, en effet, d'incorporer des éléments de la société civile au processus d'EFVP en raison des multiples avantages procurés par cette démarche[854].

Mieux mesurer les risques. D'une part, la participation de la société civile et des personnes directement concernées par le projet au processus d'EFVP permet à l'organisation de mieux mesurer et gérer les risques pouvant affecter le projet. En effet, Wright propose que :

> Engaging stakeholders, including the public, will help the assessor to discover risks and impacts that he or she might not otherwise have considered. A consultation is a way to gather fresh input on the perceptions of the severity of each risk and on possible measures to mitigate these risks[855].

La CAI propose elle aussi d'incorporer les personnes directement concernées par le projet pour la même raison. Ainsi, selon elle :

> L'implication des parties prenantes dans le cadre de cette évaluation, plus particulièrement des personnes directement concernées, est susceptible d'apporter un éclairage fort pertinent, notamment sur la nécessité de la collecte des renseignements ou les risques que suscite le projet en matière de respect de la vie privée[856].

L'incorporation de ces parties prenantes est particulièrement importante lors du développement des SIA. Tel qu'identifié, l'IA est domaine émergent dans lequel les risques posés par ces outils ne sont pas toujours bien compris et appréhendés[857].

Mesurer les principaux intérêts et préoccupations en jeu. D'autre part, la participation de ces parties prenantes facilite l'identification des principaux intérêts et préoccupations en jeu. En effet, les SIA opèrent dans un contexte où des arbitrages entre différents principes peuvent s'avérer nécessaires. Or, ces arbitrages peuvent affecter substantiellement les droits

[854] D. WRIGHT, préc., note 760, 311; D. KLOZA, A. CALVI, S. CASIRAGHI, S. MAYMIR et N. IOANNIDIS, préc., note 735, p. 2.
[855] D. WRIGHT, préc., note 701, 58.
[856] Commission d'accès à l'information du Québec, 15 mai 2020, *Enquête sur l'utilisation de la technologie d'analyse de vidéo anonyme par Ivanhoé Cambridge Inc. et Innovations Galilei 2*, préc., note 835, 7.
[857] Voir par exemple, INFORMATION COMMISSIONER'S OFFICE, préc., note 251, p. 58-60; M. KEARNS et A. ROTH, préc., note 321, p. 160 (page pdf). ; *Supra,* Chapitre 3, Section II, Sous-section C.

fondamentaux des personnes concernées. À ce titre, il apparait vital que ces individus puissent identifier quels principes devraient, selon eux, être privilégiés. Cette détermination confère, en retour, une plus grande légitimité aux choix de l'organisation responsable du projet, dans la mesure, bien sûr que les mesures et stratégies privilégiées par l'organisation reflètent les choix des parties prenantes.

Sauvegarder des ressources. Qui plus est, l'incorporation de la société civile et des personnes directement concernées par le projet permet d'évaluer rapidement si un projet devrait être réalisé. En effet, « *Engaging stakeholders is a way of testing the waters, of gauging the public's reaction to a project before it is implemented*. »[858] Bref, cette participation permet d'évaluer l'acceptabilité du projet à l'égard des principaux intéressés.

Favoriser la confiance. Aussi, il convient de se rappeler que les SIA opèrent dorénavant dans un climat de défiance[859]. Incorporer des parties prenantes de la société civile au processus d'EFVP permet de gagner la confiance du public à l'égard de l'outil. En effet, en insérant les perceptions de la société civile à l'EFVP, les organisations peuvent plus facilement accommoder différents points de vues allégeant ainsi les craintes liées à au projet[860]. Dans un contexte de tensions entre différents principes, le processus d'EFVP permet également d'informer et d'éduquer les parties prenantes relativement aux limites légales et aux risques associés à diverses mesures. Par exemple, tous ne seront pas conscients que de réclamer une parité statistique, une parité entre valeurs prédictives et une parité entre taux de faux positifs et de faux négatifs est, généralement, mathématiquement impossible. Bref, en incorporant ces parties prenantes, une organisation participe à la diffusion des connaissances, ce qui, en retour, contribue à l'élaboration d'attitudes plus réalistes sur les capacités des SIA.

Production d'une documentation utile. Finalement, la documentation produite au terme de cette participation peut être mobilisée pour défendre l'organisation si cette dernière fait l'objet d'une enquête. Par exemple, la participation d'éléments de la société civile pourrait faciliter la preuve de la nécessité de collecter ou d'utiliser certaines données. Tel qu'identifié, la CAI réclame

[858] D. WRIGHT, préc., note 701, 58.
[859] *Supra,* Introduction.
[860] D. KLOZA, A. CALVI, S. CASIRAGHI, S. MAYMIR et N. IOANNIDIS, préc., note 735, p. 2.

de démontrer, selon une interprétation du principe de nécessité, que les objectifs de la collecte sont légitimes, importants et réels[861]. Or, *a priori,* la documentation produite par la participation des parties prenantes peut faciliter la démonstration de ces éléments si, par exemple, elle permet de démontrer que ce sont des organisations de la société civile ou des personnes directement concernées par le projet qui ont demandé de collecter certains types de renseignements comme des attributs protégés. La documentation peut également être mobilisée afin de prouver la proportionnalité des mesures privilégiées puisqu'elle permet de mesurer l'importance des fins d'un renseignement personnel non seulement pour l'organisation, mais également pour le public. Bref, la documentation rattachée à la participation de groupes et de particuliers affectés par le projet peut faciliter la défense de l'organisation, ce qui peut être très utile au regard de la sévérité des sanctions prévues par la *Loi sur le privé*[862].

B. Les considérations liées à l'analyse des risques

Figure 5. – Positionnement : Analyse des risques

L'importance de considérer plus que la vie privée. Ensuite, nous proposons que, dans le cadre d'une EFVP portant sur le déploiement ou le développement d'un SIA, il importe d'aller au-delà d'une analyse qui se limite à évaluer les risques à la vie privée. Tel que démontré, les SIA présentent des risques qui dépassent largement les enjeux liés à la protection de la vie privée[863]. Qui plus est, il ne sera pas possible de tirer réellement profit de ces analyses d'impacts en tant

[861] *Supra,* Chapitre 1, Section III, Sous-section A.
[862] *Supra,* Chapitre 1, Section III, Sous-section C.
[863] *Supra,* Chapitre 1, Sections I, II et IV.

qu'instruments guidant les arbitrages entre différents principes si l'analyse se limite aux enjeux relatifs à la vie privée.

L'importance d'adopter une approche flexible. De plus, tel qu'identifié, une EFVP réclame d'évaluer la conformité du projet à la Loi. Or, les modèles d'EFVP disponibles au public ne sont pas uniformes quant à la manière d'évaluer cette conformité. Tel qu'identifié, le modèle fourni par la CNIL propose de considérer les normes légales comme des impératifs non négociables alors que la CAI semble plutôt proposer d'approcher la question de la conformité sous l'angle de la gestion des risques[864]. Bien qu'il ne soit pas nécessaire de considérer le non-respect des normes légales comme des « risques », nous désirons souligner l'importance de ne pas réduire l'évaluation de la conformité du projet à la Loi comme un exercice similaire à une « *checklist* »[865]. En effet, le régime québécois des renseignements personnels est, tel qu'identifié, particulièrement flou[866]. Il semble difficile, par exemple, d'identifier dans certaines circonstances qu'une collecte de renseignements personnels respecte assurément le critère de nécessité alors que ce même principe est interprété de façons contradictoires par la CAI[867]. À ce titre, approcher cette évaluation comme une « *checklist* » limite la capacité des évaluateurs à illustrer adéquatement le flou normatif qui peut affecter le projet étudié par l'EFVP.

L'importance d'évaluer les risques à l'organisation. Qui plus est, nous invitons à évaluer les risques pouvant affecter l'organisation responsable du projet. Nous justifions cette approche au regard du contexte législatif et institutionnel québécois. En effet, avec l'adoption de PL64, contrevenir à la *Loi sur le privé* peut coûter très cher à une entreprise[868]. Or, la loi québécoise, doit-on le rappeler, est floue et n'indique pas clairement dans quelles circonstances un principe doit être favorisé au détriment d'un autre[869]. À ce titre, il importe que l'entreprise qui choisit de

[864] *Supra,* Chapitre 3, Section I, Sous-section B(iii).
[865] *Supra,* Chapitre 3, Section II, Sous-section B.
[866] *Supra*, Chapitre 1, Section I, Sous-section C; Chapitre 1, Section II, Sous-section B; Chapitre 1, Section III, Sous-section A et Chapitre 2, Section I.
[867] *Supra*, Chapitre 1, Section III, Sous-section A.
[868] *Supra*, Chapitre 1, Section III, Sous-section C.
[869] *Supra,* Chapitre 2, Sections II et IV.

favoriser, par exemple, l'équité au sacrifice de la sécurité comprenne qu'agir ainsi l'expose à des risques légaux importants qui peuvent lui coûter très cher.

Le flou normatif renforcé par la diversité institutionnelle. Plus encore, le respect des principes ne sont pas assurés par les mêmes institutions, une situation qui complique l'application d'un droit uniforme. Alors que le mandat de la CAI l'invite surtout à évaluer la sécurité des renseignements personnels, le mandat de la CDPDJ lui réclame surtout d'évaluer les risques rattachés à la discrimination[870]. Or, parce qu'elles ne poursuivent pas des mandats identiques et que certaines mesures favorisant l'équité contreviennent parfois à la sécurité, ces deux organisations ne vont pas toujours privilégier les mêmes approches. Cette situation s'observe, entre autres, par les solutions proposées par ces institutions afin de contrecarrer les risques liés à la discrimination.

