

HEC MONTRÉAL

L'intelligence artificielle affective
Enjeux éthiques et conditions d'acceptabilité

par

Mélissa Lafond

Joé T. Martineau
HEC Montréal
Directrice de recherche

Sciences de la gestion
(Spécialisation Management)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Avril 2022
© Mélissa Lafond, 2022

Retrait d'une ou des pages pouvant contenir des renseignements personnels

Résumé

Ce mémoire vise à cerner les enjeux éthiques que soulèvent l'intelligence artificielle et plus spécifiquement l'informatique affective et à prendre le pouls des conditions d'acceptabilité pour le déploiement de tels systèmes au sein de la population. Nous explorons le sujet sur le terrain dans le contexte du déploiement d'une technologie d'informatique affective de l'entreprise en démarrage EmoScienS. Une cohorte de participants a été invitée à utiliser la technologie de détection des émotions pour une durée de 2 semaines. La méthode qualitative centrée sur l'entrevue semi-dirigée a été privilégiée pour recueillir la rétroaction de l'expérience terrain. Cette méthodologie permet d'optimiser le caractère exploratoire de la recherche. L'analyse des résultats a pris deux tangentes, une analyse des enjeux éthiques de la technologie via un cadre éthique reconnu, la *Déclaration de Montréal*, et l'analyse des critères d'acceptabilité via un modèle théorique que nous proposons pour nous ajuster au cadre spécifique de notre recherche. Ce mémoire contribue à l'enrichissement de la connaissance sur l'acceptabilité et l'éthique de l'informatique affective entre autres en proposant un modèle pour évaluer l'acceptabilité de la technologie et en mettant en lumière les enjeux sensibles aux utilisateurs.

Mots clés : Intelligence artificielle, informatique affective, enjeux éthiques, acceptabilité

Méthodes de recherche : Qualitative par entrevues semi-dirigées

Table des matières

Résumé.....	VI
Table des matières.....	VII
Liste des tableaux et liste des figures.....	XI
Remerciements.....	XII
Introduction.....	1
1 Revue de littérature.....	3
1.1 Contexte et problématique	3
1.2 Historique de l'intelligence artificielle.....	4
1.2.1 Définition	4
1.2.2 Historique	5
1.3 Survol sommaire de la technologie	7
1.3.1 Applications de l'intelligence artificielle.....	10
1.4 Enjeux de l'intelligence artificielle.....	11
1.4.1 Transformation du marché du travail	11
1.4.2 Fiabilité	12
1.4.3 Sécurité et malveillance	14
1.4.4 Biais et discrimination	14
1.4.5 Responsabilité	18
1.4.6 Transparence et explicabilité	22
1.4.7 Consentement, respect de la vie privée et autonomie	26
1.5 Encadrer le développement et le déploiement de l'intelligence artificielle	29
1.5.1 Commission européenne.....	31
1.5.2 Initiatives supranationales	32
1.5.3 Initiatives nationales.....	33
1.5.4 Initiatives universitaires	34
1.5.5 Initiatives organisationnelles.....	35
1.5.6 Force de loi	36
1.5.7 Encadrement québécois et canadien	37

1.6	Intelligence artificielle affective	38
1.6.1	Enjeux éthiques en lien avec l'intelligence artificielle affective	41
1.7	Conditions d'acceptabilité de l'IA	45
1.8	Initiatives proposées dans la littérature pour augmenter la confiance à l'égard de la technologie	49
1.8.1	Encadrement : équilibre entre l'autorégulation et la réglementation stricte	49
1.8.2	Développement responsable	51
2	Cadre d'analyse	56
2.1	Déclaration de Montréal	57
2.1.1	Objectifs de la Déclaration de Montréal	57
2.1.2	Principes de la Déclaration de Montréal	57
2.2	Conditions d'acceptabilité	63
3	Méthodologie	67
3.1	Stratégie de recherche	67
3.1.1	Justifications de la méthodologie	67
3.2	Sélection du cas et collecte de données	68
3.2.1	Le cas : La technologie de détection des émotions d'EmoSciens	68
3.2.2	Collecte de données	71
3.2.2.1	Outils de collecte de données	73
3.2.3	Analyse des données	74
3.2.4	Considérations éthiques	74
4	Analyse	75
4.1	Perception des participants par rapport aux nouvelles technologies et à l'intelligence artificielle	75
4.1.1	Intérêt, aisance et style d'adoption par rapport aux nouvelles technologies	75
4.1.2	L'intelligence artificielle aux yeux des utilisateurs	76
4.1.2.1	Perception et littératie	76
4.1.2.2	Bénéfices perçus	79
4.1.2.3	Enjeux	80
4.1.2.4	Partage des données personnelles et consentement	85
4.1.2.5	Sentiment par rapport aux conditions d'utilisation	87
4.2	Rétroaction sur la technologie d'EmoSciens utilisée dans le cadre du projet de recherche	88

4.2.1	Facilité d'utilisation	89
4.2.2	Validité des résultats	92
4.2.3	Bénéfices perçus quant à la technologie d'EmoScienS	95
4.2.4	La technologie d'EmoScienS vue à travers les principes de la Déclaration de Montréal 98	
4.2.4.1	Principe de bien-être	98
4.2.4.2	Principe de respect de l'autonomie	100
4.2.4.3	Principe de protection de l'intimité et de la vie privée	102
4.2.4.4	Principe de solidarité	104
4.2.4.5	Principe de participation démocratique.....	105
4.2.4.6	Principe d'équité	106
4.2.4.7	Principe d'inclusion de la diversité	107
4.2.4.8	Principe de prudence.....	108
4.2.4.9	Principe de responsabilité	109
4.2.5	Réflexions quant au partage d'information dans le cadre de l'utilisation d'EmoScienS	110
4.3	Critères d'acceptabilité.....	112
4.4	Perception par rapport aux initiatives et aux mesures mises en place pour encadrer l'intelligence artificielle	120
4.5	Mesures recommandées par les participants pour une IA éthique et responsable ...	122
5	Interprétation des résultats et recommandations	124
5.1	Intérêt et littératie par rapport à l'intelligence artificielle	124
5.2	Enjeux éthiques liés à l'IA	124
5.3	Enjeux éthiques soulevés par l'intelligence artificielle affective dans le contexte de l'utilisation d'EmoScienS	127
5.3.1	EmoScienS à travers les yeux de la Déclaration de Montréal	128
5.3.2	Facteurs à considérer pour une plus grande acceptabilité	129
5.4	Encadrement de l'intelligence artificielle.....	130
5.5	Recommandations.....	132
5.5.1	Établir les règles du jeu.....	132
5.5.2	Une plus grande transparence et miser sur la diversité.....	133
5.5.3	Responsabilité sociale de la profession.....	134
5.5.4	L'éducation, un levier pour l'acceptabilité	135
5.6	Limites de l'étude et avenues de recherches futures	137

Conclusion	139
Bibliographie.....	141
Annexes.....	156
Annexe A : Guide d’entrevue initial.....	157
Annexe B : Guide d’entrevue révisé	161
Annexe C : Grille de codage.....	165

Liste des tableaux et liste des figures

Liste des tableaux

Tableau 1.1 Principes de la bioéthique	31
Tableau 1.2 Recommandations du Conseil sur l'intelligence artificielle	32
Tableau 1.3 Principes de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle	34
Tableau 2.1 Objectifs de la Déclaration de Montréal	57
Tableau 2.2 Principes de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle	58
Tableau 3.1 Caractéristiques de l'échantillon	72
Tableau 4.1 Description de l'intelligence artificielle selon les participants	77
Tableau 4.2 Perception de la facilité d'utilisation de la technologie d'EmoScienS	89
Tableau 4.3 Perception des participants par rapport à la validité des résultats	92
Tableau 4.4 Bénéfices individuels et organisationnels de l'utilisation de la technologie d'EmoScienS	95
Tableau 4.5 Partage des données émotionnelles avec l'organisation	111
Tableau 4.6 Perception par rapport à l'encadrement de l'intelligence artificielle	120
Tableau 4.7 Résumé des mesures recommandées par les participants pour assurer une intelligence artificielle éthique et responsable	122

Liste des figures

Figure 2.1 Critères d'acceptabilité pour un SIA	64
Figure 4.1 Critères d'acceptabilité pour un SIA	115

Remerciements

Il y a quelques lunes, j'ai entrepris le projet ambitieux de compléter une maîtrise en management. Gonflée à bloc et déterminée à démontrer qu'avec une bonne dose de discipline et de persévérance tout est possible, j'ai avancé dans ce projet comme une funambule équilibrant une carrière en plein essor, une charge académique stimulante et une vie familiale vibrant au diapason d'un jeune garçon en soif d'aventure. Avec du recul je reconnais ma naïveté devant l'ampleur de la tâche et les sacrifices qu'elle allait me demander pour atteindre mon objectif. En pleine pandémie, j'ai quitté mon poste au sein d'une multinationale pour me dédier à 100 % à la tâche de compléter mon mémoire. Sans le soutien indéfectible de certaines personnes, j'avoue bien humblement que je n'y serais fort probablement jamais arrivée.

Je tiens à remercier premièrement Joé Martineau ma directrice de recherche qui a accepté de me superviser et de me guider au cours du plus long projet universitaire que j'aie entrepris.

Merci à Marie-Ève, Virginie et Véronique pour leur amitié inconditionnelle et leurs constants encouragements dans les situations où les obstacles me semblaient insurmontables.

Merci à Bruno et Mathieu pour tout le soutien, l'écoute et l'expérience partagée qui m'ont été d'une aide incroyablement précieuse. Dans les moments où j'avançais à petits pas dans le brouillard, vos bons conseils m'ont servi à la façon d'une boussole pour garder le cap.

Merci à tous ceux qui ont participé de près ou de loin à mon travail de recherche, votre implication a directement contribué à sa réalisation et je vous en suis profondément reconnaissante.

Merci à mes parents qui depuis que je suis toute petite travaillent avec acharnement à balayer les embûches sur mon chemin pour que je puisse aller jusqu'au bout de mes aventures les plus folles. Votre écoute, vos encouragements et votre confiance en moi me donnent des ailes. Vous avez été une pièce maîtresse à chaque étape du processus, toujours

prêts à aider là où vous le pouviez. Sans vous je n'y serais probablement jamais arrivée. Cet accomplissement est le vôtre aussi.

Enfin merci à toi Vincent, tu es la bougie d'allumage qui me donne l'énergie d'accomplir l'impossible. Durant la dernière année, tu as été incroyablement résilient face aux nombreux sacrifices que la pandémie et mon travail universitaire t'ont imposé. Tu as été patient, compréhensif et mon *fan* numéro. Tu as cru en moi avec une force qui m'a poussé à persévérer. J'espère t'avoir inspiré à poursuivre tes rêves avec passion et détermination, quels qu'ils soient.

Introduction

Depuis des siècles, l'Homme imagine la création de formes intelligentes artificielles (Gawdat, 2021). L'histoire est pleine de contes folkloriques, de légendes, de récits, de romans et plus récemment de films qui racontent chacun à leurs façons le parcours d'une créature créée par l'homme. Dans la mythologie juive, le Golem qui est fait d'argile prend une forme humanoïde, il est généralement dépourvu de libre-arbitre et est façonné pour défendre ou servir son créateur (Fauchaux, 2017). Dans son célèbre livre paru en 1818, Mary Shelley nous partage l'histoire de Victor Frankenstein un jeune scientifique qui réussit à donner vie à une créature composée de membres de cadavres rapiécés (Gawdat, 2021). En 1920, Karel Capek un auteur tchèque écrit la pièce de théâtre de science-fiction Les Robots Universels de Rossum. Il est d'ailleurs le premier à utiliser la terminologie 'robot' qui signifie en tchèque 'servitude' donc au service ou dépendant de (Gawdat, 2021). Cependant, ce n'est que dans les dernières années que la réalité a rattrapé en partie la fiction. L'intelligence artificielle (IA), propulsée tout particulièrement par les percées en apprentissage automatique de la dernière décennie a permis l'élaboration de technologies de pointes et des avancées majeures dans différents domaines. Ses sphères d'applications sont si vastes que certains font références à une 4^{ème} révolution industrielle (Madiaga, 2019). Selon Ahmad et al. (2018), ce n'est qu'une question de temps avant que la technologie soit partout dans le milieu de la santé. Des chercheurs du Massachusetts Institute of Technology (MIT) se sont intéressées aux capacités de l'intelligence artificielle pour développer une technologie d'informatique affective capable de détecter les émotions des gens par l'entremise de signaux sans fil (Mathys, 2016). De la même façon, des technologies comme deepmood s'intéressent aux émotions. Il est possible de capter des données de séquences respiratoires, d'expressions faciales, le ton de la voix ou encore des données textuelles via entre autres les médias sociaux pour en décoder l'état émotionnel des gens ainsi que leur état mental (Barrett et al., 2017). Comme le dit Mathy (2016) « l'ennui, c'est que le potentiel de dérapage est énorme ». L'éthique entourant l'intelligence artificielle, est maintenant un grand sujet de discussion dans le milieu universitaire et dans le domaine de la réglementation (Mittelstadt, 2019). La littérature sur

le sujet étant encore relativement jeune, par l'entremise de notre projet de recherche, nous avons choisis d'en étudier deux angles complémentaires qui se résument à ces questions :

Quels sont les enjeux éthiques entourant l'intelligence artificielle affective?

Quels sont les critères qui influencent l'acceptabilité de l'intelligence artificielle affective ?

Pour examiner ces questions, cette étude a utilisé le contexte de l'évaluation de la technologie d'intelligence artificielle affective de la compagnie montréalaise EmoScienS. Les individus invité à participer au projet de recherche ont utilisé une technologie de détection des émotions pour une période déterminée. Pour recueillir la rétroaction des participants, nous avons privilégié une approche qualitative centrée sur l'entrevue semi-dirigée. Au total, 14 entrevues ont été menées auprès de participants de profils différents.

Cette étude aborde les enjeux éthiques liés au développement et au déploiement d'un système d'intelligence artificielle affective et met en lumière les enjeux sensibles pour les utilisateurs. Elle met en évidence les critères qui influencent l'acceptabilité face à cette technologie et propose un modèle. De plus, elle fait la revue de pistes de solution pour augmenter l'acceptabilité face l'informatique affective.

La structure de ce travail comprend 5 chapitres : Le chapitre 1 présente une revue de la littérature en abordant l'intelligence artificielle de façon générale et plonge de façon plus spécifique au niveau de l'informatique affective. Au chapitre 2, nous détaillons le cadre d'analyse que nous avons privilégié, soit la *Déclaration de Montréal* pour les enjeux éthiques et un modèle que nous avons développé pour l'acceptabilité de l'intelligence artificielle. Le chapitre 3 décrit et justifie notre choix de méthodologie. Le chapitre 4 détaille l'ensemble des résultats récoltés au cours de nos 14 entrevues. Enfin, le chapitre 5 aborde les interprétations, discussions et recommandations qui émanent du terrain tout en exposant les limites et avenues de recherche à investiguer.

1 Revue de littérature

1.1 Contexte et problématique

Depuis quelques années, l'intelligence artificielle prend une place plus prépondérante dans nos vies et on y fait de plus en plus référence dans différentes sphères. Bien qu'elle n'ait pas encore la complexité et la subtilité de l'intelligence humaine, l'objectif de ses concepteurs est d'en imiter certains attributs. Ses performances toujours plus impressionnantes révolutionnent et frappent l'imaginaire. En 1997, Deep Blue (IBM), un système d'intelligence artificielle (SIA), remporte un tournoi d'échecs contre le champion du monde Garry Kasparov (Maclure et Saint-Pierre, 2018). En 2016, le superprogramme AlphaGo de DeepMind (une filiale de Google) marque l'histoire en remportant un match de go contre le champion mondial Lee Sedol. Établi depuis des millénaires, le jeu de go est d'une grande complexité et fait appel à des habiletés très pointues de stratégie. Pour les concepteurs de DeepMind, il n'y a pas de doute : le go était l'« Everest » de l'intelligence artificielle (AFP, 2016). Cette victoire a suscité de vives réactions face à la grandeur de l'exploit et aux promesses d'une telle technologie, mais elle a aussi soulevé des craintes par rapport à une intelligence artificielle qui puisse dépasser l'humain (Khos, 2017). L'intelligence artificielle exerce déjà une certaine forme d'influence sur les choix et les décisions de dizaines de millions de personnes mondialement via des plateformes comme Netflix, Amazon et YouTube, qui filtrent le contenu disponible et propose des options aux utilisateurs selon leur profil de préférence. Bien qu'ils puissent sembler banals au premier regard, le pouvoir prédictif et l'usage qui en est fait ont déjà des effets disruptifs dans la société, et des conséquences imprévues se font déjà sentir (Ahmad, Eckert et Teredesai, 2018). Le scandale de Cambridge Analytica démontre combien une utilisation malveillante ou cachée à des fins de manipulation peut avoir des effets pervers sur la société (Wong, 2019). L'intelligence artificielle affective, dont les algorithmes visent à détecter ou à simuler des émotions, connaît aussi un vif intérêt et s'intègre tranquillement mais sûrement à notre environnement. La technologie est déjà utilisée pour capter les émotions d'employés et de consommateurs en décodant leurs expressions faciales, en analysant leurs intonations de voix ou en prenant la mesure du mouvement de leurs yeux (Purdy, Zealley

et Maseli, 2019). La possibilité d'avoir accès aux émotions des gens via différentes plateformes et applications ouvre la porte à une panoplie d'enjeux éthiques. Considérant l'engouement pour l'intelligence artificielle et le potentiel que présente l'intelligence artificielle affective, nous nous sommes intéressées à mieux comprendre les enjeux éthiques qui entourent la technologie et les conditions d'acceptabilité de celle-ci. L'objectif de cette recherche est d'explorer les enjeux éthiques entourant la technologie et les pistes de solutions pour en assurer une grande acceptabilité dans un cadre propice à la collaboration humain-machine en toute confiance. La revue de littérature ci-dessous est présentée en sept parties : 1. L'histoire de l'intelligence artificielle; 2. Un survol sommaire de la technologie; 3. Les grands enjeux éthiques, 4. Son encadrement actuel; 5. L'intelligence artificielle affective; 6. Les conditions à mettre en place pour une plus grande acceptabilité; 7. Les initiatives proposées dans la littérature pour augmenter la confiance à l'égard de la technologie.

L'intelligence artificielle affective est un sujet qui soulève un grand engouement pour différentes disciplines. Nous avons donc concentré nos recherches principalement sur les thèmes de l'intelligence artificielle, de l'intelligence artificielle affective, des enjeux éthiques, ainsi que des conditions d'acceptabilité de la technologie.

1.2 Historique de l'intelligence artificielle

Pour saisir pleinement l'importance et l'urgence du moment actuel dans l'histoire du développement de l'intelligence artificielle, il est capital de se familiariser avec l'évolution de la recherche depuis ses débuts dans les années 1950. Avant toute chose, prenons le temps d'établir la définition de l'intelligence artificielle.

1.2.1 Définition

Il n'existe pas de définition de l'intelligence artificielle qui soit universellement adoptée par l'ensemble de la communauté scientifique. Mark Coeckelbergh définit l'IA comme une intelligence simulée par des codes ou des machines (2020 : 203, traduction libre). Charlet et Bringay la décrivent comme suit : « L'intelligence artificielle est une science dont le but est de faire faire par des machines des tâches que l'homme réalise grâce à son intelligence » (2019 : 26). Bien qu'il n'y ait pas une seule définition reconnue pour l'IA, celles que j'ai

rencontrées au fil de mes nombreuses lectures réfèrent généralement à une technologie capable de simuler des habiletés reconnues comme étant spécifiques à l'intelligence humaine. Dans le cadre de ce mémoire, nous nous en remettons à la définition de l'IA que nous retrouvons dans la *Déclaration de Montréal pour un développement responsable de l'IA* :

L'intelligence artificielle (IA) désigne l'ensemble des techniques qui permettent à une machine de simuler l'intelligence humaine, notamment pour apprendre, prédire, prendre des décisions et percevoir le monde environnant. Dans le cas d'un système informatique, l'intelligence artificielle est appliquée à des données numériques. (Université de Montréal, 2018)

1.2.2 Historique

Dès les débuts de l'intelligence artificielle, l'objectif des chercheurs dans ce domaine est de développer une machine capable de raisonner comme l'humain (Charlet et Bringay, 2019). Les scientifiques collaborant sur les projets d'IA proviennent de différentes disciplines dont, entre autres, l'informatique, les mathématiques, la statistique, la neuroscience, la psychologie, la logique et la philosophie (Coeckelbergh, 2020). Avant la Deuxième Guerre mondiale, Alan Turing, reconnu comme étant le père des sciences de l'informatique (Bowen, 2019; Gibert, 2020), a ébauché les théories de base d'une machine universelle qui modélise l'ordinateur (Bowen, 2019). Turing introduit le concept de « *Machine Learning* », aussi connu sous le nom d'« apprentissage automatique », dans le cadre d'une lecture à la London Mathematical Society le 20 février 1947 :

Let us suppose we have set up a machine with certain initial instruction tables, so constructed that these tables might on occasion, if good reason arose, modify those tables. [...] What we want is a machine that can learn from experience. (Damassino et Novelli, 2020 : 464)

C'est cependant en 1950, dans « *Computing Machinery and Intelligence* » (Turing, 1950), qu'il approfondit la possibilité qu'une machine puisse simuler des comportements humains et penser. Il fait référence aux machines intelligentes et introduit le test de Turing¹, dont

¹ Le test de Turing est aussi connu sous le nom d'« *Imitation Game* » ou de « jeu de l'imitation ». L'objectif est, pour un interrogateur, de trouver, parmi deux participants, lequel est en fait une machine. L'interrogateur se trouve dans une pièce séparée des deux participants. Le premier participant est un humain (A) et le

l'objectif est d'évaluer la performance d'une machine intelligente (Mintz et Brodie, 2019). Le terme *intelligence artificielle* est cependant utilisé pour la première fois à la conférence de Dartmouth à Hanover dans le New Hampshire en 1956 et est depuis associé à son organisateur John McCarthy (Bowen, 2019). À cette époque, deux groupes distincts travaillent sur les bases de l'intelligence artificielle selon deux angles différents : la cybernétique, qui a une approche dite connexionniste ou statistique, et l'intelligence artificielle, qui a une approche symbolique (Cardon, Cointet et Mazières, 2018). Les travaux et les recherches se sont poursuivis à travers les décennies de part et d'autre. L'approche symbolique, qui s'intéresse plus spécifiquement au raisonnement, a comme principaux fondateurs McCarthy, Minsky, Simon et Newell (Cardon, Cointet et Mazières, 2018). C'est entre 1960 et 1990 qu'elle connaîtra le plus de succès, d'intérêt, et recevra la quasi-totalité de son financement (Cardon, Cointet et Mazières, 2018). Les découvertes qui sont liées à cette approche ont apporté de belles avancées, comme les modèles experts que j'aborderai dans la prochaine section. L'approche connexionniste, dont le père est Norbert Wiener, a comme intérêt la biologie humaine et les réseaux de neurones (Cardon, Cointet et Mazières, 2018). Elle continuera durant des décennies son travail dans l'ombre. Une grande partie du travail sur les modèles de réseaux de neurones remonte aux années 1980 (Davenport *et al.*, 2019). Cependant, il faudra attendre tout près de 25 ans avant d'en voir les résultats. Les plus grandes percées dans ce domaine ont été largement dépendantes de l'accroissement de la puissance des ordinateurs, mais aussi de leurs structures et de leurs environnements logiciels (Haeich, 2020). C'est pourquoi cette méthode vient à peine de démontrer son plein potentiel, grâce à la puissance des ordinateurs actuels et des données de masse maintenant disponibles (Tual, 2015). Au début des années 2000, Geoffrey Hinton, Yann LeCun et Yoshua Bengio ont relancé le travail déjà amorcé à la fin des années 80 pour renouveler l'intérêt de la communauté pour les réseaux neuronaux. Le trio, qui se

deuxième est une machine (B). Pour l'interrogateur, les participants sont connus comme étant X et Y. En posant des questions à chacun des participants, l'interrogateur doit déterminer si X est A et Y est B, ou si X est B et Y est A. L'objectif est de valider à quel point la machine est capable de tromper un interrogateur et de passer pour un humain (Oppy, 2020). Plus la machine présente des composantes d'intelligence artificielle poussées, plus elle réussit à induire en erreur l'interrogateur.

surnommait « la conspiration du *deep learning* », a commencé à établir des records vers 2010-2012, et depuis, l'intérêt pour les possibilités entourant l'IA a véritablement explosé (Tual, 2015). Leurs avancées leur ont d'ailleurs valu le prix Turing en 2019 (Besse, Besse-Patin et Castets-Renard 2020). Cette découverte a ouvert un monde de possibilités, et les développements dans le domaine se sont alors mis à évoluer à la vitesse grand V. De 1950 jusqu'au début des années 2000, les avancées dans le domaine furent importantes, mais sans grands bouleversements. Les percées en intelligence artificielle, et tout particulièrement en apprentissage automatique, couplées aux avancées informatiques et à la quantité de données disponibles, changent les règles du jeu. En dix ans, le volume de données numériques à l'échelle mondiale est passé de 2 zettaoctets à 64 zettaoctets, se multipliant par 30 (Gaudiault, 2021). La croissance annuelle moyenne devrait se situer autour de 40 % par année d'ici 2025 (Gaudiault, 2021). Mo Gawdat, ancien dirigeant chez Google X, nous met en garde :

If you were to assume that it took us seventy-five years to get here, you might predict that it could take tens of more years to have any meaningful implication for artificial intelligence coming into our life. The rate of progress, as with every technology, was very slow at first. But it is now moving at an exponential speed. The next ten years of artificial intelligence development are expected to deliver a very unfamiliar future. (Gawdat, 2021 : 43)

Nous avons encore aujourd'hui le contrôle sur la technologie, mais certains, comme Bannister et Connolly (2020), croient qu'il pourrait venir un moment où le contrôle pourrait nous échapper. Bien que les débuts de l'intelligence artificielle aient été relativement timides, son développement exponentiel dans les dernières années nous démontre très certainement l'importance de l'étudier sous tous ses angles.

1.3 Survol sommaire de la technologie

Pour bien comprendre les enjeux entourant la technologie, il est important de s'intéresser à son essence même. À la base, l'intelligence artificielle est composée d'un algorithme ou de combinaisons algorithmiques. Principalement, un algorithme est une séquence d'instructions qui dicte à l'ordinateur, au téléphone intelligent, à une machine ou encore à un robot quoi faire (Cardon, Cointet et Mazières, 2018). L'algorithme résout des problèmes, optimise des processus, opère certaines tâches et fait des prédictions (Bannister

et Connolly, 2020). Pour y arriver, il doit traiter une grande quantité de données. L'intelligence artificielle se nourrit de données (Davenport *et al.*, 2019). Le concepteur doit programmer l'algorithme via un code pour lui dicter la façon d'analyser ces données. Il existe différentes méthodes de programmation qui offrent divers avantages et possibilités. Dans la section précédente, j'ai fait référence à l'approche symbolique et à l'approche connexionniste. Comme l'approche symbolique s'est principalement intéressée au raisonnement, elle présente entre autres des modèles à base d'arbres décisionnels ou de chaînes de raisonnements intermédiaires (Cardon, Cointet et Mazières, 2018). Cette méthode implique que le concepteur dicte au système, via la programmation, un cadre et des paramètres bien précis à l'intérieur desquels évoluer. La rationalité du calcul est codée à l'intérieur du programme (Cardon, Cointet et Mazières, 2018). Le système expert est issu de l'approche symbolique (Cardon, Cointet et Mazières, 2018). Le concepteur codifie les connaissances rationnelles de spécialistes dans un domaine, ce qui permet au programme de résoudre des problèmes spécifiques au champ d'expertise (Besse, Besse-Patin et Castets-Renard, 2020). En 1984, Buchanan et Shortliffe ont développé un système expert pour la sélection d'un antibiotique selon les critères biologiques entourant le diagnostic d'une infection bactérienne (Besse, Besse-Patin et Castets-Renard, 2020). Ce type de programme a l'avantage d'être plus transparent : il permet de retracer le raisonnement derrière le résultat.

Pour sa part, l'approche connexionniste, aussi appelée « approche statistique », s'est inspirée de la neuroscience et des réseaux de neurones. Ainsi, l'apprentissage automatique, aussi appelé « *Machine Learning* », est formé d'un réseau de neurones (des unités simples) qui communiquent à la façon d'un cerveau humain. Lorsque le réseau comporte plusieurs couches de neurones, nous faisons référence à un système d'apprentissage profond, aussi connu sous le nom de « *deep learning* » (Coeckelbergh, 2020). L'apprentissage automatique et l'apprentissage profond sont associés au concept de « boîte noire » puisque contrairement à la méthode d'arbre décisionnel, qui peut être décodé par l'humain considérant son architecture qui le rend transparent, les réseaux de neurones sont beaucoup plus opaques (Coeckelbergh, 2020). Bien que les concepteurs aient programmé l'architecture d'un réseau, il est difficile d'expliquer comment les connexions se font à

l'intérieur entre les unités de neurones de façon à émettre une prédiction ou une décision finale (Coeckelbergh, 2020).

L'apprentissage automatique est basé sur la statistique : « Étant conçus pour déceler des régularités ou des modèles (*patterns*) dans les jeux de données, les algorithmes d'apprentissage automatique arrivent à repérer les corrélations » (Maclure et Saint-Pierre, 2018 : 748). Grâce à ces corrélations, il leur est possible par la suite d'émettre des prédictions (Coeckelbergh, 2020). La tâche à accomplir est dictée par le concepteur, mais la façon dont l'apprentissage automatique arrivera à sa réponse revient au système ou au programme. Le concepteur utilisera une des trois méthodes pour l'entraîner, c'est-à-dire pour enseigner au programme comment naviguer dans les données disponibles, soit la méthode supervisée, la méthode non supervisée ou le renforcement (Besse, Besse-Patin et Castets-Renard, 2020). Dans la méthode supervisée, le concepteur lui donnera des exemples et des non-exemples pour le guider dans son processus d'apprentissage (Cardon, Cointet et Mazières, 2018). Si la méthode est non supervisée, l'algorithme fera lui-même ses recherches à travers les données massives qui l'alimentent. Il pourra en extraire ses propres apprentissages et découvrir des corrélations mathématiques entre les données (Davenport *et al.*, 2019). Cette méthode est utilisée pour explorer des domaines et des phénomènes (Mittelstadt *et al.*, 2016). Enfin, le concepteur peut favoriser l'apprentissage par renforcement, procédant par récompenses (Gibert, 2020). Le concepteur ne dicte alors pas à l'algorithme la voie à suivre, mais le système reçoit de la rétroaction, positive ou négative selon le cas, pour le guider dans son parcours (Davenport *et al.*, 2019).

L'intelligence artificielle est sensible à la quantité et à la qualité des données disponibles à traiter. C'est particulièrement important durant la phase d'apprentissage et c'est un élément clé pour le contrôle de la qualité une fois le système ou le programme implanté (Besse, Besse-Patin et Castets-Renard, 2020). J'aborderai à nouveau la qualité des données dans la section sur les enjeux éthiques liés à l'intelligence artificielle. Ce qu'il faut retenir concernant la technologie, c'est que l'intelligence artificielle se présente sous différentes formes et que toutes ces formes ne soulèvent pas les mêmes enjeux.

1.3.1 Applications de l'intelligence artificielle

L'intelligence artificielle est présente dans différentes sphères, et l'intérêt autour de celle-ci s'accroît. Dans les entreprises, on la retrouve surtout au sein de grandes organisations qui ont accès à un grand volume de données, telles que les plateformes en ligne, les services financiers, les télécommunications et les ventes au détail (Davenport *et al.*, 2019). La technologie permet d'améliorer des fonctionnalités ainsi que la performance de certains produits, aide à prendre de meilleures décisions, à créer de nouveaux produits, à optimiser les processus opérationnels internes ou externes ou encore à automatiser certaines tâches (Davenport *et al.*, 2019). Elle facilite l'accès à de nouveaux marchés, permet d'optimiser les coûts et d'offrir un service à la clientèle personnalisé. La personnalisation du contenu est un domaine déjà largement répandu au sein de différentes plateformes et profite tout particulièrement au milieu du marketing (Davenport *et al.*, 2019). La technologie offre de nombreux avantages, et beaucoup d'entreprises ont compris son importance. Plus du tiers des grandes organisations américaines disent avoir développé une stratégie basée sur l'intelligence artificielle, avoir mis en place un centre d'excellence au sein de leur entreprise et ciblé des champions dans leur équipe de management (Davenport *et al.*, 2019).

Dans le milieu de la santé, l'IA permet d'optimiser les plans de traitements et d'offrir un soutien puissant à la préparation et à la réponse à une pandémie (OMS, 2021). Le développement de la technologie apporte de l'assistance robotique en chirurgie, accélère la découverte de nouveaux médicaments ou encore favorise la réduction des coûts liés à un parcours de soins (Hague, 2019). L'IA a fait de belles percées dans différents domaines, dont la radiologie, où elle soutient grandement le travail des spécialistes en augmentant l'efficacité du département et la précision des diagnostics, ce qui permet aux radiologistes d'analyser plus rapidement les différents dossiers et de passer plus de temps sur les cas plus pointus où le degré de confiance de l'assistant intelligent est moins concluant (Hague, 2019). La technologie portable permet de personnaliser le traitement selon les symptômes que présentent les patients (Davenport *et al.*, 2019).

Le milieu financier bénéficie aussi des avantages de la technologie, notamment sa capacité rapide et précise de détection de la fraude de crédit en temps réel (Davenport *et al.*, 2019). Les forces policières et le domaine de la justice profitent de l'évaluation du risque de

récidive d'un accusé ou d'un détenu pour établir le risque que représente un individu, et de l'optimisation de la gestion des forces policières en ciblant les zones les plus à risque au plan criminel (Besse, Besse-Patin et Castets-Renard, 2020). Les bénéfices de l'IA dans diverses industries sont importants et prometteurs, mais les enjeux qui entourent son développement et son déploiement ne sont pas négligeables et peuvent ralentir son adoption dans différentes sphères.

1.4 Enjeux de l'intelligence artificielle

Bien que l'intelligence artificielle représente de belles possibilités, ses répercussions encore méconnues peuvent ralentir, voire freiner son adoption. La technologie soulève certaines préoccupations, comme la protection de la vie privée, la protection des données et la responsabilité. Dans un article de 2007, Anderson et Anderson mentionnent qu'il est possible que la crainte des gens par rapport aux machines intelligentes vienne de leurs soucis quant à la façon dont elles vont se comporter sur le plan éthique, et mettent l'accent sur l'importance d'étudier le phénomène. Depuis, les recherches sur les enjeux éthiques entourant l'intelligence artificielle se sont multipliées. Dans cette section, nous aborderons plus en détail les enjeux éthiques suivants :

- La transformation du marché du travail
- La fiabilité
- La sécurité et la malveillance
- Les biais et la discrimination
- La responsabilité
- La transparence et l'explicabilité
- La vie privée, l'autonomie et le consentement

1.4.1 Transformation du marché du travail

La perte d'emploi est une préoccupation importante. Les capacités grandissantes de la technologie laissent croire qu'elle pourra effectuer certaines tâches autrefois réservées aux humains (Lee, 2019). Le Forum économique mondial a estimé que les technologies de l'IA pourraient toucher jusqu'à 20 % des employés au Royaume-Uni et 26 % dans des pays

comme la Chine et l'Inde (Dwivedi *et al.*, 2021). D'ailleurs, Yoshua Bengio estime que les avancées en intelligence artificielle occasionneront des changements importants en matière d'emploi : « Il y a de bonnes chances que ces chamboulements créent de la souffrance chez beaucoup de gens, qui vont perdre non seulement leur gagne-pain, mais le sens qu'ils donnent à leur vie » (Mathys, 2017). On prévoit en contrepartie la création de nouveaux postes en lien avec les développements de la technologie (Dwivedi *et al.*, 2021). Cependant, aux yeux de Kai-Fu Lee (2019), elle ne compensera pas la suppression des postes due à l'arrivée de l'intelligence artificielle. Il faut donc se préparer à des changements sur le marché de l'emploi dans les années à venir.

La façon dont les différentes industries et les divers rôles seront touchés varie. Le domaine de la santé, par exemple, a ses particularités : ses tâches sont complexes, pleines de subtilités, et son processus requiert des connaissances et des interventions à chaque étape des soins (Ahmad, Eckert et Teredesai, 2018). L'intelligence artificielle n'a pas toutes les capacités nécessaires pour remplacer l'humain, elle peut le faire seulement pour certaines tâches. De façon concrète, un algorithme peut passer en revue l'ensemble des articles du *New York Times* et arriver à prédire lesquels seront les plus partagés sur les plateformes de médias sociaux, mais pour le moment, il lui est impossible d'expliquer pourquoi (Luca, Kleinberg et Mullainathan, 2016). Cette subtilité est importante et non négligeable pour nombre de tâches qui restent, pour le moment, réservées aux humains. Les avancées en intelligence artificielle et en robotique sont impressionnantes, mais la combinaison du potentiel humain à celui des nouvelles technologies est plus prometteuse encore, et cette collaboration mènera bien plus loin en matière de créativité et de productivité (Davenport *et al.*, 2019). Les études dans le domaine ont soulevé des limites dans la quête continue pour l'automatisation. Elles reconnaissent l'importance de maintenir l'humain dans le processus et de miser plutôt sur l'augmentation de ses capacités et de ses performances, et non de le remplacer (Dwivedi *et al.*, 2021).

1.4.2 Fiabilité

Un certain niveau de méfiance par rapport à l'intelligence artificielle est aussi perçu en lien avec les erreurs qui font surface dans son historique de déploiement (Hague, 2019). La technologie est faillible. Selon ses applications, une erreur algorithmique peut avoir très

peu de conséquences, comme le correcteur automatique sur les tablettes et les téléphones intelligents, qui se méprend souvent sur l'intention de l'utilisateur. Hélas, les erreurs peuvent aussi mener à des drames. En effet, un robot sur une chaîne de montage de voiture s'est trompé et a agrippé un humain par erreur au lieu d'une pièce, et cet incident a été fatal pour le travailleur (Yampolskiy, 2017). Une voiture Tesla, en mode autonome, a entraîné dans la mort son occupant en 2016 (Lee, 2019). Ces incidents peuvent bien entendu apporter leur lot de méfiance par rapport à la technologie.

Par ailleurs, une application d'analyse d'images médicales a montré ses limites lorsqu'elle n'a pas soulevé l'anomalie des images chez un patient atteint d'une très rare forme de cancer du poumon, particulièrement agressive (Oakden-Rayner *et al.*, 2020). Un algorithme d'apprentissage automatique ne peut cibler que ce qu'il connaît et a déjà rencontré (Besse, Besse-Patin et Castets-Renard, 2020), d'où l'importance des données de formation. Ici, la source de l'erreur a été le phénomène de rareté. Puisque l'algorithme n'a pas été exposé à cette forme de présentation de la pathologie, il n'a pas su la reconnaître.

La méfiance par rapport à l'intelligence artificielle peut être fondée, considérant les conséquences qui peuvent en découler. Cependant, la performance toujours grandissante des algorithmes pour certaines applications égalise, voire surpasse les performances humaines, ce qui laisse présager que dans certaines situations, l'intelligence artificielle sera plus sécuritaire que l'humain (Lee, 2019). En 2016, 1,25 million de personnes ont perdu la vie dans des accidents de la route dans le monde (Lee, 2019). Au Québec, le nombre de décès sur les routes a d'ailleurs augmenté en 2021, passant de 250 en 2020 à 262 en 2021 (Sûreté du Québec, 2021). Une grande partie de ces accidents auraient pu être évités : 24,5 % des collisions ont été causées par une conduite imprudente et des excès de vitesse, 14 % par une capacité affaiblie par l'alcool, les drogues ou la fatigue et 7 % par une inattention ou une distraction (Sûreté du Québec, 2021). Selon Kai-Fu Lee, « [à] terme, les véhicules autonomes seront beaucoup plus sûrs que ceux conduits par des humains. Leur déploiement à grande échelle va entraîner un recul spectaculaire du nombre de victimes » (Lee, 2019 : 202).

Ainsi, avec les avancées de la technologie, il est permis d’imaginer un moment où, pour certaines applications encore en développement aujourd’hui, l’intelligence artificielle sera une option plus rassurante que sa contrepartie humaine.

1.4.3 Sécurité et malveillance

Enfin, une préoccupation grandissante qui est soulevée par rapport à l’intelligence artificielle est le risque de malveillance. Il s’agit ici de la sécurité du système par rapport au piratage ou encore à la possibilité de mauvaises intentions de la part du concepteur. Ainsi, dans un monde où la conduite autonome existe ou un drone militaire peut être contrôlé à distance, les intentions malveillantes dans le milieu informatique deviennent des craintes légitimes (Coeckelbergh, 2020). Les technologies utilisant l’intelligence artificielle sont de plus en plus intrusives, comme Alexa et Siri, des assistants virtuels (Gibert, 2020) qui s’immiscent dans la sphère privée, Deepmood, qui se spécialise dans la détection des émotions (Blasimme et Vayena, 2020), ou encore les technologies de surveillance utilisées par les corps policiers (Deloitte, 2019) qui épient les faits et gestes de la communauté. Les risques de piratage, d’usage malveillant des données ou encore d’usurpation d’identité, comme dans les incidents récents dans le milieu financier avec le vol de données chez Desjardins (Boily, 2022) ou encore les faux pas dans la gestion de la sécurité des données de la RAMQ au Québec (Bélair-Cirino, 2020), peuvent alimenter la méfiance des utilisateurs. En effet, la population tend à développer une plus grande sensibilité lorsqu’elle est exposée à des brèches de confidentialité par rapport aux données (Ghafur *et al.*, 2020).