La CDPDJ semble privilégier la collecte d'attributs protégés pour contrer la discrimination. Ainsi, la CDPDJ constate, comme nous[871], la présence d'une tension entre l'application d'une approche minimaliste et la nécessité de collecter des attributs protégés afin de réduire les risques de discrimination. Par exemple, dans son rapport déposé lors des travaux portant sur PL64, la CDPDJ a souligné que : « les règlements sur les renseignements sensibles peuvent faire en sorte que des organisations ne collectent pas de telles données, ce qui complique l'identification de pratiques de discrimination par proxy »[872].

La CAI privilégie l'approche minimaliste pour contrer la discrimination. La CAI, elle, propose exactement l'inverse. Selon elle, l'approche minimaliste constitue, en fait, un moyen de lutter contre la discrimination puisque :

> limiter la collecte de renseignements personnels permet d'éviter, dans une certaine mesure, que des décisions soient prises sur la base de renseignements susceptibles d'entrainer une discrimination, les renseignements susceptibles d'engendrer ce genre de biais n'étant alors pas connus.[873]

---

[870] *Loi sur l'accès*, préc., note 12, arts. 103 et 122; *Charte des droits et libertés de la personne*, préc., note 14, arts. 57 et 71.
[871] *Supra,* Chapitre 2, Section IV.
[872] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 63.
[873] COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, préc., note 307, p. 1-2.

En fait, à la lecture du mémoire déposé par la CAI à l'occasion de la *Déclaration de Montréal pour une intelligence artificielle responsable,* il est clair que l'approche envisagée par l'institution pour répondre à la problématique de discrimination est, non pas de faciliter la collecte d'attributs protégés, mais bien de renforcer les interdits qui s'y rattachent. En effet, dans ce mémoire la CAI propose de se questionner sur l'opportunité « d'interdire la collecte de certains renseignements dans certains contextes, par exemple pour éviter la discrimination? »[874]. Bref, selon la CAI les principes de protection des renseignements personnels comme le critère de nécessité, loin de présenter des tensions avec les autres principes, contribuent à leur respect[875].

Position de la littérature. Avec égard pour la CAI, l'approche minimaliste en matière de lutte contre la discrimination a été maintes fois dénoncée par les experts en IA qui la perçoivent comme contre-productive[876]. L'approche proposée par la CDPDJ concorde beaucoup mieux avec l'état actuel de la littérature.

Des approches contradictoires. Ainsi, sur exactement le même sujet, soit l'impact des limitations imposées à la collecte des renseignements personnels sur le respect du principe d'équité, deux des principales institutions susceptibles de sanctionner les développeurs et utilisateurs de SIA, approchent la problématique de façon diamétralement opposée. Alors que la CDPDJ semble consciente des impacts des limitations imposées à la collecte de renseignements personnels sensibles, la CAI, elle, semble prompte à évaluer, au contraire, que le respect du principe d'équité passe nécessairement par l'adoption d'une approche minimaliste. Les deux approches que semblent privilégier ces deux institutions sont pourtant mutuellement exclusives.

Nécessité d'évaluer les risques à l'organisation responsable du SIA. À ce titre, les organisations désireuses d'utiliser des SIA doivent donc non seulement composer avec le caractère nébuleux

[874] *Id.*, p. 6.
[875] *Id.*, p. 1-2.
[876] Voir par exemple : M. KEARNS et A. ROTH, préc., note 321, p. 61-63 (page pdf); M. BOGEN, A. RIEKE et S. AHMED, préc., note 579, p. 1; ANYA E. R. PRINCE et DANIEL SCHWARCZ, préc., note 342, 1263; S. CORBETT-DAVIES, E. PIERSON, A. FELLER, S. GOEL et A. HUQ, préc., note 341, p. 2; S. CORBETT-DAVIES et S. GOEL, préc., note 313, p. 9; Cynthia DWORK, Moritz HARDT, Toniann PITASSI, Omer REINGOLD et Rich ZEMEL, *Fairness Through Awareness*, *ITCS 2012: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, Cambridge, Massachusetts, arXiv, 28 novembre 2011, p. 8 et 22, DOI : 10.48550/arXiv.1104.3913.

de la Loi, elles doivent en plus composer avec des institutions qui favorisent des approches parfois distinctes. Dans un tel contexte, il semble opportun pour une entreprise de considérer les risques associés aux mesures qu'elle privilégie, notamment au regard des sanctions possibles qui pourront lui être imposées par l'une ou l'autre de ces institutions.

D. Les considérations liées à l'identification des mesures

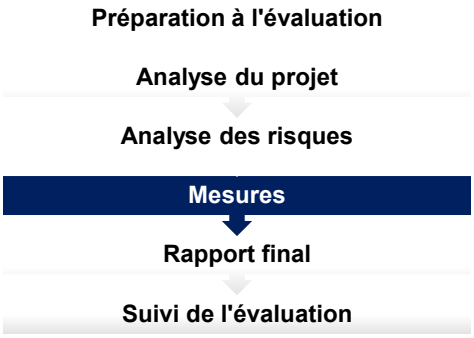

Figure 6. – Positionnement : Mesures

Expliciter les tensions et les alternatives considérées. Enfin, nous pensons qu'une EFVP réalisée pour un projet de développement ou d'usage de SIA doit reconnaitre l'existence des tensions si celles-ci se manifestent[877]. De plus, la documentation produite devrait expliciter les alternatives considérées et mesurer les risques associés à chacune d'elles. Il est impératif de documenter la démarche afin de prouver la diligence de l'entreprise au regard des différents arbitrages réalisés. Agir ainsi facilite la démonstration que les sacrifices réalisés auprès d'un principe ne pouvaient être évités sans provoquer de nouveaux risques. Cette démarche permet également à l'organisation d'identifier que les mesures privilégiées ont été choisies au terme d'une démarche réfléchie capable de considérer les différentes valeurs en jeu.

Évaluer les impacts négatifs des mesures. Finalement, l'organisation doit évaluer les risques créés par chacune des mesures qu'elle a choisi d'implémenter. En effet, la présence de tensions implique qu'une organisation ne peut pas considérer les mesures privilégiées seulement au regard des bénéfices qui lui sont associées. Des coûts sont rattachés à certaines mesures et l'EFVP

[877] GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, préc., note 95, p. 25.

doit les identifier. Ce n'est qu'en agissant ainsi qu'il devient possible pour l'organisation responsable du projet d'effectuer des arbitrages entre différents principes de façon réfléchie. C'est également ainsi que l'organisation peut documenter les raisons qui sous-tendent ses arbitrages ce qui lui permettra de mobiliser la documentation créée par l'EFVP à titre d'élément de preuve en cas d'enquête sur ses activités ou de poursuites. Notre tableau disponible en Annexe présente plusieurs mesures et stratégies évaluées dans le cadre du présent mémoire ainsi que les bénéfices et coûts associés à chacune d'entre elles.

Conclusion

Complexité née des tensions. En conclusion, assurer un développement et une utilisation des SIA qui seraient légaux au regard du droit québécois et des différents principes légaux applicables à l'IA est un exercice particulièrement complexe. Plusieurs principes, dont l'exactitude, la sécurité, l'explicabilité et l'équité doivent être considérés dans le cadre de cet exercice. Or, ces principes, malgré leur reconnaissance universelle, restent rarement bien définis. Après les avoir étudiés, nous avons démontré que l'application des mesures susceptibles de renforcer le respect de certains principes peut compromettre d'autres principes. Renforcer l'explicabilité du SIA réclame parfois de sacrifier son exactitude alors que renforcer le respect de la sécurité peut défavoriser le respect des principes d'exactitude et d'équité. De plus, les mesures susceptibles de renforcer certaines conceptions de l'équité ne peuvent être mises en œuvre qu'en sacrifiant la sécurité de certaines données. Enfin, nous avons identifié l'impossibilité mathématique d'assurer un respect conjoint d'une parité de performance et d'une parité statistique.

Un exercice de mise en balance. En conséquence, une implémentation efficace des mesures susceptibles de renforcer ces principes réclame d'approcher la problématique comme un exercice de mise en balance. Puisque certaines mesures menant à l'implémentation d'un principe peuvent en affecter un autre, il est impératif de bien mesurer les risques associés aux mesures privilégiées. Nous proposons que la rédaction d'EFVP est particulièrement utile à cet effet puisqu'elles permettent de protéger l'organisation contre les risques créés par ces tensions tout en assurant un respect de ces principes au terme d'une démarche réfléchie et rationnelle.

L'absence de protection parfaite. Toutefois, les EFVP ne doivent pas être considérées comme une mesure de protection parfaite. Elles ne pourront pas nécessairement protéger une organisation contre tous les risques juridiques. Les organisations doivent comprendre que d'utiliser ou de développer un SIA est une décision risquée en raison de la méfiance médiatique et juridique à leur égard. En effet, nous constatons, un certain double standard dans les normes que nous imposons à un traitement réalisé par un SIA et celles que nous appliquons à un décideur humain. Ce double standard s'illustre dans les dispositions législatives prévues à leur égard et

dans notre tendance à évaluer ces outils au regard d'une version idéelle de la société et non pas à l'égard des alternatives réelles à notre disposition.

La présence d'un double standard. Par exemple, tel qu'identifié, PL64 prévoit un droit de révision ainsi qu'une obligation d'expliciter les raisons, facteurs et paramètres utilisés pour prendre une décision[878]. Cependant, le projet de loi limite cette obligation aux décisions fondées sur des traitements exclusivement automatisés. En l'absence d'un processus exclusivement automatisé PL64 ne prévoit ni « la transparence du processus de traitement »[879] ni « la possibilité de contester le résultat produit »[880]. À ce titre, le projet de loi semble présumer que la simple présence d'un humain garantit nécessairement une décision plus légitime. Le traitement médiatique auprès de COMPAS est également emblématique de ce double standard. En effet, dans l'article de ProPublica l'outil était dénoncé non pas pour produire des résultats plus inéquitables que les juges, mais bien pour produire des résultats inéquitables dans l'absolu[881]. La performance du SIA était donc comparée non pas à l'égard du traitement alternatif qui serait appliqué en l'absence de l'outil, mais bien à l'égard d'un idéal paritaire. À noter qu'il n'est pas problématique, en soit, de réclamer aux SIA de poursuivre un tel idéal dans la mesure où, bien sûr, cet idéal est réalisable[882]. Le double standard naît du fait que cette poursuite n'est parfois réclamée qu'aux SIA et pas aux évaluateurs humains. Cette attitude est d'autant plus surprenante au regard du fait que les comportements problématiques des SIA ne sont très souvent que le reflet des pratiques d'une société ou d'une organisation. Ainsi, lorsque l'on observe qu'un SIA produit des résultats problématiques nous limitons souvent notre réflexion à l'acceptabilité de l'outil. Selon Crawford, lorsque l'on observe qu'un SIA produit des résultats discriminatoires, par exemple, il est assez rare « *to have a public debate about why these forms of biais and discrimination recur and whether more fondamental problems are at work than […] a poorly designed algorithm* »[883]. On se limite à blâmer l'algorithme et à réclamer qu'il soit réparé[884].

[878] *Supra,* Chapitre 1, Section I, Sous-section C et Chapitre 2, Section II, Sous-section B.
[879] J.-F. TRUDEL, A. BERWALD, M. CARPENTIER et M. FORCIER, préc., note 5, p. 64.
[880] *Id.*
[881] J. A. MATTU Jeff Larson,Lauren Kirchner,Surya, préc., note 3; J. L. MATTU Julia Angwin,Lauren Kirchner,Surya, préc., note 3.
[882] *Supra*, Chapitre 1, Section IV, Sous-section B.
[883] K. CRAWFORD, préc., note 332, p. 129.
[884] *Id.*

Toutefois, la véritable source de discrimination ne provient probablement pas de l'algorithme en soi. Tel que démontré, si un SIA est inéquitable c'est probablement parce qu'il reproduit des iniquités[885]. À ce titre, simplement réparer l'algorithme ou prohiber son utilisation ne règle en rien la véritable problématique qui tire sa source de comportements humains. Si nous étions véritablement anxieux de produire une société plus équitable, l'observation d'un comportement discriminatoire d'un SIA devrait surtout nous commander d'initier une réflexion sur les pratiques de nos organisations et de notre société plutôt que de se limiter à vilipender l'outil.

[885] *Supra,* Chapitre 1, Section IV, Sous-section B.

Références bibliographiques

Projets de loi/Propositions de règlement

Projet de Loi C-27 : Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois, Première session, quarante-quatrième législature, Première lecture le 16 juin 2022.

Proposition de règlement du Parlement européen établissant des règles harmonisées concernant l'Intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union, (2021), en ligne : <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52021PC0206> (consulté le 10 février 2022).

Documentation relative aux amendements

COMMISSION DES institutions, Amendements adoptés: Projet de loi no 64 : Loi modernisant les dispositions législatives en matière de protection des renseignements personnels, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci/mandats/Mandat-43711/index.html>;

Législation

Sources québécoises

Charte des droits et libertés de la personne, RLRQ c C-12, en ligne : <https://canlii.ca/t/6c3nj>.

Code civil du Québec, chapitre CCQ-1991, en ligne : <https://www.legisquebec.gouv.qc.ca/fr/document/lc/ccq-1991> (consulté le 10 juillet 2022).

Loi constitutionnelle de 1982, Annexe B de la Loi de 1982 sur le Canada (R-U), c 11, en ligne : <https://canlii.ca/t/q3x8>

Loi modernisant des dispositions législatives en matière de protection des renseignements personnels (anciennement Projet de Loi 64), 64 (2021).

Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels, chapitre A-2.1.

Loi sur la protection des renseignements personnels dans le secteur privé, chapitre P-39.1.

Sources canadiennes et étrangères

Loi sur le système correctionnel et la mise en liberté sous condition, L.C. 1992, ch. 20.

Loi sur la protection des renseignements personnels et les documents électroniques, LC 2000 ch 5 66, en ligne : <https://laws-lois.justice.gc.ca/PDF/P-8.6.pdf>.

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données, (2018) Journal officiel de l'Union européenne.

Jurisprudence

Décisions de la Cour suprême du Canada

R. v. Oakes, [1986] 1 SCR 103 (Cour suprême du Canada), en ligne : <https://canlii.ca/t/1ftv6> (consulté le 30 janvier 2022).

RJR-MacDonald Inc. c. Canada (Procureur général), [1995] 3 RCS 199 (Cour suprême), en ligne : <https://canlii.ca/t/1frh0> (consulté le 21 août 2022).

R. c. St-Cloud, [2015] 2 RCS 328 (Cour suprême du Canada), en ligne : <https://canlii.ca/t/ghtdb> (consulté le 15 août 2022).

Ewert c. Canada, [2018] 2 RCS 165 (Cour suprême du Canada), en ligne : <https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/17133/index.do> (consulté le 7 juillet 2022).

Egan c. Canada, [1995] 2 RCS 513 (Cour suprême du Canada), en ligne : <https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/1265/index.do> (consulté le 29 août 2022).

Ward c. Québec (Commission des droits de la personne et des droits de la jeunesse), [2021] 43 CSC (Cour suprême du Canada), en ligne : <https://canlii.ca/t/jk1tm> (consulté le 29 août 2022).

Décisions québécoises

Cour du Québec, 21 février 2003, 500-02-094423-014, *Société de transport de la Ville de Laval c. X.*

Cour du Québec, 7 décembre 2009, 14676, *A c. B*, *200-22-044658-086*, en ligne : <https://canlii.ca/t/27d0x> (consulté le 4 juillet 2022).

Cour du Québec, 17 juin 2010, 450-80-000873-090, *P.B. c. Lepage*, en ligne : <https://canlii.ca/t/2bljf> (consulté le 7 juillet 2022).

Cour supérieure, 19 septembre 2019, 500-06-000907-184, *Lévy c. Nissan Canada inc.*, en ligne : <https://canlii.ca/t/j2klc> (consulté le 9 juillet 2022).

Tribunal d'arbitrage, 31 octobre 1998, Grief n°: 97-02, *Syndicat des employées et employés professionnels et de bureau, section locale 57 et Caisse populaire St-Stanislas de Montréal*, en ligne : <https://canlii.ca/t/hnknp> (consulté le 5 juillet 2022).

Enquêtes et avis du Commissariat à la protection de la vie privée du Canada

Commissariat à la protection de la vie privée du Canada, 25 septembre 2007, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2007-389 : TJX Companies Inc./Winners Merchant International L.P.*, en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2007/tjx_rep_070925/> (consulté le 9 juillet 2022).

Commissariat à la protection de la vie privée du Canada, 22 août 2016, *Rapport de conclusions d'enquête no 2016-005*, *Enquête conjointe sur Ashley Madison menée par le commissaire à la*

protection de la vie privée du Canada et le commissaire à la protection de la vie privée/commissaire à l'information par intérim de l'Australie, en ligne : <https://canlii.ca/t/h3p5k> (consulté le 4 juillet 2022).

Commissariat à la protection de la vie privée du Canada, 7 février 2018, 2018-006, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-006 : Intrusion dans la base de données de l'Agence mondiale antidopage*, en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2018/lprpde-2018-006/> (consulté le 4 juillet 2022).

Commissariat à la protection de la vie privée du Canada, 8 janvier 2018, 2018-001, *Rapport de conclusions d'enquête en vertu de la LPRPDE no 2018-001 : Un manufacturier de jouets connectés améliore les mesures de sécurité pour protéger adéquatement les renseignements d'enfants*, en ligne : <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2018/lprpde-2018-001/> (consulté le 4 juillet 2022).

Commissariat à la protection de la vie privée du Canada, 9 décembre 2019, 2019-007, *Une agence de crédit est autorisée à se prévaloir de l'exemption au consentement pour communiquer des renseignements sur le crédit à Statistique Canada*, en ligne : <https://canlii.ca/t/j3wmq> (consulté le 2 août 2022).

Commissariat à la protection de la vie privée du Canada, 25 avril 2019, 2019-002, *Enquête conjointe du Commissariat à la protection de la vie privée du Canada et du Bureau du Commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique au sujet de Facebook*, *35606 (CVPC)*, en ligne : <https://canlii.ca/t/hzzpl> (consulté le 4 juillet 2022).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, COMMISSARIAT À L'INFORMATION ET À LA PROTECTION DE LA VIE PRIVÉE DE LA COLOMBIE-BRITANNIQUE, et COMMISSARIAT À L'INFORMATION ET À LA PROTECTION DE LA VIE PRIVÉE DE L'ALBERTA, *Conclusions en vertu de la LPRPDE no 2021-001 : Enquête conjointe sur Clearview AI, Inc. par le Commissariat à la protection de la vie privée du Canada, la Commission d'accès à l'information du Québec, le*

Commissariat à l'information et à la protection de la vie privée de la Colombie-Britannique et le Commissariat à l'information et à la protection de la vie privée de l'Alberta, 2021, en ligne : <https://priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2021/lprpde-2021-001/> (consulté le 4 juillet 2022).

Enquêtes et avis de la Commission d'accès à l'information du Québec

Commission d'accès à l'information du Québec, 25 septembre 2002, *Avis de la Commission d'accès à l'information concernant le projet de confirmation d'identité de la clientèle lors de la prestation de services de la Régie des rentes du Québec*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/480877/index.do?q=%22Analyse+d%27impact%22> (consulté le 2 août 2022).

Commission d'accès à l'information du Québec, 16 mars 2010, 08 08 85, *M.L. c. Gatineau (Ville de)*, *68*, en ligne : <https://canlii.ca/t/28r34> (consulté le 5 juillet 2022).

Commission d'accès à l'information du Québec, 9 février 2015, 111756, *Garderie Excelsiori Daycare inc.*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/351801/index.do?q=Garderie+Excelsiori+Daycare+inc> (consulté le 11 août 2022).