1.4.4 Biais et discrimination

Un enjeu important qui est souligné régulièrement dans la littérature est celui des biais qui peuvent se retrouver au cœur même d’un système apprenant. Le biais a une connotation mathématique, mais dans le cadre social, il fait plus souvent référence à l’équité (Hague, 2019). Les biais portent en général atteinte au traitement juste et équitable de la population la plus vulnérable, entre autres les minorités, les femmes et les personnes à faible revenu (O’Neil, 2016). Le biais, le préjudice et la discrimination ne sont pas nouveaux. Les médias regorgent d’histoires qui relatent le traitement biaisé de certaines couches de la société. Dernièrement, le mouvement « Black Lives Matter », qui dénonce les injustices envers la population noire, s’est enflammé à la suite de la mort de George Floyd, asphyxié par un

policier (Elkouri, 2020). Nous pouvons aussi penser au décès regrettable de Joyce Echaquan, une femme autochtone qui a fait face à des préjugés et n'a pas reçu les soins nécessaires (Josselin, 2021). La coroner Géhane Kamel a conclu que le « racisme et les préjugés auxquels Mme Echaquan a fait face ont certainement été contributifs à son décès » (Josselin, 2021). Les femmes ne font pas exception non plus, l'écart salarial au Québec s'amenuisant mais demeurant présent :

Rapporté sur une année, cet écart salarial entre les deux sexes signifie qu'à partir du 1^{er} décembre, à 10 h 26, et jusqu'à la fin de l'année, les Québécoises « travaillent gratuitement » comparativement à leurs homologues masculins, indique l'IRIS (Pavic, 2021).

Mais qu'en est-il lorsqu'un algorithme se glisse dans le processus décisionnel? La discrimination peut se présenter sous la forme d'un refus de crédit, un individu peut se voir écarté dans un processus d'entrevue ou encore relayé à la fin de la liste de priorité au triage pour une procédure dans le milieu de la santé. Les biais peuvent s'infiltrer à toutes les étapes du développement du système, que ce soit durant la conception de l'algorithme, durant sa formation via les données de masse, ou encore en cours d'utilisation (Coeckelbergh, 2020).

La conception est un moment important pour développer un outil qui présentera le moins de biais possible. Certains aspects doivent être pris en considération pour en limiter la présence. L'humain a des biais et fait des erreurs (Davenport *et al.*, 2019). Chacun d'entre nous a ses propres biais, dont nous ne sommes souvent pas conscients. Ils sont influencés par la culture dans laquelle nous grandissons, notre famille, nos expériences, et bien plus. Ainsi, le concepteur lui-même a son lot de biais, qui est possiblement un angle mort qu'il doit veiller à éviter de coder (Maclure et Saint-Pierre, 2018). De plus, les équipes de conception sont souvent homogènes (Coeckelbergh, 2020) (Caliskan, Bryson et Narayanan, 2017). Une grande proportion des scientifiques en science de la donnée et des ingénieurs sur les projets de développement sont des hommes blancs des pays de l'Ouest, âgés de 20 à 40 ans (Coeckelbergh, 2020). Aux yeux de Martin Gibert (2020), la situation est préoccupante. Les concepteurs programmeront fort probablement des systèmes alignés avec leurs intuitions morales et, conséquemment, l'intelligence artificielle sera le produit, du moins en partie, de leurs valeurs (Gibert, 2020). Selon Maclure et Saint-Pierre, « [i]l y a des biais de programmation lorsque ces valeurs reproduites par l'algorithme incluent des

croyances produisant de l'exclusion, de la stigmatisation ou de la discrimination » (2018 : 755). Le ministère allemand des Transports a publié en 2017 un rapport sur le cadre éthique pour la gestion de la conduite autonome. Pour éviter toutes formes de discrimination, le comité a statué qu'« en cas d'accident inévitable, toute forme de distinction fondée sur des caractéristiques personnelles est strictement prohibée » (Gibert, 2020 : 37). Le ministère explicite les valeurs allemandes et évite de remettre entre les mains d'une équipe de concepteurs les choix moraux de la société.

Une autre source de biais, plus complexe à repérer, est dissimulée au sein de la base de données, dans la qualité et la représentativité des données (Gibert, 2020). Il est important d'utiliser des données représentatives, c'est-à-dire qui tracent un portrait juste de l'ensemble des utilisateurs potentiels du système en question (Maclure et Saint-Pierre, 2018) en matière de morphologie dans le bassin américain (Coeckelbergh, 2020). Il en va de même pour les bases génomiques, dont les données étaient à 96 % issues de la population d'ascendance européenne blanche en 2009 et 81 % en 2016 (Besse, Besse-Patin et Castets-Renard, 2020). Il y a également un risque réel quand l'IA se base sur des données historiques dont elle va absorber les biais inhérents (Hague, 2019). Le langage utilisé peut être lui aussi biaisé par des préjugés culturels et sociaux (Caliskan, Bryson et Narayanan, 2017). Dans le cadre d'une étude utilisant le monde du Web comme base de données, il a été possible de mettre en lumière ce qui émane de cette très vaste base de données mondiale. L'étude a relevé des corrélations statistiques, entre autres que le genre féminin est associé à l'art, tandis que le genre masculin est associé au monde des sciences et des mathématiques (Caliskan, Bryson et Narayanan, 2017), ou encore que le catholicisme est associé à l'ignorance et au jugement des autres, et que la religion musulmane est associée au terrorisme et au jeûne (Baker-Brunnbauer, 2021). Ces corrélations issues des calculs mathématiques de l'algorithme ne sont pas porteuses de vérités, mais l'intelligence artificielle ne fait pas la différence (O'Neil, 2016).

Les programmes d'apprentissage automatique peuvent aussi développer des biais une fois implantés. Comme ils sont capables « d'apprendre », c'est-à-dire d'établir des corrélations à partir des données massives qu'ils traitent, il est possible que se forment en eux des biais que le concepteur n'avait pas codés au préalable. On court le risque, comme le dit Martin

Gibert (2020), de reproduire, sans même s'en rendre compte, des biais systémiques. Le système peut par lui-même développer des biais en découvrant certaines corrélations et causer des préjudices en appliquant des discriminations sur la base de ces liens (O'Neil, 2016). Tay, l'agent intelligent (*chatbot*) de Microsoft, a été actif durant 16 heures en 2016 avant d'être retiré. Tay a commencé à s'exprimer de façon raciste, misogyne et antisémite à force d'échanger avec les internautes sur Twitter. La technologie d'apprentissage automatique, exposée aux nouvelles données qui l'alimentaient au fil des échanges, s'est adaptée et a dévié de sa formation initiale (Maclure et Saint-Pierre, 2018). Xerox, qui utilisait un outil d'intelligence artificielle pour optimiser son processus de recrutement de personnel, s'est rendu compte que le système écartait les candidats sur une base géographique, c'est-à-dire par rapport à la distance et aux régions. Or, plusieurs de ces candidats ayant à voyager sur une longue distance étaient issus de populations moins bien nanties. L'organisation a apporté un changement à l'algorithme pour ne plus qu'il discrimine les candidats sur la base de leur région (O'Neil, 2016). De la même façon, dans le cadre d'une étude, un groupe de chercheurs a soumis des candidatures fictives utilisant exactement le même curriculum vitae, mais en changeant le nom des candidats. Les résultats ont démontré que les outils d'aide à la décision optimisés par l'IA dans cette étude ont discriminé les candidats sur la base de la race associée à leur nom, favorisant nettement un groupe au détriment de l'autre (Caliskan, Bryson et Narayanan, 2017). Le danger des biais se situe dans l'ampleur des préjudices qu'ils peuvent causer et le risque de creuser d'encore plus grandes inégalités (O'Neil, 2016).

Ainsi, les algorithmes peuvent être biaisés dès leur conception par leur développeur de façon non intentionnelle, ou encore, de façon malveillante, les biais peuvent être codés de façon intentionnelle dans l'algorithme (Cuttillo *et al.*, 2020) (Cardon, Cointet et Mazières, 2018). Le système peut aussi acquérir certains biais durant la phase de formation via la représentativité des données, ou encore, une fois implanté, en établissant des corrélations durant la navigation dans la base de données. Un suivi régulier des systèmes d'intelligence artificielle doit être mis en place pour éviter la présence de biais et surtout leurs conséquences (Maclure et Saint-Pierre, 2018). L'audit est une méthode privilégiée pour détecter les anomalies possibles (Maclure et Saint-Pierre, 2018). Il faut cependant faire attention de ne pas tomber dans le piège de la quête d'une intelligence artificielle sans biais,

considérant que les réels enjeux se trament au niveau des institutions qui ont elles-mêmes des biais en leur sein (Surden, 2020).

D'ailleurs, cette capacité de l'IA à faire émerger certaines corrélations peut aussi être à notre service en mettant en évidence des biais non suspectés. Elle a permis de démontrer le biais de certains juges dans leurs décisions ou de policiers dans leurs interventions (Surden, 2020). Elle peut aussi mettre en lumière des mots ou une formulation dans un texte de loi qui causent un préjudice de façon insoupçonnée, ou un processus établi qui discrimine des gens de façon insidieuse (Surden, 2020). L'intelligence artificielle met en évidence l'état actuel de la société et de ses systèmes de valeurs, un phénomène en constante évolution. Certains suggèrent d'automatiser les valeurs humaines dans l'algorithme « Ethics by Design » pour diminuer les risques de biais et de discrimination (Dignum, 2020). Cependant, les valeurs ne sont pas communes à tous les humains, elles varient et ne sont pas homogènes (Gibert, 2020). Certains sujets, comme l'avortement, les mères porteuses ou encore la criminalisation de la drogue, démontrent bien la difficulté d'avoir une seule vérité morale (Gibert, 2020). Il est ainsi difficile d'automatiser un système de valeurs universellement accepté.

Fait particulièrement intéressant, dans son article publié en 2015, Walter Frick nous explique que l'IA peut elle-même être discriminée si nous l'humanisons, entre autres si on lui donne une voix de femme ou une voix d'homme selon la tâche à laquelle elle se consacre. Les chercheurs de l'Université Soongsil en Corée du Sud ont fait une expérience avec un robot de surveillance dédié à la sécurité. Lorsque le robot était nommé John et qu'il avait une voix d'homme, il a été jugé comme étant plus utile que lorsqu'il était nommé Joan et avait une voix de femme. Bien que le robot était exactement le même et ne faisait que changer de nom et de voix, il a été jugé plus durement par rapport à ses compétences lorsqu'il prenait une voix et un nom féminins (Frick, 2015).

1.4.5 Responsabilité

La responsabilité en intelligence artificielle fait référence à l'aspect légal, mais aussi à la responsabilité morale dans le développement de nouvelles technologies qui auront une influence sur nos sociétés (Dignum, 2020). Une approche responsable de l'IA est

nécessaire pour en assurer une utilisation sécuritaire, équitable et bénéfique (Dignum, 2017).

Dans le cas de la responsabilité légale, il s'agit d'établir comment déterminer la responsabilité en cas d'incidents, c'est-à-dire « l'imputabilité pour les décisions prises par des agents artificiels » (Maclure et Saint-Pierre, 2018 : 753). Dans le cas de la voiture Tesla qui a eu un accident mortel en situation de conduite autonome, ou dans la situation du bras robotisé qui a saisi un humain par erreur, ou encore celle de l'outil d'analyse d'imagerie médicale qui n'a pas su reconnaître le cancer agressif d'un patient, à qui la faute?

Lorsque nous parlons de la responsabilité sous la forme de l'imputabilité en intelligence artificielle, un aspect important est celui de l'implication de l'humain dans le processus décisionnel. Si l'algorithme est utilisé de façon complémentaire et que l'humain est le décideur, l'humain est encore entièrement maître de la décision (Besse, Besse-Patin et Castets-Renard, 2020). L'humain exerce son jugement en mettant en perspective les différents aspects à considérer dans la balance décisionnelle, dont les données fournies par l'algorithme.

Il faut cependant faire attention aux situations où les décisions sont plus critiques et où l'humain préférerait s'en remettre à l'algorithme. Un bon exemple est le cas de COMPAS, un algorithme qui évalue les risques de récidive d'un détenu et qui agit à titre d'outil d'aide à la décision pour les juges quand vient le temps d'accorder ou non une libération (O'Neil, 2016). Les juges peuvent être tentés de s'en remettre à la prédiction de l'algorithme. S'ils autorisent la libération d'un détenu malgré le risque élevé prédit par l'algorithme de COMPAS, les conséquences pourraient être fâcheuses autant du point de vue social qu'en ce qui concerne la responsabilité possible du juge. Si, malgré un risque élevé, un juge remet en liberté un détenu et que ce dernier récidive, ce juge pourrait-il être tenu responsable de ne pas avoir suivi la recommandation? Devant cette éventualité, il pourrait donc être tenté de se protéger et de prendre des décisions largement influencées par les prédictions du système d'intelligence artificielle (O'Neil, 2016). La décision finale reste entre les mains de l'humain, mais l'influence décisionnelle est grande. Bien que la responsabilité revienne au juge, cette pression soulève un enjeu que nous aborderons plus loin, celui de l'autonomie.

Dans le domaine de la santé, Besse, Besse-Patin et Castets-Renard positionnent ainsi la responsabilité médicale :

Dès lors que le médecin reste au centre de la relation de confiance avec le patient, qu'il continue d'assumer une obligation d'information et qu'il reste maître des choix et décisions prises, la machine doit être considérée comme une simple aide à la décision qui ne remplace pas le médecin et ne modifie en rien les règles de responsabilité (2020 : 31).

Tant et aussi longtemps que le médecin reste le décideur dans le continuum, la responsabilité demeure entre ses mains. Aux yeux de McGregor (2018), lorsque l'algorithme prend la décision lui-même, mais que l'humain reste présent dans le processus pour ce qui est du suivi et de l'audit, le jugement humain peut être exercé. Tout comme précédemment, la responsabilité reste entre les mains de l'humain.

L'intelligence artificielle pourrait-elle être tenue criminellement responsable? Sur le plan criminel, il faut établir l'intention pour qu'un accusé soit tenu moralement responsable (Coeckelbergh, 2020). L'algorithme n'est pas reconnu comme étant une entité morale au sens fort du terme, ce qui le rend difficilement imputable de ses actes (Gibert, 2020), alors comment lui reconnaître une intention? Sur le plan civil, il faut démontrer la responsabilité de la personne chargée du système au moment de l'incident (Coeckelbergh, 2020). C'est un processus complexe dans le cas des algorithmes. Les algorithmes sont conçus, achetés, transformés par différentes firmes, et plusieurs intervenants participent à leur conception (Coeckelbergh, 2020). Il peut s'avérer difficile de cerner la personne responsable. Dans le cas, entre autres, d'un apprentissage automatique, un concepteur peut faire la démonstration de l'architecture d'un système, mais ne pourrait pas détailler comment l'algorithme en est arrivé à son résultat (Coeckelbergh, 2020). L'imputabilité d'une responsabilité légale à une organisation fautive est la démarche la plus probable et applicable dans le cas de l'intelligence artificielle (Coeckelbergh, 2020).

Pour assurer une traçabilité de la responsabilité, une personne devrait être nommée responsable à chaque étape du processus décisionnel et un cadre de sanction approprié devrait être mis en place pour assurer le respect de cette étape (Bannister et Connolly, 2020). Les algorithmes doivent être encadrés, et les individus et les organisations doivent être tenus responsables (Bannister et Connolly, 2020). La réflexion doit se poursuivre en

ce qui concerne le niveau de responsabilité décisionnelle que nous souhaitons reléguer aux machines apprenantes. Certains remettent en question l'utilisation d'IA opaque dans le milieu de la santé (Besse, Besse-Patin et Castets-Renard, 2020), misant sur une plus grande explicabilité et une plus grande transparence des systèmes, vu les conséquences graves que peuvent avoir les erreurs dans le milieu, l'objectif étant de pouvoir déterminer la responsabilité et la cause pour pouvoir les corriger (McGregor, 2018). Selon Floridi *et al.* (2018), la responsabilité est essentielle et devrait être un élément central dans la décision quant au type d'intelligence artificielle que nous développons. Il faut se méfier de l'idée de reléguer d'importantes tâches et décisions à des systèmes autonomes, et maintenir au minimum une supervision et l'autorité du choix de l'humain dans le continuum. Selon Maclure et Saint-Pierre, « [d]ans la majorité des cas, pour l'instant, il est envisagé d'utiliser des systèmes d'IA qui agiront en complémentarité avec l'humain, c'est-à-dire en faisant des recommandations qui devraient nécessairement être validées par des humains. » (2018 : 753) Il s'agit de trouver un équilibre entre les ambitions, les découvertes et les avancées en intelligence artificielle tout en gardant le contrôle sur son développement et ses effets (Floridi *et al.*, 2018).

La responsabilité morale réfère aux mesures qui doivent être mises en place pour assurer une intelligence artificielle qui prend en considération les valeurs, la diversité, l'inclusion et l'influence de la technologie sur notre société (Dignum, 2020). Cela implique de mettre en lumière la responsabilité de l'industrie et des concepteurs dans le développement d'une intelligence artificielle alignée avec les valeurs humaines. Certaines initiatives, comme la *Déclaration de Montréal*, la *Déclaration d'Asilomar* ou la *Déclaration de Barcelone*, ont proposé des lignes directrices sur les valeurs et les principes auxquels l'intelligence artificielle devrait adhérer (Dignum, 2020). La *Déclaration d'Asilomar* précise que « [l]es systèmes d'IA hautement autonomes devraient être conçus de manière à ce que leurs objectifs et leurs comportements s'alignent sur les valeurs humaines tout au long de leur fonctionnement. » (Gibert, 2020 : 52) Mais comme le soulève Martin Gibert, cet objectif louable et incontournable rencontre un défi de taille, « celui de se mettre d'accord sur les valeurs à implanter » (Gibert, 2020 : 53). La recherche et le développement mènent à d'incroyables avancées, mais sans cadre moral, ils peuvent mener à de graves conséquences. Ainsi, une plus grande sensibilisation par rapport à l'éthique de

l'intelligence artificielle est mise de l'avant par différents groupes. L'IEEE (Institute of Electrical and Electronics Engineers) a mis en place un programme de certification sur l'éthique pour les systèmes autonomes et les systèmes intelligents (Dignum, 2020). Ce programme vise à donner de la formation sur les méthodes et les processus permettant de créer une intelligence artificielle plus transparente, fiable et comportant moins de biais algorithmiques (Dignum, 2020). Les chercheurs et les concepteurs devraient être formés à propos des répercussions sociales potentielles de leur travail (Dignum, 2017) et ainsi avoir une plus grande sensibilité quant à leur responsabilité. Facebook, Apple et Google sont pointés du doigt pour leurs méthodes douteuses visant à obtenir le consentement des utilisateurs quant au partage de leurs données et également pour leurs fonctionnalités qui « cultivent l'usage compulsif » de leur plateforme (Braun, 2019). De tels comportements minent l'autonomie et le consentement éclairé. Des cadres réglementaires se mettent en place depuis quelques années pour éviter de laisser libre cours à l'autorégulation, qui est souvent en conflit avec la mission de rendement et de performance de l'industrie privée par rapport à ses investisseurs (Mittelstadt, 2019).

1.4.6 Transparence et explicabilité

La pression est forte pour une plus grande transparence et une plus grande explicabilité, pour aider à bâtir la confiance à l'égard de l'intelligence artificielle, pour expliquer la validité de ses prédictions et pour minimiser les erreurs et les préjudices qu'elle peut causer (Cutillo *et al.*, 2020; Coeckelbergh, 2020). Aux yeux de London (2019), lorsqu'un outil est transparent, l'utilisateur a une meilleure compréhension par rapport à ses choix, ce qui lui donne une plus grande autonomie, assure une plus grande objectivité et évite les abus d'autorité possibles. L'explicabilité est une clé notamment dans le milieu de la santé lorsqu'il s'agit de permettre à l'utilisateur de comprendre les prédictions émises par le système d'intelligence artificielle et d'avoir confiance en celles-ci (Amann *et al.*, 2020). En effet, si la recommandation va à l'encontre du jugement clinique, le clinicien peut valider le processus décisionnel de l'IA et alors décider de l'adopter ou de la rejeter (Amann *et al.*, 2020; Ahmad, Eckert et Teredesai, 2018). Il en va de même pour COMPAS, l'outil d'aide à la décision mentionné plus haut, qui, s'il était transparent et offrait une explication sur son processus décisionnel, permettrait aux juges de prendre une décision

plus informée. Le besoin de transparence et d'explicabilité n'est pas exclusivement axé sur le codage algorithmique et le processus qui mène à la prédiction. Il est aussi grandement nécessaire à l'humain pour justifier son propre parcours décisionnel et la raison pour laquelle il s'appuie sur la prédiction du système d'intelligence artificielle (Coeckelbergh, 2020).

Dans la section 1.3, nous avons détaillé sommairement la différence entre les modèles d'intelligence artificielle. Les modèles de type régressif et les arbres décisionnels offrent une grande explicabilité et une grande transparence grâce à leur architecture et à leur programmation, tandis que les modèles d'apprentissage automatique sont opaques et plus difficiles, voire impossibles à expliquer, considérant la façon dont les réseaux de neurones traitent les données (Coeckelbergh, 2020). Mais alors, pourquoi sélectionner un modèle opaque? Les modèles de type régressif et les arbres décisionnels sont souvent moins performants pour plusieurs tâches prédictives en comparaison avec des modèles plus opaques, ceux qui mobilisent l'apprentissage automatique ou profond (Ahmad, Eckert et Teredesai, 2018; Surden, 2020). La décision quant au type de modèle à utiliser dépend donc du niveau de performance souhaité et de la transparence et de l'explicabilité requises. Cette décision constitue un enjeu majeur pour les développeurs en intelligence artificielle (Amann *et al.*, 2020), qui nécessite qu'ils aient une discussion importante avec leurs clients pour guider leur décision quant au modèle retenu (Ahmad, Eckert et Teredesai, 2018). La question éthique qui se pose alors est de savoir si nous souhaitons gagner en transparence et en explicabilité au détriment de la performance (Coeckelbergh, 2020).

Certains auteurs apportent un bémol à la transparence et à l'explicabilité choisies au détriment de la performance. Comme le souligne London (2019), les professionnels de la santé prescrivent parfois des molécules sans en comprendre le mécanisme d'action; ce fut le cas pour l'aspirine pendant tout près d'un siècle et c'est toujours le cas pour le lithium. Bien que les scientifiques n'avaient pas « décodé » le mécanisme d'action de ces deux molécules, la démonstration de leur efficacité, documentée dans le cadre d'essais cliniques, a mené à leur utilisation pour traiter certaines conditions (London, 2019). Ainsi, il semble que dans certains cas, une efficacité élevée peut se substituer à la transparence et à l'explicabilité (London, 2019). Dans le milieu de l'intelligence artificielle, la détection de

la rétinopathie diabétique et le diagnostic par radiologie sont largement documentés et leurs résultats, équivalents ou supérieurs aux performances humaines, ont établi un niveau de confiance qui se passe d'une explicabilité pointue (Ahmad, Eckert et Teredesai, 2018). Ainsi, le niveau d'explicabilité requis varie en fonction du contexte (Ahmad, Eckert et Teredesai, 2018). Les auteurs ne s'entendent toutefois pas tous sur cette question de la transparence et de l'explicabilité contre la performance. Dans le milieu de la santé, London met en garde contre l'effet préjudiciable pour les patients de retarder le progrès et de préférer la transparence et l'explicabilité à une plus grande performance des soins (London, 2021), tandis que pour Amann, l'explicabilité devrait toujours être préférée à la performance là où elle est possible (Amann *et al.*, 2020). Il existe un véritable débat éthique sur le sujet.

La décision quant au choix de la technologie à développer revient cependant généralement aux concepteurs. Il est facile d'imaginer le danger que, sans cadre réglementaire, les organisations puissent favoriser la performance à l'explicabilité (Surden, 2020). Le domaine de l'intelligence artificielle étant largement dominé par le secteur privé, les intérêts que les dirigeants et les décideurs de ces entreprises doivent protéger sont ceux de leurs investisseurs et de leurs conseils d'administration (Mittelstadt, 2019). La quête de transparence et d'explicabilité dans le milieu des affaires se bute souvent au secret commercial. Dans le cadre de sa recherche auprès de gestionnaires, Baker-Brunnbauer (2021) a soulevé que les gestionnaires ont mentionné l'importance à leurs yeux de la transparence en intelligence artificielle, mais que la divulgation de secrets commerciaux entourant leurs technologies représente un obstacle important, étant donné l'environnement organisationnel compétitif dans lequel ils se positionnent. Le secret commercial et le brevet protègent les organisations de la compétition déloyale et de la copie (Surden, 2020). Nous pouvons penser aux entreprises pharmaceutiques qui investissent de vastes sommes dans la recherche et le développement de molécules. S'il n'y avait pas de loi pour protéger ces organisations contre la copie, qui entreprendrait des investissements aussi colossaux pour se faire copier la minute que le produit est enfin approuvé? Il est logique que les organisations cherchent à maintenir leur avantage commercial et à se garder une avance sur la concurrence, mais dans le cas d'une intelligence artificielle opaque, c'est un obstacle de plus à la transparence.

Quoi qu'il en soit, certains chercheurs travaillent à développer des techniques pour ouvrir la « boîte noire » (Coeckelbergh, 2020). L'audit demeure une méthode reconnue pour pallier les systèmes opaques et assurer un suivi du bon fonctionnement de l'outil (Ahmad, Eckert et Teredesai, 2018). Certains vont même un peu plus loin, avec un audit certifié, ou en programmant les systèmes pour qu'ils produisent leurs propres explications en lien avec leurs prédictions (Maclure et Saint-Pierre, 2018), ou encore en évaluant, selon une échelle de valeurs standardisée, la qualité des résultats émis pour établir un niveau de confiance (Cutillo *et al.*, 2020). Ces solutions visent à répondre au moins en partie aux enjeux que soulèvent les systèmes d'intelligence artificielle opaques. Certains prônent la mise en place d'une réglementation plus stricte qui mettrait de l'avant l'explicabilité de l'intelligence artificielle et mettrait l'accent sur le droit à l'explicabilité. Cependant, est-ce réellement toujours possible d'avoir une IA transparente? (Coeckelbergh, 2020) En réponse à ceux qui, en contrepartie, privilégient l'autorégulation pour éviter de ralentir l'innovation dans le milieu, Dignum (2020) voit plutôt la réglementation comme un défi à relever pour les chercheurs. Elle croit que ces situations sont justement riches en occasions de développement et peuvent pousser les scientifiques à trouver des pistes de solution qui allieront explicabilité et performance (Dignum, 2020).

Cathy O'Neil (2016), dans son livre *Weapons of Math Destruction*, met en garde contre les systèmes opaques qui n'offrent ni transparence ni explicabilité. Elle considère que les risques associés à un système d'intelligence artificielle sont proportionnels à son niveau d'opacité, à l'étendue de son utilisation dans la population et aux dommages qu'il peut infliger. De façon assez similaire, Bannister et Connolly (2020), pour leur part, regardent le type d'algorithme, sa transparence, son explicabilité et son effet potentiel. Selon Amann *et al.* (2020), le concepteur a intérêt à développer des techniques d'explicabilité pour son propre bénéfice en matière de contre-vérification, car cela lui vaudra très certainement une économie de coût et de temps au moment de la validation ou de la certification de l'outil intelligent, mais aussi, cela permettra de diminuer les risques associés à son produit et les poursuites potentiellement coûteuses.

1.4.7 Consentement, respect de la vie privée et autonomie

L'IA pose un risque du point de vue de la protection des données personnelles et du droit à la vie privée (Madiega, 2019). Le consentement, le respect de la vie privée et l'autonomie sont au cœur des enjeux éthiques en intelligence artificielle. L'IA se nourrit de données massives (Maclure et Saint-Pierre, 2018). Les GAFAM (Google, Apple, Facebook, Amazon et Microsoft) de ce monde ont depuis un bon moment mis à bon usage entre autres nos données personnelles de géolocalisation, notre historique d'utilisation et nos préférences d'achat (Zuboff, 2020). Comme le dit si bien Stéphane Ricoul dans son article *Au rythme des algorithmes* : « quand c'est gratuit, c'est que nous sommes le produit » (Ricoul, 2021). Il mentionne également que,

en une seule petite minute sur l'internet, nous consommons mondialement 500 heures de contenu sur YouTube, envoyons 69 millions de messages sur WhatsApp ou Messenger, partageons 695 000 *stories* sur Instagram ou téléchargeons 5000 vidéos sur TikTok. (Ricoul, 2021)

La quantité de données récoltées en une seule minute est colossale, ce qui implique que « [l]'hébergement, le partage et l'analyse des données personnelles posent des risques sur le plan de la protection de la vie privée des producteurs de données » (Maclure et Saint-Pierre, 2018 : 759). Le scandale de Cambridge Analytica a mis en lumière l'usage par des tiers des données d'utilisateurs récoltées par Facebook (Turner, 2019). Cette situation en a éveillé quelques-uns à l'utilisation secondaire qui pouvait être faite des données recueillies.

Chaque utilisateur doit consentir à ce que ses données soient amassées et utilisées. Cependant, plusieurs services apparemment « gratuits » sont offerts sur la base de contrats pratiquement cachés (Turner, 2019). Cette pratique nuit au concept de consentement. Grâce à des ententes, ou à des applications sous son empire, Facebook reçoit de l'information à caractère privé sur ses utilisateurs, comme leur rythme cardiaque, leur cycle menstruel et d'ovulation, ainsi que leurs recherches en immobilier, sans que ceux-ci en aient pleinement conscience (Zuboff, 2020). Les contrats d'utilisation sont souvent extrêmement longs, écrits en petits caractères et dans un vocabulaire complexe et technique peu familier pour une grande majorité d'utilisateurs (Turner, 2019). Ainsi, Turner (2019) émet l'opinion selon laquelle, si les contrats concernant l'IA deviennent de plus en plus fréquents pour le public général non expert, il sera nécessaire que la loi impose des limites ou un filet de sécurité

pour protéger les gens contre leur propre ignorance, en quelque sorte, pour les empêcher de donner des droits sans en comprendre la pleine portée. L'utilisateur doit veiller à la protection de ses données personnelles, cependant, de façon générale, comparativement aux entreprises qui souhaitent avoir accès à ces données, ses connaissances juridiques et techniques concernant les formulaires de consentement et la technologie sont plutôt limitées. Ainsi, comme Maclure et Saint-Pierre le soulignent, « cette asymétrie des pouvoirs et des connaissances rend quasi impossible la protection des données, et incite plusieurs analystes à prôner une réglementation juridique plus serrée de l'utilisation des données personnelles » (2018 : 760). Cette quête vise à protéger les droits de l'utilisateur du point de vue de la protection non seulement de ses données personnelles, mais aussi de plusieurs droits qui en découlent, comme celui de la vie privée, celui de l'autonomie et celui de donner un consentement libre et éclairé. Selon Maclure et Saint-Pierre,

[i]l est déjà possible, en recoupant les données partagées, même anonymisées, et l'information disponible publiquement au sujet d'une personne, de connaître plusieurs éléments de sa vie privée (état de santé, allégeance politique, habitudes d'achat, salaire, âge, lieux fréquentés, adresse, propriétés immobilières, etc.) (2018 : 759)

Dans certains contextes, ces données sont utilisées pour offrir du contenu ciblé aux utilisateurs. L'objectif est d'offrir un service toujours plus personnalisé pour que l'utilisateur baigne dans un environnement largement nourri par ses préférences. Il faut cependant prendre garde à ce type de pratique, qui a pour conséquence de limiter ce à quoi l'utilisateur est exposé et d'ainsi renforcer sa position au prix d'en faire disparaître les nuances et d'effriter l'ouverture d'esprit et la pensée critique (Maclure et Saint-Pierre, 2018). Si un algorithme décide du contenu auquel nous sommes exposés, nous perdons conséquemment notre droit de choisir et notre autonomie.

Un autre aspect à considérer est celui de notre participation à la société. Pour pouvoir participer à la vie sociale, il devient de plus en plus incontournable d'être présent dans le monde du Web et des réseaux sociaux, où les débats publics occupent une plus grande place (Maclure et Saint-Pierre, 2018). L'accès aux plateformes de médias sociaux est conditionnel à l'acceptation des termes et des conditions, ce qui inclut le partage des données personnelles. Ainsi, pour participer à la vie sociale numérique, il faut accepter de partager ses données (Maclure et Saint-Pierre, 2018). Cette pratique relègue à la marge

ceux qui ne souhaitent pas partager leurs données personnelles par conviction ou souci de protection de leur vie privée. Elle mène à l'exclusion. Un argument de plus pour la promotion d'une plus grande réglementation. Qu'en est-il des compagnies d'assurance qui offrent des rabais à ceux qui acceptent d'utiliser un GPS pour « tracker » leur bonne conduite? Cette pratique laisse entrevoir un monde où certains droits deviendraient des privilèges que seule la classe la plus aisée aurait le luxe de protéger, comme les droits à la vie privée, à la protection de ses données personnelles, à sa liberté et à son autonomie (O'Neil, 2016). En quête de tarifs plus avantageux, une partie de la population pourrait se laisser convaincre de renoncer à une partie de sa vie privée en consentant au partage de ses données.

Le consentement libre et éclairé en lien avec l'intelligence artificielle ne réfère pas seulement au consentement face à l'utilisation des données, mais aussi au consentement dans tous les milieux qui utilisent la technologie. Le milieu de la santé est un excellent exemple pour illustrer l'enjeu. Au Québec, on offre des soins centrés sur le patient, une approche qui vise à être respectueuse des souhaits et des valeurs de celui-ci (Amann *et al.*, 2020). On considère le patient comme un partenaire actif, qui est habilité à faire des choix, qui a droit à l'autonomie et au contrôle, ce qui implique d'être pleinement informé sur ses options de traitement ou de procédure, leurs bénéfices et les risques qui y sont associés (Amann *et al.*, 2020). Pour pouvoir donner un consentement éclairé et pleinement autonome, le patient doit être informé de l'utilisation d'un système d'intelligence artificielle (Coeckelbergh, 2020). Pour assurer la transparence du processus et donner le plus possible d'information pertinente au patient pour lui permettre de donner un consentement informé, le professionnel de la santé doit lui expliquer l'usage qui sera fait de la technologie. En France, dans le projet de loi bioéthique de 2019, L.4001-3, au paragraphe 1, on écrit que

[I]orsque pour des actes à visée préventive, diagnostique ou thérapeutique est utilisé un traitement algorithmique de données massives, le professionnel de santé qui communique les résultats de ces actes informe la personne de cette utilisation et des modalités d'action de ce traitement (Besse, Besse-Patin et Castets-Renard, 2020 : 31).

Cet impératif requiert cependant un plus grand accès à une explication. Pour avoir une discussion appropriée avec le patient, le clinicien doit lui-même être formé sur le

fonctionnement de la technologie (Besse, Besse-Patin et Castets-Renard, 2020). Le médecin devrait être capable d'expliquer de façon générale le type d'intelligence artificielle utilisé, son processus et ce qu'il produit comme résultats (Amann *et al.*, 2020). La question de l'explicabilité refait donc surface en lien avec le consentement éclairé du patient. L'explicabilité permet d'informer le patient afin qu'il ait une plus grande compréhension des choix qui s'offrent à lui (Amann *et al.*, 2020). Le consentement libre et éclairé est donc aussi une question de transparence et d'explicabilité.

Les enjeux éthiques entourant l'intelligence artificielle ne peuvent être examinés en silo. Les biais algorithmiques et la discrimination, l'autonomie, la protection des données personnelles, le consentement, la responsabilité, la transparence et l'explicabilité sont tous des aspects fondamentaux qui touchent les utilisateurs et qui peuvent creuser les inégalités déjà présentes dans notre société. Nous avons démontré combien l'IA se développe de façon exponentielle depuis les dernières années, comment, selon sa forme, elle soulève différents enjeux et que chacun de ces enjeux mérite une attention particulière pour éviter les débordements d'une autorégulation dans une industrie qui carbure largement aux investissements privés. Pour contribuer à la réflexion sur ces enjeux et encadrer les pratiques, certaines initiatives se sont mises en place.

1.5 Encadrer le développement et le déploiement de l'intelligence artificielle

Notre monde est en constante évolution. Nos codes moraux s'adaptent et se façonnent au rythme des bouleversements au sein de nos sociétés. Les innovations comme celles des trois grandes révolutions industrielles ont eu des répercussions indélébiles sur nos économies et ont transformé notre cadre social. Au début du siècle dernier, l'arrivée de l'électricité et du chauffage au charbon a apporté une infinité de nouvelles possibilités (O'Neil, 2016). Cependant, l'absence de normes de santé et de sécurité dans une période de croissance industrielle sans précédent a aussi apporté son lot d'abus (O'Neil, 2016). À force de pression et de revendications, un cadre réglementaire s'est mis en place pour éliminer les débordements. On a vu la naissance du droit des travailleurs, et on a mis un frein à la pratique du travail des enfants (O'Neil, 2016). En nivelant le terrain de jeu, l'encadrement des pratiques vient soutenir les efforts des entreprises consciencieuses et condamner les pratiques douteuses mises de l'avant par des entrepreneurs qui ont très peu

de scrupules et qui sont prêts à tout pour maximiser leurs marges de profit (O'Neil, 2016). Il y a quelques années, Klaus Schwab, président et fondateur du Forum économique mondial, faisait référence à une 4^e révolution industrielle en parlant des répercussions qu'auront l'automatisation intelligente et l'arrivée du numérique sur notre économie et sur nous en tant qu'êtres humains (Maclure et Saint-Pierre, 2018). Tout comme dans les révolutions industrielles précédentes, les chamboulements déjà amorcés par les nouvelles technologies demandent une adaptation constante. Schwab nous prévient :

Le leadership des systèmes signifie qu'en matière de nouvelles technologies, on ne doit pas s'arrêter à leur seule conception, mais prendre aussi en compte la façon dont elles sont régies et gouvernées, les valeurs qu'elles exposent et leur impact sur les gens de tous horizons. (Schwab, 2018)

Fortes des leçons du passé et conscientes des enjeux éthiques que nous avons exposés dans la section précédente, des initiatives se mettent en place pour encadrer le développement et le déploiement de l'intelligence artificielle. Nous proposons une revue sommaire de l'évolution du cadre de gouvernance en pleine 4^e révolution industrielle, ainsi que des principales réflexions entourant les initiatives jusqu'à maintenant.

À ce jour, plusieurs organismes ont développé des principes éthiques et des lignes directrices pour encadrer l'intelligence artificielle. Le domaine de l'éthique des « machines » est en émergence depuis au moins le début des années 2000 (Anderson et Anderson, 2007). L'organisation à but non lucratif AlgorithmWatch a compilé les initiatives entourant l'éthique de l'intelligence artificielle en recensant les cadres éthiques, les lignes directrices ainsi que les principes ébauchés partout à travers la planète. La liste n'est pas complète, mais en avril 2020, elle comptait 173 initiatives (AlgorithmWatch, 2020). Sans surprise, la liste ne contient pas les fameuses lois d'Isaac Asimov, qui sont considérées comme les premières lois de la robotique. Isaac Asimov est un prolifique écrivain de science-fiction qui, en 1942, dans sa nouvelle *Cercle vicieux*, propose trois lois dont l'objectif est d'encadrer le comportement des robots afin de servir l'intérêt des humains (Gibert, 2020). Les trois lois sont :

1. Un robot ne peut porter atteinte à un être humain ni, en restant passif, permettre qu'un être humain ne soit en danger.

2. Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi.
3. Un robot doit protéger son existence tant que cela n'entre pas en conflit avec la première ou la deuxième loi. (Gibert, 2020)

Bien entendu, ces lois sont issues de la science-fiction et n'ont jamais eu force de loi, néanmoins Asimov est le premier à avoir mis sur papier un code moral pour encadrer l'intelligence artificielle. C'est cependant les quatre principes de la bioéthique qui semblent être les lignes directrices pour les différentes initiatives mises de l'avant jusqu'à maintenant.

Tableau 1.1 Principes de la bioéthique

Principes de la bioéthique
Respect
Autonomie
Bienfaisance
Non-Malfaisance

(Beauchamp et Childress, 2001; Baker-Brunnbauer, 2021)

Les cadres se mettent en place via des initiatives supranationales, gouvernementales nationales ou régionales, et certaines initiatives sont aussi issues des milieux universitaires, de différentes industries, de regroupements de professionnels, ou encore, nous les retrouvons au sein d'organisations, tout particulièrement celles issues des technologies.

1.5.1 Commission européenne

La Commission européenne est un des principaux acteurs sur le plan mondial en ce qui concerne la gouvernance de l'intelligence artificielle (Coeckelbergh, 2020). En 2018, la Commission a mis en place un comité d'experts pour établir les lignes directrices d'une intelligence artificielle de confiance. À la fin de l'année, le groupe avait déjà une ébauche de cadre qui déterminait les droits principaux à protéger et les principes autour desquels l'IA devait s'axer. Les droits sont ainsi explicités :

- Respect de la dignité humaine;
- Liberté des individus;
- Respect de la démocratie, de la justice et de l'État de droit;
- Égalité, non-discrimination et solidarité;

- Droits des citoyens
(Groupe d'experts de haut niveau sur l'intelligence artificielle, 2019 : 13).

En décembre 2018, le groupe publie les principes éthiques qui sont au cœur de leur concertation :

- Principe de bienfaisance;
- Principe de non-malfaisance;
- Principe d'autonomie;
- Principe de justice (équité);
- Principe d'explicabilité (transparence)

(Groupe d'experts de haut niveau sur l'intelligence artificielle, 2018 : 8, traduction libre).

Nous pouvons percevoir l'influence de la bioéthique dans ce premier jet. Cependant, l'explicabilité est intégrée aux principes éthiques pour démontrer la place importante du droit à une explication et forcer un travail de réflexion quant à l'innovation du point de vue de l'ouverture de la « boîte noire » ou du choix du modèle d'intelligence artificielle (Coeckelbergh, 2020). À la base, les lignes directrices européennes veulent s'assurer d'une IA respectueuse des valeurs et des principes européens (Madiaga, 2019).

1.5.2 Initiatives supranationales

En mai 2019, l'Organisation de coopération et de développement économique a adopté des principes sur l'intelligence artificielle et émis des recommandations pour le développement de politiques publiques et de stratégies pour encadrer l'intelligence artificielle (OCDE, 2019a). Ainsi, « [l]es 36 pays membres de l'OCDE, ainsi que l'Argentine, le Brésil, la Colombie, le Costa Rica, le Pérou et la Roumanie ont adhéré aux Principes de l'OCDE sur l'intelligence artificielle » (OCDE, 2019a). Ces principes sont axés sur un développement dans le respect des droits de la personne et des valeurs démocratiques, et mettent l'accent sur la transparence et l'explicabilité, la sécurité et la responsabilité (OCDE, 2019a).

Tableau 1.2 Recommandations du Conseil sur l'intelligence artificielle

Recommandations du Conseil sur l'intelligence artificielle
Croissance inclusive, développement durable et bien-être
Valeurs centrées sur l'humain et équité
Transparence et explicabilité

Robustesse, sûreté et sécurité
Responsabilité

(OCDE, 2019b)

L'OCDE s'est d'ailleurs dotée d'un observatoire des politiques relatives à l'IA en février 2020 (OCDE, 2019a). L'Organisation mondiale de la Santé n'est pas en reste et a produit un rapport en 2021 sur ses principes directeurs relatifs à la conception et à l'utilisation de l'intelligence artificielle dans le domaine de la santé (OMS, 2021). D'ailleurs, ses principes sont très similaires à ceux de l'OCDE, en plus de nommer très clairement le principe de l'autonomie et de la dignité humaine, des valeurs fondamentales dans les soins de santé (OMS 2021).

1.5.3 Initiatives nationales

Au niveau national, plusieurs pays ont amorcé des réflexions sur l'IA et mis de l'avant des cadres éthiques. En France, Cédric Villani, mathématicien et député, a émis un rapport pour guider la stratégie française et européenne en intelligence artificielle. Il y recommande une IA qui peut être interprétée et qui est sensible à ses répercussions sur l'environnement (Villani *et al.*, 2018). La France prône l'explicabilité en IA, entre autres pour favoriser l'acceptabilité sociale (Madiaga, 2019). La Grande-Bretagne met aussi l'accent sur une IA transparente et qui puisse être interprétée pour s'assurer d'avoir une IA responsable (Coeckelbergh, 2020). En 2016, les États-Unis rendent public leur plan stratégique en IA (Lee, 2019). Ils proposent une série de mesures pour soutenir la transformation économique entourant le déploiement de la technologie et des investissements visant à stimuler le milieu (Lee, 2019). La Maison-Blanche reste à l'affût de la situation et porte un intérêt particulier aux biais, à la discrimination et à l'explicabilité; elle recommande l'autorégulation, qui à ses yeux, est du devoir des concepteurs, en rappelant que le cadre réglementaire sera adapté selon le besoin (Coeckelbergh, 2020). Seulement trois mois plus tard, Donald Trump prend le pouvoir et réduit les investissements dans la recherche en IA (Lee, 2019). Pour sa part, la Chine a publié son plan stratégique en juillet 2017 (Lee, 2019). En plusieurs points similaire à celui des États-Unis, il s'en différencie en précisant des objectifs précis à atteindre selon des échéanciers (Lee, 2019). La Chine avait un grand retard sur les meneurs dans le développement de l'IA, mais son effort de rattrapage des dernières années laisse

croire qu'elle dépassera les États-Unis dans moins de cinq ans pour certaines formes d'avancement de la technologie (Lee, 2019). La Chine est sensible au fait que l'intelligence artificielle est une technologie qui peut compromettre l'équilibre social, avoir une incidence sur les droits et porter atteinte à la vie privée et à la sécurité (Coeckelbergh, 2020).

Bien que les enjeux soulevés par les différents pays de façon générale soient assez similaires, l'approche de chacun est influencée par une culture très régionale. La stratégie européenne consiste à développer un cadre d'analyse qui met l'accent tout particulièrement sur la protection de la vie privée, des données personnelles et sur la discrimination, tandis que le modèle américain mise plus fortement sur l'autorégulation du secteur privé (Madiaga, 2019). Les États-Unis sont « à la recherche de compromis adéquats en termes de protection de la vie privée. » (Lee, 2019 : 245) En 2017, la Chine met en place sa loi pour la cybersécurité, qui prévoit des sanctions pour protéger les données privées (Lee, 2019).

1.5.4 Initiatives universitaires

De nombreuses initiatives universitaires participent aussi au développement du cadre éthique de l'intelligence artificielle. Stanford et Cambridge s'intéressent plus particulièrement à la perspective de l'éthique appliquée, tandis que l'Université de Santa Clara se concentre plus intensément sur l'éthique professionnelle concernant l'intelligence artificielle (Coeckelbergh, 2020). L'Université de Montréal a pour sa part effectué un travail de recherche qui a mené à la *Déclaration de Montréal pour une IA responsable*. Le document a été publié en 2018. La *Déclaration* est le résultat d'une vaste consultation auprès de plusieurs parties prenantes, dont les citoyens (Coeckelbergh, 2020). Les principes qui en ont émané sont proposés en tant que lignes directrices pour le développement d'une IA qui est moralement et socialement désirable (Université de Montréal, 2018). Le résultat est composé de 10 principes :

Tableau 1.3 Principes de la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*

Déclaration de Montréal en IA responsable
Bien-être

Respect de l'autonomie
Protection de l'intimité et de la vie privée
Solidarité
Participation démocratique
Équité
Inclusion de la diversité
Prudence
Responsabilité
Développement soutenable

(Université de Montréal, 2018)

Nous aborderons de nouveau la *Déclaration de Montréal* dans la section du cadre d'analyse.

1.5.5 Initiatives organisationnelles

Certaines organisations lancent leurs propres initiatives internes pour encadrer le développement de l'IA. Google, Microsoft et Accenture, entre autres, ont développé leurs propres principes (Coeckelbergh, 2020). DeepMind, IBM, Amazon, Intel, Apple, Sony et Facebook ont participé en 2015 à une collaboration des géants de l'industrie, du nom de Partnership on AI, pour une charte sur une IA au bénéfice du peuple et de la société (Besse, Besse-Patin et Castets-Renard, 2020). Bien que l'industrie mise fortement sur l'autorégulation, certains dirigeants, comme celui d'Apple, croient que la situation actuelle n'est pas suffisante et qu'une réglementation de la technologie est inévitable (Coeckelbergh, 2020). D'ailleurs, Mittelstadt (2019) met en garde contre le sentiment du devoir accompli lorsqu'une organisation met en place un code de conduite éthique pour encadrer les pratiques organisationnelles concernant l'IA. Une étude de l'Association for Computing Machinery (ACM) démontre la faible incidence de la présence d'un code de conduite sur les décisions des concepteurs (McNamara, Smith et Murphy-Hill, 2018). Pour qu'un code de conduite puisse réellement avoir une influence, il doit être intégré à la culture d'entreprise et imposé par des mesures concrètes (Mittelstadt, 2019). Ainsi, comme le mentionne Mittelstadt, il y a une différence entre développer un code d'éthique interne et l'appliquer. Il faut le rendre opérationnel.

Somme toute, de façon générale, les initiatives répertoriées par AlgorithmWatch prennent plutôt la forme de déclarations, et très peu s'attardent à donner des lignes directrices pour opérationnaliser les principes (AlgorithmWatch, 2020). Les différents cadres développés ont surtout force de recommandation. Ils amènent un consensus social sur les valeurs acceptables dans le développement et le déploiement de l'intelligence artificielle (Coeckelbergh, 2020). Les principes de transparence, d'égalité, de non-discrimination, de responsabilité (imputabilité) et de sécurité se retrouvent de façon récurrente dans une grande partie des documents, et certaines initiatives ajoutent les principes de protection des droits de la personne et d'une IA au bénéfice de la société (AlgorithmWatch, 2020). La critique principale qui est faite autour des principes et des initiatives est le manque d'applicabilité de ceux-ci et leur aspect volontaire. Il s'en dégage un manque de mécanisme de gouvernance. Souvent vagues, les principes sont plutôt généraux, et dans un environnement compétitif, il y a peu de bénéfices à les respecter, ce qui se traduit par un manque d'actions concrètes (Madiega, 2019; Mittelstadt, 2019).

1.5.6 Force de loi

Beaucoup plus rares, certaines initiatives qui ont force de loi voient le jour. Le 27 avril 2016, l'Union européenne a adopté le Règlement général sur la protection des données (RGPD), qui est applicable depuis le 23 mai 2018 sur tout le territoire des États membres de l'Union européenne, dans tout l'espace économique européen, ce qui implique aussi des répercussions hors frontières (Union européenne, 2018). La particularité du RGPD est sa portée légale (Coeckelbergh, 2020). En effet, le règlement comporte des sanctions sévères pour tout contrevenant (Union européenne, 2018). L'objectif est de protéger la population de l'Union européenne et de lui donner les moyens de s'occuper du problème de la protection de ses données (Coeckelbergh, 2020) en responsabilisant les acteurs qui traitent les données et en donnant de la crédibilité au cadre réglementaire (Union européenne, 2018). Toutes les données disponibles sur le territoire de l'Union européenne sont protégées par le RGPD. Le règlement s'impose donc à l'extérieur des frontières des États membres, car si une entreprise n'est pas physiquement sur le territoire, mais qu'elle procède à la collecte de données sur le territoire du RGPD, son cadre réglementaire doit s'appliquer (Union européenne, 2018). La Chine a récemment adopté la Personal

Information Protection Law (PIPL), un règlement similaire à celui qui est mis en place en Europe pour protéger la vie privée et les données produites sur le territoire chinois. Cette réglementation est applicable depuis le 1^{er} novembre 2021 (Cohen, 2021). Elle vise à limiter les droits de collecte, de traitement et de stockage des données, et elle impose une série de mesures pour encadrer les pratiques sur son territoire et soumettre toutes organisations à l'obligation de « rendre des comptes aux agences de surveillance du pays » (Cohen, 2021). Ces initiatives se sont dotées d'un pouvoir donnant force de loi.

1.5.7 Encadrement québécois et canadien

Le Québec n'est pas en reste : c'est la première province au Canada à se doter d'un cadre aussi strict que le RGPD pour la protection des données personnelles en mettant à jour sa *Loi sur la protection des renseignements personnels dans le secteur privé* et sa *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* par le projet de loi 64, adopté le 21 septembre 2021 (Comité Technologies de l'information, 2022). Cette réforme prévoit des sanctions financières et/ou pénales à quiconque contrevient aux nouvelles lois. Les modifications entreront progressivement en application à différents moments dans un délai de trois ans (Romano, Stoddard et Uzan-Naulin, 2021). Plusieurs autres provinces prévoient des amendements à leurs lois (Comité Technologies de l'information, 2022). Au niveau fédéral, le projet de loi C-11 a l'ambition de mettre à jour entre autres la *Loi sur la protection de la vie privée des consommateurs* et la *Loi sur le Tribunal de la protection des renseignements personnels et des données* (Romano, Stoddard et Uzan-Naulin, 2021). Le gouvernement fédéral a développé l'outil d'évaluation de l'incidence algorithmique (EIA), qui a pour objectif d'évaluer le risque potentiel d'un algorithme sur la population (Madiega, 2019; Gouvernement du Canada, 2022). De façon similaire, le RGPD requiert que le développeur fasse une analyse d'incidence sur la protection des données via son outil d'évaluation PIA (*privacy impact assessment*) (Union européenne, 2018).

Somme toute, nous pouvons constater, au cours des dernières années, une prise de conscience par rapport au développement de l'intelligence artificielle et à la nécessité de l'encadrer. Les initiatives qui ont mis de l'avant des cadres éthiques et des lignes directrices servent de fil conducteur à la mise en place de cadres plus stricts comme le RGPD, le PIPL

ou le projet de loi 64 au Québec. Un enjeu important à l'heure actuelle sera très certainement l'effet qu'aura la disparité de rigueur des cadres réglementaires étatiques sur l'innovation technologique régionale. Le manque de standardisation sur le plan mondial mènera à une compétition inégale et aura pour effet de rendre les pays les plus laxistes plus attrayants pour les organisations conceptrices en intelligence artificielle (Baker-Brunnbauer, 2021; Madiaga, 2019). Il faudra donc équilibrer les intérêts économiques nationaux et les valeurs éthiques mondiales en harmonisant les pratiques par l'entremise de consensus.

1.6 Intelligence artificielle affective

Maintenant que nous avons parcouru l'historique de l'intelligence artificielle, que nous avons survolé son fonctionnement, fait une revue de ses principaux enjeux éthiques et des initiatives qui tentent de l'encadrer, nous pouvons aborder un sous-domaine de l'IA : l'intelligence artificielle affective.

Un processus décisionnel robuste a longtemps été reconnu comme étant basé sur une logique rationnelle faisant appel à des habiletés cognitives (Richardson, 2020). La recherche en neuroscience a cependant démontré le rôle essentiel de l'émotion dans le jugement et le processus décisionnel humains (Picard, 1995). Le neuroscientifique Antonio Damasio a démontré que tout comme de trop fortes émotions peuvent altérer le jugement de quelqu'un et le porter à faire de mauvais choix, la quasi-absence d'émotions peut être tout aussi lourde de conséquences (Richardson, 2020). Considérant l'importance de l'émotion dans la prise de décision, Rosalind Picard avance ceci : « I have suggested that if computers will ever interact naturally and intelligently with humans, then they need the ability to at least recognize and express affect. » (1995 : 24) L'intelligence artificielle affective s'intéresse au développement de systèmes capables de reconnaître les émotions, d'y répondre et de les influencer (Richardson, 2020). La vie émotionnelle devient ainsi accessible pour l'analyse par l'intermédiaire de la technologie (McStay et Urquhart, 2019), la porte s'ouvre sur une nouvelle dimension de l'être humain. Les avancées en apprentissage automatique profitent largement au sous-domaine, qui est en effervescence. La reconnaissance vocale et la reconnaissance faciale qui sont utilisées dans le milieu ont connu des développements importants dans les dernières années. Le taux d'erreurs des

meilleurs systèmes de reconnaissance d'images sur la base de données ImageNet, qui est peuplée de millions de photos, est passé de plus de 30 % en 2010 à environ 4 % en 2016 (Brynjolfsson et McAfee, 2017). Les systèmes qui se développent bénéficient de ces avancées. Les algorithmes d'intelligence artificielle affective lisent les émotions et y réagissent en analysant des données provenant de différentes sources d'information, telles que le texte, la voix, la vision informatisée et/ou la biométrie via entre autres le rythme cardiaque, la température corporelle et la respiration (McStay et Urquhart, 2019; Brigham, 2017). Rosalind Picard a recensé trois types d'application pour les systèmes d'intelligence artificielle affective : la détection des émotions chez les utilisateurs, l'expression informatisée d'émotions et l'action de ressentir des émotions (Brigham, 2017).

Certains systèmes sont maintenant capables de détecter et de décoder des expressions retrouvées sur les visages humains (Daily *et al.*, 2017). Pour y arriver, les algorithmes ont dû être alimentés et formés par des bases de données au sein desquelles les émotions ont été codifiées. Selon Paul Ekman, il existe des émotions universelles de base qui sont communes à tous les humains, quels que soient leur âge, leur genre, leur appartenance ethnique ou leur culture (Richardson, 2020). En s'inspirant de ces émotions, le Facial Action Coding System (FACS) a été développé pour catégoriser et classifier les expressions faciales (Daily *et al.*, 2017). Développé entre autres par Ekman, il est basé sur les émotions universelles de base, soit la colère, la peur, la joie, la tristesse, le dégoût, le mépris et la surprise (McStay et Urquhart, 2019). L'étude des expressions faciales a permis de cerner les muscles faciaux contractés selon les différentes émotions. Le plissement sur le nez, par exemple, est associé plus étroitement au dégoût, tandis que le plissement vers le bas à la commissure des lèvres est souvent associé à la tristesse (McStay et Urquhart, 2019). Nous pouvons facilement imaginer les possibilités qu'apporte une technologie qui est capable de déceler les émotions, que ce soit analyser les communications de type courriel ou message texte, l'intonation de la voix, les expressions captées par vidéo ou photos ou pouvoir faire le lien en temps réel entre les émotions et le rythme cardiaque ainsi que la respiration.

Le développement de l'intelligence artificielle affective a suscité un vif intérêt dans différents domaines. Des chiffres avancés par Taylor Telford (2019) suggèrent que la

détection d'émotions à elle seule représenterait un marché de 20 milliards de dollars. La technologie peut permettre des avancées pour la recherche dans le milieu de la santé, du bien-être et de la psychologie. Elle peut avoir des applications dans le marketing, la communication, le divertissement et bien plus (Picard, 1995). Dans le milieu de la recherche, par exemple, on peut penser au bénéfice qu'elle représenterait pour venir en aide aux gens qui souffrent de dépression ou de douleur chronique puisque ce sont des conditions difficilement perceptibles physiquement, mais qui représentent une détresse réelle (Brigham, 2017). Dans les médias sociaux, elle permet d'analyser en temps réel une tendance émotionnelle globale (Richardson, 2020). Dans le milieu de la publicité, la possibilité de personnaliser le message aux clients potentiels selon leurs émotions est une corde de plus à son arc. Pour offrir du contenu personnalisé, Ocean Outdoor, une compagnie britannique, personnalise son contenu en fonction de l'âge, du genre et de la géolocalisation, et a eu recours à l'analyse de l'émotion (McStay et Urquhart, 2019). Dans le cadre d'une campagne de grande envergure dans Piccadilly Circus à Londres, on a capté et analysé les expressions faciales des piétons pour adapter le contenu projeté sur l'immense écran de la place publique largement affluente de la capitale (McStay et Urquhart, 2019). La technologie offre aux entreprises la possibilité de s'adapter en temps réel au pouls de sa clientèle. C'est d'ailleurs un aspect important en organisation. Depuis un moment maintenant, plusieurs organisations comprennent l'importance de prendre la mesure du bien-être des employés en entreprise. Il y a donc un vif intérêt dans l'intelligence artificielle affective pour optimiser l'environnement de travail (McStay, 2020). La vision informatisée est utilisée pour prendre le pouls de l'attitude et de l'engagement des employés (McStay et Urquhart, 2019). Des compagnies comme Unilever et IBM l'utilisent dans leur processus d'entrevue pour analyser le comportement du candidat pendant la rencontre (Zetlin, 2018). Elle est utilisée aussi dans les salles de cours : en Chine, des caméras ont été placées de sorte à pouvoir balayer l'ensemble des étudiants aux 30 secondes pour détecter le niveau d'attention du groupe. Le professeur peut être informé si un étudiant semble moins éveillé à la matière (Richardson, 2020).

La recherche et le développement de cette technologie attirent plusieurs entreprises, comme Facebook, IBM, Microsoft, Google, Apple, et des plus petites compagnies, comme Sensing Feeling, Affectiva (McStay, 2020), ou EmoScienS, une entreprise en démarrage

montréalaise. Les nombreux développements dans le domaine pourront très certainement aider à résoudre certains enjeux. Prenons les accidents de la route, par exemple. Comme le soulève Richardson (2020), le niveau d'attention, la fatigue et l'état émotionnel représentent des facteurs de risque importants et des causes d'accidents souvent rapportées. Certaines compagnies en intelligence artificielle affective commencent à développer des solutions grâce à la détection des émotions pour alerter et sensibiliser les conducteurs à leur état (Richardson, 2020). Cependant, si cette technologie est utilisée dans un tout autre contexte, elle peut au contraire soulever de nouveaux enjeux. McStay et Urquhart (2019) nous donnent l'exemple de son utilisation par des compagnies d'assurance : que penser si les compagnies utilisaient un lecteur d'émotions intégré au rétroviseur du véhicule, qui pourrait déceler la rage au volant présente chez un conducteur juste avant un accident?

Tout comme l'intelligence artificielle soulève des enjeux éthiques, ce sous-domaine apporte aussi son lot d'enjeux. Le développement et le déploiement de telles technologies nous poussent à réfléchir entre autres à la protection de la vie privée, à la sécurité, à l'autonomie, à la fiabilité et aux biais qui peuvent mener à de la discrimination.

1.6.1 Enjeux éthiques en lien avec l'intelligence artificielle affective

La protection de la vie privée est un enjeu important. Les émotions étant personnelles et privées, on ne veut pas nécessairement qu'elles soient accessibles et encore moins qu'elles puissent être utilisées contre nous, par exemple dans un contexte commercial (Daily *et al.*, 2017). En rendant l'émotion visible, on envahit l'espace où chacun conçoit sa propre identité (McStay et Urquhart, 2019). Les villes intelligentes sont sujettes à l'intrusion dans la vie privée et, considérant les abus quant aux données personnelles par le passé, il faudra avancer avec précaution pour respecter et protéger la vie privée des utilisateurs et des citoyens (McStay et Urquhart, 2019). En effet, la surveillance est un aspect délicat dans le contexte de villes intelligentes. Le fait que l'intelligence artificielle affective puisse déceler les émotions de la population soulève un tout autre degré d'invasion dans la vie privée des citoyens. Il en va de même pour l'employé ou le consommateur qui ne veut pas nécessairement qu'on ait accès à son intimité. Le conducteur ne veut possiblement pas non plus être épié en permanence par sa compagnie d'assurances.

La sécurité devient aussi un enjeu incontournable. Il faut développer des logiciels sécuritaires qui laissent les utilisateurs décider qui pourra avoir accès à leurs informations (Daily *et al.*, 2017). Maintenir la sécurité pour chaque individu sur des réseaux potentiellement moins sécuritaires, c'est un défi pour toutes les technologies (Daily *et al.*, 2017). Nous avons déjà abordé l'importance de protéger les données contre des accès non autorisés, du piratage et du vol de données dans la section 2.4.3. Cependant, lorsqu'on parle d'informations aussi sensibles que le profil émotionnel de clients, d'employés, de citoyens, quels pourraient être les répercussions d'une brèche dans de tels systèmes? Le risque d'abus par la manipulation, par l'influence passive est une crainte réelle (Daily *et al.*, 2017). Les systèmes peuvent aussi être faillibles ou avoir des effets pervers que les concepteurs n'avaient pas prévus. Considérant que ce type de technologie opère au niveau des émotions, il faut faire attention que son utilisation n'influence pas négativement les bonnes émotions (Daily *et al.*, 2017) ou plus largement le bien-être de la personne. De façon non intentionnelle, une fonctionnalité peut, par inadvertance, avoir un effet sur l'humeur d'un utilisateur. La malveillance pouvant vraisemblablement entrer dans l'équation, une utilisation abusive qui prendrait entre autres la forme de la manipulation pourrait avoir un effet sur le bien-être des individus (Beavers et Slattery, 2017).

L'autonomie n'est pas en reste et représente un enjeu dont il faut tenir compte quand il s'agit d'intelligence artificielle affective. Le profilage passif et l'utilisation d'une telle technologie pour influencer inconsciemment les utilisateurs touchent à une corde sensible et peuvent miner l'intégrité humaine (McStay, 2020) et son autonomie. La personnalisation de l'expérience en magasin par l'analyse des émotions des clients amène à un tout autre niveau, qui peut facilement être perçu comme de la manipulation (McStay et Urquhart, 2019). Les notions de transparence et de consentement deviennent particulièrement importantes (McStay et Urquhart, 2019). Un consommateur qui n'est pas conscient des données et des techniques utilisées pour le convaincre perd de l'autonomie dans son processus décisionnel et il ne peut recourir pleinement à son esprit critique. Ainsi, il est primordial de maintenir informées toutes les personnes qui sont sujettes à l'utilisation d'une telle technologie et d'obtenir leur consentement (McStay et Urquhart, 2019). Il en va de même pour les systèmes visant à améliorer le bien-être par la voie de la sensibilisation ou de l'éducation émotionnelle. L'utilisation de ce type de système peut miner

l'intelligence émotionnelle des utilisateurs qui s'appuieraient indûment sur la technologie (Beavers et Slattery, 2017). Elle doit être utilisée de façon responsable et demeurer une forme d'outil pour améliorer le bien-être (Daily *et al.*, 2017) et minimiser les formes de dépendances.

La fiabilité représente un enjeu de taille également. C'est particulièrement délicat lorsqu'on parle de dépendance et d'intelligence émotionnelle. Les algorithmes sont sujets à des données mal étiquetées, à des circonstances contextuelles changeantes et à toutes sortes d'éléments de la sorte (Landowska, 2019). Les utilisateurs devraient connaître le niveau de fiabilité des systèmes d'intelligence artificielle affective pour mettre en perspective leurs recommandations et leurs résultats (Landowska, 2019). Selon McStay (2020), l'expression faciale prise hors contexte peut ne pas être traduite de façon appropriée, car il existe une similitude subtile entre différentes catégories d'émotions (McStay, 2020), et sans le contexte, il est possible de déduire à seulement 20 à 30 % la bonne émotion à partir de l'expression faciale à elle seule (Richardson, 2020). Il est clair que l'information contextuelle liée à l'expression faciale est un facteur important pour déterminer une émotion (McStay et Urquhart, 2019). Les mêmes limites ont été retrouvées dans la détection d'émotions par l'analyse de textes (Richardson, 2020). L'humain fait d'ailleurs face, lui aussi, à ces limites : il n'est pas rare de rencontrer des malentendus qui émanent d'échanges de courriels ou de messages textes. Souvent, il manque certains éléments contextuels, comme l'intonation, pour bien cerner l'émotion derrière le message. Cependant, pour prendre en compte le contexte, il va falloir faire une plus grande collecte de données, en faisant appel à différentes sources, telles que la caméra et les technologies portables, et capter plus de mesures, comme le rythme cardiaque, les conditions météorologiques ou la circulation routière, donc être bien plus invasif dans la vie privée des gens (McStay et Urquhart, 2019). Ce sont ici deux enjeux qu'il faudrait équilibrer pour obtenir une plus grande fiabilité tout en respectant la vie privée des gens. Enfin, la fiabilité passe aussi par la qualité des banques de données utilisées par le système. C'est un élément controversé au sein du sous-domaine puisque bien qu'Ekman ait établi que les émotions et les expressions faciales sont universelles, J.A. Russell a pour sa part trouvé des failles dans l'étude d'Ekman et remet en question ses conclusions : aux yeux de Russell, les émotions sont bien plus subtiles et sujettes à différentes interprétations (McStay et Urquhart, 2019).

De la même façon, Feldman Barrett *et al.* (2019) soutiennent que la manière dont les gens communiquent leurs émotions varie selon leur culture, le contexte et même au sein d'un groupe d'individus placés dans la même situation. Pour le moment, il existe un risque élevé de faux positif dans la détection d'émotions, ce qui signifie que la technologie conclut à un état psychologique chez un utilisateur, tandis que ce dernier ne reconnaît pas avoir expérimenté une telle émotion (Bullington, 2005). Landowska (2019) recommande fortement que les développeurs qui utilisent la détection émotionnelle rapportent la marge d'erreur de leur système, considérant les décisions qui peuvent être prises en lien avec leurs résultats.

L'enjeu de **biais et de discrimination** est aussi à considérer en lien avec ce type de technologie. En effet, la catégorisation des émotions est basée depuis longtemps sur une vue largement centrée sur l'Occident (McStay et Urquhart, 2019). Les bases de données et l'étiquetage sont donc largement influencés par la culture qui y est représentée. Brigham (2017) pointe du doigt le manque de considération pour la variabilité culturelle de l'expression des émotions. L'éducation émotionnelle se fait énormément à travers notre environnement et nos interactions sociales (Brigham, 2017). Nous apprenons ce qui est socialement acceptable au sein de notre culture. La variabilité de l'expression de nos émotions selon notre culture peut rendre les systèmes d'intelligence artificielle affective moins fiables au sein de populations moins étudiées et moins représentées dans la base de données utilisée (Brigham, 2017) et créer des biais non suspectés. Cette situation soulève des « *red flags* » du point de vue de l'applicabilité de systèmes optimisés par l'intelligence artificielle affective au sein de cultures non représentées dans la base de données.

Les risques d'**étiquetage et de stigmatisation** par la catégorisation des individus peuvent mener à un traitement inégal au sein de la population (McStay et Urquhart, 2019). Il faut prendre garde à ne pas causer de préjudices en oubliant que l'émotion est bien plus subtile et complexe que sa simple réduction à des mesures (McStay et Urquhart, 2019). De plus, considérant le manque de consensus scientifique sur ce que lit réellement l'intelligence affective en termes d'émotions, il faut prendre nos précautions quant à des contextes d'application qui peuvent avoir une implication sérieuse et pourraient causer du tort à

certaines individus (McStay et Urquart, 2019). Son utilisation par des forces policières, par exemple, peut mener à la stigmatisation de certains individus.

Enfin, Richardson (2020) met en garde contre **la démocratisation** et la privatisation de l'intelligence artificielle. La démocratisation réfère à l'absence de grandes barrières à l'entrée et à l'accessibilité dans le développement d'outils d'intelligence artificielle. Il n'y a pas de prérequis très rigoureux en termes de formation ou d'habileté de programmation pour pouvoir développer et mettre sur le marché de nouvelles technologies d'IA (Richardson, 2020). Pour l'intelligence artificielle affective, cette porte grande ouverte pose de sérieux risques quant à la fiabilité, à la protection de la vie privée, à l'éthique et à la gouvernance des données (Richardson, 2020). Considérant le développement de technologies de nature aussi délicate, plusieurs souhaitent voir la mise en place d'un code de conduite pour les concepteurs, un peu comme le serment d'Hippocrate pour les médecins (Richardson, 2020).

Beavers et Slattery (2017) mettent en perspective pour nous le fait que l'intelligence artificielle affective est moralement neutre. Elle n'est pas mal intentionnée ou dangereuse, c'est son contexte de développement, d'application et d'utilisation qui soulève des enjeux, c'est l'humain derrière son développement et son déploiement qui a un cadre de valeurs et non le système en soi (Beavers et Slattery, 2017). Considérant que la technologie se développe plus rapidement que la compréhension humaine et la législation (Beavers et Slattery, 2017), il est important de continuer à l'étudier et à tracer des lignes directrices claires pour l'encadrer au niveau international (Brigham, 2017).

1.7 Conditions d'acceptabilité de l'IA

Nous avons maintenant établi que l'intelligence artificielle est une technologie pleine de promesses et qu'elle connaît actuellement un intérêt et un développement sans précédent. En réponse aux enjeux éthiques qui l'entourent, un cadre réglementaire se met en place pour donner des lignes directrices aux concepteurs et aux organisations qui l'adoptent. Il reste que le déploiement de la technologie dans différentes sphères, et tout particulièrement en santé, reste timide (Asan, Bayrak et Choudhury, 2020). Il est donc pertinent de se questionner sur les facteurs qui influencent l'acceptabilité des systèmes d'intelligence artificielle et ultimement les systèmes d'intelligence artificielle affective.

Chaque individu a ses propres critères d'acceptabilité, mais de façon générale, une norme sociale émerge et des attentes communes prennent forme. Caron, Bernardi et Nicolini rapportent que « [l]e concept d'acceptabilité sociale a fait son apparition au cours des années 2000 et [que] depuis, il est présent au sein de plusieurs discours publics. » (2021 : 7) Il ne semble pas y avoir une définition uniforme du concept (Caron, Bernardi et Nicolini, 2021). L'acceptabilité sociale renvoie à l'évaluation du coût-bénéfice d'une acceptation. Il s'agit d'évaluer non seulement la perception quant à la technologie, mais aussi les normes sociales interpellées dans le processus de réflexion (Bobillier-Chaumon et Dubois, 2009).

Plusieurs facteurs influencent le processus d'acceptabilité sociale, tels que l'environnement de l'utilisateur, les caractéristiques de la technologie et les caractéristiques de l'utilisateur (Bobillier-Chaumon et Dubois, 2009). L'environnement de l'utilisateur peut être son contexte culturel tout comme son environnement professionnel. Les caractéristiques de la technologie peuvent faire référence à un système d'intelligence artificielle transparent ou opaque, ou encore à la facilité d'utilisation du système. Tandis que, pour leurs parts, les caractéristiques de l'utilisateur peuvent faire référence entre autres à l'âge ou encore aux variables de personnalité. Le poids des variables de personnalité, par exemple, peut influencer le processus d'acceptabilité (Terrade *et al.*, 2009). Un utilisateur peut être plus naturellement enclin à prendre des risques et à rechercher les sensations fortes, ce qui pourrait l'influencer favorablement à utiliser certaines nouvelles technologies (Terrade *et al.*, 2009), tandis qu'au contraire, si l'utilisateur est plus sensible à la stabilité, à la familiarité et à la sécurité, il pourrait être réfractaire à la nouveauté.

Un autre facteur qui peut influencer l'acceptabilité et ultimement l'adoption d'une technologie est la perception de l'individu par rapport à celle-ci. Von Walter, Kremmel et Jäger (2022) se sont intéressés à l'influence de la perception par rapport à l'IA sur l'adoption des conseils algorithmiques. Ils ont étudié la perception de non-experts par rapport à « l'intelligence » de l'IA en comparaison avec celle qu'ils avaient par rapport à l'intelligence humaine. Ils ont conclu que plus un individu percevait positivement les aptitudes de l'IA en regard de celles de l'humain, plus il était ouvert à écouter les conseils de l'IA. Ils ont aussi conclu que plus la recommandation algorithmique fait appel à des habiletés quantitatives et complexes, plus élevée est la confiance quant au conseil de l'IA

en comparaison avec la confiance accordée à celui de l'humain. Tandis que si les habiletés requises sont plus de l'ordre du qualitatif et de l'intuitif, la confiance penche plus du côté de l'humain.

Manikonda et Kambhampati (2018), de leur côté, ont pris le pouls des communications au sujet de l'intelligence artificielle. Ils ont analysé deux millions de « *tweets* » émis par 40 000 usagers sur la plateforme de médias sociaux *Twitter*. Ils se sont intéressés aux sentiments à l'égard de l'intelligence artificielle et ont différencié les commentaires émis par les non-experts et ceux des experts, et ceux des femmes et ceux des hommes. Leurs conclusions ont démontré que les non-experts ont un discours plus positif que les experts, que les femmes sont plus positives que les hommes au sujet de l'IA et que les femmes ont un intérêt plus marqué pour les enjeux éthiques entourant l'IA. De façon générale, les perceptions quant à l'IA sont optimistes et positives (Manikonda et Kambhampati, 2018). Ces perceptions peuvent avoir une influence sur l'acceptabilité des utilisateurs quant à des technologies d'IA qui leur sont proposées.

Selon Portales et Castañeda (2016), le concept d'acceptabilité sociale comporte trois éléments centraux qui peuvent influencer positivement une plus grande acceptabilité : la légitimité, la crédibilité et la confiance. Baron, Bernardi et Nicolini (2021) citent Baba et Raufflet (2015) en décrivant la légitimité comme étant le respect des normes établies au sein de la société. Ainsi, une technologie sera légitime si elle est en accord avec les normes sociales de la communauté au sein de laquelle elle souhaite se déployer. La crédibilité, pour sa part, émane de la mise en place d'un cadre clair et d'accords écrits (Caron, Bernardi et Nicolini, 2021). Portales et Castañeda (2016) ont étudié l'acceptabilité sociale dans le cadre d'un projet de mine en évaluant l'acceptabilité de la communauté quant au projet. Ils se sont intéressés à la Licence Sociale pour Opérer (SLO), introduite dans les années 90. La SLO, qui s'inspire de la licence légale à obtenir pour pouvoir mettre sur le marché un produit, doit être acquise pour pouvoir déployer, dans les conditions optimales, des projets délicats. La SLO est une acceptation non écrite, informelle et intangible de la communauté par rapport au projet (Portales et Castañeda, 2016). Elle comporte quatre niveaux : le rejet, l'acceptation, l'approbation et l'identification psychologique avec le projet (Thomson et Boutilier, 2011). Elle est changeante au gré des événements qui peuvent l'influencer.

La confiance est un élément soulevé par Portales et Castañeda (2016) dans le processus d'acceptabilité, et d'autres auteurs, comme Asan, Bayrak et Choudhury (2020), ainsi que Dawar (2018), la mentionnent aussi en tant que facteur clé. Elle peut être influencée par l'expérience actuelle ou passée de l'utilisateur, son éducation, ses biais, la réputation qu'il associe à la technologie et sa perception quant à l'automatisation (Asan, Bayrak et Choudhury, 2020). Le niveau de confiance à l'égard de l'intelligence artificielle influence la manière dont les utilisateurs vont s'appuyer sur ses résultats (Asan, Bayrak et Choudhury, 2020). Les propriétés du système ont une influence importante sur le niveau de confort de l'utilisateur par rapport à la technologie. La transparence, la complexité du modèle et les risques qui y sont associés sont quelques-uns des facteurs qui peuvent influencer ce niveau de confort, tout comme la fiabilité du système, c'est-à-dire la précision de ses prédictions et la constance de ses résultats (Asan, Bayrak et Choudhury, 2020). Dawar (2018) soulève lui aussi l'importance de la confiance en lien avec l'acceptabilité de l'intelligence artificielle. Tout comme Asan, Bayrak et Choudhury, il met l'accent sur l'importance d'établir un lien de confiance en étant transparent sur ses pratiques. De plus, il fait référence à la fiabilité de l'algorithme, tout particulièrement en ce qui concerne la précision des prédictions algorithmiques en lien avec les préférences des utilisateurs (Dawar, 2018). Est-ce que l'algorithme vise constamment dans le mille? Enfin, il soulève que la confiance à l'égard des plateformes d'intelligence artificielle est étroitement liée à la perception de respect de la vie privée (Dawar, 2018). Il doit donc se créer un équilibre entre la collecte de données massives nécessaire à la performance des systèmes d'intelligence artificielle et le besoin de protection des données personnelles. Pour favoriser le sentiment de confiance en lien avec le respect de la vie privée, certaines plateformes offrent aux utilisateurs la possibilité de personnaliser les paramètres en lien avec l'accès aux données personnelles (Dawar, 2018). Cette pratique permet de donner le contrôle au consommateur sur le niveau d'intrusion qu'il est à l'aise d'accepter. La confiance est donc un facteur non négligeable en lien avec l'acceptabilité de l'intelligence artificielle, et ce, sous différents angles.

Davis (1989), pour sa part, a proposé le modèle d'acceptation de la technologie TAM (*Technology Acceptance Model*), qui prend en considération des aspects proposés par un autre modèle d'intention, la théorie de l'action raisonnée (TAR). Il inclut dans son modèle

des facteurs tels que l'attitude générale de l'utilisateur face à un choix, l'influence sociale et la motivation d'un individu à se conformer à son environnement social (Bobillier-Chaumon et Dubois, 2009). L'influence sociale fait référence à un jugement de valeur sur la conduite à tenir. Il s'agit de prendre le pouls de ce que les gens font de façon majoritaire (Terrade *et al.*, 2009). Davis ajoute deux éléments centraux largement liés aux intentions des individus : l'utilité perçue et la facilité d'utilisation perçue. L'utilité perçue est le degré auquel l'individu perçoit que la technologie répond à un besoin ou améliore sa situation (Bobillier-Chaumon et Dubois, 2009). « Il s'agit de la correspondance, partielle ou totale, entre les fonctionnalités du système et les besoins actuels ou futurs de l'utilisateur » (Terrade *et al.*, 2009 : 385). La facilité d'utilisation réfère à la perception de l'individu du degré de complexité de la nouvelle technologie (Bobillier-Chaumon et Dubois, 2009).

Nous avons donc pu observer que la théorie sur l'acceptabilité de l'intelligence artificielle s'attarde à la prédisposition de l'utilisateur par rapport à cette technologie, aux caractéristiques de la technologie et aux mesures mises en place par les organisations qui développent ou déploient des systèmes d'IA. Au fil de nos nombreuses lectures, nous avons pu relever certaines initiatives qui peuvent être mises en place pour augmenter la confiance à l'égard de l'intelligence artificielle et donc possiblement influencer positivement l'acceptabilité de la technologie en lien avec les enjeux éthiques. Nous avons cerné deux grands thèmes : l'encadrement et le développement responsable.

1.8 Initiatives proposées dans la littérature pour augmenter la confiance à l'égard de la technologie

1.8.1 Encadrement : équilibre entre l'autorégulation et la réglementation stricte

Pour favoriser l'acceptabilité et l'adoption en toute confiance de l'intelligence artificielle, il faut mettre en place des balises claires. Ainsi, « [i]l sera primordial d'équilibrer le besoin d'un cadre réglementaire et le besoin d'innovation » (Da Silva *et al.*, 2021 : 9). Étant donné que le milieu carbure largement à l'expérimentation, il ne faut pas étouffer l'innovation avec un cadre réglementaire trop strict. Le travail amorcé depuis quelques années pour mettre en place des principes éthiques et développer des réglementations à différents niveaux permet de donner des lignes directrices au développement et au déploiement de

l'intelligence artificielle et de réfléchir aux enjeux qui l'entourent, comme la protection des données personnelles.

Les processus politiques et gouvernementaux sont souvent lents dans leur développement et leur déploiement, ils mettent un certain temps à aboutir et peinent à suivre la vitesse fulgurante des découvertes et des avancées survenues au cours de la dernière décennie (Baker-Brunnbauer, 2021). Un élément additionnel est alors la mise en place de codes d'éthique organisationnels pour encadrer la pratique interne. Ils peuvent s'inspirer de cadres éthiques bien établis, comme la *Déclaration de Montréal*, ou plus spécifiques à leur milieu, comme les principes de l'OMS pour le domaine de la santé. Les organisations qui mettent en place de telles initiatives bénéficient d'un avantage double selon Floridi *et al.* (2018) : non seulement l'organisation démontre sa conscience sociale et l'importance qu'elle accorde à une conception ou à une implantation éthique de l'intelligence artificielle, mais elle pourra aussi bénéficier d'économies dans son processus de conception en prévenant des erreurs potentiellement coûteuses (Baker-Brunnbauer, 2021). Cependant, pour avoir une réelle influence, ces codes d'éthique doivent être écrits, doivent être concrets, clairs dans leur application et intégrés à la culture d'entreprise, c'est-à-dire mis en action et vécus au quotidien (Mittlestadt, 2019).