Commission d'accès à l'information du Québec, 28 septembre 2016, 061063, *Banque Nationale du Canada*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/351226/index.do?q=%22atteinte+minimale%22> (consulté le 11 août 2022).

Commission d'accès à l'information du Québec, 14 mars 2018, 111494, *X c. Ministère de la Santé et des Services sociaux et Bibliothèque des archives nationales du Québec et Société de généalogie canadienne-française et Société de généalogie de Québec et Institut généalogique Drouin*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/350960/index.do> (consulté le 7 juin 2022).

Commission d'accès à l'information du Québec, 15 mai 2020, 1019951-S, *Enquête sur l'utilisation de la technologie d'analyse de vidéo anonyme par Ivanhoé Cambridge Inc. et Innovations Galilei 2*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/484279/index.do?q=%C3%A9valuation+des+facteurs+%C3%A0+la+vie+priv%C3%A9e> (consulté le 1 août 2022).

Commission d'accès à l'information du Québec, 12 mai 2021, 1019622-S, *Enquête à l'égard de 9055-4635 Québec inc. (Alimentation Larouche)*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/500003/index.do> (consulté le 30 janvier 2022).

Commission d'accès à l'information, 30 septembre 2021, 1015556-S, *Enquête à l'égard de Bruneau Électrique inc.*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/514817/index.do>.

Commission d'accès à l'information, 6 janvier 2022, *Enquête à l'égard du Centre intégré universitaire de santé et services sociaux de l'Estrie et du Centre hospitalier universitaire de Sherbrooke et Ministère de la Santé et des Services sociaux*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519701/index.do> (consulté le 7 juin 2022).

Commission d'accès à l'information, 6 janvier 2022, *Plainte à l'endroit de la Fondation Vipassana de l'Est du Canada*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/500008/index.do> (consulté le 7 juin 2022).

Commission d'accès à l'information du Québec, 12 janvier 2022, 1011867-S, *X. c. MRC des Collines-de-l'Outaouais - Commission d'accès à l'information du Québec*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519698/index.do> (consulté le 26 août 2022).

Commission d'accès à l'information du Québec, 14 janvier 2022, 1016217-S, *Enquête à l'égard de Compagnie Selenis Canada*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/519705/index.do> (consulté le 10 août 2022).

Commission d'accès à l'information, 5 avril 2022, 1011672-S, *Services financiers Globex 2000 inc.*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520899/index.do>.

Commission d'accès à l'information, 7 avril 2022, *L'Auberge du lac Sacacomie inc.*, *1014137-S*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520898/index.do>.

Commission d'accès à l'information du Québec, 13 mai 2022, 1014505-S et 1016098-S, *Plainte à l'égard de l'Association québécoise des transports et de la Société de l'assurance automobile du Québec*, en ligne : <https://decisions.cai.gouv.qc.ca/cai/ss/fr/item/520910/index.do> (consulté le 26 août 2022).

Autre

Cour fédérale du Canada, 20 décembre 2010, T-246-10, *Nammo c. TransUnion of Canada Inc.*, *RCF* 3.600, en ligne : <https://canlii.ca/t/2f3n8> (consulté le 27 août 2022).

Doctrine et autres

Articles de revues

ADADI, A. et M. BERRADA, « Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) », (2018) 6 *IEEE Access* 52138-52160, DOI : 10.1109/ACCESS.2018.2870052.

ANYA E. R. PRINCE et DANIEL SCHWARCZ, « Proxy Discrimination in the Age of Artificial Intelligence and Big Data », (2020) 105-1257 *Iowa Law Review*, en ligne : <https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data/> (consulté le 5 juin 2022).

AUGER, P., T. M. DEVINNEY et J. J. LOUVIERE, « Global segments of socially conscious consumers: Do they exist? », dans *Global Challenges in Responsible Business*, Royaume-Uni, Cambridge University Press, 2010, en ligne : <https://opus.lib.uts.edu.au/handle/10453/14347> (consulté le 15 août 2022).

BAMBAUER, J. et T. ZARSKY, « The Algorithm Game », (2018) 94-1 *Notre Dame Law Review* 1.

BARREDO ARRIETA, A., N. DÍAZ-RODRÍGUEZ, J. DEL SER, A. BENNETOT, S. TABIK, A. BARBADO, S. GARCIA, S. GIL-LOPEZ, D. MOLINA, R. BENJAMINS, R. CHATILA et F. HERRERA, « Explainable Artificial Intelligence (XAI):

Concepts, taxonomies, opportunities and challenges toward responsible AI », (2020) 58 *Information Fusion* 82-115, DOI : 10.1016/j.inffus.2019.12.012.

BENAIM, A. R., R. ALMOG, Y. GORELIK, I. HOCHBERG, L. NASSAR, T. MASHIACH, M. KHAMAISI, Y. LURIE, Z. S. AZZAM, J. KHOURY, D. KURNIK et R. BEYAR, « Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison From Five Observational Studies », (2020) 8-2 *JMIR Medical Informatics* e16492, DOI : 10.2196/16492.

BIRD, E., J. FOX-SKELLY, N. JENNER, R. LARBEY, E. WEITKAMP, A. WINFIELD, EUROPEAN PARLIAMENT, et DIRECTORATE-GENERAL FOR PARLIAMENTARY RESEARCH SERVICES, *The ethics of artificial intelligence: issues and initiatives.*, 2020, en ligne : <https://op.europa.eu/publication/manifestation_identifier/PUB_QA0119779ENN> (consulté le 19 mai 2022).

BONCHI, F., S. HAJIAN, B. MISHRA et D. RAMAZZOTTI, « Exposing the probabilistic causal structure of discrimination », (2017) 3-1 *Int J Data Sci Anal* 1-21, DOI : 10.1007/s41060-016-0040-z.

BRCIC, M. et R. V. YAMPOLSKIY, « Impossibility Results in AI: A Survey », 2022, DOI : 10.48550/arXiv.2109.00484.

BRODY, R. G., S. KERN et K. OGUNADE, « An insider's look at the rise of Nigerian 419 scams », (2020) 29-1 *Journal of Financial Crime* 202-214, DOI : 10.1108/JFC-12-2019-0162.

BURRELL, J., « How the machine 'thinks': Understanding opacity in machine learning algorithms », (2016) 3-1 *Big Data & Society* 2053951715622512, DOI : 10.1177/2053951715622512.

CHOULDECHOVA, A., « Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments », (2017) 5-2 *Big Data* 153-163, DOI : 10.1089/big.2016.0047.

CORBETT-DAVIES, S. et S. GOEL. *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, arXiv, 14 août 2018.

WRIGHT, D., « Making Privacy Impact Assessment More Effective », (2013) 29-5 *The Information Society* 307-315, DOI : 10.1080/01972243.2013.825687.

DÉZIEL, P.-L., « Est-ce bien nécessaire ? Le principe de limitation de la collecte face aux défis de l'intelligence artificielle et des données massives », (2019) 465 *Développements récents en droit à la vie privée* 22.

DÉZIEL, P.-L., « L'utilisation de renseignements personnels dans le contexte de la justice prédictive : le cas des outils actuariels d'évaluation des risques de récidive », (2018) 60-1 *Archives de philosophie du droit* 253-269, DOI : 10.3917/apd.601.0268.

DIAKOPOULOS, N., « Algorithmic Accountability Reporting: On the Investigation of Black Boxes », *Tow Center for Digital Journalism Publications* 2014, DOI : 10.7916/D8ZK5TW2.

DRESSEL, J. et H. FARID, « The accuracy, fairness, and limits of predicting recidivism », (2018) 4-1 *Science Advances*, DOI : 10.1126/sciadv.aao5580.

FERNANDO, M.-P., F. CÈSAR, N. DAVID et H.-O. JOSÉ, « Missing the missing values: The ugly duckling of fairness in machine learning », (2021) 36-7 *International Journal of Intelligent Systems* 3217-3258, DOI : 10.1002/int.22415.

FLORIDI, L., J. COWLS, M. BELTRAMETTI, R. CHATILA, P. CHAZERAND, V. DIGNUM, C. LUETGE, R. MADELIN, U. PAGALLO, F. ROSSI, B. SCHAFER, P. VALCKE et E. VAYENA, « AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations », (2018) 28-4 *Minds & Machines* 689-707, DOI : 10.1007/s11023-018-9482-5.

GOYAL, M., T. KNACKSTEDT, S. YAN et S. HASSANPOUR, « Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities », (2020) 127 *Computers in Biology and Medicine* 104065, DOI : 10.1016/j.compbiomed.2020.104065.

HAGENDORFF, T., « The Ethics of AI Ethics: An Evaluation of Guidelines », (2020) 30-1 *Minds & Machines* 99-120.

HERZOG, C., « On the risk of confusing interpretability with explicability », *AI Ethics* 2021, DOI : 10.1007/s43681-021-00121-9.

HU, H., Z. SALCIC, L. SUN, G. DOBBIE, P. YU et X. ZHANG, « Membership Inference Attacks on Machine Learning: A Survey », *ACM Computing Surveys* 2022, DOI : 10.1145/3523273.

JOHNSON, K. D., D. P. FOSTER et R. A. STINE, « Impartial Predictive Modeling and the Use of Proxy Variables », dans *17th International Conference, iConference 2022*, Évènement virtuel, 2022, DOI : 10.48550/arXiv.1608.00528.

KLEINBERG, J., H. LAKKARAJU, J. LESKOVEC, J. LUDWIG et S. MULLAINATHAN, « Human Decisions and Machine Predictions », (2018) 133-1 *The Quarterly Journal of Economics* 237-293, DOI : 10.1093/qje/qjx032.

LANG, M., « Reviewing Algorithmic Decision Making in Administrative Law », (2021) 26-2 *Lex Electronica* 195.

LIU, X., L. XIE, Y. WANG, J. ZOU, J. XIONG, Z. YING et A. V. VASILAKOS, « Privacy and Security Issues in Deep Learning: A Survey », (2021) 9 *IEEE Access*, DOI : 10.1109/ACCESS.2020.3045078.