À un autre niveau, Baker-Brunnbauer (2021) discute de la création d'une certification ISO pour standardiser les pratiques organisationnelles entourant l'IA. Cette certification serait bénéfique de trois façons : 1. Elle donnerait des lignes directrices claires aux organisations conceptrices, et elle standardiserait les meilleures pratiques; 2. Elle pourrait servir de guide de revue interne pour les gestionnaires au sein d'organisations conceptrices et d'entreprises en processus d'implantation; 3. La certification ISO serait un outil promotionnel important pour démontrer un standard élevé en termes de rigueur et inspirer la confiance aux utilisateurs potentiels (Baker-Brunnbauer, 2021).

Dans certains milieux, l'approbation par des entités spécialisées ou encore la certification du produit selon des critères de production et de performance sont requises pour permettre la commercialisation de certains produits. C'est notamment le cas dans le transport, l'alimentation, ou la santé. Au Canada, « le Bureau des matériels médicaux (BMM) de Santé Canada est responsable de l'autorisation de la mise en marché d'environ 35 000

dispositifs à risque moyen ou élevé et il rend environ 13 000 décisions réglementaires chaque année. » (Les Instituts de recherche en santé du Canada en collaboration avec Santé Canada, 2019) Voyant l'augmentation du nombre de produits et de services numériques offerts, le BMM a créé une division d'examen pour les technologies numériques, dont l'intelligence artificielle, pour faciliter leur accès au marché. (Les Instituts de recherche en santé du Canada en collaboration avec Santé Canada, 2019). La certification de la technologie démontre qu'elle répond aux exigences et qu'elle peut être performante dans le cadre hautement hétérogène du terrain (Amann *et al.*, 2020). La mise en place d'un processus d'approbation spécifique au monde numérique permet d'établir des normes qui déterminent le cadre à respecter pour sa commercialisation. Bannister et Connolly (2020) recommandent la création d'un organisme indépendant dont le mandat est de procéder à la certification des systèmes d'intelligence artificielle et à des audits réguliers une fois ces systèmes déployés.

Tous les systèmes d'intelligence artificielle ne requièrent cependant pas la même attention. Ainsi, une attention plus pointue devrait être donnée aux systèmes d'intelligence artificielle qui présentent le plus de risques. Pour classer le niveau de risque, des outils d'auto-évaluation sont déjà obligatoires dans le cadre de certaines réglementations, comme mentionné plus haut, dont au Canada avec l'évaluation de l'incidence algorithmique et en Europe avec le PIA.

Les initiatives réglementaires visant à encadrer les technologies numériques à risque moyen à élevé permettent à l'utilisateur de naviguer avec plus de confiance dans un environnement dans lequel il a généralement peu de littératie.

1.8.2 Développement responsable

La phase de conception est une étape importante qui peut favoriser une plus grande acceptabilité face à la technologie. Certaines initiatives peuvent renforcer le niveau de confiance et ainsi favoriser l'adoption de systèmes d'intelligence artificielle.

Le choix du modèle est un aspect important. La discussion tourne souvent autour de l'équilibre des inconvénients liés à la performance et de ceux qui sont liés à la transparence et à l'explicabilité. L'intelligence artificielle basée sur les modèles d'apprentissage

automatique ou d'apprentissage profond est reconnue comme étant beaucoup plus performante mais aussi bien plus opaque. Certains, comme Besse, Besse-Patin et Castets-Renard (2020), s'entendent pour dire que l'utilisation d'une technologie opaque dans le milieu de la santé devrait être interdite, considérant les risques liés à une technologie apprenante dont on ne peut comprendre les rouages décisionnels. Nous pourrions dire la même chose pour le milieu du transport, l'armée ou la détection d'émotions. Ceci étant dit, London (2019) et le rapport *IA et soins de santé* (Da Silva *et al.*, 2021), pour leurs parts, insistent plutôt sur l'importance de faire bénéficier à la société des avancées de la technologie lorsqu'elle peut améliorer les conditions de vie et de miser plutôt sur une transparence et une explicabilité adaptées. Ainsi, le choix du modèle, s'il est basé sur une décision responsable quant à son objectif, peut influencer positivement l'acceptabilité.

Pour augmenter la confiance par rapport à un modèle opaque, il faut mettre en place des initiatives qui viseront à protéger l'équité, la transparence et la robustesse (Asan, Bayrak et Choudhury, 2020). Qui recherche l'équité vise à minimiser les biais (Asan, Bayrak et Choudhury, 2020). Certaines mesures permettent de limiter les biais présents au sein des algorithmes. La base de données peut être une source importante de biais. Ainsi, le nettoyage des données (Asan, Bayrak et Choudhury, 2020) et une grande représentativité du bassin populationnel sont deux éléments non négligeables (Cuttillo *et al.*, 2020). La détection d'émotions est particulièrement sensible à la représentativité. Bullington (2005), Brigham (2017) et McStay et Urquhart (2019) ont soulevé le fait que la base de données qui alimente les algorithmes est largement occidentalocentrée. Cuttillo *et al.* (2020) recommandent donc d'obtenir une validation via différentes bases de données. Si, malgré tout, un biais est soulevé, une correction devrait être apportée. Faute de données représentatives disponibles, une plus grande transparence quant aux limites de l'outil permettra aux utilisateurs de mieux cibler l'utilisation qui en sera faite (Asan, Bayrak et Choudhury, 2020). Dans le milieu de la santé, les hommes blancs sont surreprésentés dans les données utilisées pour former les algorithmes (Blassime et Vayena, 2020), et c'est un facteur à prendre en compte et à corriger pour diminuer les biais potentiels.

La robustesse, pour sa part, réfère à la fiabilité du modèle, c'est-à-dire à une constance en termes de résultats (Asan, Bayrak et Choudhury, 2020). Cette rigueur permet à l'utilisateur

de s'appuyer sur les conseils algorithmiques. Cutillo *et al.* (2020) suggèrent de standardiser une échelle pour évaluer et coter la qualité des données. Ainsi, un concepteur pourrait rendre publique la cote de qualité de sa base de données comme élément pour démontrer la robustesse de son système d'intelligence artificielle, ce qui pourrait en favoriser l'acceptabilité (Cutillo *et al.*, 2020). Landowska (2019) et Besse, Besse-Patin et Castets-Renard (2020) se penchent sur une façon de traiter l'enjeu de la transparence en lien avec la fiabilité des systèmes. Comme nous l'avons dit, en ce qui concerne l'intelligence artificielle et plus spécifiquement l'intelligence artificielle affective, les algorithmes ne sont pas toujours très performants, ayant des taux d'erreurs variables. Considérant que certaines décisions peuvent être prises en lien avec les prédictions issues des systèmes d'IA, il serait capital de rendre public la marge d'erreur ou le taux de confiance (Landowska, 2019; Besse, Besse-Patin et Castets-Renard, 2020). Une telle pratique standardisée serait un indicateur de confiance pour les utilisateurs. De plus, les concepteurs devraient laisser savoir aux utilisateurs potentiels dans quel cadre leur technologie performe le mieux (Landowska, 2019). La détection d'émotions est sensible à l'angle du visage et à la luminosité, donc faire connaître certains critères pour optimiser la performance serait une belle valeur ajoutée (Landowska, 2019).

L'équipe de conception est aussi un facteur à considérer pour influencer positivement l'adoption de l'intelligence artificielle. Comme nous l'avons exposé, les biais des concepteurs peuvent s'infiltrer par mégarde dans l'algorithme lors de la conception (Gibert, 2020; Coeckelbergh, 2020). Pour diminuer l'incidence de biais instillés par le concepteur, une plus grande diversité au sein de l'équipe de conception peut aider à réduire au minimum cet enjeu (Dignum, 2020).

Pour augmenter la confiance des utilisateurs, les concepteurs devraient les impliquer dans le processus de conception, ce qui leur permettrait, en plus, de profiter de l'expertise terrain de ceux-ci, de bien cerner les enjeux propres à leurs milieux et de faciliter l'adoption de la technologie (Cutillo *et al.*, 2020; Ahmad, Eckert et Teredesai, 2018). De cette manière, ils pourraient aussi mieux comprendre les besoins en termes d'explicabilité spécifique au contexte professionnel de l'utilisateur (Ahmad, Eckert et Teredesai, 2018). Les concepteurs

doivent prendre le pouls du milieu et pouvoir coder ses valeurs et ses besoins spécifiques (Amann *et al.*, 2020).

Mittlestadt (2019), pour sa part, apporte une suggestion qui n'a pas beaucoup de vent dans les voiles, mais qui soulève un aspect éthique valable. Les concepteurs dans le milieu de l'intelligence artificielle sont principalement issus du secteur privé et donc du milieu des affaires (Mittelstadt, 2019). Les objectifs des gestionnaires sont généralement la croissance, l'optimisation des processus et non le coût social (Baker-Brunnbauer, 2020). Le concepteur a une responsabilité morale à l'égard de la société lorsqu'il développe des systèmes à haut risque, comme c'est le cas pour le domaine de la santé, la surveillance ou l'armée. Le concepteur doit se sentir libre de mettre en place les meilleures pratiques éthiques, il doit pouvoir agir sans crainte de représailles (Mittelstadt, 2019). Le souci éthique ici est la pression de l'entreprise qui peut peser sur l'équipe de conception. Mittlestadt (2019) suggère de créer une licence et une profession formelle pour les concepteurs de systèmes d'intelligence artificielle dont la portée représente de plus grands risques pour la société. De la même façon, Cathy O'Neil (2016) et Richardson (2020) recommandent qu'ils s'engagent à respecter un code de conduite tout comme les médecins font le serment d'Hippocrate. Il serait alors permis de penser qu'une plus grande indépendance de l'équipe de conception par rapport aux enjeux de l'entreprise pourrait influencer positivement l'acceptabilité.

Enfin, un suivi soutenu postconception par l'entremise, par exemple, de reproductions des résultats sur de nouveaux échantillons à intervalle régulier (Cuttillo *et al.*, 2020), ainsi que des audits pour assurer la fiabilité continue du système et examiner l'apparition potentielle de biais (Maclure et Saint-Pierre, 2018), est un facteur important. De plus, une analyse systématique à la suite d'un incident est capitale pour mieux comprendre ce qui l'a causé et mettre en place les corrections nécessaires pour prévenir d'autres incidents dans le futur (Shneiderman, 2016).

La meilleure offensive est une bonne défensive. En d'autres termes, mettre en place les outils et les processus nécessaires pour prévenir les erreurs et les incidents liés à l'intelligence artificielle fait partie d'un développement responsable.

L'intelligence artificielle, et plus précisément l'intelligence artificielle affective, est relativement jeune. L'ampleur que prendront son développement et son déploiement reste encore un objet de spéculations, mais est largement prometteuse. De nombreuses disciplines et de nombreux experts s'y intéressent et l'étudient sous différents angles. Au fil de nos recherches, nous avons aussi tenté de déterminer ce qui se dégage de la perception de la population générale puisqu'elle est la plus répandue dans l'utilisation de la technologie en termes de proportion. Nous avons cependant remarqué que la littérature est beaucoup moins étoffée en ce qui concerne plus spécifiquement la perception de la population générale quant aux enjeux que la technologie soulève et les conditions qu'elle juge acceptables pour l'utiliser. Nous avons donc formulé la question suivante :

Quels sont les enjeux éthiques liés au développement et au déploiement d'un système d'intelligence artificielle affective et quelles en sont les conditions d'acceptabilité?

Pour explorer la question, nous avons recruté des participants à qui nous avons fait utiliser une technologie de détection des émotions pour une période donnée. Par la suite, nous avons passé en entrevue chaque participant pour recueillir plus d'information sur leur perception quant à l'intelligence artificielle, aux nouvelles technologies, aux bénéfices et aux dangers qu'elle soulève à leurs yeux, et nous avons pris le pouls de leur expérience en lien avec l'intelligence artificielle affective et sondé leur sentiment quant à l'acceptabilité d'un tel outil de détection des émotions. Nous avons développé un cadre d'analyse sur deux volets, que nous vous présenterons dans la section suivante.

2 Cadre d'analyse

Afin d'explorer les enjeux éthiques et l'acceptabilité d'une technologie d'IA affective, un sujet en pleine évolution, nous avons choisi de travailler avec un cadre d'analyse flexible et empli de nuances. Comme présenté dans le chapitre précédent, au cours des dernières années, de nombreuses initiatives en éthique de l'intelligence artificielle ont vu le jour. Elles émanent de travaux locaux, nationaux ou internationaux, et ce, dans différentes sphères, que ce soit aux niveaux politique, académique ou issues des milieux professionnels ou corporatifs. Pour notre part, nous avons choisi de mobiliser la Déclaration de Montréal pour notre cadre d'analyse. La Déclaration est le fruit d'une grande consultation auprès de la population et d'experts du milieu, dont entre autres, Yoshua Bengio une sommité de l'apprentissage automatique. Ce cadre est pertinent pour notre travail de recherche considérant la reconnaissance qu'il a récolté au niveau de la communauté internationale, étant cité dans de nombreux travaux. De plus, il est adapté au contexte canadien ayant été élaboré ici même à Montréal. Nous avons donc jugé opportun de l'utiliser comme cadre de référence pour notre projet. Nous nous sommes d'ailleurs inspirées de la Déclaration pour alimenter notre canevas d'entrevue.

Pour analyser les conditions d'acceptabilité de l'intelligence artificielle et plus spécifiquement l'informatique affective, nous avons choisi de proposer notre propre modèle en nous basant sur la théorie soulevée dans la littérature. Notre objectif est de valider sur le terrain si les critères retenus représentent bien les facteurs qui influencent l'acceptabilité des utilisateurs et comment la technologie d'EmoScienS répond à ces critères.

Il est opportun de décrire plus en détail les éléments qui composeront notre cadre d'analyse. Ainsi nous aborderons de façon sommaire les objectifs et principes de la Déclaration de Montréal ainsi que les éléments retenus qui sont spécifiques à l'acceptabilité dans les sous-sections ci-dessous.

2.1 Déclaration de Montréal

La Déclaration de Montréal, un projet scientifique de l'Université de Montréal, est le résultat de travaux de co-constructions rassemblant plus de 500 citoyens, experts et parties prenantes de différents horizons qui ont pris part à une quinzaine d'ateliers, de journées de co-constructions et de tables rondes sur une période de 3 mois s'échelonnant de février à avril 2018 (Université de Montréal, 2018).

La Déclaration de Montréal est une œuvre collective qui a pour objectif de mettre le développement de l'IA au service du bien-être de tout un chacun, et d'orienter le changement social en élaborant des recommandations ayant une forte légitimité démocratique. (Université de Montréal, 2018)

Les avancées ont été majeures dans le milieu de l'intelligence artificielle dans les dernières années, tout particulièrement en matière d'apprentissage automatique. Il y a cependant encore peu de connaissances sur les impacts sociaux de ces nouvelles technologies. Ainsi les chercheurs de l'Université de Montréal ont jugé qu'un moment de réflexion éthique, politique, juridique, sociologique et psychologique s'imposait.

2.1.1 Objectifs de la Déclaration de Montréal

Les chercheurs avaient trois grands objectifs que nous reprenons et présentons dans le tableau ci-dessous :

Tableau 2.1 Objectifs de la Déclaration de Montréal

Élaborer un cadre éthique pour le développement et le déploiement de l'IA.
Orienter la transition numérique afin que tous puissent bénéficier de cette révolution technologique.
Ouvrir un espace de dialogue national et international pour réussir collectivement un développement inclusif, équitable et écologiquement soutenable de l'IA.

(Université de Montréal, 2018)

2.1.2 Principes de la Déclaration de Montréal

La Déclaration de Montréal contient 10 principes. Ils ne sont pas placés en ordre d'importance et un principe n'a pas préséance sur un autre. La Déclaration est un outil visant à ouvrir la discussion sur l'éthique et le développement responsable de l'intelligence

artificielle, les auteurs sont conscients des nuances, tout particulièrement au niveau international considérant les différences sociales et culturelles.

Tableau 2.2 Principes de la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*

Déclaration de Montréal IA responsable
1. Bien-être
2. Respect de l'autonomie
3. Protection intimité et vie privée
4. Solidarité
5. Participation démocratique
6. Équité
7. Inclusion de la diversité
8. Prudence
9. Responsabilité
10. Développement soutenable

(Université de Montréal, 2018)

Principe de bien-être

« Le développement et l'utilisation des systèmes d'intelligence artificielle (SIA) doivent permettre d'accroître le bien-être de tous les êtres sensibles. » (Université de Montréal, 2018)

Ce principe mise non seulement sur l'élaboration mais aussi l'utilisation de SIA dans l'optique d'améliorer le bien-être et la qualité de vie que ce soit en termes de conditions de vie, de santé ou de conditions de travail. Il est spécifié dans la Déclaration que « L'utilisation des SIA ne devrait pas contribuer à augmenter le stress, l'anxiété et le sentiment de harcèlement liés à l'environnement numérique. » (Université de Montréal, 2018). L'intelligence artificielle ne devrait pas non plus être une source de mal-être en créant des pressions indues de toute sorte sur les utilisateurs.

Principe de respect de l'autonomie

« Les SIA doivent être développés et utilisés dans le respect de l'autonomie des personnes et dans le but d'accroître le contrôle des individus sur leur vie et leur environnement. » (Université de Montréal, 2018)

L'intelligence artificielle ne doit pas enfreindre l'autodétermination des individus. Elle doit s'assurer au contraire de donner aux utilisateurs plus de contrôle de façon générale dans tous les aspects de leur vie. Les SIA doivent permettre de s'accomplir selon ses propres objectifs. Il ne faut pas que l'intelligence artificielle pave la voie à suivre en influençant les individus par des mécanismes d'incitation ou de surveillance. Le développement des SIA devrait prendre en considération et éviter les aspects de la technologie qui pourraient mener à de la dépendance. Enfin, l'autonomie passe par l'éducation face aux nouvelles technologies pour permettre aux utilisateurs d'être aiguisés en la matière et ainsi être maîtres de leurs choix et décisions.

Principe de protection de l'intimité et de la vie privée

« La vie privée et l'intimité doivent être protégées de l'intrusion de SIA et de systèmes d'acquisition et d'archivage des données personnelles (SAAD). » (Université de Montréal, 2018)

La ligne est parfois mince en ce qui concerne la protection de l'intimité et de la vie privée, ainsi la Déclaration a bien détaillé cette section bien qu'elle laisse toujours place à une évolution au sein de la société. L'intelligence artificielle peut être intrusive alors il est important de s'outiller pour déterminer d'un cadre social acceptable et des limites à ne pas franchir. Ainsi, il faut protéger entre autres la pensée, les émotions, l'accès aux données personnelles et la connectivité dans la sphère privée. Le texte souligne l'importance de protéger la pensée et les émotions envers les SIA qui pourraient causer du tort. De plus il spécifie que les utilisateurs devraient toujours avoir le choix de leur niveau de connectivité au monde numérique dans leur vie privée. Aussi, la collecte, l'usage et la dissémination des données personnelles comme entre autres les informations relatives aux préférences des individus devraient être soumis à un consentement libre et éclairé et l'utilisateur devrait toujours en avoir le contrôle. Enfin, il mentionne que l'accès au monde numérique ne devrait pas être conditionnel à l'abandon de la propriété de ses données personnelles. De

façon sommaire, le principe de protection de l'intimité et de la vie privée vise à maintenir le contrôle entre les mains des utilisateurs face à toutes formes d'intrusion ou de conditions excessives au niveau de leurs données personnelles, connectivité au monde numérique, émotions et pensées.

Principe de solidarité

« Le développement de SIA doit être compatible avec le maintien de liens de solidarité entre les personnes et les générations. » (Université de Montréal, 2018)

Ce principe vise à mettre en lumière l'importance des relations humaines de toutes sortes. Les SIA devraient favoriser le développement et le maintien des relations entre les humains et veiller à réduire l'isolement. De la même façon, ils devraient encourager la collaboration entre les humains et les machines et celles entre humains. Un SIA ne devrait pas être développé ou utilisé dans le but de remplacer un humain dans une tâche qui est sensible à une relation de qualité, il devrait plutôt miser sur la complémentarité ou le support. Ainsi, le développement des SIA devrait se faire dans l'esprit d'accroître les interactions propices aux développements de relations humaines.

Principe de participation démocratique

« Les SIA doivent satisfaire les critères d'intelligibilité, de justifiabilité et d'accessibilité, et doivent pouvoir être soumis à un examen, un débat et un contrôle démocratique. » (Université de Montréal, 2018)

Le principe de participation démocratique fait référence entre autres aux enjeux d'explicabilité et de transparence. Ainsi, un concepteur doit être capable d'expliquer le fonctionnement de son SIA, tout comme les algorithmes doivent être accessibles aux instances nécessaires et que l'utilisateur doit pouvoir comprendre le raisonnement d'une décision algorithmique qui l'affecte personnellement. De la même façon, un utilisateur doit savoir lorsqu'il interagit avec un SIA et doit être informé lorsqu'une décision le concernant ou l'affectant a été prise par un tel système. La transparence est un facteur important et englobe de façon très large le développement et l'utilisation de l'intelligence artificielle, elle appelle à une vigilance et proactivité de la part des concepteurs.

Principe d'équité

« Le développement et l'utilisation des SIA doivent contribuer à la réalisation d'une société juste et équitable. » (Université de Montréal, 2018)

La protection contre la discrimination et une mission claire pour une justice pour tous sont l'essence du principe d'équité. Il s'agit d'encadrer les pratiques en ce qui concerne, d'une part, le développement libre de biais dans la conception et formation d'algorithmes et d'autre part, l'utilisation juste et équitable des SIA. Le développement et l'utilisation de l'intelligence artificielle doit viser une accessibilité à tous, veiller à la réduction d'inégalités et assurer des conditions de travail décentes tout au long du continuum de son développement.

Principe d'inclusion de la diversité

« Le développement et l'utilisation de SIA doivent être compatibles avec le maintien de la diversité sociale et culturelle et ne doivent pas restreindre l'éventail des choix de vie et des expériences personnelles. » (Université de Montréal, 2018)

Le principe vise à faire la promotion de la diversité sociale et culturelle au niveau du développement et de l'utilisation de SIA, et ce, autant au stade de conception des algorithmes qu'en matière de représentativité au sein des équipes de recherche ou de l'industrie. Les SIA doivent permettre la libre expression de la diversité d'opinions et doivent veiller à ne pas enfermer les utilisateurs dans un profil bien précis, ce qui pourrait leur limiter l'accès à une variété d'informations.

Principe de prudence

« Toutes les personnes impliquées dans le développement des SIA doivent faire preuve de prudence en anticipant autant que possible les conséquences néfastes de l'utilisation des SIA et en prenant des mesures appropriées pour les éviter. » (Université de Montréal, 2018)

Bien que le développement et l'utilisation d'une technologie se fassent avec de bonnes intentions, il est capital de se questionner et de prévoir les effets néfastes que cette dernière peut avoir entre de mauvaises mains ou dans une situation où le SIA venait qu'à faire défaut. Des mesures devraient être mises en place pour réduire les risques au maximum ou restreindre son utilisation si nécessaire. Ainsi, avant d'être mis sur le marché un SIA devrait

être scrupuleusement examiné, évalué et testé pour en assurer une performance optimale et la plus grande sécurité. La transparence est de mise face aux enjeux de sécurité d'un SIA qui présente un danger au niveau individuel ou social. Il va de soi que la protection des données personnelles est un incontournable pour éviter un usage malintentionné ou les abus.

Principe de responsabilité

« Le développement et l'utilisation des SIA ne doivent pas contribuer à une déresponsabilisation des êtres humains quand une décision doit être prise. » (Université de Montréal, 2018)

De façon générale, la responsabilité ultime reste et doit demeurer avec l'humain. L'utilisation d'un SIA dans le but de commettre un crime ou de tuer ne déresponsabilise en aucun temps l'humain face au geste posé. Il reste maître de la décision et en est responsable. De la même façon lorsqu'un SIA est utilisé dans le cadre d'une décision qui a des effets sur la qualité de vie ou réputation d'une personne, la décision finale devrait rester entre les mains de l'humain.

Principe de développement soutenable

Considérant que nous n'avons pas exploré les enjeux du développement durable en lien avec l'intelligence artificielle, nous n'utiliserons pas ce principe dans le cadre de notre analyse. Nous avons cependant tenu à l'inclure dans cette section pour y exposer le lecteur.

« Le développement et l'utilisation de SIA doivent se réaliser de manière à assurer une soutenabilité écologique forte de la planète. » (Université de Montréal, 2018)

Le monde du numérique ne fait pas exception et il contribue à la pollution de l'environnement. Ainsi, il est important que l'industrie se commette à mettre les efforts pour réduire au maximum son empreinte écologique et participe par des gestes concrets à une planète en santé pour les générations à venir.

Notre cadre d'analyse reprendra ces principes pour valider sur le terrain quels sont les enjeux éthiques importants aux yeux des utilisateurs et comment ils s'appliquent à l'évaluation d'une technologie d'intelligence artificielle affective. Nous sommes

conscients qu'il existe un plus grand éventail de chartes, de déclaration et de cadre éthique, cependant nous croyons qu'il est le plus approprié dans notre contexte.

2.2 Conditions d'acceptabilité

Pour mieux comprendre les gestes à poser, les pratiques, politiques et mécanismes à mettre en place de façon concrète pour le développement et déploiement d'une technologie éthique et responsable, il faut aussi investiguer les conditions sensibles à l'utilisateur qui en influencent l'acceptabilité. Ainsi, il faut aussi comprendre les conditions gagnantes pour une IA socialement acceptable aux yeux des utilisateurs. Suite à la revue de la littérature, il est clair que de nombreux éléments peuvent influencer l'acceptabilité d'un individu face à une nouvelle technologie.

Dans la section 2.7 sur les conditions d'acceptabilité, nous avons abordé certains facteurs qui ont été liés à l'acceptabilité par différents chercheurs. Nous avons abordé les caractéristiques de l'utilisateur telles que ses traits de personnalité (Terrade *et al.* 2009). La perception de l'individu face à la technologie aura elle aussi une influence directe sur l'acceptabilité de l'utilisateur envers l'outil d'IA proposé. Von Walter, Kremmel et Jäger (2021) ont démontré que lorsque l'utilisateur perçoit les aptitudes de l'IA favorablement face à celles de l'humain, il aura une plus grande tendance à se rallier derrière ses résultats.

À des fins d'analyse pour notre travail de recherche, nous avons choisi de mobiliser des éléments mis de l'avant par différents modèles en lien avec l'acceptabilité de technologies et proposons le cadre d'analyse suivant :

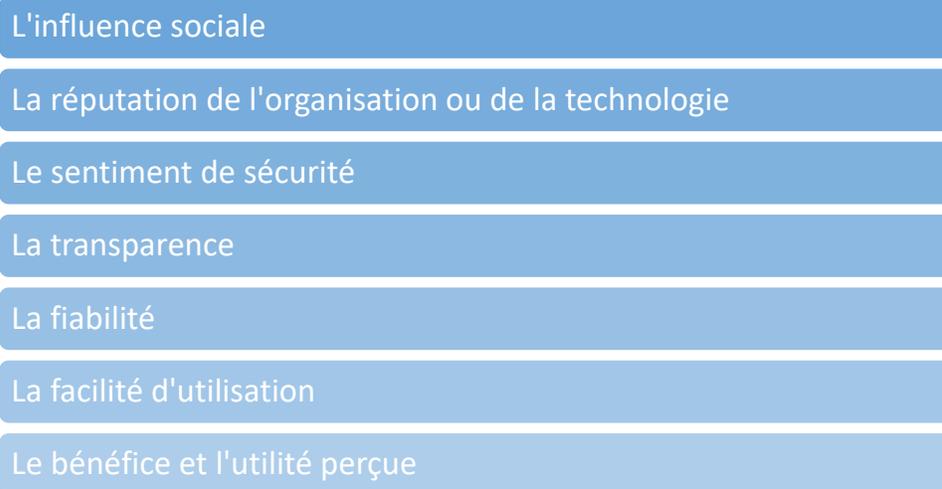


Figure 2.1 Critères d'acceptabilité pour un SIA

L'influence sociale issue de la théorie de l'intention raisonnée (TAR) fait référence à l'influence qu'a l'environnement social sur le choix de l'utilisateur. Si la technologie est largement utilisée ou si une personne de confiance l'a déjà adoptée, l'utilisateur sera influencé en conséquence à accepter la nouvelle technologie et conforté à aller de l'avant avec son utilisation.

La réputation de l'organisation ou de la technologie est un aspect soulevé entre autres par Asan *et al.* (2020). La réputation de l'organisation derrière une nouvelle technologie d'intelligence artificielle ou la réputation d'une certaine forme de technologie comme l'informatique affective peut influencer l'acceptabilité de l'utilisateur. Certaines compagnies qui sont reconnues comme étant très éthiques auront plus facilement la confiance du public qu'une organisation dont la réputation a été entachée par un scandale quelconque ou qui n'est pas du tout connu du public. Une nouvelle technologie bénéficie d'être déployée par une compagnie connue et respectée.

Le sentiment face à la sécurité de la technologie en lien avec l'enjeu de la protection de la vie privée et des données personnelles a été souligné entre autres par Dawar (2018). La démonstration que la plateforme ou que le SIA est sécuritaire est un facteur qui influence l'acceptabilité d'une nouvelle technologie en intelligence artificielle. Une organisation qui a des standards élevés en termes de mécanismes de protection des données personnelles,

de la confidentialité, et de la vie privée offre un cadre sécuritaire à l'utilisateur. Aussi, lorsqu'une organisation offre à l'utilisateur l'option de personnaliser facilement ses paramètres de partage de données personnelles, c'est un facteur qui joue favorablement sur le sentiment de protection de la vie privée et de confidentialité.

La transparence est un élément mentionné par Asan *et al.* (2020) et Dawar (2018). Une organisation qui partage ses valeurs, la façon dont elle utilise les données personnelles des utilisateurs, le fonctionnement de sa technologie et les risques qui y sont associés influence favorablement l'acceptabilité envers sa technologie.

La fiabilité est un facteur soulevé par Asan *et al.* (2020) et Dawar (2018). Un SIA dont la précision et la constance des prédictions sont élevées insuffle un niveau de confiance envers la technologie et permet à l'utilisateur de s'y fier avec assurance. Dawar (2018) ajoute l'élément de précision de la prédiction en lien avec les préférences de l'utilisateur. Ainsi, plus une plateforme offre des options personnalisées juste à ses consommateurs, plus ses utilisateurs en verront la pertinence. Plus grande est la fiabilité de la prédiction de l'algorithme aux yeux de l'utilisateur, plus on acceptabilité sera facilitée.

La facilité d'utilisation et d'intégration au mode de vie est un facteur qui peut peser dans un contexte organisationnel tout comme personnel. Si la technologie requiert une longue courbe d'apprentissage et ajoute du temps de gestion dans une journée déjà bien remplie, l'utilisateur risque d'avoir une plus grande résistance à l'intégrer à son environnement et à sa routine. L'utilisateur doit sentir qu'il a les compétences requises pour l'adopter et que la technologie demandera très peu d'effort de sa part.

Les bénéfices et l'utilité perçus font référence à l'adéquation de la technologie à répondre aux besoins immédiats et futurs de l'utilisateur. Il doit percevoir que la technologie lui permettra d'améliorer sa situation. Lorsque l'utilisateur décide d'adopter un nouveau système, il fera le balancier entre le coût en terme d'effort, de temps et possiblement d'argent et les bénéfices qu'il retirera de l'utilisation de la technologie. En organisation, l'employé doit percevoir un bénéfice personnel à l'adoption du système et non un bénéfice pour l'entreprise.

Dans le cadre de notre recherche, nous voulons investiguer si ces facteurs sont au cœur des réflexions d'acceptabilité en lien avec une nouvelle technologie et comment la technologie d'EmoScienS répond à ces critères.

3 Méthodologie

Les entreprises qui se lancent dans le milieu de l'intelligence artificielle doivent composer avec un environnement qui évolue rapidement. N'ayant que très peu de points de repère bien établis il peut être difficile de s'y retrouver. Cette étude a donc 3 principaux objectifs :

- Faire une revue des enjeux éthiques liés au développement et au déploiement d'un système d'intelligence artificielle affective et mettre en lumière les enjeux sensibles pour les utilisateurs;
- Comprendre quels sont les critères qui influencent l'acceptabilité face aux systèmes d'intelligence artificielle affective;
- Explorer les pistes de solution pour augmenter l'acceptabilité face à la technologie mobilisant l'informatique affective.

Considérant que l'intelligence artificielle a connu un essor important au cours de la dernière décennie, tout particulièrement au niveau des techniques d'apprentissage automatique, l'essor de technologies comme l'informatique affective offre un terrain de recherche foisonnant. Pour ne pas nous limiter et permettre de dégager des apprentissages sur le sujet nous avons opté pour la méthode de recherche qualitative de type exploratoire.

3.1 Stratégie de recherche

En prenant en compte nos objectifs de recherche et le contexte dans lequel évolue notre problématique, nous avons choisi de procéder par étude de cas d'une technologie d'intelligence artificielle mobilisant l'informatique affective, développée par l'entreprise en démarrage québécoise EmoScienS. Comme l'a définie Yin, l'étude de cas est une méthode empirique qui investigate un phénomène contemporain en profondeur dans un contexte non-contrôlé et réel, elle est tout particulièrement de mise quand la ligne entre le phénomène et le contexte est floue (Yin, 2018 traduction libre : 15).

3.1.1 Justifications de la méthodologie

Nous avons fait une revue de littérature en profondeur sur la technologie d'intelligence artificielle et plus spécifiquement sur l'informatique affective. Nous avons utilisé des mots

clés telles que intelligence artificielle, intelligence artificielle affective, informatique affective, apprentissage automatique, enjeux éthiques, acceptabilité informatique et fait des recherches sur les bases de données telles que Google Scholar et les nombreuses bases de données via EBSCOhost. Nous avons aussi investigué les références pertinentes dans les différents textes que nous avons revus. Voulant bénéficier d'une méthode de recherche qui nous permette une investigation large du terrain, nous avons opté pour la méthode qualitative via l'étude de cas de type exploratoire. Nous visions à décrire et à explorer le terrain pour en extraire des réflexions sur un phénomène émergent, ainsi que des pistes de recommandations et de développement. L'étude de cas est de mise lorsque la question tente de répondre aux « comment » ou « pourquoi », lorsque nous avons peu ou pas de contrôle sur les événements et lorsque l'accent de l'étude est un phénomène contemporain et non historique (Yin, 2018 voir si traduction libre p13). Enfin, nous avons choisi de faire une étude de cas unique et non à cas multiples. L'étude de cas unique peut être un choix judicieux si notre cas représente une des situations suivantes : le cas représente une situation (1) critique; (2) exceptionnelle ou rare; (3) commune; (4) révélatrice; ou encore (5) longitudinale (Yin, 2018). Notre cas représente une situation commune. Il met les participants en situation d'acceptabilité face à une nouvelle technologie à utiliser. Une situation qui n'est pas exceptionnelle de nos jours, puisqu'il n'est pas inhabituel d'utiliser une nouvelle plateforme, une nouvelle technologie, un nouveau logiciel dans le contexte actuel. De plus, le participant utilise une application de détection des émotions qui présente des caractéristiques standards pour une technologie d'intelligence artificielle affective. L'objectif d'utiliser un tel cas, selon Yin, est de capturer les circonstances et conditions d'une situation de tous les jours pour les leçons qu'elle pourrait nous donner (Yin, 2018).

3.2 Sélection du cas et collecte de données

3.2.1 Le cas : La technologie de détection des émotions d'EmoSciens

Dans le cadre de notre travail de recherche terrain, nous avons eu l'opportunité de faire tester à nos participants la technologie d'intelligence artificielle affective de l'entreprise en

démarrage montréalaise EmoScienS. Fondée par 3 scientifiques en novembre 2019, la compagnie a pour mission « d'améliorer la santé émotionnelle dans les organisations, de façon éthique et bienveillante » (www.emosciens.com). Les 3 cofondateurs ont combiné leurs expertises pour développer une technologie permettant à l'utilisateur de gérer son bien-être mental. Pierrich Plusquellec est professeur à l'École de psychoéducation de l'Université de Montréal, codirecteur du Centre d'études sur le stress humain et cofondateur du Centre d'études en sciences de la communication non verbale. Nathe François est spécialiste du transfert de connaissances dans le domaine de la santé mentale et pour sa part, Ted Hill est un expert dans le domaine de la reconnaissance automatique d'émotions. La technologie d'EmoScienS mobilise l'informatique affective afin de monitorer l'état émotionnel des utilisateurs et de leur offrir un « *coach émotionnel* » numérique.

Le logiciel d'EmoScienS prend des photos des utilisateurs à intervalles réguliers via la webcam. Les expressions faciales sur les photos sont analysées par un algorithme d'intelligence artificielle affective et compilées pour dresser un tableau de bord émotionnel journalier. Les données ne sont visibles que par l'utilisateur, à moins que ce dernier ne souhaite partager l'information (par exemple avec son gestionnaire, lors d'un déploiement en entreprise). Dans un tel cas, elle peut être dépersonnalisée et partagée. Les utilisateurs ont non seulement accès à leur tableau de bord mais aussi à des outils pour mieux gérer leurs émotions. La technologie est accompagnée d'un portail de ressources pour permettre non seulement de bénéficier d'un portrait émotionnel mais aussi permettre la mise en place de mesures concrètes pour agir sur les résultats.

L'entreprise vise à utiliser la technologie pour donner à l'utilisateur le contrôle sur ses données : « Parce que VOS données doivent d'abord et avant tout VOUS servir. » (www.emosciens.com). L'utilisateur a accès aux photos qui sont partagées avec le serveur d'EmoScienS et il en a le contrôle. D'ailleurs la compagnie refuse de se mouler au modèle actuel qui veut que les données numériques soient monnayables. Les données sont utilisées pour et par les utilisateurs. EmoScienS a à cœur l'éthique et la responsabilité sociale. D'ailleurs, les photos sont facilement accessibles et peuvent être gérées par l'utilisateur. En mode actif d'enregistrement, chaque photo prise est à la vue de l'utilisateur. Elle

disparaît de l'écran une fois qu'une nouvelle expression faciale est captée. De plus, EmoScienS offre la gestion personnalisée des témoins (*cookies*) en permettant à l'utilisateur de tout refuser, de tout accepter ou de personnaliser sa sélection selon ses préférences pour activer ou désactiver les cookies marketing, fonctionnels et analytiques. L'information est accessible facilement et courte à lire.

Le marché visé par la technologie est le milieu organisationnel. L'entreprise qui l'adopte peut alors l'offrir à ses employés et prendre le pouls de son équipe. Avec l'autorisation de ses employés, l'organisation peut avoir accès aux données dépersonnalisées et agglomérées au sein d'un groupe de 10 travailleurs. L'objectif est de donner aux gestionnaires une mesure du climat émotionnel. Cet outil permet de voir l'impact de différentes initiatives organisationnelles sur le moral des troupes.

La technologie peut être utilisée sur MAC ou PC sous la forme d'un logiciel. De cette façon, la prise de photo est constante à intervalle régulier, quelle que soit l'activité de l'utilisateur. La version logiciel offre l'option de préprogrammer l'horaire d'activation de la technologie ce qui en simplifie l'utilisation régulière. Le logiciel est compatible avec les plateformes de vidéoconférence telles que Zoom, Teams ou Google Meet. Pour les utilisateurs qui ne peuvent pas utiliser la version logicielle, EmoScienS est aussi disponible via une plateforme Web. Cependant, cette version ne permet pas la programmation d'un horaire d'utilisation ou la prise de photo lorsque l'utilisateur est en vidéoconférence sur son PC ou son MAC. Cette version plus limitée ne peut donc pas prendre en compte toute l'expérience émotionnelle de l'utilisateur si une partie de sa journée se déroule en vidéoconférence.

Les participants à notre recherche ont utilisé la technologie pour une période de deux semaines. L'objectif était de donner une expérience équivalente à tous les participants avec l'utilisation d'une technologie d'intelligence artificielle affective durant une période circonscrite. Les participants n'ont pas eu accès au portail de ressources ou à des outils pour compléter l'information détaillée dans leur tableau de bord.

Suite à une période test équivalente pour l'ensemble des participants, nous pouvions alors recueillir la perception de chacun sur les enjeux éthiques que soulève l'IA, ainsi que sur les conditions d'acceptabilité d'une telle technologie d'informatique affective. Ainsi, cette

utilisation terrain de la technologie visait aussi à cerner pourquoi les participants avaient accepté d'utiliser une telle technologie et qu'est-ce qui les motive généralement à accepter d'utiliser des nouveautés dans le monde du numérique.

3.2.2 Collecte de données

Nous avons mobilisé pour cette recherche une méthodologie qualitative, réalisant des entretiens semi-structurés avec des utilisateurs de la technologie étudiée. La collecte de données s'est déroulée sur une période de 3 semaines soit du 26 février au 17 mars 2022.

Les entretiens semi-dirigés ont eu lieu en mode virtuel. Elles ont eu une durée de 45 à 65 minutes. Pour guider les rencontres, un canevas d'entretien (annexe A) a été développé. Ce canevas a subi une itération suite à la réalisation des deux premières entretiens (annexe B) et avons utilisé la version amendée pour le reste des entretiens. Nous avons utilisé un échantillonnage de convenance, recrutant 15 participants qui ont pris part au processus. Le choix des répondants s'est fait de façon non aléatoire, donc par sélection pour obtenir une certaine représentativité de la population en termes de sexe, groupe d'âge et variété d'occupations. L'échantillon d'une recherche qualitative varie en général entre 7 et 12 participants. Il s'agit de recueillir de l'information pouvant aider à mieux comprendre un phénomène (Mongeau, 2011). Un échantillon de qualité atteint une saturation, c'est-à-dire un point où l'information ajoutée par les participants additionnels ne vient que confirmer ce que les participants antérieurs ont partagé et ainsi, les informations additionnelles n'apportent plus de valeur ajoutée (Mongeau, 2011).