MCCOY, L. G., C. T. A. BRENNA, S. S. CHEN, K. VOLD et S. DAS, « Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based », (2022) 142 *Journal of Clinical Epidemiology* 252-257, DOI : 10.1016/j.jclinepi.2021.11.001.

MEHRABI, N., F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, « A Survey on Bias and Fairness in Machine Learning », (2021) 54-6 *ACM Comput. Surv.* 115:1-115:35, DOI : 10.1145/3457607.

NGUYEN, G., S. DLUGOLINSKY, M. BOBÁK, V. TRAN, Á. LÓPEZ GARCÍA, I. HEREDIA, P. MALÍK et L. HLUCHÝ, « Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey », (2019) 52-1 *Artif Intell Rev* 77-124, DOI : 10.1007/s10462-018-09679-z.

PAYROVNAZIRI, S. N., Z. CHEN, P. RENGIFO-MORENO, T. MILLER, J. BIAN, J. H. CHEN, X. LIU et Z. HE, « Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review », (2020) 27-7 *J Am Med Inform Assoc* 1173-1185, DOI : 10.1093/jamia/ocaa053.

RANKIN, D., M. BLACK, R. BOND, J. WALLACE, M. MULVENNA et G. EPELDE, « Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing », (2020) 8-7 *JMIR Med Inform* e18910, DOI : 10.2196/18910.

ROBBINS, S., « A Misdirected Principle with a Catch: Explicability for AI », (2019) 29-4 *Minds & Machines* 495-514, DOI : 10.1007/s11023-019-09509-3.

ROCHER, L., J. M. HENDRICKX et Y.-A. DE MONTJOYE, « Estimating the success of re-identifications in incomplete datasets using generative models », (2019) 10-1 *Nat Commun* 3069, DOI : 10.1038/s41467-019-10933-3.

RUDIN, C. et J. RADIN, « Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition », (2019) 1-2 *Harvard Data Science Review*, DOI : 10.1162/99608f92.5a8a3a3d.

SANTONI DE SIO, F. et J. VAN DEN HOVEN, « Meaningful Human Control over Autonomous Systems: A Philosophical Account », (2018) 5 *Frontiers in Robotics and AI*, en ligne : <https://www.frontiersin.org/article/10.3389/frobt.2018.00015> (consulté le 13 avril 2022).

SARAVANAKUMAR, K. K., « The Impossibility Theorem of Machine Fairness -- A Causal Perspective », (2021) arXiv, 2021 *arXiv:2007.06024*, DOI : 10.48550/arXiv.2007.06024.

SELBST, A. D., « An Institutional View of Algorithmic Impact Assessments », (2021) 35-1 *Harvard Journal of Law and Technology* 75.

Seung S. H., Myoung S. K., Woohyung L., Gyeong H. P., Ilwoo P. et Sung E. C., « Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm », (2018) 138-7 *Journal of Investigative Dermatology* 1529-1538, DOI : 10.1016/j.jid.2018.01.028.

STADLER, T., Bristena OPRISANU et Carmela TRONCOSO, « Synthetic Data -- Anonymisation Groundhog Day », *arXiv: Learning* 2022, DOI : 10.48550/arXiv.2011.07018.

THOMPSON, M. Q., O. THEOU, G. R. TUCKER, R. J. ADAMS et R. VISVANATHAN, « FRAIL scale: Predictive validity and diagnostic test accuracy », (2020) 39-4 *Australasian Journal on Ageing* e529-e536, DOI : 10.1111/ajag.12829.

TRUDEL, P., « Quel droit et quelle régulation dans le cyberespace ? », (2000) 32-2 *Sociologie et sociétés* 21.

VEALE, M., R. BINNS et L. EDWARDS, « Algorithms that remember: model inversion attacks and data protection law », (2018) 376-2133 *Philos Trans A Math Phys Eng Sci* 20180083, DOI : 10.1098/rsta.2018.0083.

VEMOU, K. et M. KARYDA, « Evaluating privacy impact assessment methods: guidelines and best practice », (2019) 28-1 *ICS* 35-53, DOI : 10.1108/ICS-04-2019-0047.

WACHTER, S. et B. MITTELSTADT, « A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI », (2019) 2019-2 *Columbia Business Law Review* 494-620, DOI : 10.7916/cblr.v2019i2.3424.

WACHTER, S., B. MITTELSTADT et L. FLORIDI, « Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation », (2017) 7-2 *International Data Privacy Law* 76-99, DOI : 10.1093/idpl/ipx005.

WALONOSKI, J., M. KRAMER, J. NICHOLS, A. QUINA, C. MOESEL, D. HALL, C. DUFFETT, K. DUBE, T. GALLAGHER et S. MCLACHLAN, « Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record », (2018) 25-3 *Journal of the American Medical Informatics Association* 230-238, DOI : 10.1093/jamia/ocx079.

WEN, D., S. M. KHAN, A. JI XU, H. IBRAHIM, L. SMITH, J. CABALLERO, L. ZEPEDA, C. DE BLAS PEREZ, A. K. DENNISTON, X. LIU et R. N. MATIN, « Characteristics of publicly available skin cancer image datasets: a systematic review », (2022) 4-1 *The Lancet Digital Health* e64-e74, DOI : 10.1016/S2589-7500(21)00252-1.

WIJNHOVEN, F. et J. VAN HAREN, « Search Engine Gender Bias », (2021) 4 *Frontiers in Big Data*, en ligne : <https://www.frontiersin.org/articles/10.3389/fdata.2021.622106> (consulté le 14 août 2022).

WRIGHT, D., « The state of the art in privacy impact assessment », (2012) 28 *Computer Law and Security Review: The International Journal of Technology and Practice* 54-61.

YING, X., « An Overview of Overfitting and its Solutions », (2018) 1168-022022 *Journal of Physics: Conference Series* 7.

Articles de conférences

JOHNSON, K. D., D. P. FOSTER et R. A. STINE, « Impartial Predictive Modeling and the Use of Proxy Variables », dans *17th International Conference, iConference 2022*, Évènement virtuel, 2022, DOI : 10.48550/arXiv.1608.00528.

BENJAMINS, R., A. BARBADO et D. SIERRA, « Responsible AI by Design in Practice », dans *Proceedings of the Human-Centered AI: Trustworthiness of AI Models & Data (HAI) track at AAAI Fall Symposium*, Washington, États-Unis, 2019, DOI : 10.48550/arXiv.1909.12838.

BOGEN, M., A. RIEKE et S. AHMED, « Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination », dans *FAT* '20: Conference on Fairness, Accountability, and Transparency*, Barcelone, Espagne, 2020, DOI : 10.1145/3351095.3372877.

CABAÑAS, J. G., Á. CUEVAS et R. CUEVAS. *Facebook Use of Sensitive Data for Advertising in Europe*, 27th USENIX Security Symposium (2018) 479-495, Boston, États-Unis, 14 février 2018.

CORBETT-DAVIES, S., E. PIERSON, A. FELLER, S. GOEL et A. HUQ, « Algorithmic Decision Making and the Cost of Fairness », dans *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, coll. KDD '17, New York, NY, USA, Association for Computing Machinery, 2018, p. 797-806, DOI : 10.1145/3097983.3098095.

DAM, H. K., T. TRAN et A. GHOSE, « Explainable Software Analytics », *International Conference on Software Engineering'18 New Ideas and Emerging Results* 2018, en ligne : <http://arxiv.org/abs/1802.00603> (consulté le 17 février 2022).

DANKS, D. et A. J. LONDON, « Algorithmic Bias in Autonomous Systems », dans *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence AI and autonomy track (IJCAI-17)*, Melbourne, Australia, 2017, p. 6, en ligne : <https://www.ijcai.org/proceedings/2017/654> (consulté le 14 juin 2022).

DEY, A., VELAY, M., FAUVELLE, J-P., et NAVERS, S., « Adversarial vs behavioural-based defensive AI with joint, continual and active learning: automated evaluation of robustness to deception, poisoning and concept drift », *European Cyber Week- C&ESAR/IAD Conference- Artificial Intelligence and Defence*, Rennes, France, hal-02432377, 2019, p. 2.

DWORK, C., M. HARDT, T. PITASSI, O. REINGOLD et R. ZEMEL, « Fairness Through Awareness », dans *ITCS 2012: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, Cambridge, Massachusetts, arXiv, 2011, DOI : 10.48550/arXiv.1104.3913.

FREDRIKSON, M., S. JHA et T. RISTENPART, « Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures », dans *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, coll. CCS '15, New York, NY, USA, Association for Computing Machinery, 2015, p. 1322-1333, DOI : 10.1145/2810103.2813677.

HELLSTRÖM, T., V. DIGNUM et S. BENSCH, « Bias in Machine Learning -- What is it Good for? », dans *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), 2020. CEUR Workshop Proceedings 2659*, Santiago de Compostella, Espagne, arXiv, 2020, DOI : 10.48550/arXiv.2004.00686.

KEARNS, M., S. NEEL, A. ROTH et Z. S. WU, « Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness », dans *Proceedings of the 35th International Conference on Machine Learning : International Conference on Machine Learning*, 80, PMLR, 2018,

p. 2564-2572, en ligne : <https://proceedings.mlr.press/v80/kearns18a.html> (consulté le 24 août 2022).

KHALID, S., T. KHALIL et S. NASREEN, « A survey of feature selection and feature extraction techniques in machine learning », dans *2014 Science and Information Conference*, Londres, Royaume-Uni, 2014, p. 372-378, DOI : 10.1109/SAI.2014.6918213.

KLEINBERG, J., « Inherent Trade-Offs in Algorithmic Fairness », dans *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, coll. SIGMETRICS'18, New York, NY, USA, Association for Computing Machinery, 2018, p. 40, DOI : 10.1145/3219617.3219634.