Le recrutement de participants a été complexifié pour 2 raisons. Premièrement, la technologie n'est pas adaptée aux téléphones ou aux tablettes intelligentes. Elle requiert un portable ou ordinateur Mac ou PC. Ainsi, cette particularité a limité le bassin populationnel disponible. Deuxièmement, la technologie requiert l'accès à la caméra de l'appareil. Les individus qui travaillent sur un appareil corporatif ont dû demander l'approbation de leur entreprise pour pouvoir participer au projet de recherche. Considérant l'aspect invasif de la technologie, certains de ces participants n'ont pas obtenu l'approbation nécessaire. De plus, la technologie peut être utilisée sous deux différentes formes, via la plateforme logicielle ou la version internet. La version internet n'est pas compatible avec les plateformes de rencontres virtuelles comme Zoom, Teams ou Google Meet, cet enjeu

complexifiant le recrutement dû au temps considérable en rencontres virtuelles durant la période de pandémie de Covid. Enfin, une participante qui a connu des ennuis techniques a dû se retirer du projet de recherche, réduisant à 14 participants au projet de recherche dans son entièreté.

Pour assurer la confidentialité des participants, nous avons anonymisé la liste de participants. Le tableau 3.1 présente les caractéristiques de notre échantillon.

Tableau 3.1 Caractéristiques de l'échantillon

Nom	Sexe	Âge	Occupation	Éducation complétée
Participante A	Femme	45-49 ans	Gestionnaire de produit Enseignante, université	Maîtrise en recherche
Participante B	Femme	35-39 ans	Designer d'intérieur Entrepreneure	Attestation d'études collégiales
Participante C	Femme	50-54 ans	Étudiante (Psychologie)	Maîtrise
Participante D	Femme	45-49 ans	Directrice d'école	DESS
Participante E	Femme	30-34 ans	Directrice, département comptabilité conseil	Maîtrise, comptabilité publique
Participante F	Femme	25-29 ans	Étudiante (Maîtrise) Assistante de recherche	Maîtrise, finance
Participante G	Femme	25-29 ans	Directrice en stratégie, firme-conseil	Maîtrise, stratégie
Participante H	Femme	20-24 ans	Étudiante (Bacc. Enseignement)	Diplôme études collégiales
Participant I	Homme	45-49 ans	Ingénieur en sport automobile	Maîtrise, advance motorsport engineering
Participant J	Homme	40-44 ans	Architecte Entrepreneur	Bacc. Architecture
Participant K	Homme	40-44 ans	Directeur en création	Diplôme d'études professionnelles
Participant L	Homme	20-24 ans	Étudiant (Bacc. microbiologie)	Diplôme études collégiales

Participant M	Homme	40-44 ans	Entrepreneur Développement des affaires	Bacc. Marketing
Participant N	Homme	30-34 ans	Étudiant (Maîtrise, innovation sociale)	Bacc. Environnement
Participant O	Homme	25-29 ans	Développement des affaires	Maîtrise

* La participante C a dû abandonner le projet à cause d'un problème technique durant l'utilisation de la technologie

Il y a eu 3 participants qui ont utilisé le logiciel et 11 participants qui ont utilisé la plateforme internet.

Nous ne partageons pas d'autres informations sur les participants pour éviter la réidentification des sujets de recherche.

Dans le cadre des entrevues semi-dirigées, l'information collectée a été enregistrée avec l'approbation de chacun des participants et par la suite retranscrite verbatim dans un délai de 48 h pour assurer le maximum de validité dans le processus. L'information retranscrite a été codée en utilisant le logiciel d'analyse MaxQDA.

3.2.2.1 Outils de collecte de données

Pour encadrer la collecte de données, nous avons développé un guide pour les entrevues individuelles semi-dirigées. L'objectif de créer un tel outil est de s'assurer que la problématique étudiée dans le cadre de la recherche soit abordée de façon relativement standard, que tous les thèmes soient investigués et permettent au lecteur de connaître notre processus exploratoire.

Le Canevas d'entrevue (annexe A) explore 4 thématiques : 1. La perception des participants face à l'intelligence artificielle en général; 2. La perception des participants face à l'utilisation de la technologie d'EmoScienS; 3. La perception des participants face aux enjeux éthiques liés à l'intelligence artificielle, et plus précisément à l'IA affective et 4. La perception des participants face aux conditions d'acceptabilité de l'intelligence artificielle affective. Dans la majorité des cas, nous avons utilisé des questions ouvertes

pour laisser au participant la place à élaborer sur différentes avenues et ainsi possiblement laisser émerger de nouvelles réflexions.

3.2.3 Analyse des données

Avant de commencer notre collecte de données, nous avons mis en place une stratégie d'analyses préliminaire nous assurant ainsi une bonne structure tout au long du processus. Ayant l'étape d'analyse en tête, nous avons pu prendre en considération les pistes d'hypothèses possibles et ainsi étoffer notre investigation en amont sur le terrain. Nous avons aussi pris des notes sur nos réflexions tout au long de la collecte de données. L'objectif de l'analyse est la réduction des données et leur interprétation. Ainsi, une fois la collecte de données terminée et la transcription complétée, l'information prête à être analysée a été codée. Nous avons élaboré une grille de codage selon les thèmes et sous-thèmes liés à notre cadre d'analyse et informé de notre revue de littérature (annexe C). Pour nous assister dans le processus de codage, nous avons utilisé le logiciel MaxQDA. Nous avons codé jusqu'à ce que toute l'information pertinente soit traitée. À la suite de cette étape, nous avons pu extraire les principales thématiques, enjeux et réflexions, et ainsi en dégager des résultats pertinents et des conclusions riches que nous élaborons dans le prochain chapitre.

3.2.4 Considérations éthiques

Cette recherche a été approuvée par le comité d'éthique de la recherche de HEC Montréal. Nous avons mis en place toutes les mesures applicables pour protéger au mieux la confidentialité des participants. Pour nous assurer de respecter le droit à la protection de la vie privée, nous avons anonymisé les données et nous avons nommé les participants à l'aide d'une lettre au hasard tout en gardant confidentielles toutes informations pouvant mener à leur réidentification. Nous leur avons aussi partagé l'objectif de la recherche et les risques anticipés à y participer et nous nous sommes assurés de les informer de leur droit de se retirer en tout temps du projet de recherche. Nous avons obtenu de la part de chaque participant une autorisation écrite à l'aide d'un document conçu à cet effet. Le consentement s'est fait de façon libre et éclairé.

4 Analyse

Dans ce chapitre, nous vous présenterons les résultats des entrevues que nous avons menées auprès des 14 participants du projet de recherche. Nous aborderons cinq grands thèmes. Premièrement, nous nous intéresserons à la perception des participants par rapport aux technologies d'intelligence artificielle en général. Par la suite, il sera question de la rétroaction des participants sur la technologie d'EmoScienS. Nous détaillerons les conditions d'acceptabilité d'EmoScienS et de l'IA affective. Enfin, nous verrons la perception des participants quant aux mesures actuellement en place pour encadrer l'IA et leur recommandation pour assurer le développement d'une IA éthique et responsable.

4.1 Perception des participants par rapport aux nouvelles technologies et à l'intelligence artificielle

4.1.1 Intérêt, aisance et style d'adoption par rapport aux nouvelles technologies

En ce qui concerne **l'intérêt** pour les nouvelles technologies, les participants ont démontré de façon générale un bel enthousiasme pour ces dernières : 12 participants sur 14 se sont décrits comme ayant un intérêt ou un grand intérêt pour les nouvelles technologies, et seulement 2 se sont décrits comme n'ayant pas du tout d'intérêt ou ayant moyennement de l'intérêt pour celles-ci. Parmi les participants intéressés par les nouvelles technologies, 3 d'entre eux ont cependant spécifié que leur intérêt est beaucoup plus marqué pour leur vie professionnelle que pour leur vie personnelle. Enfin, un participant a mentionné que son intérêt pour les nouvelles technologies décline drastiquement lorsque la technologie essaie de se substituer aux relations sociales.

Le groupe est relativement à l'aise avec les nouvelles technologies. Dix participants ont eu un discours dans lequel ils se jugeaient à l'aise avec les nouvelles technologies, avec des commentaires allant de « Oui, quand même » à « Très confortable, très à l'aise », ou à « Ça, je n'ai pas de problème, très confortable ». Quatre participants ont pour leur part soit démontré que c'est moins naturel pour eux, qu'ils ont besoin d'un coup de pouce parfois ou encore que ça requiert un effort de leur part, avec des commentaires comme : « Je mets

l'effort », « Oui, si on m'explique comment ça marche », « Je me débrouille, mais j'ai de besoin de conseil parfois » (participante D).

Le style d'adoption fait référence à la courbe d'adoption de nouveaux produits techno. Chacun d'entre nous a sa propre personnalité lorsque vient le temps d'emboîter le pas à la nouveauté et au changement. Nous avons demandé aux participants de décrire leur personnalité lorsque vient le temps d'adopter une nouvelle technologie. Est-ce qu'ils se situent parmi les premiers à vouloir essayer la nouvelle technologie ou plutôt parmi ceux qui préfèrent observer la vague et se joindre au groupe lorsqu'ils le jugeront opportun? La grande majorité de nos participants a exprimé qu'elle adopte habituellement une nouvelle technologie une fois qu'un bon nombre d'utilisateurs l'ont adoptée, qu'elle a été évaluée et testée, et donc qu'elle a fait ses preuves. D'autres facteurs, comme le contexte, le prix et la conscience écologique, ont aussi été mentionnés comme susceptibles d'influencer le comportement d'adoption. Une participante a précisé que dans sa vie personnelle, elle tarde à adopter une nouvelle technologie, mais que sur le plan professionnel, considérant son rôle, elle est souvent plutôt parmi les premières à essayer les nouveautés, ce qui nous indique que le contexte d'utilisation n'est pas négligeable. Deux participants ont mentionné que les frais importants reliés aux nouvelles technologies sont un frein pour eux et un pensez-y-bien quant à l'adoption des produits les plus récents sur le marché. De plus, un des participants précise qu'il va « attendre la version 2.0 pour être sûr que les *glitches*, les petits *bugs* ont été réparés et qu'une certaine quantité de la population l'a testé » (participant J). Un autre facteur qui a été mentionné est le poids écologique relié à l'adoption des nouvelles technologies. Le participant qui a évoqué ce facteur mentionne qu'il aimerait bien se procurer les plus récentes technologies, mais qu'avant de faire un tel choix, il met en perspective non seulement le prix, mais aussi son empreinte écologique reliée à la consommation.

4.1.2 L'intelligence artificielle aux yeux des utilisateurs

4.1.2.1 Perception et littératie

Lorsque nous parlons d'intelligence artificielle, il est intéressant d'avoir le portrait de la perception des gens par rapport à la technologie. Pour cerner le niveau de littératie et la perception par rapport à la technologie, nous avons posé deux questions :

Quel est le premier mot qui vous vient à l'esprit quand on vous dit « intelligence artificielle »?

Quelle serait votre définition de l'intelligence artificielle dans vos mots?

Le tableau ci-dessous présente les réponses des participants.

Tableau 4.1 Description de l'intelligence artificielle selon les participants

Participants	Premier mot qui vient à l'esprit quand on pense à « intelligence artificielle »?	Votre définition de l'intelligence artificielle
A	Algorithme	« L'intelligence artificielle c'est un moyen qui nous permet d'aller capturer des datas, de les analyser et par la suite de proposer des solutions ou des alternatives pour la personne. »
B	Avancement technologique	« En fait c'est une façon de développer et d'améliorer la technologie pour améliorer nos vies. »
D	E.T. Back to the Future	« L'intelligence artificielle je te dirais, c'est la robotique qui remplace l'humain, c'est l'utilisation de la technologie pour répondre aux besoins de la société. »
E	Robot	« J'ai l'impression que c'est un processus à travers lequel tu regroupes tellement d'information que c'est possible de reconnaître des <i>patterns</i> qui amènent une machine à traverser des processus de réflexion et de penser pratiquement comme un humain, déduire des informations, un processus de réflexion dans le fond. »
F	Technologie	« Pour moi, ce serait une technologie mais vraiment poussée au maximum donc qui est capable de répliquer ce que le cerveau humain pourrait anticiper. »
G	Interconnectivité	« Je dirais que c'est la collecte, le traitement et l'analyse et éventuellement le déploiement de données stockées de façon informatique/numérique pour permettre justement l'automatisation, et je n'irais pas jusqu'à la responsabilisation, mais le fait que ça se fasse de manière autonome, l'automatisation peut-être de la donnée de

		façon numérique, que ça se nourrit par ce que ça collecte et que ça évolue par ce que ça se nourrit. »
H	Cookies	« Dans mes mots, je dirais que c'est un système informatique qui utilise tes données, donc les recherches que tu fais, tes photos, tes réseaux sociaux, ta vie informatique on va dire, donc qui prend ça et qui les utilise justement pour faire des moyennes, envoyer des annonces, des publicités, ils font des analyses sur ce que tu fais. »
I	Rapidité	« Ça serait l'utilisation de la technologie pour faciliter un processus décisionnel ou pour remplacer certaines fonctions cognitives jusqu'à fonctionnelle de l'être humain, à tous les niveaux. Donc essayer de remplacer ou faciliter le processus, un processus quelconque humain qui passerait du <i>psyche</i> jusqu'à l'action. »
J	Ordinateur	« L'intelligence artificielle est un programme qui analyse un certain nombre d'informations dans un champ d'activité très particulier et qui transforme cette information soit en actions ou en données supplémentaires. »
K	Steven Spielberg	« L'intelligence artificielle pour moi c'est un système informatique programmé pour fonctionner de manière autonome et résoudre des problèmes de manière autonome. »
L	Robot et Isaac Asimov	« Ça serait... L'ensemble des... C'est dur à définir, honnêtement! L'ensemble des... technologies et connaissances... qui sont... parce qu'il y a le point du fait qu'il faut que ça soit un minimum autorégulé, c'est-à-dire que si c'est juste l'homme qui l'utilise là ça devient un outil. C'est vraiment difficile à définir, honnêtement je ne saurais pas te mettre une définition dessus là comme ça, il faudrait que j'y pense. »
M	Web 3.0	« C'est le Metavers et l'automatisation de ce qu'on ne croyait pas être automatisé un jour, et c'est où l'automatisation de plusieurs processus comme les voitures électroniques

		par exemple qui font un mixte de ce que l'humain uniquement était capable de contrôler. »
N	Peur	« Ça serait une forme, un programme qui apprendrait par ses propres moyens, on ne serait pas obligé de lui donner de l'information, mais il se fabriquerait sa propre information par lui-même. Et il pourrait avoir une certaine autonomie justement d'action, de décision. »
O	Machine Learning	« Je dirais que c'est vraiment un mélange de toutes les données que tu vas transmettre aux applications, aux logiciels que tu utilises et après ça ce qu'ils vont en faire, c'est-à-dire comme je te disais des références plus personnalisées et après ça changer ton parcours digital, donc pour moi l'intelligence artificielle c'est toute la mécanique qui est derrière. »

4.1.2.2 Bénéfices perçus

Aux yeux des participants, les bénéfices que peut apporter l'intelligence artificielle sont grands. Nous avons demandé aux participants : « Quelle est votre opinion générale sur l'intelligence artificielle et plus spécifiquement, quels en sont les bénéfices à vos yeux? » Les commentaires recueillis s'attardent entre autres sur les avantages tels que la rapidité d'analyse et de dissémination de l'information, l'optimisation, l'efficacité, la productivité, l'aide à la prise de décision et l'amélioration de l'expérience client grâce à la personnalisation de l'offre. Certains participants ont aussi parlé de l'automatisation des procédés, qui peut entre autres aider à la pénurie de main-d'œuvre et remplacer l'humain pour plusieurs tâches. Fréquemment, les participants ont fait référence à « aider » et à « dépasser les limites humaines ». Ainsi, il y a une perception selon laquelle l'intelligence artificielle peut apporter une contribution à l'humanité. Ils ont d'ailleurs approfondi cet aspect. Ils ont abordé les bénéfices potentiels pour le milieu de la recherche, comme la conquête de l'espace, la faim dans le monde ou le milieu de la santé. Ils ont parlé de l'aide que peut apporter l'intelligence artificielle pour poser des diagnostics, découvrir des

remèdes, fabriquer rapidement des vaccins et soigner des maladies. Les participants perçoivent que la technologie peut aider l'humain et la société à évoluer.

L'ensemble des participants a cependant une opinion ambivalente sur l'intelligence artificielle. Autant l'ensemble des participants ont pu nous indiquer de nombreux bénéfices de la technologie, autant ils ont mentionné être mitigés sur le sujet. Ils ont d'ailleurs dit des choses comme « je pense qu'on en a de besoin, qu'on est rendu là, surtout pour faire avancer la technologie, l'humanité, on est rendu là » (participante B), « il y a du bon et du mauvais mais ça permet quand même de faire certaines avancées notamment au niveau de la science, de la santé, ça a quand même beaucoup d'apport sur comment comprendre certaine maladie » (participante E), « Moi je pense que c'est nécessaire, elle est là où on est rendu en société elle est nécessaire l'intelligence artificielle parce qu'on a atteint un certain nombre de limite qu'on n'est pas capable de franchir et que seul l'intelligence artificielle pourra nous aider à avancer » (participant J). Cependant, l'ensemble a aussi émis des commentaires comme : « il y a le côté éthique qui me questionne beaucoup » (participante D), « dès qu'il y a une personne mal intentionnée derrière et bien partout ça peut créer des problèmes » (participant F), « Tout dépendant de quelle manière c'est développé et avec quel motif. Ça peut être soit très positif soit très négatif » (participante G) ou « autant je trouve ça impressionnant, autant ça me fait peur » (participante H). Ainsi, malgré que nous en voyions très certainement les avantages, l'intelligence artificielle soulève des questionnements, des réflexions, des enjeux.

4.1.2.3 Enjeux

Pour mettre en lumière les enjeux éthiques que l'intelligence artificielle soulève aux yeux des participants, nous leur avons demandé de nous exprimer leur appréciation générale de l'intelligence artificielle et la mise en garde qu'ils feraient par rapport à celle-ci de façon générale. À des fins de codage et d'analyse, nous avons classé les commentaires récoltés selon les enjeux que nous avons retrouvés dans la littérature : sécurité, biais et discrimination, responsabilité, transparence et explicabilité, vie privée et autonomie. Le consentement est une question que nous aborderons dans la sous-section suivante.

Sécurité

L'enjeu de sécurité est celui qui a été le plus mentionné par les participants à notre recherche. L'intelligence artificielle suscite des craintes par rapport aux fuites de données, au piratage ou à l'hameçonnage. Des événements récents, comme la fuite de données chez Desjardins ou le partage de données des utilisateurs de Facebook, ont généré chez eux une certaine méfiance quant à la sécurité des SIA :

« La fuite de données qu'on a eue chez Desjardins » (participant N);

« Facebook qui *drop*, partage, vend les données et ça, pratiquement personne ne le sait » (participant N);

« Déjà, avec Desjardins, moi j'ai fait partie des milliers de personnes fraudées » (participante G).

Les mauvaises intentions sont un élément dont les participants ne se sentent pas à l'abri :

« À partir du moment où on a accès à mes profils, à partir de là, on peut voler nos identités, c'est plutôt ça qui m'inquiète. » (participant K);

« Ce qui me fait le plus peur, c'est malheureusement le *dark Web*, un jour, qu'il puisse gagner accès à toutes ces données et qu'on perde le contrôle de façon assez *rough* » (participant M).

De façon générale, les craintes des participants se sont concentrées autour du vol d'identité et de la fraude. Cependant, un participant a soulevé aussi les risques quant à l'armement : « Le fait que les armements, le nucléaire, c'est quasiment rendu faible comparativement à tout ce qui est possible de faire dans l'avenir » (participant L). Un autre a verbalisé le risque de piratage des voitures autonomes ou des avions (participant M). Au moins 11 des 14 participants ont abordé clairement l'enjeu de la sécurité. C'est un sujet important à leurs yeux. Bien que les participants voient beaucoup de potentiel dans la technologie, leurs craintes se fondent surtout sur les mauvaises intentions. Une participante a très bien résumé le sentiment en disant : « Là où il y a de l'homme, il y a de l'hommerie, ça veut dire qu'il va toujours y avoir quelqu'un qui va trouver une faille et bifurquer de l'intention première du projet » (participante D).

La superintelligence, c'est-à-dire une intelligence artificielle générale, a aussi été abordée comme enjeu en lien avec la sécurité. Un participant a soulevé la possibilité qu'une intelligence artificielle qui nous dépasserait pourrait nous voir comme une nuisance. Tandis qu'un autre a dit :

Ils sont des millions de fois plus intelligents que nous avec du *processing power* beaucoup plus fort que nous. La seule chose, c'est que c'est nous qui les contrôlons, donc ça fait peur que ça soit quelques lignes de codes, finalement, qui vont nous protéger des IA dans l'avenir (participant L).

Vie privée

En deuxième place vient l'enjeu du respect de la vie privée. Certains participants mettent en garde contre l'intrusion dans la vie privée. Certaines fonctionnalités sont jugées trop intrusives, comme lorsqu'elles requièrent la géolocalisation ou qu'elles font un profilage de l'individu (participant L). Un participant donne un exemple : « me scanner à l'entrée d'un magasin, et qu'on me dise que j'ai de besoin d'un jeans bleu 34x34, je me dis que c'est *too much* » (participant J).

Les technologies d'assistance virtuelle ont été abordées en ce sens. Une participante a mentionné avoir Alexa (technologie d'Amazon) à la maison et ressentir un sentiment de malaise parfois. Elle perçoit que la technologie prend beaucoup d'information sur sa vie personnelle et elle se questionne sur ce qui est fait avec l'information par la suite. Elle a comparé son sentiment à celui de se sentir espionnée (participante E).

Certains ont aussi mentionné ne pas avoir le choix de partager leurs informations personnelles. Une participante a souligné : « maintenant, c'est rare que tu peux ne pas mettre d'informations personnelles pour t'inscrire à quelque chose, donc tu n'as pas vraiment le choix non plus. À quel point je suis à l'aise? pas tant que ça » (participante H). Elle a un sentiment d'intrusion dans sa vie privée et l'impression de ne pas avoir réellement d'autre choix.

La reconnaissance faciale a été abordée dans deux contextes, celui de la surveillance dans certaines villes pour diminuer le taux de criminalité (participant O) et celui du crédit social en Chine (participante F). Dans les deux cas, il y a eu une référence à l'atteinte à la vie privée.

De façon générale, les participants ont mentionné « Big Brother », « un œil sur la population », « surveillance », « confidentialité de la vie privée », « tout savoir ce que tu es en train de faire à la minute près ».

Neuf participants ont fait référence à l'enjeu du respect de la vie privée lorsque questionnés sur les principaux enjeux ou problématiques liés à l'IA en général.

Autonomie

Trois participants ont parlé de l'enjeu de l'autonomie. Un participant a souligné le danger que « les gens aient le sentiment de perdre un peu leur pouvoir décisionnel puisqu'ils se font tout offrir tout le temps, tout est vraiment biaisé [...] ça s'en vient vraiment de la surexposition » (participant O). Dans la même veine, un autre participant a mentionné :

Mes intérêts personnels, je n'aime pas qu'on les observe pour être capable de me suggérer quelque chose, j'aime bien faire ma propre recherche moi-même, trouver mon chemin de croix, comme on dit, tout seul, au lieu de me faire suggérer quelque chose par intérêt commercial. (participant I)

Il précise également qu'il se pose cette question quand il regarde une nouvelle technologie : « Est-ce que je garde le contrôle? » (participant I). Enfin, une participante a parlé de l'usage de l'intelligence artificielle dans le cadre de la surveillance faite en Chine en lien avec le crédit social (participante F).

Trois participants ont fait référence à l'autonomie en abordant la dépendance aux technologies. Un participant se demande ce qu'il arriverait si, pour une raison quelconque, la technologie cessait de fonctionner à cause d'une catastrophe naturelle ou d'un piratage qui mettrait K.O. un réseau complet (participant M). Un autre spécifie que l'engouement pour les écrans de façon générale, les dépendances qui se créent, c'est un véritable problème (participant K), et une participante mentionne : « On dirait que je me dis, l'intelligence artificielle, on va-tu devenir tellement dépendant et on ne sera plus capable de faire ce qu'on fait dans notre quotidien? » (participante E)

Cinq participants ont spécifiquement fait référence à l'autonomie.

Marché de l'emploi

Quatre participants ont parlé de la possibilité que l'intelligence artificielle puisse remplacer l'humain pour certaines tâches. Ils ont cependant mentionné entre autres qu'il ne faut pas que « la pensée critique de l'humain soit remplacée » (participante A). La participante A met en garde contre la tentation de remplacer l'humain pour certaines tâches, comme diagnostiquer, en nous rappelant que l'humain a encore un 6^e sens, que l'intelligence

artificielle n'a pas. Un participant nous prévient que le remplacement des employés en entrepôt, qui peut avoir beaucoup de sens d'un point de vue financier, peut être moins bon d'un point de vue humain parce que « tu fais perdre des emplois à des gens » (participant O).

Transparence

Trois participants ont fait référence à la transparence, entre autres avec des commentaires comme : « Je n'ai pas l'impression que les compagnies sont toujours *up front* sur la façon dont ils utilisent les données » (participant L) et « Il y a sûrement des endroits où je donne mes données et je ne le sais pas encore » (participant N).

Cependant, la transparence sera abordée plus en détail en lien avec le consentement et les conditions d'acceptabilité.

Responsabilité

Un participant a fait référence à la responsabilité. Il a abordé la question des situations où l'intelligence artificielle peut prendre une décision à la place de l'humain, comme en cas d'accident avec une voiture autonome (participant K). Est-ce qu'on laisse la technologie prendre la décision? Est-ce possible de réagir assez rapidement pour reprendre le contrôle? Aux yeux du participant, c'est un enjeu important et il faut savoir le maîtriser.

Autres enjeux

D'autres enjeux ont été mentionnés, celui du **respect de l'intégrité humaine** : « Le *down side*, c'est vraiment juste de faire attention comment on va respecter l'humain derrière ça, comment on va respecter l'individu, son intégrité » (participante A). Il faut faire attention aux dérapages, par exemple avec CRISPR sur le plan de la manipulation génétique (participant L).

La **perte de contact humain** a aussi été abordée : « Les gens sont beaucoup plus attirés et cylindrés dans leur monde interconnecté et virtuel » (participant M). Un participant a mentionné ne pas être favorable à toute technologie qui tente de remplacer le tissu social. Enfin, un dernier enjeu a été soulevé : « Est-ce qu'on a de besoin d'avoir quelque chose

pour nous aider à aller plus vite, réfléchir plus vite, agir plus vite, sentir plus vite? C'est encore très jeune comme situation » (participant I).

Toutefois, aucun participant n'a fait référence aux enjeux de biais et de discrimination ou encore à l'explicabilité en lien avec l'intelligence artificielle dans le cadre de mises en garde par rapport à la technologie.

4.1.2.4 Partage des données personnelles et consentement

La question du partage des données personnelles et du consentement est un enjeu éthique délicat concernant lequel il a été difficile d'avoir des réponses claires et tranchées. Elles étaient plutôt dans la nuance, c'est-à-dire qu'à la question « êtes-vous à l'aise de partager vos données personnelles? », bien que certains aient répondu « Oui, vraiment », l'ensemble a nuancé ses réponses avec leurs zones d'inconfort. Aucun participant n'a répondu être à 100 % à l'aise, tandis que 4 participants ont mentionné sans nuance être « Pas à l'aise », « moyennement à l'aise », « pas très à l'aise » ou méfiant. Les zones d'inconfort mentionnées sont :

- Lorsqu'on demande de s'inscrire en utilisant son nom d'utilisateur et son mot de passe d'une autre plateforme (ex. : le nom d'utilisateur Google ou Facebook);
- La géolocalisation;
- L'obligation d'accepter toutes les conditions de partage pour avoir accès à la plateforme, au site ou à la technologie;
- Le partage de données bancaires;
- Le profilage par l'accumulation de données personnelles sur l'utilisateur.

Le contexte du partage de données, c'est-à-dire à quelles fins les données seront utilisées, pèse aussi dans la balance. Certains participants ont indiqué être plus favorables si les données sont utilisées à des fins de recherches ou d'amélioration de la technologie. La fréquence d'utilisation de la plateforme ou de la technologie a aussi été mentionnée comme un facteur influençant le choix du partage de données. Si l'utilisation est quotidienne, il y a un bénéfice perçu à y partager ses données parce que l'utilisateur pourra avoir

possiblement du contenu plus personnalisé à force d'utiliser la plateforme ou le site. Le niveau de sécurité de la plateforme ou de la technologie quant à la protection des données, ainsi que le pays d'origine, sont aussi des aspects qui ont été mentionnés. Le type de données concernées fait aussi une différence dans le choix ou non de les partager. En effet, plus l'information est considérée comme sensible, privée, personnelle, moins le participant est prêt à la partager. Ainsi, un utilisateur spécifie qu'il est à l'aise avec le partage de ses données professionnelles, mais moins avec celui de ses données privées. Un participant qui se montre prudent par rapport au partage de données spécifie cependant : « ça ne me dérange pas que YouTube, au bout de la ligne, sache que j'ai été voir trois fois une vidéo de Van Halen » (participant I).

Les craintes par rapport au partage de données personnelles qui ont été soulevées sont principalement les fuites de données, la fraude, le vol d'identité, la manipulation et la non-protection de son image numérique.

Enfin, plusieurs participants ont dit avoir abdicqué à ce sujet et avoir le sentiment de ne pas avoir le choix de partager leurs données. Nous avons d'ailleurs demandé aux participants : « Sentez-vous que vous avez le contrôle sur vos données personnelles dans l'univers du numérique? » Des 14 participants, 11 ont catégoriquement répondu « Non » à la question. Un participant a mentionné croire qu'il pourrait avoir le contrôle, mais qu'il ne passe pas son temps à y faire attention. Il croit que tout le monde le peut, mais que personne ne prend le temps, soit par manque d'éducation ou manque d'intérêt. Un participant a dit avoir un sentiment de contrôle sur ses données quand on lui demande d'accepter ou de refuser le partage de données, et enfin, une dernière participante a répondu « pas toujours, tu sais, parfois, ils nous le demandent, mais je me dis qu'il doit y en avoir qui ne nous le demandent pas » (participante D).

Les participants ont mentionné avoir une plus grande ouverture à utiliser des technologies lorsqu'ils sentent qu'ils ont une forme de contrôle sur leurs données personnelles. Ils n'apprécient pas se faire imposer de créer des comptes où ils doivent y inscrire de nombreux détails ou l'obligation d'accepter de partager leurs données personnelles pour utiliser une plateforme ou une technologie. Plusieurs le font à reculons, faute de temps et d'alternatives. Ainsi, les organisations qui offrent l'option de « rejeter » le partage de

témoins (cookies) ou qui offrent d'utiliser une technologie en tant qu'invité et non en tant que membre donnent une forme de contrôle aux utilisateurs.

4.1.2.5 Sentiment par rapport aux conditions d'utilisation

Pour aller plus en profondeur dans le sentiment de contrôle sur les données personnelles dans l'univers du numérique, nous avons décidé de questionner les participants sur leur gestion générale du processus de prise de décision en lien avec les formulaires de consentement aux conditions d'utilisation lorsqu'ils accèdent à une nouvelle technologie. Nous leur avons donc demandé :

Lorsque vous utilisez différents sites ou téléchargez de nouvelles applications, prenez-vous le temps de lire le consentement des conditions d'utilisation? Pourquoi?

La grande majorité des participants (11 sur 14) ont répondu « non » ou « rarement » à la question. Le texte qui détaille les conditions d'utilisation et l'information relative aux données est trop long, écrit trop petit et dans un « jargon » qui est technique et compliqué pour le commun des mortels. Certains perçoivent d'ailleurs cela comme relevant d'une tactique pour obtenir un consentement et y glisser des clauses que les utilisateurs ne prendront pas le temps de lire et qu'ils n'ont pas les compétences requises pour analyser et comprendre. Trois participants ont mentionné qu'ils apprécieraient que l'information soit présentée sous la forme d'un court résumé, en « *bullet point* », et que ça soit clair, « *straightforward* » et transparent.

Un participant a dit ne pas particulièrement faire confiance à ces documents. Il va généralement plutôt rechercher sur le Web des articles qui concernent le développeur, la technologie ou le site, ainsi que des rétroactions d'utilisateurs, pour obtenir de l'information qu'il considère moins biaisée sur la réputation de la technologie et de l'organisation derrière celle-ci.

Enfin, une participante a émis ce commentaire intéressant par rapport à la situation actuelle : « Les gens sont prêts à laisser aller bien des choses, j'ai l'impression, pour être sur les médias sociaux » (participante E). Donc, à ses yeux, pour pouvoir participer à la vie

numérique, les utilisateurs acceptent d'emblée les conditions d'utilisation des réseaux sociaux.

Par la suite, nous avons posé la même question, mais cette fois spécifiquement en lien avec la technologie d'EmoScienS :

Lorsque vous avez commencé à utiliser EmoScienS, avez-vous pris le temps de lire l'information reliée au consentement des conditions d'utilisation sur la plateforme? Pourquoi?

La moitié des participants ont pris le temps de lire l'information reliée au consentement avant d'accepter. Les participants qui ont lu l'information ont dit l'avoir fait pour voir ce qui allait être fait avec les photos, où elles allaient se retrouver « parce que c'est quand même tes propres données », ou encore, ils ont lu l'information parce que c'était court.

Les participants qui ne l'ont pas lue ont majoritairement fait référence à la confiance soit en la personne qui les a invités à participer au projet ou au contexte d'un projet de recherche, qui passe par un comité d'éthique et qui est rigoureusement encadré. Un participant, pour sa part, a dit ne pas l'avoir lue parce que « pour [lui], ça n'a pas vraiment de valeur, les formulaires de consentement » (participant N).

4.2 Rétroaction sur la technologie d'EmoScienS utilisée dans le cadre du projet de recherche

Pour amorcer la discussion sur l'utilisation de la technologie d'EmoScienS, nous avons demandé aux participants quelles étaient leurs perceptions par rapport à la technologie avant même de l'utiliser : quels ont été leurs premiers sentiments par rapport à EmoScienS? La grande majorité des participants avaient des sentiments initialement positifs quant à l'utilisation d'EmoScienS; seulement une participante a mentionné avoir ressenti un malaise à l'idée de se faire prendre en photo et a dit que ça la « stressait un peu, honnêtement » (participante H). Sinon, les participants ont dit avoir pensé que ça semblait intéressant à essayer, avoir été curieux, avoir pensé que « c'était génial comme idée », avoir eu « hâte de l'essayer », avoir été « enthousiaste » ou encore, une participante précise : « j'étais contente dans le sens que je me suis dit : “bon, on va analyser les émotions que je peux avoir durant mon travail” » (participante A).

Quelques participants ont spécifié ne pas être angoissés, ne pas avoir d’appréhension ou d’inquiétude et être même très à l’aise avec cet essai. Cinq participants ont parlé de l’aspect intrusif de la prise de photo comme étant un facteur à considérer, cependant seulement un de ces participants a ressenti personnellement un stress par rapport à ceci, comme mentionné plus haut. Les autres ont mentionné que c’était un point qui donne matière à réflexion, et l’un d’entre eux a spécifié que les mesures mises en place pour assurer la protection des données, telles que mentionnées par la compagnie, l’ont rassuré. Enfin, une personne a dit avoir commencé le projet avec une certaine fermeture d’esprit par rapport à la technologie : « je vais le faire, mais peu importe ce qu’elle me dit, je ne pense pas qu’elle va me faire réaliser quelque chose » (participante K). Tandis qu’une participante a mentionné s’être souciée de ne pas avoir les habiletés technologiques suffisantes pour apporter assez de richesse au projet de recherche.

4.2.1 Facilité d’utilisation

Pour synthétiser l’information et la présenter de façon claire, nous avons choisi de vous la présenter sous la forme d’un tableau.

Tableau 4.2 Perception de la facilité d’utilisation de la technologie d’EmoScienS

PARTICIPANTS	FACILITÉ D’UTILISATION	COMMENTAIRES
A	Facile	« J’oubliais de le partir parfois (version Web). Ça reste facile d’utilisation ».
B	Facile	« <i>User friendly</i> , facile à utiliser, comme si ça n’existait pas, vraiment facile d’utilisation, c’est assez simple et non complexe à utiliser et non invasif. »
D	Facile	« Très facile, très convivial, c’est très bien fait »
E	Facile	« J’ai trouvé ça vraiment facile, j’ai trouvé ça le <i>fun</i> , tu te connectes avec ton tableau de bord en deux secondes »

F	Facile (+/-)	La participante a eu des difficultés avec la détection de son visage et n'a eu accès à son tableau de bord qu'après avoir terminé la période d'utilisation. Cependant, elle a jugé qu'outre les problèmes techniques qu'elle a rencontrés, la plateforme est facile d'utilisation.
G	Facile (+/-)	La participante a dit : « pour accéder aux données, j'ai trouvé ça facile », cependant pour reprendre l'enregistrement, l'accès n'est pas intuitif, il faut faire rejouer le tutoriel, « il manque un bouton direct ».
H	Facile (+/-)	La participante juge qu'en général, c'était facile : « tu te connectais et tu partais la caméra ». Cependant, elle a rencontré plusieurs difficultés techniques en lien avec la caméra, elle devait donc la surveiller régulièrement.
I	Facile (+/-)	Le participant a trouvé que c'était facile de cliquer sur l'onglet. Il a cependant mentionné qu'à partir du tableau de bord, il n'y avait rien (bouton) pour retourner du côté opérationnel. De plus, il a rencontré certains défis pour la détection de son visage.
J	Difficile	« Écoute, je vais être très honnête avec toi, j'ai trouvé la chose très emmerdante, j'ai trouvé que c'était très <i>time consuming</i> , peut-être parce que j'ai eu plein de <i>glitches</i> informatiques, et la caméra fonctionnait une fois sur deux ».

K	Facile	Le participant a mentionné que c'était une technologie « <i>hands-off</i> », c'est-à-dire qui ne requiert pas une participation active : « On le démarre et on le laisse aller », « C'est devenu un automatisme, ça se passait dans le <i>background</i> ».
L	Facile (+/-)	Le participant a dit avoir rencontré parfois des problèmes de détection de visage s'il était mal placé. Sinon, il a dû travailler avec un collègue, donc a dit avoir eu ce défi pour s'assurer d'être le seul devant l'écran.
M	Facile	Le participant a eu des problèmes avec la version logicielle, mais une fois qu'il a passé à la version Web, « c'était parfait, ça fonctionne super bien ». Le participant a eu un problème technique pour avoir accès aux résultats dans son tableau de bord, mais une fois que ça a été réglé, « c'était très facile à consulter et c'était très clair. »
N	Facile	Le participant considère que la facilité d'utilisation est « quand même élevée, c'est accessible »
O	Facile (+/-)	« J'ai trouvé ça vraiment facile à utiliser ». Cependant, le participant a trouvé que pour rouvrir la caméra, ce n'était pas intuitif, il fallait faire rejouer le tutoriel chaque fois.

La majorité des utilisateurs ont mentionné avoir trouvé la plateforme particulièrement facile à utiliser. D'ailleurs, malgré ses ennuis techniques, le participant M a précisé qu'une fois ceux-ci réglés, « c'était parfait » et que « ça fonctionn[ait] super bien ». Cela dit, certains aspects pourraient être ajustés pour améliorer la fluidité d'utilisation, comme

l'accès pour revenir à l'enregistrement des données à partir du tableau de bord, un point qui a été soulevé par plus qu'un participant.

4.2.2 Validité des résultats

Nous avons choisi de vous présenter la perception des participants par rapport à la « validité des résultats » sous la forme d'un tableau.

Tableau 4.3 Perception des participants par rapport à la validité des résultats

PARTICIPANTS	VALIDITÉ DES RÉSULTATS	COMMENTAIRES
A	Validité (-)	« Quand je me suis mis à regarder mes résultats, j'étais un peu déçue parce que je me suis rendu compte que j'étais souvent en colère, et pourtant je n'étais pas en colère. » « Je ne me rallie pas aux résultats ».
B	Validité (+/-)	« Oui, très juste », « J'ai été vraiment triste cette semaine et il y avait beaucoup de tristesse, et ça disait que j'avais de la tristesse, donc ça a même vu ça, même si j'essayais que ça ne paraisse pas ». Cependant, la participante n'était pas certaine de la validité des heures travaillées présentées dans son compte rendu
D	Validité (+)	« Oui, je sens que c'est représentatif »
E	Validité (-)	La participante a le sentiment que le tableau n'est pas nécessairement représentatif.
F	Validité (+/-)	La participante considère que le système a relativement bien cerné le pourcentage de temps avec un visage neutre considérant qu'elle travaille beaucoup devant son ordinateur sans interaction. Elle considère cependant que la technologie a surestimé son niveau de joie.

G	Validité (-)	Dès que la participante a vu que ses rencontres clients (Zoom) ne seraient pas prises en compte dans les données récoltées, elle a considéré que le portrait ne serait pas « <i>accurate</i> ». Ce sont les moments où elle pense avoir le plus d'émotions perceptibles. Elle n'a pas été convaincue.
H	Validité (+/-)	La participante considère que le compte rendu ne reflète pas comment elle se sentait, mais en regardant ses photos, elle aussi aurait décelé ce type d'émotions. La participante dit qu'elle a un problème dans la vie, que de façon générale, son visage au repos peut avoir l'air bête. Ainsi, la participante ne croit pas que c'est représentatif de ses émotions, mais peut comprendre pourquoi le système en est arrivé à de telles conclusions.
I	Validité (-)	« Non, <i>niet</i> », « ça apparaissait tout le temps triste, ce qui n'était pas le cas », « non, j'étais en période où tout coulait comme de l'eau et j'avais aucune raison d'être triste. Au bout de la ligne, <i>looking back to those days, they were just amazing to be at work</i> ». Le participant dit avoir été élevé dans une culture qui n'encourage pas l'expression d'émotion, il dit avoir une « <i>poker-face</i> ».
J	Validité (-)	« Non, je dirais même pas du tout », « Je n'ai pas compris pourquoi il me montrait en colère tout le temps », « Je n'étais pas en colère pendant 24 heures »
K	Validité (+)	Pas de surprise.