LIPTON, Z. C., « The Mythos of Model Interpretability », dans *2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016),* New York, NY, USA, 2017, en ligne : <http://arxiv.org/abs/1606.03490> (consulté le 11 février 2022).

LIPTON, Z., A. CHOULDECHOVA et J. MCAULEY, « Does mitigating ML's disparate impact require disparate treatment? », dans *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada, 2018.

MCNAMARA, A., J. SMITH et E. MURPHY-HILL, « Does ACM's code of ethics change ethical decision making in software development? », dans *Conference: the 2018 26th ACM Joint Meeting*, New York, États-Unis, 2018, p. 729-733, DOI : 10.1145/3236024.3264833.

MEHRABI, N., F. MORSTATTER, N. SAXENA, K. LERMAN et A. GALSTYAN, « A Survey on Bias and Fairness in Machine Learning », (2021) 54-6 *ACM Comput. Surv.* 115:1-115:35, DOI : 10.1145/3457607.

MILLI, S., L. SCHMIDT, A. D. DRAGAN et M. HARDT, « Model Reconstruction from Model Explanations », dans *Proceedings of the Conference on Fairness, Accountability, and Transparency*, coll. FAT* '19, New York, NY, USA, Association for Computing Machinery, 2019, p. 1-9, DOI : 10.1145/3287560.3287562.

MITTELSTADT, B., C. RUSSELL et S. WACHTER, « Explaining Explanations in AI », dans *CM FAT* Conference 2019*, Atlanta, 2019, DOI : 10.1145/3287560.3287574.

MOORE, C., « Detecting Ransomware with Honeypot Techniques », dans *2016 Cybersecurity and Cyberforensics Conference (CCC)*, Amman, Jordanie, 2016, p. 77-81, DOI : 10.1109/CCC.2016.14.

PLEISS, G., M. RAGHAVAN, F. WU, J. KLEINBERG et K. Q. WEINBERGER, « On Fairness and Calibration », dans *Advances in Neural Information Processing Systems*, 30, Los Angeles, États-Unis, Curran Associates, Inc., 2017, en ligne : <https://papers.nips.cc/paper/2017/hash/b8b9c74ac526fffbeb2d39ab038d1cd7-Abstract.html> (consulté le 22 juin 2022).

SANDVIG, C., K. HAMILTON, K. KARAHALIOS et C. LANGBORT, « Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms », dans *Data and Discrimination: Converting Critical Concerns into Productive Inquiry, a preconference at the 64th Annual Meeting of the International Communication Association*, Seattle, États-Unis, 2014, p. 23, en ligne : <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>.

TSCHANTZ, M. C., « What is Proxy Discrimination? », dans *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency*, Seoul, Republic of Korea, arXiv, 2022, p. 1993-2003, DOI : 10.48550/arXiv.2205.05265.

YEOM, S., A. DATTA et M. FREDRIKSON, « Hunting for Discriminatory Proxies in Linear Regression Models », dans *32nd Conference on Neural Information Processing Systems (NIPS 2018)*, Montréal, Canada., 2018, p. 11.

ZHANG, B. H., B. LEMOINE et M. MITCHELL, « Mitigating Unwanted Biases with Adversarial Learning », dans *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society.*, New York, 2018, p. 6.

ZHANG, M., « Affirmative Algorithms: Relational Equality as Algorithmic Fairness », dans *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, Seoul, South Korea, Association for Computing Machinery, 2022, p. 13, DOI : https://doi.org/10.1145/3531146.3533115.

ZHOU, J., F. CHEN, A. BERRY, M. REED, S. ZHANG et S. SAVAGE, « A Survey on Ethical Principles of AI and Implementations », dans *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, Australia, 2020, p. 8.

Articles de journaux

CHAN, R., « LinkedIn is using AI to make recruiting diverse candidates a no-brainer », *Business Insider*, en ligne : <https://www.businessinsider.com/linkedin-new-ai-feature-increase-diversity-hiring-2018-10> (consulté le 22 juin 2022).

JAMES COOK, « Amazon patents new Alexa feature that knows when you're ill and offers you medicine », *The Telegraph* (2018), en ligne : <https://perma.cc/V346-HFWE> (consulté le 4 juin 2022).

LECAVALIER, C., « Données personnelles des Québécois: Le chien de garde réclame « des bras » et son indépendance », *La Presse*, sect. Politique (29 avril 2022), en ligne : <https://www.lapresse.ca/actualites/politique/2022-04-29/donnees-personnelles-des-quebecois/le-chien-de-garde-reclame-des-bras-et-son-independance.php> (consulté le 5 juin 2022).

MATTU, J. L., Julia Angwin,Lauren Kirchner,Surya, « How We Analyzed the COMPAS Recidivism Algorithm », *ProPublica* (2016), en ligne : <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=J5k3utYNqmWvBjTaTBs4TylpiUAiFx2o> (consulté le 13 avril 2022).

MATTU, J. A., Jeff Larson,Lauren Kirchner,Surya, « Machine Bias », *ProPublica*, en ligne : <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (consulté le 13 avril 2022).

SAM CORBETT-DAVIES, EMMA PIERSON, AVI FELLER, et SHARAD GOEL, « A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. », *Washington Post* (17 octobre 2016), en ligne :

<https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> (consulté le 24 juin 2022).

« Google apologises for Photo app's racist blunder », *BBC News,* sect. Technology (1 juillet 2015), en ligne : <https://www.bbc.com/news/technology-33347866> (consulté le 29 août 2022).

Monographies

A. I. PUBLISHING, *Deep Learning Crash Course for Beginners with Python: Theory and Applications of Artificial Neural Networks, CNN, RNN, LSTM and Autoencoders usingTensorFlow 2*, AI Publishing LLC, 2020.

BAROCAS, S., M. HARDT et A. NARAYANAN, *Fairness and Machine Learning : Limitation and opportunities*, fairmlbook.org, New York, 2019, en ligne : < https://fairmlbook.org/> (consulté le 24 juin 2022).

CRAWFORD, K., *Atlas of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, Yale University Press, 2021.

CONERY, J. S., *Explorations in Computing: An Introduction to Computer Science*, CRC Press, Boca Raton, FL, CRC Press, 2010.

FERGUSON, A. G., *The Rise of Big Data Policing*, New York, New York University Press, 2017.

GÉRON, A., *Deep Learning Avec Keras et Tensorflow 2ed.*, Paris, Dunod, 2020.

KEARNS, M. et A. ROTH, *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*, New York, Oxford University Press, 2019.

WRIGHT, D. et P. DE HERT (dir.), *Privacy impact assessment*, coll. Law, governance and technology series, volume 6, Dordrecht ; New York, Springer, 2012.

Rapports

ARTICLE 29 DATA PROTECTION WORKING PARTY et DIRECTORATE C (FUNDAMENTAL RIGHTS AND UNION CITIZENSHIP) OF THE EUROPEAN COMMISSION, *Opinion 05/2014 on Anonymisation Techniques*, WP216, 0829/14/EN, Brussel, 2014.

BARREAU DU QUÉBEC, *Mémoire du Barreau du Québec - Projet de loi no 64 — Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, CI– 031M C.P. – PL 64 Protection des renseignements personnels, Québec, Barreau du Québec, 2020.

BORGESIUS, FREDERIK ZUIDERVEEN, *Discrimination, intelligence artificielle et décisions algorithmiques*, coll. Publication de la Direction générale de la Démocratie, Strasbourg, Conseil de l'Europe, 2018.

COMMISSION D'ACCÈS À L'INFORMATION, *Guide d'accompagnement : Réaliser une évaluation des facteurs relatifs à la vie privée*, Commission d'accès à l'information, 2021.

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Intelligence artificielle et protection des renseignements personnels - Retour sur la consultation sur les principes en intelligence artificielle*, Québec, 2021, en ligne : <https://www.cai.gouv.qc.ca/publications-et-documentation/documents-de-reflexion-et-danalyse/>.

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Pour un développement responsable de l'intelligence artificielle qui respecte le droit à la vie privée et responsabilise tous les acteurs impliqués - Présenté à la Déclaration de Montréal pour une intelligence artificielle responsable*, Montréal, Commission d'accès à l'information du Québec, 2018, en ligne : <https://www.cai.gouv.qc.ca/publications-et-documentation/documents-de-reflexion-et-danalyse/>.

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, *Projet de loi no 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels - Mémoire de la Commission d'accès à l'information*, Québec, 2020

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE AU CANADA, *Rapport annuel au Parlement 2015-2016 concernant la Loi sur la protection des renseignements personnels et les documents électroniques*

et la Loi sur la protection des renseignements personnels: Le temps est venu de moderniser les outils du 20e siècle, Gatineau (Québec), Commissariat à la protection de la vie privée au Canada, 2016.

COMMISSION D'ACCÈS À L'INFORMATION, *Fiche d'information sur les pièces d'identité : entreprises*, en ligne : <https://www.cai.gouv.qc.ca/documents/CAI_FI_pieces_identite_entreprises.pdf> (consulté le 20 janvier 2022).

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - La méthode*, Commission nationale informatique et libertés, 2018.

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - Les bases de connaissances*, Commission nationale informatique et libertés, 2018.

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Analyse d'impact relative à la protection des données - Privacy impact assessment (PIA) - Les modèles*, Commission nationale informatique et libertés, 2018.

CRAWFORD, K., R. DOBBE, T. DRYER, G. FRIED, B. GREEN, E. KAZIUNAS, A. KAK, V. MATHUR, E. MCELROY, A. N. SÁNCHEZ, D. RAJI, J. L. RANKIN, R. RICHARDSON, J. SCHULTZ, S. M. WEST et M. WHITTAKER, *AI Now 2019 Report*, New York, AI Now Institute, New York University, 2019.

DÉZIEL, P.-L., K. BENYEKHLEF et E. GAUMOND, *Repenser la protection des renseignements personnels à la lumière des défis soulevés par l'IA*, Montréal, Observatoire international sur les impacts sociétaux de l'IA et du numérique et le Laboratoire de cyberjustice, 2020.