L	Validité (+/-)	Le participant considère que son portrait en général est représentatif, mais il se questionne tout de même sur la représentativité des émotions avec une prise de photo aussi espacée (aux cinq minutes).
M	Validité (-)	« Je ne crois pas m’être senti méprisant ou triste depuis deux semaines ». Le participant aimerait comprendre comment l’intelligence artificielle analyse les photos. D’ailleurs, il a vu certaines photos qu’il a jugées douteuses pour émettre un jugement quant à une émotion.
N	Validité (-)	Le participant ne croit pas avoir été aussi joyeux que ce que son compte rendu reflète. Il considère que le logiciel n’est pas super fiable encore.
O	Validité (+)	Le participant considère que son compte rendu est relativement juste. Les trois quarts de ses émotions sont de la joie. Il considère que c’est normal étant donné qu’il a utilisé la technologie entre autres pendant ses rencontres clients.

Les participants ont formulé des recommandations qui pourraient potentiellement améliorer la fiabilité de la technologie :

- La technologie pourrait faire une analyse du visage du participant pour pouvoir s’adapter à sa physiologie et ainsi être plus personnalisée (participante A);
- La technologie pourrait faire quelques exercices avec l’utilisateur au début pour établir une base de données plus personnalisée (participante E);
- Il pourrait y avoir une meilleure représentativité au sein de la base de données, car deux participants se sont demandé si la plateforme s’appliquait à eux, considérant le fait que leur visage n’exprime pas nécessairement de la même façon les émotions déjà à la base (participants H et I);

- Une explication de la façon dont la technologie détermine l'émotion serait intéressante afin de mieux comprendre les résultats (participante H);
- Un contexte optimal d'utilisation de la technologie pourrait être déterminé, car certains participants ont considéré que la technologie s'appliquait mieux en situation de travail au bureau ou de rencontre client plutôt que dans un cadre de travail seul devant son écran (participantes A et E);
- Certains participants ont suggéré une prise de photos plus régulière pour avoir un compte rendu plus représentatif, recommandant une fréquence d'une minute ou même en continu (participants K, L, N et O);
- Un participant a suggéré que la plateforme puisse être utilisée sur les appareils intelligents, comme le téléphone, qui est utilisé dans un contexte professionnel et social, de façon très intégrée à la vie de tous les jours (participant N).

4.2.3 Bénéfices perçus quant à la technologie d'EmoScienS

Les participants ont testé la technologie dans un contexte d'utilisation individuelle et non dans un contexte de déploiement en organisation. Nous leur avons donc demandé de nous dire quels sont, selon eux, les bénéfices que la technologie d'EmoScienS peut apporter à une personne qui utiliserait régulièrement la plateforme (bénéfices individuels). Nous avons aussi recueilli la perspective des participants sur les bénéfices que peut apporter la technologie d'EmoScienS à une organisation si cette dernière la déploie au sein de son entreprise (bénéfices organisationnels).

Tableau 4.4 Bénéfices individuels et organisationnels de l'utilisation de la technologie d'EmoScienS

PARTICIPANTS	BÉNÉFICES INDIVIDUELS	BÉNÉFICES ORGANISATIONNELS
A	Ne voit pas de bénéfice individuel.	Prendre le pouls du bien-être des employés

B	Prise de conscience, remise en perspective.	Permet de voir comment ça se passe, de cibler si certains employés ou départements auraient besoin d'aide. Améliorer la situation au travail et de la même façon, la productivité.
D	Prise de conscience pour, par la suite, mettre en place des actions si nécessaire.	Améliorer les conditions de travail et permettre de cibler les efforts pour augmenter le bien-être.
E	La participante n'arrive pas à voir la pertinence dans son cas, considérant le temps qu'elle passe devant des fichiers. Elle n'est pas certaine que ce soit très représentatif, donc c'est difficile pour elle de faire ressortir des bénéfices.	Voit un bénéfice pour prendre le pouls de l'équipe; c'est un aspect important au sein de son entreprise.
F	Prise de conscience	Apprécier l'humeur globale de l'entreprise et détecter les personnes qui auraient des émotions négatives, voir l'effet sur le travail et la performance. Ça permet de mettre en place des actions, comme un programme de soutien psychologique.
G	Prise de conscience. Bénéfice si c'est accompagné de recommandations, de piste de solutions pour les émotions rencontrées.	Utile possiblement en lien avec une formation RH. Voit le bénéfice pour quelqu'un en retour d'arrêt de maladie pour dépression, par exemple, pour avoir un suivi.
H	Aucun bénéfice à ses yeux (lien avec la validité des résultats).	Aucun bénéfice à ses yeux (lien avec la validité des résultats).
I	Prise de conscience et optimisation de la productivité.	Prise de conscience et outil pour mesurer l'effet d'actions ou d'annonces en organisation, optimisation de la productivité.
J	Voit le potentiel d'associer les émotions à certains événements et d'en retirer des prises de conscience.	Comprendre un peu la dynamique des gens, leur humeur selon leurs conditions et environnement de travail
K	Ne voit pas le bénéfice à l'utilisation de l'outil sur le plan individuel.	Évaluer le moral dans la compagnie. Le participant, cependant, perçoit que la technologie peut être utilisée comme un outil de surveillance.

L	Prise de conscience.	Prise de conscience du bonheur des employés et amélioration des conditions de travail. Permettre d'ajuster les tâches en fonction des résultats et d'augmenter l'efficacité de la compagnie.
M	Donner un retour sur comment je me sens.	Perception que la technologie peut aider à améliorer les conditions de travail. Le participant voit entre autres un bénéfice pour les PME ou les entreprises en démarrage qui n'ont pas de département de RH. Considère que l'information peut être moins falsifiée que les questionnaires papier sur le bien-être en organisation.
N	Prise de conscience et possiblement encourager à changer certaines habitudes.	Le participant voit le potentiel organisationnel pour augmenter le bien-être et la joie dans le milieu de travail, mais voit l'intérêt organisationnel axé surtout sur le bénéfice de l'entreprise entourant la productivité pour engranger du profit.
O	Prise de conscience	Prendre le pouls de ses employés, obliger la prise de conscience de l'organisation et voir les problèmes s'il y a lieu.

Sur le plan individuel, les participants, de façon générale, considèrent la prise de conscience face à sa situation émotionnelle comme étant le principal bénéfice de l'utilisation d'EmoScienS pour les utilisateurs. Certains mentionnent que cette prise de conscience permet potentiellement de poser des actions concrètes pour agir sur ce qui a été constaté dans les résultats. Les participants n'ont pas eu accès au portail de ressources ou à des informations complémentaires en lien avec leur tableau de bord, mais quelques participants ont mentionné que si l'information du tableau de bord était accompagnée de recommandations et de piste de solutions, ce serait certainement une valeur ajoutée.

Sur le plan organisationnel, on voit comme bénéfice principal le bien-être organisationnel. On considère que la technologie peut aider à prendre le pouls de l'humeur globale et du bonheur d'une équipe. Elle peut aider à comprendre l'effet de certaines mesures, notamment en ce qui concerne les conditions de travail, sur le moral des troupes. Un

participant mentionne qu'elle peut obliger l'organisation à prendre conscience de la réalité sur le terrain (participant O). Cependant, un participant est quant à lui méfiant quant à l'utilisation réelle d'une telle technologie, qui pourrait, à ses yeux, plutôt être déployée pour tenter d'augmenter la productivité chez les employés et donc, de la même façon, la rentabilité.

4.2.4 La technologie d'EmoScienS vue à travers les principes de la Déclaration de Montréal

Nous avons choisi d'utiliser la *Déclaration de Montréal* comme cadre d'analyse pour la technologie d'EmoScienS pour évaluer les enjeux éthiques qu'elle soulève selon le point de vue des participants. Nous avons choisi d'exclure le principe de développement soutenable de l'analyse, puisqu'il ne s'appliquait pas dans le cadre du projet de recherche.

4.2.4.1 Principe de bien-être

Nous avons demandé aux participants : « Trouvez-vous qu'EmoScienS a amélioré votre bien-être, votre qualité de vie? » La quasi-totalité (13 sur 14) a répondu par la négative, soit parce qu'ils n'ont pas ressenti de bénéfice (1); parce que les résultats obtenus étaient négatifs (2); qu'ils perçoivent que la plateforme ne peut pas apporter de bien-être, que ce sont les actions concrètes de l'utilisateur qui peuvent influencer son niveau de bonheur (3); ou encore parce qu'ils ne trouvaient pas que leur tableau de bord était représentatif de leur profil émotionnel de la période (4). Cependant, certains utilisateurs ont mentionné voir le potentiel si le logiciel était plus performant et qu'il était utilisé sur une plus longue période : « À long terme, oui, si c'était plus performant et sur mon téléphone » (participant N).

Bien que les participants n'aient pas trouvé que leur utilisation de la technologie d'EmoScienS sur une courte période de deux semaines a amélioré leur bien-être ou leur qualité de vie, plusieurs considèrent qu'elle peut être un outil pertinent dans la quête de bien-être et d'amélioration de la qualité de vie :

« Si je vois que je suis toujours triste après le souper, ça peut m'encourager à changer mes habitudes » (participant N);

« Définitivement pour l'ergonomie au travail, les performances au travail, le taux de rétention, pour toutes sortes d'éléments comme ça, définitivement, oui, ça peut devenir un outil pertinent » (participant J);

« Au bout de la ligne, c'est des outils qui sont fabuleux, on peut voir si *our guys are happier now for work, if our people we give them the free reigns to the company*, on leur donne des actions, est-ce que c'est des gens qui vont être plus heureux dans la compagnie? », (participant I);

« C'est un diagnostic que ça *pop*, c'est capable de dégager une tendance, donc oui, ça peut améliorer ton bien-être » (participante D);

« Si ça peut permettre de détecter en avance que oui, il y en a qui ne vont vraiment pas bien, des signes de dépression, tu peux intervenir et offrir de l'aide, ne serait-ce que la personne puisse voir son propre bilan et dire : "oui, finalement, je ne file pas, peut-être que j'en fais trop" » (participante B).

L'utilisation d'EmoScienS a aussi généré un certain stress pour certains utilisateurs quant au sentiment d'être observé. Un participant a justement expliqué :

La petite lumière de caméra allumée, ça me tombait sur les nerfs, mais à un point immense. Ça m'empêchait de me concentrer parfois sur mes cours parce que le fait de la voir allumée, ça me dit : "Oh mon Dieu, tu es filmé en ce moment", ça me créait un petit malaise (participant L).

Une autre participante nous a confié : « Honnêtement, ça ne me faisait pas triper de me faire prendre en photo par je ne sais trop qui et ce qu'il allait faire avec ma face » (participante H). Une autre a émis cette opinion : « Pas certaine que c'est assez commun dans les pratiques d'être monitoré comme ça, un peu à la Big Brother, où tout le monde te surveille. Je sens que ça pourrait soulever ce genre de craintes là » (participante G). Bien que ce commentaire soit directement en lien avec le contrôle, l'autonomie et l'intimité, la participante fait référence à la crainte et donc à l'angoisse que de telles situations pourraient possiblement générer au sein d'une équipe.

D'autres participants, entre autres les participants O, L, E, D et B, ont parlé du sentiment de surveillance et de micromanagement, qui est un stress ajouté à un environnement de travail. C'est un poids additionnel qui peut affecter le bien-être des utilisateurs. Bien que ça soulève clairement un enjeu d'atteinte à l'intimité et à la vie, privé, nous voyons que les enjeux ne fonctionnent pas en silo : le sentiment de surveillance peut causer de l'angoisse, donc un stress chez les utilisateurs et donc porter atteinte au principe de bien-être.

Enfin, certains participants ont mentionné le fait que pour certains utilisateurs plus vulnérables, recevoir un portrait négatif de leurs émotions et de leur temps de travail, « ça pourrait être démobilisant » (participante G).

Ainsi, de façon générale, les participants ont considéré que la technologie d'EmoScienS a été développée pour améliorer le bien-être des utilisateurs, autant du point de vue individuel que dans un cadre organisationnel. Certains participants ont cependant soulevé que le contexte du déploiement de la technologie peut miner le bien-être. Si l'utilisation est faite dans un contexte où l'objectif est d'augmenter la productivité ou de faire de la supervision de proximité, le déploiement de la technologie au sein d'une organisation peut susciter au contraire des craintes, de l'anxiété et du stress, ce qui va à l'encontre du principe de bien-être et peut mener à une démotivation. Aussi, si la technologie présente des tableaux de bord négatifs de façon régulière à des utilisateurs qui se sentent au contraire heureux, il y a un risque que cela affecte leur bien-être. En résumé, le développement de la technologie est aligné avec le principe de bien-être, mais le contexte de son déploiement doit être encadré pour être conforme aux valeurs mises de l'avant par EmoScienS.

4.2.4.2 Principe de respect de l'autonomie

Nous avons recensé des commentaires en lien avec le respect de l'autonomie lorsque les participants ont mentionné de différentes façons le fait que la technologie d'EmoScienS pourrait faciliter la prise de conscience des utilisateurs face à certains éléments en lien avec leur profil émotionnel. Cette prise de conscience mène alors à la possibilité d'apporter des changements et de prendre le contrôle sur leur expérience émotionnelle.

Une participante a aussi souligné le fait qu'un tel outil peut amener du conflit dans un contexte organisationnel. Si un employé n'a pas envie de parler de ses émotions et qu'ils sont disponibles sur son rapport, ça peut causer des frictions (participante H).

Le participant I, pour sa part, spécifie que la présence de la technologie lui a fait prendre conscience de l'importance de ses émotions au travail :

Écoute, je dois te dire qu'EmoScienS m'a mis conscient de l'importance des émotions dans le travail, je me suis dit : « oui, c'est vrai, effectivement, les émotions doivent affecter le travail. » Juste le fait d'avoir eu la question et d'être dans la situation m'a fait réfléchir à certaines choses, effectivement (participant I).

La technologie ne change pas son niveau de bien-être, mais la prise de conscience lui permet d'agir en conséquence.

La participante F donne le même type d'argument en expliquant que la technologie ne lui a pas apporté du bien-être en tant que tel, mais qu'il revient au participant de prendre connaissance de son profil émotionnel et de mettre en place des actions concrètes selon la teneur des résultats.

La participante D croit que l'outil pose un diagnostic, ce qui permet à l'utilisateur de poser des gestes concrets en conséquence. Le tableau de bord donne une forme de contrôle à l'utilisateur sur la gestion de son bien-être.

Les commentaires en lien avec l'autonomie dans le contexte organisationnel, quant à eux, vont plutôt dans le sens d'une certaine forme de perte d'autonomie en lien avec un sentiment de surveillance et de micromanagement. La participante E a déclaré : « On veut que les gens performant, il y a toujours tout plein de discussions autour de ça, là tu as ton temps moyen d'écran, les pauses d'écran, j'ai l'impression que les gens se sentiraient micromanagés ». Tandis que la participante G a fait référence à un sentiment d'obligation à se soumettre à une telle technologie en organisation : « Ton patron qui t'oblige à avoir ça, je suis pas mal certaine qu'il y aurait du mécontentement quelque part ».

Certains participants sont allés encore plus loin et ont évoqué la possibilité d'une utilisation abusive et malveillante de la technologie par la voie de la manipulation émotionnelle via la technologie. Le participant N l'a verbalisé comme ceci : « On va essayer de maximiser tes émotions pour engranger du profit » et « Tu donnes beaucoup d'informations, donc tes supérieurs vont avoir plus de facilité à te contrôler ». Le participant I a soulevé le même type de préoccupations par rapport à de l'abus et à de la manipulation « La technologie peut amener à de l'abus par rapport à l'information émotionnelle. Si le patron était capable de voir que tu étais de plus en plus malheureux, *he could keep up the pressure, if he's putting more and more pressure on you to make you quit* ».

Pour sa part, la participante A a soulevé un aspect technique de la technologie qui respecte l'autonomie des utilisateurs : « le système était bien parce qu'il était en sourdine, qu'il n'est pas dérangent et qu'il ne t'envoie pas de *pop-up* durant ta journée ». C'est en effet un aspect de la technologie qui respecte l'autonomie de ses utilisateurs puisqu'elle n'a pas de technique de captation d'attention.

4.2.4.3 Principe de protection de l'intimité et de la vie privée

Le principe de protection de l'intimité et de la vie privée a été largement discuté. La nature de la technologie EmoScienS invite bien entendu à aborder un tel sujet.

De façon générale, les participants se sont dits à l'aise avec la technologie en lien avec la protection de leur intimité et de leur vie privée. Deux participants ont répondu considérer la technologie comme étant intrusive (participante D) ou avoir ressenti un malaise à être observé (participant L). Les autres participants se sont dits généralement à l'aise, faisant des commentaires comme « je pense qu'on est filmé partout maintenant, ça fait partie de la vie en général » (participant A), « Je sais que c'est confidentiel, et c'est utilisé à des fins de recherche, donc c'est pour une bonne cause » (participante B) et « on dirait que ça ne me dérangeait pas de me faire filmer en travaillant » (participante E). D'ailleurs, plusieurs participants ont fait référence au fait de ne pas se sentir dans leur intimité et dans leur vie privée, à la base, lorsqu'ils travaillent. Pour la participante H, au début, elle trouvait le processus intrusif, mais avec le temps et avec le fait de voir les photos disparaître, ce qui lui donnait le sentiment qu'une fois la photo analysée, cette dernière était détruite, elle s'est sentie rassurée (participante H). Le participant I s'est assuré d'avoir un environnement assez neutre et il a jugé qu'« au bout de la ligne, c'est quand même minimalement intrusif », et à quelques nuances près, cette opinion trouvait écho chez les autres participants. Bien qu'il n'ait pas senti que la technologie était intrusive dans sa vie privée, le participant O a cependant dit avoir eu le sentiment ici et là qu'il était tout de même vulnérable devant son écran : étant dans l'intimité de son domicile, il pouvait parfois passer en sous-vêtements devant son écran, ce qui l'a fait réfléchir. Le participant I met en garde contre l'utilisation malveillante : « Lorsqu'on a le profil visuel, émotionnel, les habitudes de vie de quelqu'un, si quelqu'un s'en sert à mauvais escient, ça peut être catastrophique ».

Enfin, la participante D a apporté une nuance à sa réponse en parlant du contexte dans lequel la technologie est utilisée, tout comme la participante B. Toutes deux ont abordé la question de l'utilisation de la technologie en organisation. Elles ont spécifié que leur réponse serait différente si l'utilisation était faite dans le cadre de leur organisation.

Lorsque nous avons abordé la question plus spécifiquement de l'utilisation d'EmoScienS dans un cadre organisationnel, les participants se sont montrés sensibles à l'aspect de la

protection de leur intimité et de leur vie privée. Dans le tableau 5.5, dans la section 5.2.6, nous aborderons le confort des participants quant au partage de leurs informations avec leur patron. Des 14 participants, 9 ont répondu ne pas vouloir partager leurs informations émotionnelles ou encore vouloir mettre des conditions à ce partage. Lorsque l'information est anonymisée, 3 participants refusent toujours de partager l'information avec leur patron. Ainsi, la question du contexte influe grandement sur le sentiment de protection de l'intimité et de la vie privée en lien avec la technologie. Il y a une forme de gradation dans l'intrusion. Si la plateforme est utilisée dans un cadre individuel hors organisation, de façon générale, les individus sont à l'aise. Si c'est utilisé en organisation, mais que les données sont anonymisées, la majorité des participants sont à l'aise. Cependant, si les données non anonymisées peuvent être partagées avec leur patron, il y a un fort sentiment d'intrusion dans la vie privée.

Dans le milieu organisationnel, certains participants mentionnent que les gens ont le droit de garder privées leurs émotions et de choisir de les exposer ou non selon les contextes. La terminologie en lien avec la plateforme dans le milieu organisationnel passe à « intrusif », « surveillance de proximité », « invasif », « contrôle » ou « voyeurisme ». L'anonymisation des données et leur agglomération au sein d'un groupe sont donc importantes pour tempérer le sentiment d'intrusion de chacun.

La participante E met en évidence le fait que l'utilisation d'une telle technologie dans la sphère organisationnelle ne peut prendre que le pouls du bien-être spécifique à l'environnement de travail ou aux conditions en entreprise. Elle peut être invasive dans la sphère personnelle et privée aussi puisqu'il est difficile de faire la part des émotions de la vie privée et de la vie professionnelle dans une journée. Nous pouvons traîner des difficultés familiales qui nous affectent durant nos journées sans que ce soit relié à notre environnement de travail. Bien qu'elle soit privée, cette sphère est exposée malgré tout dans le cadre de l'utilisation d'une technologie de détection des émotions. Certains participants soulignent le côté intime de l'émotion : « l'émotion, c'est quelque chose de très personnel » (participant I) et « l'émotion, c'est très privé » (participante H).

Ainsi, bien que la technologie d'EmoScienS soit perçue comme étant respectueuse de la protection de la vie privée et de l'intimité dans un contexte d'utilisation individuelle de la

technologie, dès qu'il est question d'utiliser cette dernière dans un contexte organisationnel, le groupe ressent un malaise. Le fait que les données soient anonymisées et agglomérées au sein d'un groupe d'employés est rassurant. Cependant, certains participants restent méfiants quant à la possible réidentification ou à la possibilité de fuites d'informations.

4.2.4.4 Principe de solidarité

Concernant le principe de solidarité, les deux aspects principalement soulevés durant les entrevues avec les participants sont les risques de miner la confiance au sein d'une relation gestionnaire-employé et celui d'une perte de contact humain liée à l'utilisation de la technologie.

Le sentiment de supervision de proximité et de micromanagement peut miner la confiance dans la relation entre l'employé et son gestionnaire. De plus, la crainte d'une utilisation malveillante, comme la manipulation de la part de certains patrons mal intentionnés qui pourraient utiliser les informations au détriment de certains employés (participant I), pourrait engendrer de la méfiance au sein des utilisateurs par rapport à leur patron.

La perte de contact humain, pour sa part, est exprimée entre autres par rapport à la prise de mesures de tout et dans ce contexte-ci, à la prise de mesures des émotions. Le participant O l'exprime ainsi : « Je suis convaincu que c'est un des grands débats en intelligence artificielle, tu perds un peu le côté humain dans la gestion, ça devient vraiment des décisions qui sont toutes basées sur des données ». La participante H le dit ainsi : « Je préférerais le contact humain, c'est toujours mieux ». La participante A, qui est elle-même gestionnaire, dit qu'elle prend chaque semaine du temps avec ses employés et qu'une des questions qu'elle leur pose est justement « Comment ça va? », mais que plusieurs de ses collègues peinent à rencontrer leurs employés régulièrement et encore plus à leur demander « Comment ça va? ». Est-ce que la technologie aiderait à ouvrir la discussion ou l'enlèverait des mains du patron?

Le principe de solidarité n'a pas été un élément très développé par les participants à l'exception de ce sentiment d'intrusion dans le milieu professionnel quand leurs informations ne sont pas anonymisées. De façon générale, l'outil a été perçu comme étant

un complément pour aider à l'analyse, à la prise de conscience de la situation émotionnelle d'un individu ou d'un environnement de travail. Le participant K dit : « Je pense que c'est une méthode complémentaire qui peut effectivement aider à adopter des pratiques corporatives qui vont augmenter le moral de ses employés ».

4.2.4.5 Principe de participation démocratique

Les participants ont généralement compris comment fonctionnait la technologie au sens large, c'est-à-dire le principe de prendre des photos, de les analyser et de transmettre les données analysées dans un tableau de bord selon les catégories identifiées. Cette portion était simple et particulièrement bien détaillée dans le tableau de bord.

Certains participants ont cependant remis en question la validité des résultats. Ainsi, ils auraient aimé mieux comprendre le processus d'analyse et les critères qui permettent de catégoriser les différentes photos selon la gamme d'émotions de la plateforme. Les algorithmes d'apprentissage automatique n'étant pas transparents, il est plus difficile d'obtenir une interprétabilité ou une explicabilité des données obtenues. Un bon nombre de participants ont cependant mentionné qu'ils auraient apprécié avoir accès à une explication sur le processus d'analyse de l'expression faciale :

« J'aurais aimé ça avoir un petit peu une idée de [...] qu'est-ce qu'ils définissent comme de la colère » (participant O);

« J'aurais aimé comprendre la mécanique derrière » (participant O);

« Je me dis : "À quel point elle est capable de décerner de la tristesse, par exemple?" » (participant K);

« Je n'ai pas compris pourquoi il me montrait en colère tout le temps » (participant J);

« Démontrer, peut-être à partir de la photo : "Regarde, ton œil est un peu plus haut, ta bouche est un peu plus ouverte; nous, on analyse que tu es super content. Peut-être, encore une fois, d'accompagner l'utilisateur à travers le processus d'analyse et lui expliquer ce qui se passe » (participant J);

« Au bout de la ligne, je serais très curieux de voir l'algorithme, qu'est-ce qui se passe derrière tout ça » (participant I);

« Peut-être que l'interprétation de ma face de colère et de ma face concentrée est inexacte, ça rapporte que j'étais en colère, alors que j'étais juste concentrée » (participante E).

Les participantes A et E parlent toutes les deux du fait qu'il aurait été souhaitable que la technologie prenne le temps de personnaliser son analyse en prenant des informations plus spécifiques sur l'utilisateur. Elles suggèrent une personnalisation avec un balayage du visage de l'utilisateur, comme le font les outils de reconnaissance faciale qui nous permettent maintenant d'ouvrir nos téléphones intelligents. Aussi, elles croient qu'un profil plus précis de l'utilisateur, avec des éléments comme son âge, pourrait être une idée importante à prendre en considération :

Il n'y a pas eu d'analyse de mon visage avant, à savoir mon âge, est-ce que j'ai des rides, qu'est-ce que j'ai l'air quand je souris, qu'est-ce que j'ai l'air quand je suis fâchée. Probablement que ça aurait aidé l'intelligence artificielle à personnaliser l'analyse (participante A).

Ainsi, certains participants souhaitaient avoir un peu plus d'explications sur le processus d'analyse des données et sur les résultats.

Il y a eu aussi des préoccupations concernant la transparence, cependant elles étaient plus dirigées vers les organisations qui utiliseront la technologie, pour s'assurer qu'elles expliquent clairement l'objectif de leur démarche en implantant la technologie d'EmoScienS. À ce sujet, le participant K souligne qu'il faudrait « [e]xpliciter l'objectif de la démarche, donc pas simplement juste rassurer les gens, mais leur dire la raison pour laquelle ils font ça ». En ce qui concerne EmoScienS, lorsqu'on a fait référence à la transparence dans les entrevues, c'était en lien avec l'utilisation des données et la confidentialité. Les participants étaient confiants quant à la protection de leurs données et à la confidentialité en grande partie par le fait qu'ils étaient dans un contexte d'étude de recherche et en situation de confiance avec la chercheuse, qui leur avait fourni de l'information sur les mesures en place pour les protéger dans le cadre de leur utilisation de la technologie. Ils ont mentionné que plus grand est le niveau de transparence de la part de l'organisation conceptrice quant à l'utilisation des données et au niveau de confidentialité, plus grande est leur confiance à l'égard de la technologie.

4.2.4.6 Principe d'équité

Les principaux éléments qui sont ressortis en lien avec le principe d'équité sont la relation de pouvoir et la possibilité d'une stigmatisation en lien avec le profil émotionnel. C'est dans le cadre organisationnel que les participants voient davantage les enjeux. Certains

participants ont perçu que si leur patron avait accès à leur profil émotionnel, ceci aurait un effet potentiellement négatif sur la relation de pouvoir en leur défaveur :

« J'ai le sentiment que si ces informations-là étaient transmises à mon boss, je perdrais un petit peu cette barrière-là qui me permet d'avoir une certaine façade quand tu n'es pas d'accord avec un de ses propos ou une de ses décisions » (participant O);

« C'est une forme de paternalisme très intrusive de garder un œil sur ses employés, ça devient de la surveillance de proximité *one on one* qui, selon moi, est abusive » (participant N);

« Je ne pense pas que c'est forcément une bonne idée de partager avec son supérieur hiérarchique direct parce que du coup, ça va influencer la relation de travail, ça va influencer la manière dont la personne nous voit parce qu'on se présente toujours d'une certaine façon devant son supérieur hiérarchique » (participant F).

Certains participants ont aussi souligné le risque de stigmatisation en lien avec les résultats :

« Ça pourrait stigmatiser, étiqueter certaines personnes » (participant N);

« Je pense que dans la culture actuelle, c'est partout comme ça : le moins que tu déplaces d'air, le moins que tu demandes, tu es OK, *you're fine*. Tu coches la case et on passe à une autre étape. Dès que tu remets en question des choses, dès que tu as une opinion qui est un petit peu différente, dès que tu ne te sens pas à ton maximum ou que tu ne performs pas à ton maximum, on te remet en question » (participant A).

4.2.4.7 Principe d'inclusion de la diversité

De façon générale, l'ensemble des participants a considéré que la technologie d'EmoScienS était facile d'utilisation, simple et accessible. Donc, sous cet angle, le principe d'inclusion nous a semblé respecté. Cependant, quelques participants ont soulevé des questions en lien avec la fiabilité des résultats au sein de certains groupes, ce qui a suscité des réflexions sur le principe d'inclusion de la diversité.

La participante F s'est questionnée sur la performance de la technologie en lien avec les minorités :

C'est une question que je me suis posée quand moi, j'utilisais la plateforme parce qu'au bout de quelque temps d'utilisation, je voyais que ça ne s'actualisait pas, donc je me disais : « eh bien, mince, pourquoi ça ne s'actualise pas, alors que je fais tout ce qu'on me demande de faire pour

l'utiliser? » Et je me suis dit : « Est-ce que le logiciel a un problème à identifier mes émotions à cause de ma couleur de peau? » C'est une possibilité. (participante F)

La participante H, pour sa part, s'est questionnée sur l'applicabilité d'une telle technologie à différents types de personnalité. Elle explique que tout le monde ne démontre pas ses émotions de la même façon :

Des personnes comme moi, qui ont encore plus des émotions neutres. C'est pas pratique parce que leurs émotions ne seraient pas bien détectées par la machine, parce qu'il y a du monde qui ont des airs bêtes bien pires que le mien et il n'y aurait peut-être même pas de joie, alors qu'ils sont très heureux dans la vie (participante H).

Le participant I, tout comme la participante H, considère ne pas exprimer ses émotions comme un livre ouvert. Il se questionne sur l'adaptabilité de la technologie aux différentes cultures :

La technologie ne peut peut-être pas s'appliquer ou analyser les gens de la même manière parce que cette technologie-là doit avoir, si on prend quelqu'un de nord-américain qui a vécu toute sa vie à Montréal, qui a une intelligence émotionnelle relativement éduquée, au bout de la ligne, lorsque cette personne-là a des sentiments, elle va avoir des réactions qui vont être dans un certain standard, mais il va toujours y avoir des gens, d'après-moi 20 % de la population, qui sont à l'extérieur de ce standard-là, où la technologie ne peut pas s'appliquer à eux (participant I).

Enfin, la participante A, pour sa part, se questionne sur la capacité de la technologie à percevoir la différence entre une ride de vieillesse et une expression quelconque : « Je serais curieuse de voir une personne âgée, vraiment âgée, pleine de rides, qui est de bonne humeur, qui fait de petits sourires, mais qui a l'air fâchée » (participant A).

Les commentaires des participants portent à réfléchir sur la capacité de la technologie à saisir ces subtilités et à être inclusive de la diversité. Ce sont des aspects intéressants soulevés par les participants et que nous retrouverons dans la littérature sur l'informatique affective.

4.2.4.8 Principe de prudence

De façon générale, dans sa forme actuelle, la technologie n'inquiète pas les participants. Les mesures en place pour la protection des données et le fait qu'il s'agit de photos prises dans un cadre défini semblent des éléments rassurants. C'est plutôt dans son application

que les participants ont mentionné qu'il faut faire preuve de prudence, c'est-à-dire dans l'utilisation qui peut être faite de la technologie.

Les enjeux qui ont été mentionnés et qui méritent un certain niveau de prudence de la part des concepteurs et des organisations sont les risques de piratage et les risques d'utilisation malveillante de la technologie, comme la manipulation émotionnelle, la surveillance ou toute forme d'abus. Des participants (participants A, I, E, H et G) ont aussi mentionné le risque de démobilité chez certains utilisateurs plus sensibles à des résultats négatifs (colère, tristesse, mépris, dégoût). Une certaine prudence est de mise par rapport aux conséquences sur l'état émotionnel que peuvent avoir de tels résultats pour ces utilisateurs plus vulnérables. Aussi, le risque de stigmatisation et d'étiquetage de certains utilisateurs a été soulevé par des participants, tout comme une mise en garde quant à la dépendance ou à la confiance aveugle en de telles technologies.

Un participant (N) recommande la mise en place d'informations sur la marge d'erreur pour mettre en perspective la fiabilité des données, ce qui pourrait avoir un effet bénéfique pour contrer autant la démobilité, la stigmatisation que la confiance aveugle dans les résultats.

4.2.4.9 Principe de responsabilité

Dans le cas de la technologie d'EmoScienS, il s'agit d'une technologie passive qui ne prend pas de décision. Elle est un complément d'information. Nous n'avons pas relevé beaucoup d'allusions, dans le cadre des entrevues, à la responsabilité en lien avec la technologie. Les commentaires qui s'y rapportent sont sous trois formes.

D'abord, la participante B soulève le fait que certains utilisateurs peuvent développer une dépendance aux technologies. Elle met en garde contre le fait de se fier à 100 % au bilan d'EmoScienS et souligne la nécessité de garder un jugement critique. Ainsi, ce à quoi la participante fait référence est un transfert de responsabilité dans l'analyse de ses propres émotions et le danger de mettre notre jugement critique de côté pour se fier à la technologie. Connaître le pourcentage ou la marge d'erreur serait particulièrement bénéfique en matière de responsabilité ici.

De plus, l'utilisation de cette technologie apporte une nouvelle forme de responsabilité aux gestionnaires. La participante D, qui est une gestionnaire, a exprimé ses préoccupations quant au fait que l'implantation d'EmoScienS dans son environnement de travail lui ajouterait une nouvelle tâche, celle du suivi du tableau de bord de la situation émotionnelle de son équipe. Elle a aussi souligné qu'on pourrait développer des attentes envers elle : on pourrait attendre d'elle qu'elle agisse lorsqu'elle voit la condition émotionnelle précaire de certains de ses employés. Elle mentionne alors l'importance de donner le droit aux gestionnaires aussi de choisir ou non de recevoir des données personnalisées ou anonymisées. Pour sa part, elle préférerait avoir accès à un tableau de l'état émotionnel général du groupe plutôt que celui de chaque membre de son équipe.

Enfin, toujours dans le milieu organisationnel, la participante A a abordé le fait que certains de ses collègues consacrent très peu de temps à prendre le pouls de leurs employés et à parler du « Comment ça va? ». L'arrivée d'une telle technologie pourrait potentiellement transférer la responsabilité du gestionnaire vers la technologie. La participante mentionne l'importance de ce contact humain pour échanger avec ses employés, mais est consciente que cela ne fait pas partie des priorités de tous les gestionnaires.

La question de la déresponsabilisation des êtres humains n'a été abordée que par la participante A, quand elle parlait de ses collègues gestionnaires. En effet, la participante a exposé un scénario possible où certains gestionnaires moins enclins d'emblée à prendre le pouls de leurs employés quant à leur bien-être au travail pourraient être tentés de s'en remettre à la révision du tableau de bord et de transférer cette responsabilité à la technologie d'EmoScienS. Sinon, la technologie n'a pas soulevé d'autres enjeux en lien avec le principe de responsabilité.

4.2.5 Réflexions quant au partage d'information dans le cadre de l'utilisation d'EmoScienS

La plateforme d'EmoScienS prévoit que l'utilisateur pourra donner accès aux résultats de son profil émotionnel pour permettre à, par exemple, un patron ou à un département de ressources humaines de prendre le pouls des employés au travail. Nous avons donc questionné les participants sur leur niveau de confort à partager ces informations avec leur

patron. Nous leur avons demandé, dans un deuxième temps, si la possibilité d’anonymiser les informations partagées avec leur patron au sein d’un groupe influencerait leur décision.

Tableau 4.5 Partage des données émotionnelles avec l’organisation

PARTICIPANTS	Patron	ANONYMISATION
A	Oui	Oui
B	Court terme oui, long terme non	Oui
D	Réflexion	Oui
E	Non	Non
F	Non	Oui
G	Oui	Oui
H	Non	Non
I	Oui	Oui
J	Oui	Oui
K	Non	Oui
L	Non	Oui
M	Oui	Oui
N	Non	Non
O	Non	Oui

Les participants qui ont indiqué qu’ils accepteraient de partager leurs informations avec leur patron ont vu le bénéfice de le faire dans l’optique de possiblement améliorer leurs conditions de travail. Une participante s’est dit à l’aise parce qu’elle ne croit pas qu’un tel partage peut menacer son emploi ou sa performance étant donné qu’elle croit qu’ultimement que ce qui compte aux yeux de son patron, c’est que les résultats soient là. Un autre participant a dit qu’il serait à l’aise à ce que son patron ait accès à ses résultats émotionnels pour améliorer ses conditions de travail, mais pas pour pousser ses performances.

Les participants qui refuseraient le partage de leurs données avec leur patron ont mentionné ne pas voir la pertinence ou l’intérêt que celui-ci ait accès à leur profil émotionnel, disent que c’est trop intrusif, et un participant a clairement dit : « Ce n’est pas de ses affaires »

(participant L). Il y a une crainte par rapport au micromanagement et au sentiment d'une atteinte à la vie privée des gens.

Lorsque l'option d'anonymiser les résultats a été offerte aux participants, la moitié de ceux qui étaient originalement réfractaires au partage des données ont indiqué qu'ils l'autoriseraient dans ces conditions. Certains restent toutefois méfiants quant au risque de réidentification de leurs données. Lorsque les informations sont anonymisées au sein d'un groupe, on voit la possibilité de prendre le pouls du groupe et d'en faire ressortir un bénéfice pour l'environnement de travail.

Enfin, 3 participants sont restés réfractaires au partage de ce type d'information même si elle est anonymisée, n'y voyant pas de bénéfice potentiel, considérant qu'il s'agit d'une intrusion dans la vie privée ou encore ayant la crainte d'une fuite d'information.

4.3 Critères d'acceptabilité

Dans le cadre des entrevues, nous avons demandé aux participants pourquoi ils ont accepté d'utiliser la technologie d'EmoScienS et quels sont les critères qui les motivent à accepter d'utiliser une nouvelle technologie de façon générale.

Nouvelles technologies

Lorsque nous avons demandé aux participants les facteurs importants à leurs yeux pour accepter d'utiliser une nouvelle technologie, la protection de la vie privée, de l'intimité et de la confidentialité a été mentionnée par 10 participants sur 14. Le sentiment de sécurité par rapport à ses données personnelles et à sa vie privée est exprimé ainsi : « Tant que c'est sécuritaire et que ça ne brise pas mon intimité. Ce sont mes deux points majeurs » (participant L). Ensuite, 8 participants ont mentionné que l'utilité ou le bénéfice que peut avoir la technologie dans leur vie va influencer leur décision. Plus la technologie répond à un besoin, plus ils seront enclins à accepter de l'utiliser. La participante E spécifie clairement qu'elle se questionne sur « [c]e que ça va [lui]apporter de nouveau, est-ce que ça répond à un de [s]es besoins ». La facilité d'utilisation est aussi un élément important. La participante B dit qu'il faut que

ça soit facile d'utilisation, facile d'intégration dans [s]on mode de vie. Sinon, c'est sûr qu[']elle n'y adhère pas quand c'est compliqué et il faut que ça [lui] serve à quelque chose, il faut que ça ait un bénéfice potentiel. Peu

importe pour quoi [elle] veu[t] l'utiliser, il faut qu'il y ait un but, que ça soit utile, quoi

La moitié des participants ont dit que l'acceptabilité sociale de la technologie est aussi un facteur qui guide leur décision. Ainsi, lorsque la technologie a fait ses preuves auprès de la masse, qu'elle est recommandée par une personne de confiance, qu'elle est adoptée par un grand nombre d'utilisateurs ou qu'elle est grandement démocratisée publiquement, les participants auront tendance à être plus à l'aise de l'utiliser. La participante E présente ainsi ses critères d'acceptabilité : « Est-ce que je connais des gens qui l'utilisent? C'est quoi leurs *feedbacks*? Un petit *screening* des conditions d'utilisation. Si je me suis fait recommander plus, plus, plus, puis que beaucoup de monde l'utilise, on dirait que je vais un peu baisser ma garde ». Pour le participant M, ça pèse aussi beaucoup dans la balance : « c'est vraiment si la technologie est acceptée et reconnue par plusieurs personnes. Ça revient à ce qu'on disait tantôt, ça va automatiquement me mettre en confiance. C'est un gros, gros, gros facteur, je te dirais ». Un autre participant le mentionne comme critère d'acceptabilité, parmi d'autres facteurs :

Je m'assure qu'elle ne compromet pas mon système, à partir du moment où c'est une technologie qui est adoptée par un nombre critique de gens, qui a fait ses preuves, essentiellement qui est connue, qui a une réputation, je sais que ça remplit les critères de base qui, moi, vont me satisfaire (participant K).

Des facteurs comme la réputation de l'organisation qui a commercialisé la technologie, la facilité d'utilisation, le pays d'origine et le niveau de transparence ont aussi été mentionnés par les participants comme ayant une influence sur leur décision quant à l'utilisation d'une nouvelle technologie. La confiance est un élément clé qui est nourri par plusieurs de ces facteurs.

EmoScienS

Deux critères principaux ont motivé les participants à utiliser la technologie d'EmoScienS. Dans un premier temps, la recommandation ou la demande de la part d'une personne de confiance les a influencés positivement. La participante D a spécifié : « Je vais être bien honnête, c'est parce que c'est toi qui me l'as proposé, et que j'ai confiance en toi ». Dans un deuxième temps, mais de façon tout aussi importante, les participants ont été motivés par le fait que l'utilisation se faisait dans le cadre d'un projet de recherche, donc avec

l'objectif noble de faire avancer la recherche dans un environnement rassurant, avec un cadre éthique et strict. Le participant N a mentionné : « parce que c'est dans un cadre de recherche. Genre si c'était Apple qui m'avait demandé de faire ça, j'aurais dit non, c'est sûr ». La participante F, pour sa part, le formule ainsi : « Je pense que c'était la facilité d'utilisation et le fait que ça soit utilisé dans un cadre de recherche, ça m'a un peu plus rassurée, je dirais ». Le contexte est un facteur important, et une participante le dit très clairement :

Que c'était confidentiel, et c'était pour une bonne cause, c'était pour la recherche, mais c'est clair que si mon employeur me disait « Il faut que tu mettes ça pour travailler dans le jour », je vais dire : « Mais pourquoi que je ferais ça, pourquoi est-ce que j'accepterais que tu me filmes pour que tu voies mes émotions? » Il faut qu'il y ait un but, une raison, une étude où on veut analyser une période de temps, c'est sûr, sinon je ne l'utiliserais pas (participante B).