Dillon REISMAN, Jason SCHULTZ, Kate CRAWFORD et Meredith WHITTAKER, *Algorithmic impact assessment*, AiNow, 2018,

EID, P., J. MAGLOIRE et M. TURENNE, *Profilage racial et discrimination systémique des jeunes racisés: rapport de la consultation sur le profilage racial et ses conséquences*, Québec, Commission des droits de la personne et des droits de la jeunesse, 2011.

DATATILSYNET, *Artificial intelligence and privacy*, Oslo, Norvège, The Norwegian Data protection Authority, 2018, en ligne : <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>.

Dariusz KLOZA, Alessandra CALVI, Simone CASIRAGHI, Sergi MAYMIR et Nikolaos IOANNIDIS, *Analyse d'impact relative à la protection des données dans l'Union européenne : élaboration d'un modèle de rapport du processus d'analyse*, d. pia.lab Note de Politique N° 1/2020, Bruxelles, Belgique, Laboratoire bruxellois de l'analyse d'impact relative à la protection des données et de la vie privée (d.pia.lab), 2020, en ligne : <https://hal.archives-ouvertes.fr/hal-03332455/>.

ENISA, *Pseudonymisation techniques and best practices: recommendations on shaping technology according to data protection and privacy provisions.*, Union européenne, European Union Agency for Network and Information Security-Publications Office, 2019, en ligne : <https://data.europa.eu/doi/10.2824/247711> (consulté le 30 janvier 2022).

GAUTRAIS, VINCENT et N. AUBIN, *Modèle d'évaluation des facteurs relatifs à la circulation des données*, Montréal, Chaire L.R. Wilson sur le droit des technologies de l'information et du commerce électronique de l'Université de Montréal et l'Observatoire international sur les impacts sociétaux de l'IA et du numérique, 2022, en ligne : <https://www.gautrais.com/conferences/modele-devaluation-des-facteurs-relatifs-a-la-circulation-des-donnees/> (consulté le 10 juillet 2022).

GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE DE LA COMMISSION EUROPÉENNE, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, Bruxelles, Commission européenne, 2019.

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *A definition of AI: Main capabilities and Disciplines*, Brussels, European Commission, 2019.

HOLDREN JOHN P., AFUA BRUCE, ED FELTEN, TERAH LYONS, et MICHAEL GARRIS, *Preparing for the Future of Artificial Intelligence*, Washington, D.C., Executive Office of the President of The United States National Science and Technology Council Committee on Technology, 2016.

IBM SECURITY, *Cost of a Data Breach : Report 2022*, IBM, 2022, en ligne : <https://www.ibm.com/>.

INFORMATION COMMISSIONER'S OFFICE, *Guidance on AI and data protection*, Information Commissioner's Office, 2020, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/>.

INFORMATION COMMISSIONER'S OFFICE, *How should we assess security and data minimisation in AI?*, ICO, 2021, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/how-should-we-assess-security-and-data-minimisation-in-ai/> (consulté le 22 août 2022).

INFORMATION COMMISSIONER'S OFFICE, *Privacy Impact Assessment Handbook*, Royaume-Uni, Privacy Impact Assessment Handbook, 2009.

INFORMATION COMMISSIONER'S OFFICE et THE ALAN TURING INSTITUTE, *Explaining decisions made with Artificial Intelligence*, Royaume-Uni, Information Commissioner's Office, 2020, en ligne : <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/annexe-2-algorithmic-techniques/> (consulté le 12 août 2022).

MEREDITH WHITTAKER, KATE CRAWFORD, ROEL DOBBE, GENEVIEVE FRIED, ELIZABETH KAZIUNAS, VAROON MATHUR, SARAH MYERS WEST, RASHIDA RICHARDSON, JASON SCHULTZ, et OSCAR SCHWARTZ, *AI Now 2018 Report*, New York, AI Now Institute, New York University, 2018.

MICROSOFT, *Microsoft responsible AI principles*, en ligne : <https://www.microsoft.com/en-us/ai/our-approach> (consulté le 17 juillet 2022).

MOSS, E., E. A. WATKINS, R. SINGH, M. C. ELISH et J. METCALF, *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest*, 3877437, Rochester, NY, Data & Society, 2021, en ligne : <https://papers.ssrn.com/abstract=3877437> (consulté le 29 juillet 2022).

NICHOLAS DIAKOPOULOS, SORELLE FRIEDLER, MARCELO ARENAS, SOLON BAROCAS, MICHAEL HAY, BILL HOWE, H. V. JAGADISH, KRIS UNSWORTH, ARNAUD SAHUGUET, SURESH VENKATASUBRAMANIAN, CHRISTO WILSON, CONG

YU, et BENDERT ZEVENBERGEN, *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, Fairness, Accountability, and Transparency in Machine Learning, en ligne : <https://www.fatml.org/resources/principles-for-accountable-algorithms> (consulté le 17 juillet 2022).

RUGGIE, J., *Rapport du Représentant spécial du Secrétaire général chargé de la question des droits de l'homme et des sociétés transnationales et autres entreprises - Principes directeurs relatifs aux entreprises et aux droits de l'homme: mise en œuvre du cadre de référence «protéger, respecter et réparer»*, A/HR/17/31, Nations Unies, Conseil des droits de l'homme, 2011.

TRUDEL, J.-F., R. AVILA, G. ST-LAURENT et R. AVILA, *Mémoire à la Commission d'accès à l'information sur le document de consultation « Intelligence artificielle »*, Cat. 2.412.133, coll. Document adopté à la 681e séance de la Commission, tenue le 15 mai 2020, par sa résolution COM-681-4.2.1, Québec, Commission des droits de la personne et des droits de la jeunesse, 2020.

TRUDEL, J.-F., A. BERWALD, M. CARPENTIER et M. FORCIER, *Mémoire à la Commission des institutions de l'Assemblée nationale - Projet de loi n° 64, Loi modernisant les dispositions législatives en matière de protection des renseignements personnels*, Québec, Commission des droits de la personne et des droits de la jeunesse, 2020.

VICTOR ARMONY, MARIAM HASSAOUI, et MASSIMILIANO MULONE, *Les interpellations policières à la lumière des identités racisées des personnes interpellées : Analyse des données du Service de Police de la Ville de Montréal (SPVM) et élaboration d'indicateurs de suivi en matière de profilage racial (Rapport final remis au SPVM)*, Montréal, Canada, Équipe Armony-Hassaoui-Mulone (Chercheurs indépendants), en ligne : <https://spvm.qc.ca/upload/Rapport_Armony-Hassaoui-Mulone.pdf>.

WILLIAM DIETERICH, CHRISTINA MENDOZA, et TIM BRENNAN, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity : Performance of the COMPAS Risk Scales in Broward County*, Northpointe Inc., Northpointe Inc. Research Department, 2016.

Consultations et études du projet de loi 64

ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », (23 septembre 2020) 45-96 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-200923.html> (consulté le 4 juin 2022).

ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », (29 septembre 2020)) 45-96 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-200929.html> (consulté le 30 janvier 2022).

ASSEMBLÉE NATIONALE DU QUÉBEC, « Consultations particulières et auditions publiques sur le projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », (17 mars 2021) 45-126 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210317.html> (consulté le 6 juin 2022).

ASSEMBLÉE NATIONALE DU QUÉBEC, « Étude détaillée du projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », (17 février 2021) 45-120 *Journal des débats de la Commission des institutions*, en ligne : <http://www.assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210217.html> (consulté le 30 janvier 2022).

ASSEMBLÉE NATIONALE DU QUÉBEC, « Étude détaillée du projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels », *Journal des débats de la Commission des institutions* 45-132 (31 mars 2021), en ligne : <http://assnat.qc.ca/fr/travaux-parlementaires/commissions/ci-42-1/journal-debats/CI-210331.html> (consulté le 21 janvier 2022).

Thèse de doctorat

MOUCHARD, E., *L'accountability ou le principe de responsabilité en matière de protection des renseignements personnels*, Thèse de doctorat, Université de Montréal et Université Paris-Sud XI, 2018.

Dictionnaires

« Définitions : facteur », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/dictionnaires/francais/facteur/32600> (consulté le 17 août 2022).

« Définitions : raison », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/dictionnaires/francais/raison/66270> (consulté le 17 août 2022).

« Intelligence artificielle », dans *Dictionnaire de français Larousse*, Éditions Larousse, en ligne : <https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257> (consulté le 21 juillet 2022).

« Proxy Variable », dans *A Dictionary of Statistics*, Oxford University Press, 2014, en ligne : <https://www.oxfordreference.com/view/10.1093/acref/9780199679188.001.0001/acref-9780199679188-e-1315> (consulté le 25 juillet 2022).

« Predictive validity », dans *American Psychological Association Dictionary of Psychology*, en ligne : <https://dictionary.apa.org/> (consulté le 26 août 2022).

Sites internet et logiciels

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Privacy Impact Assessments (PIAs) » (16 septembre 2019), en ligne : <https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/> (consulté le 28 juillet 2022).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Communiqué : L'utilisation par la GRC de la technologie de reconnaissance faciale de Clearview AI contrevenait à la Loi sur la protection des renseignements personnels, selon une enquête » (10 juin 2021), en ligne :

<https://www.priv.gc.ca/fr/nouvelles-du-commissariat/nouvelles-et-annonces/2021/nr-c_210610/> (consulté le 5 juin 2022).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Expectations: OPC's Guide to the Privacy Impact Assessment Process » (13 octobre 2011), en ligne : <https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/gd_exp_202003/> (consulté le 28 juillet 2022).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Lignes directrices pour l'obtention d'un consentement valable » (24 mai 2018), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/collecte-de-renseignements-personnels/consentement/gl_omc_201805/> (consulté le 4 novembre 2021).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Nos attentes : Guide du Commissariat au sujet du processus d'évaluation des facteurs relatifs à la vie privée » (13 octobre 2011), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/evaluations-des-facteurs-relatifs-a-la-vie-privee/gd_exp_202003/> (consulté le 28 juillet 2022).