La personne de confiance et le contexte d'utilisation, c'est-à-dire ici l'utilité perçue, soit faire avancer la recherche ou avoir accès à son profil émotionnel, ont largement influencé la décision des participants.

Par la suite, des facteurs comme la facilité d'utilisation, la simplicité et la facilité d'intégration au mode de vie ont été soulevés :

Purement au niveau de la technologie, c'était vraiment un aspect pratique de *hands free*, un truc que je pouvais intégrer à ma vie active professionnelle sans impact, ça n'allait pas m'empêcher d'avoir des réunions Zoom, ça n'allait pas ralentir mon ordinateur ou quoi que ce soit (participant K).

La confidentialité des données et l'origine montréalaise de la plateforme sont également des aspects qui ont influencé certains participants : « Quand j'ai lu les conditions d'utilisation et je me suis renseignée sur EmoScienS et j'ai vu que c'était à Montréal, donc si ça avait été dans n'importe quel autre pays, je ne sais pas. Je sais qu'au Canada, le respect de la vie privée, c'est un petit peu plus surveillé » (participante H). Enfin, la curiosité pour la nouvelle technologie a aussi stimulé les participants à aller de l'avant avec l'utilisation d'EmoScienS. Un des participants a grandement apprécié le fait d'essayer une nouveauté, mais la protection des données était très importante dans son acceptabilité :

Juste le fait que je ne savais pas que ça existait, et de pouvoir l'essayer, je trouvais ça vraiment le *fun*. Et le fait qu'il n'y aurait pas un être humain qui allait être capable de regarder mes images. C'est tout ce que ça me prenait

pour l'essayer. Si tu m'avais dit que quelqu'un aurait pu voir mes images, je ne pense pas que j'aurais voulu » (participant L).

Compte tenu des réponses spécifiques des participants en lien avec la question sur l'acceptabilité, nous avons considéré que le modèle que nous avons proposé comme cadre d'analyse était représentatif des éléments soulevés. Nous avons donc évalué l'acceptabilité d'EmoScienS par rapport à notre cadre d'analyse, résumé dans cette figure que nous reprenons ci-dessous.

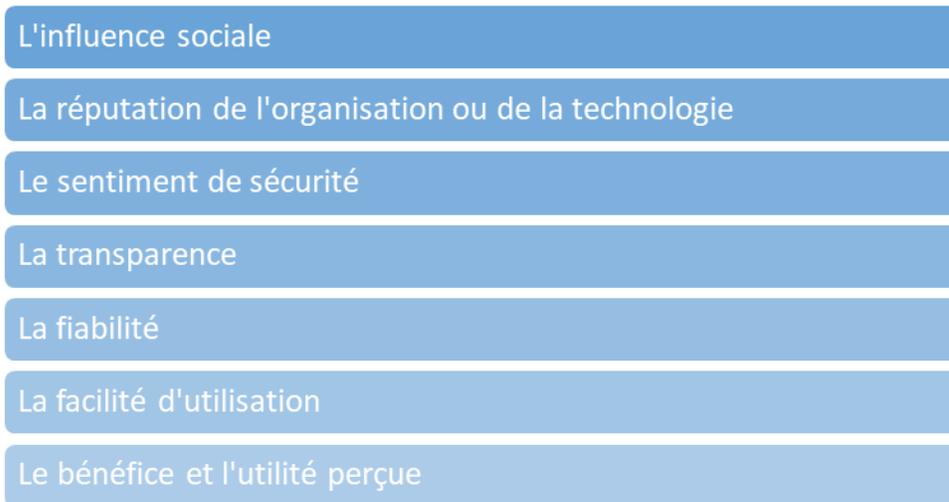


Figure 4.1 Critères d'acceptabilité pour un SIA

L'influence sociale

Les participants ont souligné régulièrement l'influence qu'une personne de confiance ou une utilisation répandue de la technologie a sur leur sentiment d'acceptabilité par rapport à une nouvelle technologie. Devant la possibilité d'utiliser une nouvelle technologie, l'utilisateur cherche des repères pour le conforter dans sa décision. Le fait qu'une personne de confiance lui recommande la technologie, l'utilisation largement répandue de celle-ci ou sa démocratisation publique sont des facteurs qui influencent positivement l'acceptabilité de la nouvelle technologie. Dans le cas d'EmoScienS, les participants avaient relativement peu de repères en ce qui concerne l'informatique affective, ayant été peu exposés à ce type de technologie, mais le fait qu'une personne de confiance, soit la chercheuse dans le cadre de ce projet, les invite à essayer la technologie était rassurant et a

constitué un facteur important pour les influencer positivement. En effet, 8 participants ont fait référence spécifiquement au fait qu'une personne de confiance leur a offert d'essayer la technologie, ce qui les a plus facilement convaincus d'accepter.

La réputation de l'organisation ou de la technologie

La réputation du concepteur est un facteur qui influence l'acceptabilité de la technologie. Le nom, le « *branding* » associé à une nouvelle technologie, lui donne automatiquement l'aura de confiance associée à l'organisation qui le produit. Dans le cas d'EmoScienS, ni la technologie ni l'organisation n'ont jusqu'à présent une réputation bien établie dans la population générale. Ainsi, la référence par une personne de confiance a compensé le facteur de la réputation. Certains participants ont cependant fait leur devoir et exploré le site Internet de l'organisation pour y repérer certains points susceptibles d'augmenter leur acceptabilité, tels que le pays d'origine de la plateforme. Les participantes E, F et H ont mentionné que le fait que l'entreprise soit à Montréal est rassurant. La réputation de la ville et du pays d'origine en matière de normes a joué sur l'acceptabilité de la technologie. De plus, la participante E a considéré rassurant le profil des dirigeants de l'entreprise, qui sont principalement des professeurs et des scientifiques. Cet aspect en lien avec la réputation associée à la profession à ses yeux a augmenté le niveau d'acceptabilité. La conscience éthique de l'organisation a aussi influencé positivement sa réputation et donc le degré d'acceptabilité qui y est associé.

Le sentiment de sécurité

Les utilisateurs souhaitent avoir le sentiment que la plateforme, le site ou la technologie est sécuritaire. Ils sont sensibles à la protection de leurs données personnelles, tout particulièrement quant à la possibilité de fuites de données, de vols d'identité ou de toute autre forme de piratage. L'utilisateur veut se sentir à l'abri de l'éventualité d'une utilisation malveillante de la technologie et avoir le sentiment que sa vie privée est protégée. L'informatique affective peut être plus facilement considérée comme intrusive dans la vie privée, considérant la nature de son rôle. Étant invasive dans l'espace intime des utilisateurs, elle peut éveiller plus facilement un sentiment d'insécurité. Toutefois, une forte majorité de participants ont mentionné s'être sentis à l'aise avec la technologie du point de vue de leur vie privée. Le fait que la technologie prévoit une protection stricte des

données et qu'elle ne prenne des photos que dans un cadre contrôlé est rassurant. Bien que la technologie soit perçue comme étant sécuritaire, le contexte de son utilisation, les intentions des gestionnaires dans un milieu organisationnel et les craintes quant à une réidentification possible mettent un léger bémol au sentiment de sécurité lié à la technologie. Ainsi, le déploiement d'une telle technologie requiert un cadre clair et strict pour maintenir à un bon niveau la confiance liée à celle-ci.

La transparence

Les participants ont indiqué avoir de la méfiance à l'égard des organisations en ce qui concerne la protection et l'utilisation de leurs données personnelles, ayant le sentiment d'avoir peu, voire pas de contrôle sur leurs données dans l'univers du numérique. Certains participants ont communiqué clairement l'importance qu'ils accordent à la transparence des organisations sur la gestion des données personnelles, leurs valeurs et la façon dont les données sont utilisées. Dans le contexte d'utilisation d'EmoScienS, une plus grande proportion des participants ont répondu avoir pris le temps de lire les conditions d'utilisation que pour des nouvelles technologies en général. L'information reçue à ce sujet a été perçue positivement : l'organisation y fait part de l'importance de la protection des données personnelles, ce qui est rassurant pour les utilisateurs. De plus, l'organisation offre aux utilisateurs l'option d'accepter, de refuser ou de personnaliser la gestion des témoins (*cookies*), ce qui est considéré comme étant un facteur important pour le sentiment de protection de la vie privée. La communication transparente de la part d'EmoScienS sur l'utilisation et la protection des données personnelles a été un facteur apprécié. De plus, le fait d'avoir accès aux données (photos) et de pouvoir les gérer a été mentionné comme un élément rassurant. Certains participants ont cependant mentionné qu'ils auraient aimé avoir plus de détails sur le processus d'analyse des photos qui mène aux résultats pour comprendre ce qui détermine les différentes émotions. Les utilisateurs ont exprimé rechercher une plus grande transparence en terme d'explicabilité, ces commentaires ont principalement été en lien avec la validité perçue des émotions aux tableaux de bord et donc de la fiabilité de la technologie.

La fiabilité

Cet aspect est un facteur clé pour déterminer la pertinence de la technologie aux yeux de l'utilisateur. Nous avons remarqué que la fiabilité de la technologie d'EmoScienS est un élément qui a eu un effet important sur le niveau d'intérêt et d'engagement pour certains participants. Seulement 3 participants ont considéré que l'outil a été capable de cerner le portrait de leurs émotions quotidiennes avec justesse. Une participante explique ainsi cet effet :

rapidement, je suis tombée dans le désintérêt quand j'ai compris que ça n'allait pas m'apporter tant que ça et c'était très redondant comme résultat, et c'était pas trop représentatif de ce que je vivais réellement dans ma journée, je sentais qu'il y avait peut-être certaines émotions qui n'étaient pas prises en compte par l'intelligence artificielle, justement (participante G).

Le participant J a exprimé un sentiment similaire : « je n'ai pas compris pourquoi il me montrait en colère tout le temps. Bon, autant je trouvais que l'interface était intéressante, j'étais déçu. Je n'y ai pas porté plus d'attention parce que mon analyse s'est arrêtée au résultat colère et je n'ai pas investigué plus ». Il a précisé qu'« [il] pense qu'[il] aurai[t] apprécié si l'analyse des expressions faciales avait été plus juste (participant J). Ainsi, nous avons remarqué que plus le participant s'identifiait aux résultats, plus il était ouvert à poursuivre son utilisation de la technologie.

La fiabilité n'est pas un critère suffisant pour adopter la technologie, mais c'est un facteur qui influence la confiance à l'égard de la technologie et son acceptabilité. Dans le cas d'EmoScienS, les utilisateurs ont aimé le concept et ont même donné des idées pour participer à l'amélioration continue de la précision de ses résultats. D'ailleurs, certains participants ont aussi tenté d'améliorer leur résultat en portant attention à leur posture ainsi qu'à l'éclairage dans leur environnement. Cependant, nous avons noté que pour les participants dont les résultats n'étaient vraiment pas représentatifs de l'expérience émotionnelle perçue, l'intérêt à poursuivre l'utilisation chutait de façon significative.

La facilité d'utilisation et d'intégration au mode de vie

La technologie doit être simple à apprendre et à intégrer. Elle doit être la plus subtile possible. C'est un des aspects que la grande majorité des participants ont appréciés de la

technologie d'EmoSciens et qu'ils ont mentionnés à de nombreuses reprises en lien avec l'acceptabilité d'une nouvelle technologie. Une vaste majorité des participants ont souligné la facilité d'utilisation de la technologie d'EmoSciens et la facilité de l'intégrer à son quotidien. La version logicielle est plus simple que la version Web en raison de son aspect programmable, mais dans les deux cas, un commentaire récurrent est la subtilité du système qui fonctionnait en « *background* ». Certains petits aspects pourraient être améliorés pour augmenter encore plus la fluidité, cependant, dans l'ensemble, les participants ont trouvé que la technologie était très simple et conviviale.

Le bénéfice et l'utilité perçue

Dans un monde où il existe énormément d'options, l'utilisateur potentiel d'une technologie doit percevoir que cette dernière répond à un de ses besoins pour s'y intéresser. Les gens sont sollicités de partout, donc pour décider d'intégrer une nouveauté à leur vie, ils doivent en voir l'avantage et la pertinence dans leur quotidien. La nouvelle technologie doit combler un besoin ou y répondre mieux qu'une autre option déjà utilisée. C'est la même logique qui s'applique dans le contexte organisationnel. La nuance principale, c'est que pour avoir une plus grande acceptabilité de la technologie en entreprise, l'employé doit en voir le bénéfice pour lui et non pour l'organisation. Ainsi, un outil de travail qui demande beaucoup de temps et d'effort pour son intégration à la vie de l'employé sans qu'il lui apporte directement de bénéfice personnel risque de ne pas être accueilli de la même façon qu'un outil qui permet à l'employé de bénéficier directement de ses retombées. Dans le cadre d'EmoSciens, les participants ont perçu des bénéfices personnels et organisationnels. Le tableau 4.4 en fait le résumé. Ils ont toutefois mis un bémol à son utilisation dans un cadre organisationnel, ressentant un certain inconfort quant au risque de supervision de proximité qu'elle soulève. Cependant, ils y ont vu de beaux bénéfices pour améliorer le bien-être en entreprise, une quête positive pour tous. Au point de vue individuel, on perçoit cette technologie comme un outil pertinent pour permettre une certaine prise de conscience émotionnelle et pour favoriser la mise en place de mesures concrètes pour améliorer son bonheur.

De façon générale, la technologie d'EmoScienS bénéficie d'un bon niveau d'acceptabilité, cependant les participants ont soulevé qu'elle pourrait être encore peaufinée en ce qui concerne la précision de ses résultats.

4.4 Perception par rapport aux initiatives et aux mesures mises en place pour encadrer l'intelligence artificielle

Nous avons demandé aux participants leur perception par rapport aux mesures et aux initiatives mises en place jusqu'à maintenant pour encadrer l'intelligence artificielle. Nous voulions avoir un portrait du niveau de connaissances et du sentiment de sécurité des participants par rapport au cadre réglementaire de la technologie.

Tableau 4.6 Perception par rapport à l'encadrement de l'intelligence artificielle

PARTICIPANTS	PERCEPTIONS PAR RAPPORT À L'ENCADREMENT DE L'IA
A	Il n'y a pas assez de réglementation, « c'est très, très <i>free</i> ». Les réglementations en place sont-elles mises en application? Est-ce surveillé? C'est quoi les conséquences? Il y a très peu de jurisprudence, donc c'est flou. « En tout cas, je ne me sens pas protégé en tant que citoyen ».
B	<ul style="list-style-type: none"> – C'est difficile de mettre des balises en place, cependant la participante est confiante qu'elles vont être établies. – Considère qu'il n'y a pas beaucoup d'encadrement en ce qui concerne la sécurité des données. – C'est un secteur très vaste. – Sentiment que c'est le « <i>free for all</i> », et tout particulièrement pour les gros joueurs, qui peuvent faire ce qu'ils veulent : « on a rarement notre mot à dire sur leur façon de fonctionner ».
D	« On en sait peu de choses, c'est comme toute nouveauté ».
E	La participante réalise qu'elle ne sait pas à quel point c'est réglementé. Elle a cependant « l'impression qu'il y en a plein de coupables, qui ne sont pas réglo du tout ». Elle se questionne sur la limite entre éthique et non-éthique, mais considère que « vu qu'on est tellement dépendant à la technologie, on embarque dans le panneau quand même ».
F	Mitigée sur l'effet des réglementations déjà en place (fait référence au RGPD). Elle se doute que les gros trouvent les « <i>loopholes</i> » pour les contourner. Elle considère que les mesures en place ne sont pas suffisantes, mais qu'ils font toujours le maximum puisqu'« on est toujours en retard sur les technologies au final ».

G	Considère qu'elle n'est pas assez informée pour répondre à cette question.
H	« En général, je trouve que ce n'est vraiment pas géré. Ils ont trop de laisser-aller » La participante mentionne toutefois la complexité de réglementer sur le plan international, ce qui, à ses yeux, mène à des règles un peu passe-partout sans contraintes réelles. Elle considère que la réglementation n'est pas suffisante, mais se questionne sur ce qui peut être fait de plus, considérant la complexité.
I	Le participant considère que « c'est difficile de légiférer rapidement sur des choses comme ça », mais que les gouvernements sont sur la bonne voie : « <i>politicians will never be able to be as quick as business and technology</i> », il faut donc donner un cadre dans lequel évoluer en posant des limites claires. Les mesures ne sont pas suffisantes.
J	Aux yeux du participant, pour l'usage qu'il en fait, il pense que c'est bien encadré. Il considère cependant que c'est un milieu très vaste qu'il faut encadrer de façon très sérieuse pour éviter les dérapages, la perte de contrôle et surtout l'exploitation des individus.
K	Le participant considère qu'il ne peut pas répondre à cette question parce qu'il n'en connaît pas suffisamment. À ses yeux, on en est aux balbutiements et on n'a pas encore « laissé la bête sortir ».
L	Le participant juge qu'il ne connaît pas les mesures pour encadrer l'intelligence artificielle. Il considère que plus les années vont passer, plus il y aura de règles en place. Les applications doivent cependant nous demander la permission en lien avec certaines activités, par exemple : ai-je le droit d'utiliser ta localisation?
M	« Je crois que c'est trop récent et j'ai juste peur que le goût et le désir d'avancer côté technologique fasse en sorte qu'on ne grandisse pas parallèlement, que les mesures préventives ne grandissent pas aussi rapidement en parallèle. »
N	Le participant considère qu'il faut être beaucoup plus proactif que réactif. La législation, c'est généralement lent. Cependant, il va falloir être beaucoup plus proactif avec l'intelligence artificielle avant qu'elle nous dépasse. « Ce n'est pas assez réglementé, c'est ce que je pense ». Il croit que la population générale n'est pas très informée sur l'encadrement de l'intelligence artificielle.
O	Le participant considère que c'est un pas dans la bonne direction. Il ne pense pas être assez « calé » sur le sujet pour pouvoir se prononcer cependant. Il aimerait voir une législation beaucoup plus stricte, avec des conséquences plus importantes. Il a le sentiment qu'il est difficile de contrôler la protection des données.

Les participants ont pour la plupart mentionné n’avoir que très peu ou pas de connaissances en la matière, mais tout de même avoir le sentiment que ce qui est fait est insuffisant. Il y a la perception que c’est très peu réglementé, que c’est « très, très *free* », « le *free for all* » et qu’il y a « trop de laisser-aller ». Certains participants ont mentionné être conscients que c’est très complexe à réglementer et que c’est particulièrement difficile de mettre en place un cadre en amont, considérant la vitesse exponentielle de développement de la technologie et la lenteur des processus législatifs. Un participant a aussi mentionné le fait que la population générale n’est pas très informée sur l’encadrement de l’intelligence artificielle. Le sentiment général est qu’il faut continuer à mettre en place un cadre clair pour éviter les abus et les dérapages.

4.5 Mesures recommandées par les participants pour une IA éthique et responsable

Nous avons demandé aux participants : « Quelles mesures devraient être mises en place afin de s’assurer que les entreprises développent des technologies éthiques et responsables en intelligence artificielle? »

Tableau 4.7 Résumé des mesures recommandées par les participants pour assurer une intelligence artificielle éthique et responsable

Réglementer le développement avec la mise en place de standards – pour le moment, n’importe qui peut créer et rendre accessibles de nouvelles technologies.
Mettre de l’avant la bienfaisance et proscrire la malfaisance.
Établir une réglementation sévère avec des conséquences.
Encadrer les technologies avec un code d’éthique que chaque entreprise devrait s’assurer de diffuser, d’implanter, en plus d’avoir un suivi d’adhésion auprès des employés.
Exiger de meilleurs contrôles par la voie d’audits pour assurer le respect des règles (et établir les règles).
Faire collaborer les entreprises avec les régulateurs et les utilisateurs dans le développement de nouvelles technologies – s’assurer d’être dans les règles et de participer au développement de nouvelles normes.
Augmenter le nombre d’experts en cybersécurité et la formation sur le sujet.

Mettre en place des agences de vérification pour certifier la conformité des technologies, comme ISO pour les standards internationaux (organisme régulateur).
Établir un code de conduite universel applicable à tous.
Exiger une plus grande transparence.
Influencer le changement à partir des bancs d'école, donc formation en éthique de l'intelligence artificielle pour les ingénieurs et les spécialistes du milieu.

De façon générale, les participants ont soulevé le besoin de mettre en place un cadre plus strict, plus sévère et surtout avec des conséquences, pour en assurer un plus grand respect. Ils ont parlé de législations, de règles, de code de conduite universel ou de code d'éthique. Ils veulent plus de transparence et de collaboration de la part du milieu pour avoir une plus grande confiance en l'industrie. Ils ont soulevé l'importance d'avoir des experts en cybersécurité pour un suivi plus étroit au sein des organisations. Certains participants ont suggéré la création d'agences de vérification indépendantes pour assurer la conformité des technologies pour la mise en marché et pour un suivi par des audits réguliers pour le respect continu des normes et pour déceler des anomalies possibles. Il a aussi été mentionné qu'il serait important d'avoir un département au sein des gouvernements pour gérer l'encadrement de l'industrie, qui est particulièrement vaste et complexe. Certains participants ont souligné l'importance de l'éducation pour s'assurer que nos valeurs, la bonne conduite et l'éthique de l'intelligence artificielle soient enseignées aux développeurs de demain.

5 Interprétation des résultats et recommandations

Les entrevues et l'analyse de l'information recueillie portent à de nombreuses réflexions. Dans cette section, nous élaborerons sur différentes constations en lien avec l'intelligence artificielle, les enjeux éthiques et les critères d'acceptabilité et plus spécifiquement sur l'informatique affective. Nous ferons le pont avec la littérature, la technologie d'EmoScienS et la Déclaration de Montréal et nous vous partagerons une série de recommandation.

5.1 Intérêt et littératie par rapport à l'intelligence artificielle

Les participants se positionnent généralement de façon favorable par rapport aux nouvelles technologies pour ce qui est de leur intérêt, tout particulièrement sur le plan professionnel. Cependant, une forte majorité du groupe s'est identifiée comme faisant partie des utilisateurs qui adopteront une nouvelle technologie lorsqu'elle aura fait ses preuves, qu'elle aura été adoptée par un grand nombre de personnes et que les premiers « *glitches* » auront été réglés. Nous avons demandé aux participants de nous définir l'intelligence artificielle dans leurs mots. Bien que des références à la technologie, notamment l'automatisation, la collecte et l'analyse de données et la reconnaissance de *patterns*, aient été faites, nous avons pu soulever le flou qu'il existe sur ce que représente l'intelligence artificielle aux yeux du citoyen.

Les participants ont été relativement ambivalents par rapport à la technologie. Ils ont évoqué des bénéfices organisationnels, tels que l'optimisation, l'efficacité, la productivité et l'avantage que représente l'automatisation en situation de pénurie de main-d'œuvre. Ils ont vu les possibilités qu'offre la technologie pour aider l'humanité en lien avec différents enjeux tels que la conquête de l'espace, la faim dans le monde ou les développements dans le milieu de la santé. Cependant, ils ont aussi exprimé clairement leur méfiance par rapport à la manière dont elle peut être utilisée.

5.2 Enjeux éthiques liés à l'IA

L'intelligence artificielle suscite des sentiments mitigés. Les participants ont abordé plusieurs des enjeux présentés dans la revue de littérature. En particulier, les enjeux de

sécurité et de protection de la vie privée ont soulevé beaucoup de discussions, tandis que les enjeux tels que la fiabilité, la transformation du marché de l'emploi, la transparence, l'autonomie et la responsabilité ont été mentionnés de façon beaucoup plus variable au sein du groupe. Il est à noter que les participants n'ont pas fait référence aux enjeux de biais et de discrimination ou d'explicabilité lorsque nous leur avons demandé de nous exposer les mises en garde qu'ils auraient à faire par rapport à l'intelligence artificielle.

La sécurité et la protection de la vie privée sont les deux enjeux les plus abordés quand il est question de mises en garde par rapport à l'intelligence artificielle. Les craintes soulevées touchent principalement les fuites de données, le vol d'identité, ou les mauvaises intentions, qui mènent par exemple à du piratage dans le monde du numérique. L'intelligence artificielle est aussi associée au sentiment de surveillance et d'intrusion via entre autres des technologies comme la reconnaissance faciale ou les assistants virtuels, mais elle est aussi associée au sentiment d'obligation de partager des données personnelles dans certains contextes.

La population québécoise a récemment été ébranlée par un incident important en lien avec la sécurité numérique lorsque des millions de clients de Desjardins ont été victimes d'un vol de données au sein de la coopérative financière. L'étendue de cette brèche de sécurité a été majeure et, comme le disent Ghafur *et al.* (2020), lorsqu'elle est exposée à un événement en lien avec la confidentialité des données, la population tend à développer une plus grande sensibilité par rapport à cet enjeu. Dans le cadre de ce projet de recherche, nous pouvons valider ce point puisque certains participants ont mentionné clairement des incidents liés à des fuites de données personnelles pour expliquer la méfiance qu'ils ont à l'égard de la technologie et des questions de sécurité entourant la gestion de leurs données.

Certains enjeux ont été abordés plus que d'autres, cependant nous ne serions pas prêts à établir le niveau d'importance des enjeux proportionnellement au nombre de fois qu'ils ont été mentionnés. Nous nous questionnons plutôt sur le niveau de littératie et d'exposition de la population aux enjeux de l'intelligence artificielle. Le groupe de participants étant relativement homogène en matière de profils culturels et issu en majorité d'une population largement représentée au sein des banques de données, il est possible qu'il soit moins sensible à des enjeux tels que les biais et la discrimination. Ceci étant dit, une population

plus susceptible de subir de la discrimination pourrait tout de même ne pas soulever ces enjeux si elle n'est pas renseignée quant à l'existence des biais algorithmiques et à leurs conséquences.

Une absence notable dans les mises en garde soulevées par les participants est la protection de l'environnement. L'enjeu n'a pas été abordé dans le cadre de notre revue de littérature, cependant nous avons remarqué qu'aucun participant n'y a fait référence au moment d'énumérer ses mises en garde en lien avec l'intelligence artificielle. Une fois de plus, nous ne pouvons conclure qu'il y a un manque d'intérêt de la part des participants pour cet enjeu, mais il serait intéressant de valider la littérature de la population entourant les répercussions environnementales des technologies du numérique.

Le consentement est un enjeu important quand on parle d'intelligence artificielle. Nous avons interrogé les participants pour savoir s'ils prenaient le temps de lire les conditions d'utilisation avant de les accepter pour accéder à une plateforme numérique ou utiliser une nouvelle technologie. Une vaste majorité des participants ont répondu ne pas lire ou lire très rarement et en diagonale les informations disponibles avant d'accepter. C'est un aspect troublant dans l'ère actuelle de l'intelligence artificielle, qui est encore à la phase de développement de son cadre réglementaire. Les participants ont expliqué qu'ils ne prennent pas le temps de les lire parce que les textes sont trop longs, écrits trop petits, dans un langage technique qu'ils qualifient de jargon. Ils ont le sentiment que les textes sont relativement tous du pareil au même de toute façon. Cependant, comme le dit Turner (2019), considérant l'ampleur de l'utilisation de ce type de contrat auprès du public non expert, il est capital de créer un filet de sécurité pour protéger les utilisateurs afin d'éviter qu'ils abandonnent des droits dont ils ne comprennent pas la pleine portée. En effet, les participants ont exprimé quelle était la limite de leur compréhension quant au langage technique de ces documents. Il y a une asymétrie de pouvoir entre l'organisation qui offre le service et établit ses conditions, et l'utilisateur qui veut bénéficier des services, mais ne comprend pas forcément l'engagement qu'il prend en acceptant les conditions. Certains participants ont émis l'opinion selon laquelle les textes sont présentés de la sorte de façon délibérée pour « assommer » les utilisateurs et glisser des conditions sous le radar. Ainsi, nous pouvons constater qu'ils perçoivent un manque de transparence intentionnel de la part

des organisations qui ont une telle pratique et que cela génère un sentiment d'impuissance et de résignation chez les utilisateurs dans de telles situations. Il y a une désensibilisation par rapport à ces documents, qui sont maintenant monnaie courante. À force d'être exposé à une telle pratique, l'utilisateur baisse sa garde et accepte de façon automatique la situation. Il y a aussi le sentiment de ne pas avoir le choix d'accepter pour avoir accès à l'information ou au site en question. La *Déclaration de Montréal pour un développement responsable de l'IA* se prononce ainsi sur ce point : « L'utilisation par des particuliers de SIA et de services numériques ne peut être conditionnée à l'abandon de la propriété de ses données personnelles. » (Université de Montréal, 2018)

5.3 Enjeux éthiques soulevés par l'intelligence artificielle affective dans le contexte de l'utilisation d'EmoSciens

Les enjeux éthiques en lien avec l'informatique affective soulevés dans la littérature ont tous été évoqués dans le cadre des entrevues : la fiabilité des résultats, la discrimination, la stigmatisation, l'étiquetage, le risque de causer du tort aux utilisateurs, la possibilité de manipulation, l'autonomie et l'aspect intrusif dans la vie privée.

La fiabilité des résultats est un enjeu important en intelligence artificielle affective. Comme le mentionnent Feldman Barrett *et al.* (2019) et Landowska (2019), le « décodage » de l'émotion par l'analyse faciale des expressions est sujet à de nombreuses erreurs. Il existe une grande subtilité dans l'expression des émotions d'un individu à l'autre (Bullington, 2005) et, comme Brigham (2017) le mentionne, une variabilité culturelle. Les participants ont justement soulevé ces enjeux en évoquant ces subtilités en matière de variabilité de l'expression des émotions au sein de différentes cultures, d'un individu à l'autre ou même selon l'âge. L'interprétation des expressions faciales est pleine de subtilités; un froncement de sourcils ne signifie pas automatiquement de la colère. La perception de la fiabilité dans le contexte de l'informatique affective est liée non seulement aux résultats, mais aussi à la transparence en ce qui concerne le contexte optimal de performance et l'applicabilité à différentes populations, et est aussi liée à un certain niveau d'explicabilité sur les critères d'analyse des données. L'indication de la marge d'erreur, recommandée par certains chercheurs et un participant, représente une belle solution dans un contexte d'intelligence artificielle opaque.

Von Walter, Kremmel et Jäger (2022) ont statué que plus les habiletés requises de la part d'un système d'intelligence artificielle sont de l'ordre du qualitatif et de l'intuition, moins l'humain aura tendance à faire confiance au résultat, contrairement à si les habiletés requises sont plutôt quantitatives et complexes. Ainsi, il est possible qu'un utilisateur doute d'emblée qu'une technologie puisse évaluer avec justesse les émotions, considérant que la subjectivité et la subtilité relèvent plus de l'intuition humaine.

Nous avons pu noter de la confusion de la part de certains utilisateurs quant à leur propre état affectif, un questionnement. C'est d'ailleurs une inquiétude de Beavers et Slattery (2017) que l'utilisation de l'intelligence artificielle affective puisse amener les individus à remettre en question leur intelligence émotionnelle. C'est un aspect à prendre en considération, qui peut non seulement affecter l'autonomie, mais aussi causer du tort pouvant aller jusqu'à affecter l'humeur, un enjeu soulevé autant par Daily *et al.* (2017) que certains utilisateurs.

Les enjeux de sécurité et de protection de la vie privée ont été abordés par rapport à la manipulation et à la surveillance. La technologie d'informatique affective d'EmoScienS n'a pas été perçue comme étant intrusive ni pouvant porter à une intention de manipulation, cependant le contexte dans lequel elle peut être utilisée est sujet à de l'abus. Dans le cadre organisationnel, on se méfie de la surveillance, du micromanagement, de la manipulation, de l'intrusion dans la vie privée et de la discrimination par la voie, entre autres, de la stigmatisation et de l'étiquetage. L'idée de surveiller la sphère émotionnelle de ses employés dans un cadre professionnel laisse une impression de paternalisme intrusif. Il s'en dégage une perception de surveillance de proximité, qui peut être abusive aux yeux du personnel. Ainsi, bien que l'intention des concepteurs soit moralement juste puisqu'elle vise le bien-être en entreprise et l'utilisation des données par et pour les utilisateurs, il existe une méfiance à l'égard de la visée première des gestionnaires qui déploient de tels systèmes. Les systèmes de détection des émotions devraient avoir un cadre éthique et réglementaire strict prévenant les abus et rassurant les utilisateurs sur leur portée.

5.3.1 EmoScienS à travers les yeux de la Déclaration de Montréal

L'utilisation d'un cadre éthique tel que la *Déclaration de Montréal* pour évaluer une technologie permet de mettre en perspective les enjeux sensibles que le SIA peut soulever.

Elle est un outil très pertinent, autant pour le concepteur qui souhaite offrir une technologie éthiquement responsable que pour une organisation qui souhaite déployer un système en minimisant les enjeux auprès de ses employés et/ou de ses consommateurs. Dans le cadre de notre analyse de la technologie d'EmoScienS, nous avons utilisé la *Déclaration de Montréal*. Le document se prête relativement bien à l'évaluation des enjeux éthiques d'une nouvelle technologie d'intelligence artificielle. La *Déclaration* est écrite d'une telle façon qu'elle peut cibler de nombreuses situations et porte ses utilisateurs à réfléchir. Large et inclusive, elle permet de couvrir un grand éventail d'enjeux. Cependant, elle peut tout de même être limitante dans l'énumération des différents aspects compris sous chaque principe. Dans le cas de la technologie d'EmoScienS, nous avons remarqué qu'un enjeu qui a été soulevé ne semblait pas cadrer spécifiquement sous un principe. Si une organisation venait à faire appel à une technologie de détection des émotions et qu'elle pouvait, avec le consentement des utilisateurs, avoir accès au profil émotionnel de ses employés ou même de ses consommateurs, que devons-nous penser de la responsabilité qui lui reviendrait si elle percevait la détresse chez quelqu'un? Aurait-elle la responsabilité d'agir d'une quelconque façon? Qu'en est-il des gestionnaires qui recevraient le rapport sur l'état émotionnel de leurs employés, devront-ils agir? Le principe de responsabilité ne fait pas allusion à ce type de situation. Le développement de technologies d'informatiques affectives soulève de tels enjeux éthiques de responsabilité sociale, qui sont en conflit avec l'autonomie et le respect de la vie privée.

Nous croyons que la *Déclaration de Montréal* offre un cadre rigoureux qui porte à réfléchir en profondeur pour toutes les entreprises qui souhaitent assurer une gestion éthique de l'intelligence artificielle au sein de leur organisation.

5.3.2 Facteurs à considérer pour une plus grande acceptabilité

Comme nous l'avons soulevé dans la revue de littérature, la recherche s'intéresse à l'acceptabilité des nouvelles technologies sous différents angles. Bobillier-Chaumon et Dubois (2019) ont abordé des facteurs comme l'environnement de l'utilisateur, les caractéristiques de la technologie ainsi que les caractéristiques de l'utilisateur. Von Walter, Kremmel et Jäger (2022), pour leur part, se sont intéressés à la perception de l'individu par rapport à l'intelligence artificielle, tout comme Manikonda et Kambhampati (2018), qui se

sont intéressés aux sentiments à l'égard de l'IA. Portales et Castañeda (2016) ont étudié le concept d'acceptabilité sociale de la communauté pour des projets publics plus controversés. Tandis qu'Asan, Bayrak et Choudhury (2020) ont, pour leur part, soutenu que les propriétés du système, telles que la transparence, la complexité du modèle et les risques associés, sont des facteurs importants, tout comme la fiabilité et la constance des résultats. De notre côté, nous avons validé, dans le cadre de notre recherche sur le terrain, les critères d'acceptabilité que les utilisateurs sont plus enclins à évaluer avant d'aller de l'avant avec l'utilisation d'une nouvelle technologie d'intelligence artificielle, et plus spécifiquement d'informatique affective. De plus, nous avons formulé un modèle en rassemblant des éléments que nous avons puisés au sein de la théorie dans la littérature. L'influence sociale, la réputation de l'organisation ou de la technologie, le sentiment de sécurité, la transparence, la fiabilité des résultats, la facilité d'utilisation et le bénéfice et l'utilité perçue ont été soulevés par les participants et sont les éléments centraux de notre modèle sur l'acceptabilité de nouvelles technologies d'intelligence artificielle. Ces critères représentent une ligne directrice dans les éléments importants aux yeux des utilisateurs. L'organisation qui vise à développer ou à déployer une nouvelle technologie peut faire la revue de ces critères en lien avec sa propre technologie et sa propre situation organisationnelle pour faciliter l'acceptabilité d'un système d'informatique affective.

Les enjeux éthiques soulevés dans le cadre de ce projet de recherche, tels que la fiabilité, la transparence, la protection de la vie privée et la sécurité, sont des critères d'acceptabilité spécifiquement nommés par les utilisateurs. La responsabilité, l'autonomie et la discrimination sont des enjeux qui n'ont pas été spécifiquement nommés en terme de critères d'acceptabilité, mais qui sont fortement liés à la réputation de la technologie ou de l'organisation. La façon dont les organisations vont mettre en place des mesures pour répondre aux enjeux éthiques va influencer leur réputation et par le fait même, l'acceptabilité de leur technologie.

5.4 Encadrement de l'intelligence artificielle

Force est de constater que la population générale est peu informée sur l'encadrement de l'intelligence artificielle. Les utilisateurs considèrent que l'intelligence artificielle est encore relativement jeune et qu'un cadre se met en place. C'est un domaine vaste et

complexe à réglementer et tout comme le mentionne Baker-Brunnbauer (2021), la vitesse à laquelle la technologie évolue constitue un défi pour une législation gouvernementale dont le processus est long et lent. Il faut être plus proactif que réactif et travailler en amont plutôt qu'en aval. Nous sommes à la traîne d'une technologie qui se développe plus rapidement que ce que nous pouvons anticiper. Nous remarquons, dans la population, un sentiment d'impuissance, une impression que malgré les mesures en place, les géants trouveront toujours le moyen de contourner la réglementation en place, qu'ils mènent la barque. Les gros joueurs ont le contrôle et il est très difficile de les encadrer. Enfin, nous constatons une perception de ne pas être protégé en tant que citoyen dans le cadre actuel des choses. Les mesures ne semblent pas suffisantes, la population aimerait voir une législation beaucoup plus stricte, avec des conséquences pour les contrevenants. Une réglementation adaptée tout aussi applicable aux plus petits, comme les entreprises en démarrage, qu'aux géants de ce monde.

Nous croyons en effet qu'un encadrement plus serré des technologies d'intelligence artificielle devrait être une priorité. En ce sens, les initiatives en cours sont encourageantes, notamment la toute dernière qui a été annoncée. Le 23 avril dernier, l'Union européenne a adopté le *Digital Services Act* (DSA) : « Ce sera désormais aux géants du Web de démontrer qu'ils déploient suffisamment d'efforts pour modérer efficacement les contenus des utilisateurs, tout en rendant plus accessible le fonctionnement de leurs algorithmes. » (Le Monde, 2022) Cette législation, la première du genre au monde, « vient consacrer le principe selon lequel ce qui est illégal hors ligne doit également être illégal en ligne » (AFP, 2022). La réglementation imposera entre autres un audit une fois par an par des organismes indépendants, et des amendes substantielles pouvant aller jusqu'à 6 % des ventes annuelles dans le cas d'infractions répétées (AFP, 2022). Elle comporte des clauses spécifiques aux GAFAM de ce monde, avec des obligations imposées aux plateformes de plus de 45 millions d'utilisateurs actifs sur le territoire de l'Union européenne (AFP, 2022). Ce que chaque pays n'aurait pu faire seul face aux très grandes plateformes, la mobilisation de l'Union européenne peut l'accomplir grâce à ses 27 États membres tout en inspirant d'autres à faire de même. Cette réglementation arrive à un moment opportun, considérant le sentiment d'impuissance et de vulnérabilité que nous avons recensé sur le terrain.

5.5 Recommandations

Dans le cadre de notre revue de littérature, nous avons exposé différentes pistes de solutions en lien avec les enjeux éthiques des systèmes d'intelligence artificielle. Notre recherche terrain nous a permis pour sa part de mettre en lumière de façon concrète la perception des utilisateurs. Nous vous présentons nos recommandations inspirées de la littérature et du terrain. Il est clair à nos yeux que pour augmenter l'acceptabilité de l'informatique affective et de l'intelligence artificielle en général, il faut répondre aux enjeux éthiques qu'elles soulèvent et miser sur l'éducation citoyenne.

5.5.1 Établir les règles du jeu

À nos yeux, la réglementation est la voie à suivre, autant pour permettre un développement éthique et responsable de l'IA au sein de la communauté que pour donner confiance à la population. En ce moment, le consommateur n'a pas le sentiment d'être protégé. Ainsi, des initiatives comme le projet de loi 64 au Québec, le Règlement général de protection des données (RGPD) et le nouveau *Digital Services Act* (DSA) de l'Union européenne sont des pas dans la bonne direction pour donner un cadre clair aux concepteurs et aux utilisateurs. Pour que, avec l'autorégulation, les entreprises ne succombent pas à la tentation du gain et à la pression de la concurrence, cela requiert une conscience sociale et une éthique forte. La collaboration internationale doit continuer de frayer le chemin vers un encadrement concerté afin d'éviter les inégalités en matière d'occasions de développement sur le plan mondial et la naissance de zones d'innovation très peu réglementées, comme nous connaissons le parallèle dans le monde de la finance avec les paradis fiscaux.

Toutes les entreprises qui conçoivent ou implantent un système d'intelligence artificielle devraient se doter d'un code d'éthique organisationnel relatif aux enjeux de l'intelligence artificielle. Ce code d'éthique devrait être applicable de façon concrète dans le quotidien, et chaque employé devrait en faire la révision annuellement par la voie d'une formation.