COMMISSARIAT À LA PROTECTION DE LA VIE PRIVÉE DU CANADA, « Position de principe sur la publicité comportementale en ligne » (4 décembre 2015), en ligne : <https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/technologie/protection-de-la-vie-privee-en-ligne-surveillance-et-temoins/pistage-et-publicite/bg_ba_1206/> (consulté le 4 novembre 2021).

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « La collecte de renseignements personnels », en ligne : <https://www.cai.gouv.qc.ca/la-collecte-de-renseignements-personnels/> (consulté le 10 mars 2022).

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « Pointage électronique des employés par géolocalisation », en ligne : <https://www.cai.gouv.qc.ca/pointage-electronique-employes-par-geolocalisation-regles-a-respecter/> (consulté le 8 juin 2022).

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC, « Évaluation des facteurs relatifs à la vie privée », en ligne : <https://www.cai.gouv.qc.ca/espace-evolutif-modernisation-lois/thematiques/evaluation-facteurs-relatifs-vie-privee/> (consulté le 28 juillet 2022).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Paramètre (IA) », en ligne : <https://www.cnil.fr/fr/definition/parametre-ia> (consulté le 17 août 2022).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (CNIL), « Facial recognition: the CNIL orders CLEARVIEW AI to stop reusing photographs available on the Internet » (16 décembre 2021), en ligne : <https://www.cnil.fr/en/facial-recognition-cnil-orders-clearview-ai-stop-reusing-photographs-available-internet> (consulté le 7 août 2022).

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Caractéristique », en ligne : <https://www.cnil.fr/fr/definition/caracteristique> (consulté le 17 août 2022).

DAVID WEINBERGER, « Playing with AI Fairness », en ligne : <https://pair-code.github.io/what-if-tool/ai-fairness.html> (consulté le 10 juillet 2022).

DANIEL DIAMOND, LEWIS BAKER, MARK WARD, DAVID WEINBERGER, VISHAL MURALI, et JOE NASO. *Pymetrics. 2019. audit-AI (Python library).*, Python, 2020.

ÉCOLE DE LA FONCTION PUBLIQUE DU CANADA, « L'intelligence artificielle est à nos portes : Savoir quand et comment utiliser l'IA au gouvernement (vidéo) » (24 mai 2022), en ligne : <https://www.csps-efpc.gc.ca/video/artificial-intelligence-here-series/deciding-when-fra.aspx> (consulté le 29 juillet 2022).

EUROPEAN DATA PROTECTION SUPERVISOR, « Synthetic Data », en ligne : <https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_fr> (consulté le 27 juin 2022).

FAIRLEARN 0.7.0, « Frequently asked questions », en ligne : <https://fairlearn.org/v0.7.0/faq.html> (consulté le 22 mai 2022).

FAIRLEARN 0.7.0, « Mitigation », en ligne : <https://fairlearn.org/v0.7.0/user_guide/mitigation.html> (consulté le 27 juin 2022).

FAIRLEARN 0.7.0, « fairlearn.preprocessing package », en ligne : <https://fairlearn.org/v0.7.0/api_reference/fairlearn.preprocessing.html#fairlearn.preprocessing.CorrelationRemover> (consulté le 29 août 2022).

FAIRLEARN 0.8.0, « fairlearn.postprocessing package », en ligne : <https://fairlearn.org/main/api_reference/fairlearn.postprocessing.html#fairlearn.postprocessing.ThresholdOptimize> (consulté le 8 juillet 2022).

FEDERAL BUREAU OF INVESTIGATION, « 2019 Crime in the United States - Clearances », *FBI* (2020), en ligne : <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/clearances> (consulté le 24 juin 2022).

GOUVERNEMENT DU QUÉBEC, « Décision fondée exclusivement sur un traitement automatisé », en ligne : <https://www.quebec.ca/gouvernement/travailler-gouvernement/services-employes-etat/conformite/protection-des-renseignements-personnels/technologie-et-droit-a-la-protection-des-renseignements-personnels/decision-traitement-automatise> (consulté le 20 janvier 2022).

GOUVERNEMENT DU QUÉBEC, « Renseignements personnels sensibles », en ligne : <https://www.quebec.ca/gouvernement/travailler-gouvernement/services-employes-etat/conformite/protection-des-renseignements-personnels/consentement/renseignements-personnels-sensibles> (consulté le 4 novembre 2021).

INFORMATION COMMISSIONER'S OFFICE, « ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted » (31 mai 2022), en ligne : <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/> (consulté le 7 août 2022).

GOOGLE, « AI at Google: our principles », *Google* (7 juin 2018), en ligne : <https://blog.google/technology/ai/ai-principles/> (consulté le 12 avril 2022).

GOOGLE AI BLOG, « The What-If Tool: Code-free probing of machine learning models », en ligne : <https://pair-code.github.io/what-if-tool/> (consulté le 21 août 2019).

IBM, « AI Fairness 360 toolkit », en ligne : <https://github. com/IBM/AIF360> (consulté le 21 août 2019).

OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - facteur (statistique) » (1975), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8477014> (consulté le 17 août 2022).

OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - facteur (philosophie) » (1979), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8462443> (consulté le 17 août 2022).

OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - mode (statistique) » (2012), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=17563886> (consulté le 24 août 2022).

OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE, « Grand dictionnaire terminologique - intelligence artificielle » (2018), en ligne : <https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26552508> (consulté le 30 janvier 2022).

UNIVERSITÉ DE MONTRÉAL, « La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », *declarationiaresp*, en ligne : <https://www.declarationmontreal-iaresponsable.com/la-declaration-2> (consulté le 10 février 2022).

UNIVERSITY OF CHICAGO, CENTER FOR DATA SCIENCE AND PUBLIC POLICY, « Aequitas - Bias and Fairness Audit Toolkit » (2018), en ligne : <http://aequitas.dssg.io/audit/n5uppdoy/compas_for_aequitas/report-1.html> (consulté le 7 juin 2022).

ZHONG, Z., « A Tutorial on Fairness in Machine Learning », *Toward data science* (19 juin 2020), en ligne : <https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb> (consulté le 24 juin 2022).

Recherches sur des bases de données

CANLII, « Ewert; citing Loi sur la protection des renseignements personnels dans le secteur privé, RLRQ c P-39.1 », en ligne : <https://www.canlii.org/fr/#search/text=Ewert&origin1=/fr/qc/legis/lois/rlrq-c-p-39.1/derniere/rlrq-c-p-39.1.html&section1=11&linkedNoteup=> (consulté le 27 août 2022).

CANLII, « Ewert; citing Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels, RLRQ c A-2.1 ».

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC- SECTION SURVEILLANCE, « Recherche avancée - "Analyse d'impacts" », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=%22Analyse+d%27impacts%22&ref=&d1=&d2=&p=&col=162&or=> (consulté le 2 août 2022).

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC- SECTION SURVEILLANCE, « Recherche avancée - "Évaluation des facteurs relatifs à la vie privée" », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=%22%C3%89valuation+des+facteurs+relatifs+%C3%A0+la+vie+priv%C3%A9e%22&ref=&d1=&d2=&p=&col=162&or=> (consulté le 2 août 2022).

COMMISSION D'ACCÈS À L'INFORMATION DU QUÉBEC - SECTION SURVEILLANCE, « Recherche avancée : Ewert », en ligne : <https://decisions.cai.gouv.qc.ca/cai/fr/d/s/index.do?cont=ewert&ref=&d1=&d2=&p=&col=162&or=> (consulté le 27 août 2022).

SOQUIJ, « Recherche Juridique : Ewert c. Canada (C.S. Can., 2018-06-13), 2018 CSC 30, SOQUIJ AZ-51502366, 2018EXP-1629, [2018] 2 R.C.S. 165 et Accès aux documents des organismes publics et sur la protection des renseignements personnels (Loi sur l'), (RLRQ, c. A-2.1) », en ligne : <https://soquij.qc.ca/portail/recherchejuridique/Rechercher/5283005> (consulté le 27 août 2022).

Annexe

Mesures ou stratégies	Explicabilité	Exactitude	Sécurité	Équité		
				Traitement	Performance	Statistique
Fournir une explication détaillée du modèle de SIA au public dans un contexte dynamique.	☑	☒	☒*	☐	☐	☐
Utiliser un SIA dynamique dans un contexte dynamique.	☒	☑	☒	☐	☐	☐
Interpréter restrictivement le critère de nécessité. (par e.g. Se limiter aux renseignements personnels dits « essentiels ».)	☐	☒	☑	☑	☒	☒
Collecter un nombre élevé de variables et d'instances à des fins d'entrainement.	☐	☑	☒	☐	☐	☐
Favoriser un SIA qui utilise un nombre élevé de variables.	☒	☑*	☒	☐	☐	☐
Favoriser des SIA simples et transparents.	☑	☒*	☐	☐	☐	☐
Collecter des attributs protégés :	☐	☐	☒	☒	☑	☑
Afin de détecter l'absence d'une parité statistique.	☐	☐	☒	☒	☐	☑
Afin de détecter l'absence d'une parité de performance.	☐	☐	☒	☒	☑	☐

Afin d'utiliser des jeux d'entrainement diversifiés.	☐	☐	☒	☒	☑	☐
Afin d'utiliser des algorithmes de mitigation favorisant une plus grande parité de performance.	☐	☒*	☒	☒	☑	☒
Afin d'utiliser des algorithmes de mitigation ou d'appliquer des seuils favorisant une plus grande parité statistique.	☐	☒	☒	☒	☒	☑

☑ : Peut favoriser un meilleur respect.

☒ : Présence de risques.

☐ : Non évalué dans le cadre de la présente étude.

*** : Tensions ou bénéfices contextuels**

Tableau 2. – Les tensions identifiées