À des fins d'encadrement, toutes les entreprises qui conçoivent des systèmes d'intelligence artificielle devraient avoir :

- Une personne-ressource spécialisée en éthique de l'intelligence artificielle au sein de leur organisation ou une firme-conseil

spécialisée dans le domaine qui veille à la révision de leur code d'éthique et de leurs pratiques en entreprise;

- Un ou plusieurs expert(s) en cybersécurité pour assurer la sécurité de leur technologie, veiller à la conformité de la réglementation sur la protection des données et résoudre toutes brèches ou infractions internes ou externes.

Les auteurs dans la revue de littérature et les participants ont mentionné l'importance de la mise en place d'agences de vérification pour certifier la conformité des technologies. Cette recommandation est déjà en application dans certains domaines, comme la santé. La mise en place d'une certification internationalement reconnue de type ISO a aussi été proposée, autant dans la littérature qu'au sein des entrevues, pour certifier la conformité des systèmes d'intelligence artificielle selon des standards reconnus. De telles pratiques augmenteraient l'acceptabilité de ces technologies en influençant positivement leur réputation. Cette certification, qui démontrerait un haut standard de conformité éthique, pourrait très certainement être un levier marketing promotionnel pour l'organisation certifiée.

De la même façon, des audits devraient être effectués par des agences externes indépendantes à intervalles réguliers sur les systèmes d'intelligence artificielle par la voie de tests et d'une révision des rapports pour valider la conformité en matière de sécurité et de normes applicables, et ce cadre deviendra d'ailleurs obligatoire au sein du *Digital Services Act*, qui devrait entrer en vigueur en 2023.

Enfin, il est souhaitable de voir la réglementation évoluer dans un cadre de collaboration et de consultations des parties prenantes entre les concepteurs, les régulateurs et les utilisateurs pour établir un cadre rigoureux pour un développement éthique et responsable qui répond aux enjeux de tous.

5.5.2 Une plus grande transparence et miser sur la diversité

Pour augmenter l'acceptabilité de l'informatique affective, les organisations conceptrices devraient palier à des systèmes opaques en mettant en place des mécanismes pour en augmenter l'explicabilité. Elles devraient entre autre rendre accessibles les informations sur la fiabilité de leur technologie par la voie de la marge d'erreur, communiquer le contexte optimal pour son utilisation et les limites de leur algorithme. Connaître la marge d'erreur permettrait aux utilisateurs d'ajuster leurs attentes et leurs décisions en

conséquence. De plus, la marge d'erreur devrait indiquer dans quel contexte elle performe le mieux, avec le taux correspondant à cette performance, pour que l'utilisateur puisse reproduire des conditions similaires ou mettre en perspective les résultats selon sa situation particulière. Enfin, les organisations devraient être transparentes sur la représentativité de leur bassin populationnel au sein de leur base de données pour que l'utilisateur puisse prendre en considérant l'information lorsqu'il utilise la technologie en mettant en perspective son propre profil.

Les organisations devraient avoir des discussions ouvertes avec leurs clients sur le choix du type de modèle algorithmique en expliquant clairement les enjeux selon le modèle. Il faut revoir les besoins du client, qui sera possiblement plus sensible à la performance ou à l'importance de la transparence et de l'explicabilité entourant les prédictions algorithmiques. La collaboration des parties prenantes, telles que les utilisateurs dans le développement de la technologie, aura pour effet d'en augmenter l'acceptabilité.

Aussi, pour veiller à diminuer les biais issus de l'algorithme, tout particulièrement en ce qui concerne l'informatique affective, les organisations devraient faire la promotion de la diversité au sein de l'équipe de conception. Une diversité de genres, de formations, d'âges et de cultures permettrait de diminuer les biais qui sont inconsciemment codés au sein des algorithmes. Des formations sur le sujet, une sensibilisation ou la validation de l'algorithme via différentes bases de données pourraient pallier les problèmes actuels et permettre d'offrir des technologies toujours plus performantes et plus équitables au sein de plus grands bassins populationnels.

5.5.3 Responsabilité sociale de la profession

La profession d'ingénieur en conception de systèmes d'intelligence artificielle implique une responsabilité de plus en plus grande dans la société dans laquelle elle évolue. Les algorithmes sont des outils puissants qui peuvent être utilisés en sourdine de façon insidieuse si le concepteur est malveillant. Ils peuvent aussi être utilisés par des gouvernements ou des groupes pour faire de la propagande, comme dans le cas de l'assaut au Capitole en janvier 2021 (AFP, 2022) ou dans le contexte de la guerre en Ukraine

(Zakrzewski, 2022). Le concepteur doit donc prendre conscience des répercussions que son travail peut avoir dans la communauté.

Il devrait y avoir un encadrement plus rigoureux du métier d'ingénieur en conception de systèmes d'intelligence artificielle. Mittelstadt (2019) a proposé la création d'une licence pour les ingénieurs qui travaillent sur des systèmes d'intelligence artificielle qui présentent des risques plus importants pour la population. De cette façon, il serait possible d'assurer un encadrement plus strict pour le développement de systèmes d'intelligence artificielle potentiellement plus lourds de conséquences.

Landowska (2019) et O'Neil (2016) abordent quant à elles la création d'un code de conduite auquel les ingénieurs pourraient prêter serment de la même façon que les médecins prêtent le serment d'Hippocrate. L'objectif est de donner un code de valeurs comme ligne directrice à respecter dans le développement d'algorithmes.

Nous croyons que d'un point de vue éthique, il faudrait assurer une indépendance entre l'équipe de concepteurs et l'équipe des ventes. L'équipe de concepteurs mise sur le développement d'une technologie performante, sécuritaire et éthiquement responsable. Tandis que l'équipe des ventes mise généralement sur des objectifs de gain de marché et de profitabilité. L'équipe de conception devrait être indépendante pour éviter tout conflit éthique. Un bon exemple est le milieu de la santé, qui régleme très strictement les relations d'influence entre l'industrie et les professionnels.

5.5.4 L'éducation, un levier pour l'acceptabilité

Pour apporter des changements au sein des pratiques des concepteurs en intelligence artificielle et les sensibiliser à l'importance des enjeux éthiques, nous croyons fermement que l'éducation est un élément central de la solution. Il faudrait intégrer de façon rigoureuse, dans les universités et les programmes de formation, l'enseignement d'un cursus couvrant les aspects fondamentaux du développement éthique et responsable des systèmes d'intelligence artificielle. De plus, un nombre de crédits obligatoires en formation ou en conférence touchant l'éthique de l'intelligence artificielle devrait être requis de façon annuelle pour tous les concepteurs pour qu'ils puissent se maintenir à jour sur les avancées et les développements sur le sujet.

Une éducation citoyenne est également nécessaire, non seulement en ce qui concerne l'informatique affective, mais aussi l'intelligence artificielle de façon générale. Pour mieux naviguer dans l'ère numérique, la population doit être sensibilisée aux enjeux que la technologie soulève et aux mesures qui sont actuellement en place pour l'encadrer. L'intelligence artificielle étant une technologie complexe et relativement nouvelle aux yeux des non-experts, une éducation citoyenne continue devrait être déployée et mise de l'avant de façon plus prépondérante. Une telle initiative devrait être travaillée conjointement par le gouvernement, les différents organismes et les milieux scolaires et universitaires pour donner aux citoyens une meilleure compréhension de la technologie, de sa conception, des éléments qui l'encadrent, de leurs droits et des mesures à prendre pour se protéger dans le monde numérique. De plus, il devrait y avoir une plus grande visibilité des ressources disponibles pour répondre à leurs questions. La fondation Le CIEL (Centre pour l'intelligence émotionnelle en ligne), récemment mise sur pied, est un bel exemple d'éducation citoyenne. Pour sa part, elle se concentre sur le bien-être numérique. Elle donne des conférences, entre autres dans le milieu scolaire, et offre de l'information pratique autour de certains enjeux.

En ce qui concerne les conditions d'utilisation liées au partage de données personnelles, comme le recommande Turner (2019), un travail constant à la mise en place d'un filet de sécurité devrait être effectué sur le plan législatif pour éviter que les utilisateurs abandonnent des droits sans en comprendre pleinement la portée. Une plus grande sensibilisation par l'éducation citoyenne devrait être faite pour aider la population à comprendre les subtilités de ces conditions et à être plus alerte. Les organisations devraient faire preuve d'une plus grande transparence et collaborer au développement de la littératie dans le domaine de la protection des données en mettant au point des documents courts, dans un langage familier et clair. Nous sommes conscients qu'une lourde charge légale pèse également dans la balance lorsque les organisations produisent ces documents. Il y a cependant des organisations qui misent sur la simplicité et offrent plus d'options de contrôle à leurs utilisateurs. EmoScienS, par exemple, offre à ses utilisateurs des options : « Nous utilisons des cookies pour comprendre comment vous interagissez avec notre site. En acceptant, vous consentez à notre utilisation de ces cookies. » L'utilisateur peut « Tout refuser » ou « Accepter », et il peut aussi décider d'ajuster ses préférences par des

paramètres clairement accessibles. Ces pratiques sont alignées avec ce qui est recherché par les gens sur le terrain.

Grâce à ce projet de recherche, nous avons pu démontrer que l'acceptabilité est largement influencée par les enjeux éthiques au cœur des préoccupations de la population. Les enjeux prédominants en ce qui concerne l'informatique affective tournent autour de la sécurité sous la forme de la protection de la vie privée, des données personnelles et la mise en place de mécanisme pour minimiser les risques d'utilisation malveillante de la technologie. La fiabilité et la constance des résultats tout comme la transparence des organisations et l'explicabilité de la technologie sont des enjeux qui jouent sur le niveau de confiance de l'utilisateurs. Il faut miser sur une éducation citoyenne, un encadrement rigoureux et une responsabilisation des organisations et des concepteurs pour augmenter la confiance de la population envers la technologie.

5.6 Limites de l'étude et avenues de recherches futures

Il est important de mentionner certaines limites à notre recherche. Dans le cadre de notre travail terrain, la durée d'utilisation de deux semaines n'a pas permis aux utilisateurs d'avoir accès à la pleine expérience de la technologie d'EmoScienS. En effet, les participants n'ont pas eu d'informations personnalisées ou de ressources en lien avec leur tableau de bord.

De plus, certains participants n'ont pas pu utiliser la version logicielle de la technologie d'EmoScienS. Nous avons remarqué que c'est une limite importante, et tout particulièrement en période de pandémie, pour tous les participants dont une bonne partie de la journée se passait en rencontres virtuelles. La version Internet n'étant pas compatible avec les plateformes de vidéoconférence telles que Zoom, Google Meet ou Teams, l'expérience émotionnelle de ces participants n'a pas pu être captée complètement.

L'échantillon populationnel était relativement homogène sur le plan culturel. Bien que la recherche n'ait pas eu comme objectif d'étudier de façon spécifique cet aspect, il aurait été intéressant d'avoir une plus grande diversité.

Notre projet de recherche a soulevé de nouvelles pistes pour la recherche. Nous avons pris le pouls de la littérature de façon générale en ce qui concerne les enjeux que soulèvent l'intelligence artificielle et l'informatique affective, mais il serait intéressant d'établir à quel point la population est sensibilisée à l'empreinte environnementale du numérique et aux gestes concrets qui sont posés individuellement pour la minimiser. Aussi, nous avons remarqué un malaise par rapport à l'informatique affective de la part des gestionnaires qui ont participé à notre projet de recherche. Ils ont mentionné des enjeux de responsabilité et d'accès à la vie privée. Il serait pertinent de s'intéresser à la perspective des gestionnaires dans l'utilisation d'informatique affective au sein des équipes de travail en entreprise afin de mieux comprendre les bénéfices et les enjeux qu'une telle technologie soulève dans leur rôle. Enfin, il serait intéressant de faire une étude similaire en utilisant la même version logicielle, mais en ayant un échantillon populationnel hétérogène au niveau culturel pour examiner la perception de l'informatique affective au sein des différents milieux. Le terrain de recherche dans le milieu de l'intelligence artificielle, de l'informatique affective et de l'éthique est riche et fertile. Comme ce milieu est encore relativement à ses balbutiements devant les possibilités de la technologie et de son plein déploiement en cette ère numérique, la collaboration et la recherche sont des voies prometteuses à privilégier.

Conclusion

Dans les dernières années, nous avons vu l'arrivée des téléphones intelligents, des réseaux sociaux, des voitures autonomes, des assistants virtuels comme Siri et Alexa et tout récemment de nombreuses options de réalité virtuelle et réalité augmentée. Pour certains d'entre nous, la recherche est passée des livres, encyclopédies et journaux scientifiques dans les établissements scolaires et universitaires, aux banques de données électroniques mondiales. L'accès à l'information est plus grand et plus rapide, l'intelligence artificielle nous permet d'optimiser des processus, de synthétiser un nombre incroyable de données et de réaliser des progrès inatteignables auparavant dans différentes sphères, dont les sciences de la vie. La technologie soulève énormément d'intérêts et de curiosité, et tout particulièrement l'informatique affective. Sur le terrain nous avons pu remarquer à quel point elle intrigue les utilisateurs qui se questionnent sur la capacité d'une machine à lire les émotions. Comme nous l'avons décrit dans le chapitre 2, le déploiement de la technologie ne se fait pas sans heurts et soubresauts, il soulève des enjeux. A nos yeux, comme l'a dit Kaifu Lee « Nous ne sommes pas les spectateurs passifs de l'histoire de l'IA; nous en sommes les acteurs. » (2019 : 420). Il est donc de notre devoir à tous de prendre part à notre façon au développement de la technologie en harmonie avec notre société. Pour cette raison, nous avons entamé il y a un an cette recherche sur l'intelligence artificielle. Nous avons canalisé nos énergies plus spécifiquement vers l'intelligence artificielle affective, ses enjeux éthiques et ses conditions d'acceptabilité. Ce projet de recherche nous a permis de contribuer à la communauté scientifique de différentes façons :

Il nous a été possible de mettre en lumière la perception actuelle de la population face aux enjeux éthiques de l'intelligence artificielle, de soulever ce qui est au cœur des préoccupations citoyennes, et de façon plus pointue de celles en lien avec l'informatique affective.

Nous avons contribué à la recherche sur l'acceptabilité des nouvelles technologies en élaborant un modèle composé de sept critères influençant l'acceptabilité et en le validant comme cadre d'analyse auprès d'utilisateurs d'une technologie d'informatique affective.

De plus, nous avons recensé une liste de recommandations extraites de la littérature et de nos entrevues pour assurer le développement d'une IA éthique et responsable, et nous avons partagé ce qui à nos yeux devrait être adopté dans un futur rapproché.

Nous avons par ailleurs pu soulever la vulnérabilité de la population dans un contexte d'asymétrie de connaissances et de pouvoir face aux concepteurs d'algorithmes et nous militons pour une plus grande éducation citoyenne et un travail continue à étoffer la législation pour encadrer les pratiques dans le milieu de l'IA.

L'IA sera à l'image de la communauté au sein de laquelle elle se développera. Dans notre quête pour une technologie alignée avec nos valeurs, il faut revenir aux sources et valider quelle est la nature profonde de l'Humain. Le développement de l'informatique affective sera très certainement un élément essentiel pour créer des machines capables de faire preuve de jugement, d'y instiller des valeurs et pourquoi pas, des sentiments. Kaifu-Lee nous met en garde contre une IA développée selon un modèle de conquêtes économiques et de profit et prône plutôt l'optique d'une technologie pour nous libérer des tâches de performance et d'optimisation et offrir à l'humanité l'opportunité de s'épanouir pleinement (Lee, 2019). L'histoire est en train de s'écrire, à nous d'en établir les règles du jeu et de déterminer quelles sont les valeurs qui sont au cœur de la société que nous souhaitons préserver et surtout retrouver au sein des algorithmes de demain. L'éducation et la consultation citoyenne sont incontournables pour construire ensemble un futur en harmonie avec une technologie qui est représentative de ce à quoi on aspire bâtir.

Bibliographie

AFP (2016, 14 mars). « La victoire d'Alphago contre le champion du jeu de go restera dans l'Histoire », *Les Affaires*, section Technologies de l'information. Récupéré de <https://www.lesaffaires.com/techno/technologie-de-l-information/la-victoire-d-alphago-contre-le-champion-du-jeu-de-go-restera-dans-l-histoire/585999>

AFP (2022, 23 avril). « L'Union européenne achève une réforme historique contre la jungle numérique », *Radio-Canada*, section Politiques internationales. Récupéré de <https://ici.radio-canada.ca/nouvelle/1878418/ue-reforme-historique-numerique-gafam>

Agrawal, Ajay, Joshua Gans et Avi Goldfarb (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston, Harvard Business Press.

Ahmad, Muhammad Aurangzeb, Carly Eckert et Ankur Teredesai (2018). « Interpretable Machine Learning in Healthcare », présenté au *2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Washington, 29 août au 1^{er} septembre, p. 559-560.

AlgorithmWatch (2020). *AI Ethics Guidelines Global Inventory*, AlgorithmWatch. Récupéré de <https://inventory.algorithmwatch.org>

Amann, Julia, Alessandro Blasimme, Effy Vayena, Dietmar Frey et Vince I. Madai (2020). « Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective », *BMC Medical Informatics and Decision Making*, vol. 20, n° 1, p. 1-9.

Anderson, Michael et Susan Leigh Anderson (2007). « Machine Ethics: Creating an Ethical Intelligent Agent », *AI Magazine*, vol. 28, n° 4, p. 15-26.

Asan, Onur, Alparslan Emrah Bayrak et Avishek Choudhury (2020). « Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians », *Journal of Medical Internet Research*, vol. 22, n° 6, p. 1-7.

Baba, Sofiane et Emmanuel Raufflet (2015). « L'acceptabilité sociale : une notion en consolidation », *Management international / International Management / Gestión Internacional*, vol. 19, n° 3, p. 98-114.

Baker-Brunnbauer, Josef (2021). « Management Perspective of Ethics in Artificial Intelligence », *AI Ethics*, vol. 1, n° 2, p. 173-181.

Bannister, Frank et Regina Connolly (2020). « Administration by Algorithm: A Risk Management Framework », *Information Polity*, vol. 25, n° 4, p. 471-490.

Barrett, Paddy M, Steven R Steinhubl, Evan D Muse et Eric J Topol (2017). « Digitising the mind », *The Lancet*, vol. 389, no 10082, p. 1877.

Beauchamp, Tom L. et James F. Childress (2001). *Principles of Biomedical Ethics*, New York, Oxford University Press.

Beaudouin, Valérie, Isabelle Bloch, David Bounie, Stéphan Cléménçon, Florence d'Alché-Buc, James Eagan *et al.* (2020). « Identifying the "Right" Level of Explanation in a Given Situation », présenté au *First International Workshop on New Foundations for Human-Centered AI (NeHuAI)*, Santiago de Compostella, 4 septembre, p. 63.

Beavers, Anthony F. et Justin P. Slattery (2017). « On the Moral Implications and Restrictions Surrounding Affective Computing », dans Myonghoon Jeon (dir.), *Emotions and Affect in Human Factors and Human-Computer Interaction*, London, Elsevier, p. 143-161.

Bélaïr-Cirino, Marco (2020, 4 juin). « La vérificatrice générale s'inquiète de la gestion des données confidentielles », *Le Devoir*, section Politique. Récupéré de <https://www.ledevoir.com/politique/quebec/580092/ramq-et-retraite-quebec-la-vg-s-inquiete-de-la-gestion-des-donnees-confidentielles>

Besse, Philippe, Aurèle Besse-Patin et Céline Castets-Renard (2020). « Implications juridiques et éthiques des algorithmes d'intelligence artificielle dans le domaine de la santé », *Statistique et Société*, vol. 8, n° 3, p. 21-53.

Blasimme, Alessandro, Effy Vayena (2020). « The Ethics of AI in Biomedical Research, Patient Care, and Public Health », dans Markus D. Dubber, Frank Pasquale et Sunit Das (dir.), *The Oxford Handbook of Ethics of AI*, New York, Oxford University Press. p.703-718.

Bobillier-Chaumon, Marc-Eric et Michel Dubois (2009). « L'adoption des technologies en situation professionnelle : quelles articulations possibles entre acceptabilité et acceptation? », *Le travail humain*, vol. 72, n° 4, p. 355-382.

Boily, Daniel (2022, 23 mars). « Vol massif de données : Desjardins suspectait cinq autres employés », *La Presse*, section Finance et Assurances. Récupéré de <https://ici.radio-canada.ca/nouvelle/1870938/vol-massif-donnees-desjardins-suspectait-cinq-autres-employes>

Boutilier, Robert G. et Ian Thomson (2011). *Modelling and Measuring the Social License to Operate: Fruits of a Dialogue between Theory and Practice*, Social License. Récupéré de <http://sociallicense.com/publications/Modelling%20and%20Measuring%20the%20SLO.pdf>

Bowen, Jonathan P. (2019). « The Impact of Alan Turing: Formal Methods and Beyond », dans Jonathan P. Bowen, Zhiming Liu et Zili Zhang (dir.), *Engineering Trustworthy Software Systems: 4th International School, SETSS 2018, Chongqing, China, April 7–12, 2018, Tutorial Lectures*, Cham, Springer International Publishing, p. 202-235.

Braun, Élisabeth (2019, 17 avril). « Enfants et géants : les manipulations des géants du Web dans le viseur des régulateurs », *Le Figaro*, section Tech & Web. Récupéré

de <https://www.lefigaro.fr/secteur/high-tech/enfants-et-ecrans-les-manipulations-des-geants-du-web-dans-le-viseur-des-regulateurs-20190417>

Brigham, Tara J. (2017). « Merging Technology and Emotions: Introduction to Affective Computing », *Medical Reference Services Quarterly*, vol. 36, n° 4, p. 399-407.

Brynjolfsson, Erik et Andrew McAfee (2017, 18 juillet). « The Business of Artificial Intelligence: What It Can and Cannot Do for Your Organization », *Harvard Business Review*, section Technology and Analytics. Récupéré de <https://hbr.org/2017/07/the-business-of-artificial-intelligence>

Bullington, Joseph (2005). « 'Affective' Computing and Emotion Recognition Systems: The Future of Biometric Surveillance? », communication présentée au *InfoSecCD '05: 2nd Annual Conference on Information Security Curriculum development*, Kennesaw, 23-24 septembre.

Caliskan, Aylin, Joanna J. Bryson et Arvind Narayanan (2017). « Semantics Derived Automatically from Language Corpora Contain Human-Like Biases », *Science*, vol. 356, n° 6334, p. 183-186.

Cardon, Dominique (2018). « Le pouvoir des algorithmes », *Pouvoirs*, no 164, p. 63-73.

Cardon, Dominique, Jean-Philippe Cointet, et Antoine Mazières (2018). « La revanche des neurones : l'invention des machines inductives et la controverse de l'intelligence artificielle », *Réseaux*, vol. 211, n° 5, p. 173-220.

Caron, Daniel J., Sara Bernardi et Vincent Nicolini (2021). « L'acceptabilité sociale du partage des données de santé: revue de la littérature », *Chaire de recherche en exploitation des ressources informationnelles, École nationale d'administration publique*.

Charlet, Jean et Sandra Bringay (2019). « Intelligence Artificielle et Santé : une analyse rétrospective depuis 2010 », communication présentée aux 30^{es} Journées Francophones d'Ingénierie des Connaissances, Toulouse, Juillet 2019, AFIA, p. 26-42.

Coeckelbergh, Mark (2020). *AI Ethics*, Cambridge, The MIT Press.

Cohen, Claudia (2021, 20 août). « Données personnelles : la Chine adopte son RGPD », *Le Figaro*, section Médias & Publicité. Récupéré de <https://www.lefigaro.fr/medias/donnees-personnelles-la-chine-adopte-son-rgpd-20210820>

Comité Technologies de l'information (CTI) (2022). *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*. Récupéré de <https://www.caij.qc.ca/dossier/projet-de-loi-n-64-loi-modernisant-des-dispositions-legislatives-en-matiere-de-protection-des-renseignements-personnels>

Cutillo, Christine M., Karlie R. Sharma, Luca Foschini, Shinjini Kundu, Maxine Mackintosh, Kenneth D Mandl *et al.* (2020). « Machine Intelligence in Healthcare— Perspectives on Trustworthiness, Explainability, Usability, and Transparency », *NPJ digital medicine*, vol. 3, n° 47, p. 1-5.

Daily, Shaundra B., Melva T. James, David Cherry, John J. Porter III, Shelby S. Darnell, Joseph Isaac, *et al.* (2017). « Affective Computing: Historical Foundations, Current Applications, and Future Trends » dans Myonghoon Jeon (dir.), *Emotions and Affect in Human Factors and Human-Computer Interaction*, London, Elsevier, p. 213–231.

Damassino, Nicola et Nicholas Novelli (2020). « Rethinking, Reworking and Revolutionising the Turing Test », *Minds and Machines*, vol. 30, n° 4, p. 463-468.

Da Silva *et al.* (2021). « La fusion du droit et de la science : atelier sur la fusion du droit et de la science ». Récupéré de <https://cifar.ca/wp-content/uploads/2021/05/IA-et-soins-de-sante-La-fusion-du-droit-et-de-la-science-II.pdf>

Davenport, Thomas H., Erik Brynjolfsson, Andrew McAfee et H James Wilson (2019). *Artificial Intelligence: The Insights You Need from Harvard Business Review*, Boston, Harvard Business Review Press.

Davis, Fred D. (1989). « Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology », *MIS Quarterly*, vol. 13, n° 3, p. 319-340.

Dawar, Niraj (2018). « Marketing in the Age of Alexa », *Harvard Business Review*, section Marketing. Récupéré de <https://hbr.org/2018/05/marketing-in-the-age-of-alexa>

Deloitte (2019). *Impératif de l'IA au Canada : surmonter les risques, instaurer la confiance*, Deloitte. Récupéré de <https://www2.deloitte.com/ca/fr/pages/deloitte-analytics/articles/canadas-ai-imperative.html>

Deloitte (2021). *Surveillance and Predictive Policing Through AI*, Deloitte. Récupéré de <https://www2.deloitte.com/global/en/pages/public-sector/articles/urban-future-with-a-purpose/surveillance-and-predictive-policing-through-ai.html>

Dignum, Virginia (2017, 25 septembre). « Responsible Artificial Intelligence: Designing AI for Human Values », *ITU Journal: ICT Discoveries*, vol. 2018, n° 1, p. 1-8.

Dignum, Virginia (2020). « Responsibility and Artificial Intelligence », dans Markus D. Dubber, Frank Pasquale et Sunit Das (dir.), *The Oxford Handbook of Ethics of AI*, New York, Oxford University Press. p. 215-231.

Dwivedi, Yogesh K, Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick *et al.* (2021). « Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy », *International Journal of Information Management*, vol. 57, récupéré de <https://www-sciencedirect-com.proxy2.hec.ca/science/article/pii/S026840121930917X>

Elkouri, Rima (2020, 20 octobre). « Black Lives Matter et nous », *La Presse*, section Actualité. Récupéré de <https://www.lapresse.ca/actualites/20-20/black-lives-matter-et-nous.php>

Faucheux, Michel (2017). *Norbert Wiener, le Golem et la Cybernétique, Éléments de Fantastique Technologique*, Editions du Sandre.

Feldman Barrett, Lisa, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez et Seth D. Pollak (2019). « Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements », *Psychological Science in the Public Interest*, vol. 20, n° 1, p. 1-68.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum *et al.* (2018). « AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations », *Minds and Machines*, vol. 28, n° 4, p. 689-707.

Frick, Walter (2015). « When Your Boss Wears Metal Pants », *Harvard Business Review*, vol. 93, n° 6, p. 84-89.

Gaudiault, Tristan (2021, 19 octobre). « Le Big Bang du Big Data », *Statista*, section Données numériques. Récupéré de <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/>

Gawdat, Mo (2021) *Scary Smart: The Future of Artificial Intelligence and How You Can Save Our World*, London, Bluebird.

Ghafur, Saira, Jackie Van Dael, Melanie Leis, Ara Darzi et Aziz Sheikh (2020). « Public Perceptions on Data Sharing: Key Insights from the UK and the USA », *The Lancet Digital Health*, vol. 2, no 9, p. 444-446.

Gibert, Martin (2020). *Faire la morale aux robots : une introduction à l'éthique des algorithmes*, Montréal, Atelier 10.

Gouvernement du Canada (2022). *Évaluation de l'incidence algorithmique*. Récupéré de <https://canada-ca.github.io/aia-eia-js/?lang=fr>

Groupe d'experts de haut niveau sur l'intelligence artificielle (2019). *Lignes directrices en matière d'éthique pour une IA digne de confiance*. Récupéré de <https://www.justice-ia.com/files/sites/181/2019/10/EthicsguidelinesfortrustworthyAI-FRpdf.pdf>

Hague, Douglas C. (2019). « Benefits, Pitfalls, and Potential Bias in Health Care AI », *North Carolina Medical Journal*, vol. 80, n° 4, p. 219-223.

Haiech, Jacques (2020). « Parcourir l'histoire de l'intelligence artificielle, pour mieux la définir et la comprendre », *Médecine sciences*, vol. 36, n° 10, p. 919-923.

Josselin, Marie-Laure (2021, 1^{er} octobre). « Joyce Echaquan : le racisme est une des causes de la mort », *Radio-Canada*, section Société. Récupéré de <https://ici.radio-canada.ca/amp/1828421/coroner-joyce-echaquan-atikamekw-rapport-gouvernement-racisme-lacunes>

Khos, Greg (real.) (2017). « AlphaGo » [film], États-Unis, Robert Fernandez et Robert Levinson (prod.).

Landowska, Agnieszka (2019). « Uncertainty in Emotion Recognition », *Journal of Information, Communication and Ethics in Society*, vol. 17, n° 3, p. 273-291.

Lee, Kai-Fu, (2019). *IA la plus grande mutation de l'histoire : comment la Chine devient le leader de l'intelligence artificielle*, traduit de l'anglais par Élise Roy, Paris, Les Arènes.

Le Monde (2022). « Régulation du Web : l'Europe en pointe », *Le Monde*, section Débats/Libertés numériques. Récupéré de https://www.lemonde.fr/idees/article/2022/04/26/regulation-du-web-l-europe-en-pointe_6123698_3232.html

London, Alex John (2019). « Artificial Intelligence and Black-Box Medical Decisions: Accuracy Versus Explainability », *Hastings Center Report*, vol. 49, n° 1, p. 15-21.

Luca, Michael, Jon Kleinberg et Sendhil Mullainathan (2016). « Algorithms Need Managers, Too », *Harvard Business Review*, vol. 94, n° 1, p. 20.

Maclure, Jocelyn et Marie-Noëlle Saint-Pierre (2018). « Le nouvel âge de l'intelligence artificielle : une synthèse des enjeux éthiques », *Les cahiers de propriété intellectuelle*, vol. 30, n° 3, p. 741-765.

Madiega, Tambiama André (2019, 19 septembre). « EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation ». *Policy Commons*, section Rapport. Récupéré de <https://policycommons.net/artifacts/1337743/eu-guidelines-on-ethics-in-artificial-intelligence/1945725/>

Manikonda, Lydia et Subbarao Kambhampati (2018). « Tweeting AI: Perceptions of Lay Versus Expert Twitterati », présenté au *Twelfth International AAAI Conference on Web and Social Media*, Stanford, 25 au 28 juin, Arizona State University.

Mathys, Catherine (2016, 23 septembre). « Un signal sans fil pour détecter vos émotions », *Radio-Canada*, section Techno. Récupéré de <https://ici.radio-canada.ca/nouvelle/804802/wifi-emotions-mit>

Mathys, Catherine (2017, 10 novembre). « Le rebelle de l'intelligence artificielle », *L'Actualité*, section Techno. Récupéré de <https://lactualite.com/techno/le-rebelle-de-lintelligence-artificielle/>

McGregor, Lorna (2018). « Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti's Foreword », *European Journal of International Law*, vol. 29, n° 4, p. 1079-1085.

McNamara, Andrew, Justin Smith et Emerson Murphy-Hill (2018). « Does ACM's Code of Ethics Change Ethical Decision Making in Software Development? », présenté au 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Lake Buena Vista, 4-9 novembre, p. 729-733.

McStay Andrew et Lachlan Urquhart (2019). « 'This Time with Feeling?' Assessing EU Data Governance Implications of Out of Home Appraisal Based Emotional AI », *First Monday*, vol 24, n° 10. Récupéré de <https://firstmonday.org/ojs/index.php/fm/article/view/9457/8146>

McStay, Andrew (2020). « Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy », *Big Data & Society*, p. 1-12.

Mintz, Yoav et Ronit Brodie (2019). « Introduction to Artificial Intelligence in Medicine », *Minimally Invasive Therapy & Allied Technologies*, vol. 28, n° 2, p. 73-81.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter et Luciano Floridi (2016). « The Ethics of Algorithms: Mapping the Debate », *Big Data & Society*, vol. 3, n° 2. récupéré de <https://journals.sagepub.com/doi/10.1177/2053951716679679>

Mittelstadt, Brent (2019). « Principles Alone Cannot Guarantee Ethical AI », *Nature Machine Intelligence*, vol. 1, n° 11, p. 501-507.

Mongeau, Pierre (2011). *Réaliser son mémoire ou sa thèse : côté jeans et côté tenue de soirée*, Québec, Presses de l'Université du Québec.

Oakden-Rayner Luke, Jared Dunnmon, Gustavo Carneiro et Christopher Ré (2020), « Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging », présenté au *ACM Conference on Health, Inference, and Learning 2020*, Toronto, 2-4 avril, p. 151-159.

OCDE (2019a). *Présentation des principes sur l'IA*. Récupéré de <https://oecd.ai/fr/ai-principles>

OCDE (2019b). *Recommandation du Conseil sur l'intelligence artificielle*. Récupéré de <https://oecd.ai/fr/ai-principles>

OMS (2021, 28 juin). *Éthique et gouvernance de l'intelligence artificielle pour la santé : résumé analytique*. Récupéré de <https://www.who.int/fr/publications/i/item/9789240037403>

O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1^e édition, New York, New York Crown.

Pavic, Clémence (2021, 1^{er} décembre). « À partir du 1^{er} décembre, les Québécoises travaillent gratuitement », *Le Devoir*, section Économie. Récupéré de <https://www.ledevoir.com/economie/651000/remuneration-a-partir-du-1er-decembre-les-quebecoises-travaillent-gratuitement>

Picard, Rosalind W. (1995). *Affective Computing*, Rapport technique n° 321, Cambridge, MIT Media Laboratory Perceptual Computing. Récupéré de <http://www.macs.hw.ac.uk/~yjc32/project/ref-social%20media%20campaign/1995-affective%20computing.pdf>

Portales, Luis et Scherezada Romero Castañeda (2016). « Incohérences et limites de l'acceptabilité sociale: le cas de l'exploitation minière mexicaine », *Éthique publique* :

revue internationale d'éthique sociétale et gouvernementale, vol. 18, no 1. Récupéré de <https://journals.openedition.org/ethiquepublique/2552>

Purdy, Mark, John Zealley et Omaro Maseli (2019, 18 novembre). « The Risks of Using AI to Interpret Human Emotions », *Harvard Business Review*, section Gender. Récupéré de <https://hbr.org/2019/11/the-risks-of-using-ai-to-interpret-human-emotions>

Richardson, Sharon (2020). « Affective Computing in the Modern Workplace », *Business Information Review*, vol. 37, n° 2, p. 78-85.

Ricoul, Stéphane (2021, 8 août), « Au rythme des algorithmes », *La Presse*, section Opinion. Récupéré de <https://www.lapresse.ca/debats/opinions/2021-08-08/opinion/au-rythme-des-algorithmes.php#>

Romano, Mathilde, Jennifer Stoddard et Julie Uzan-Naulin (2021, 23 septembre). « Le début d'un temps nouveau pour le secteur privé : le projet de loi 64 sur la protection des renseignements personnels est adopté », *Fasken Martineau Dumoulin LLP*. Récupéré de <http://www.fasken.com/fr/knowledge/projet-de-loi-64/2021/09/23-debut-temps-nouveau-secteur-prive-pl-64-adopte>

Shneiderman, Ben (2016). « Opinion: The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight », *Proceedings of the National Academy of Sciences*, vol. 113, n° 48. Récupéré de <https://www.pnas.org/doi/10.1073/pnas.1618211113>

Schwab, Klaus (2018, 16 janvier). « Klaus Schwab : comment façonner la quatrième révolution industrielle », *La Tribune*, section Opinions. Récupéré de <https://www.latribune.fr/opinions/tribunes/klaus-schwab-comment-faconner-la-quatrieme-revolution-industrielle-764814.html>

Surden, Harry (2020). « Ethics of AI in Law », dans Markus D. Dubber, Frank Pasquale et Sunit Das (dir.), *The Oxford Handbook of Ethics of AI*, New York, Oxford University Press. p. 719-736.

Sûreté du Québec (2022, 13 janvier). *Dévoilement du bilan routier 2021*. Récupéré de <https://www.sq.gouv.qc.ca/communiqués/devoilement-du-bilan-routier-2021/>

Telford, Taylor (2019, 31 juillet). « “Emotion Detection” AI Is a \$20 Billion Industry: New Research Says It Can’t Do What It Claims », *The Washington Post*, section Business. Récupéré de <https://www.washingtonpost.com/business/2019/07/31/emotion-detection-ai-is-billion-industry-new-research-says-it-cant-do-what-it-claims/>

Terrade, Florence, Hélène Pasquier, Juliette Reerinck-Boulanger, Gérard Guingouain et Alain Somat (2009). « L’acceptabilité sociale : la prise en compte des déterminants sociaux dans l’analyse de l’acceptabilité des systèmes technologiques », *Le travail humain*, vol. 72, n° 4, p. 383-395.

Thomson, Ian et Robert G. Boutilier (2011). dans Peter Darling (dir.), *SME Mining Engineering Handbook*, Colorado, Society for Mining, Metallurgy, and Exploration, p. 673-690.

Tual, Morgane (2015, 24 juillet). « Comment le “deep learning” révolutionne l’intelligence artificielle », *Le Monde*, section Pixels. Récupéré de https://www.lemonde.fr/pixels/article/2015/07/24/comment-le-deep-learning-revolutionne-l-intelligence-artificielle_4695929_4408996.html

Turing, Alan M. (1950). « Computing Machinery and Intelligence », *Mind*, vol. 59, n° 236, p. 433-460.

Turner, Jacob (2019). « Responsibility for AI », dans *Robot Rules*, Springer, p. 81-132.

Union européenne (2018). *Le règlement général sur la protection des données – RGPD*. Récupéré de <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>

Université de Montréal (2018). *Déclaration de Montréal pour un développement responsable de l'IA*. Récupéré de <https://www.declarationmontreal-iaresponsable.com/>

Villani, Cédric, Yann Bonnet, Charly Berthet, François Levin, Marc Schoenauer, Anne Charlotte Cornut *et al.* (2018). *Donner un sens à l'intelligence artificielle: pour une stratégie nationale et européenne*, Conseil national du numérique.

von Walter, Benjamin, Dietmar Kremmel et Bruno Jäger (2022). « The Impact of Lay Beliefs about AI on Adoption of Algorithmic Advice », *Marketing Letters*, vol. 33, n° 1, p. 143-155.

Wolford, Ben (2020). *Data Protection Impact Assessment (DPIA)*. Récupéré de <https://gdpr.eu/data-protection-impact-assessment-template/>

Wong, Julia C. (2019, 18 mars). « The Cambridge Analytica Scandal Changed the world – but It Didn't Change Facebook », *The Guardian*, section Technology. Récupéré de <https://www.theguardian.com/technology/2019/mar/17/the-cambridge-analytica-scandal-changed-the-world-but-it-didnt-change-facebook>

Yampolskiy, Roman V. (2017, 10 avril). « What Will Happen When Your Company's Algorithms Go Wrong? », *Harvard Business Review*, section Technology and Analytics. Récupéré de <https://hbr.org/2017/04/what-will-happen-when-your-companys-algorithms-go-wrong>

Yin, Robert K. (2018). *Case Study Research and Applications: Design and Methods*, VI^e édition, Los Angeles, SAGE.

Zetlin, Minda (2018, 28 février). « AI Is Now Analyzing Candidates' Face Expressions During Job Interviews: Unilever, IBM, Dunkin Donut and Many Others Already Use This Technology », *Inc.*, section Innovate. Récupéré de <https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facial-expressions-during-video-job-interviews.html>

Zakrzewski, Cat (2022, 22 avril). « Europe to Slap New Regulations on Big Tech, Beating U.S. to the Punch », *The Washington Post*, section Tech Policy. Récupéré de <https://www.washingtonpost.com/technology/2022/04/22/european-lawmakers-digital-services/>

Zuboff, Shoshanna (2020, 24 janvier). « You Are Now Remotely Controlled: Surveillance Capitalist Control Science and the Scientists, the Secret and the Truth », *The New York Times*, section Opinion. Récupéré de <https://nyti.ms/2NVh1eH>

Annexes

Annexe A : Guide d'entrevue initial

Guide d'entrevue semi-dirigée Février 2021

Bonjour,

Suite à la période d'utilisation de la technologie d'EmoScienS qui vient de se terminer, vous avez été invité à réaliser une entrevue dans le cadre du projet de recherche pour en apprendre davantage sur votre expérience. Je tiens à vous remercier de votre participation à cette entrevue qui devrait durer d'environ 45 minutes. Je vais vous poser différentes questions pour nous permettre de mieux comprendre votre appréciation de la technologie EmoScienS et de l'intelligence artificielle en général.

Je tiens à reconfirmer que vous êtes à l'aise avec l'enregistrement de cette entrevue pour permettre la transcription dans le cadre de l'analyse des données. Sentez-vous aussi à l'aise d'arrêter la rencontre en tout temps ou de ne pas répondre à certaines questions.

Questions pour établir le profil du participant pour démontrer l'hétérogénéité au sein de notre échantillon :

Votre rôle actuel?

Âge?

Éducation complétée?

Nous aimerions entamer l'entrevue avec quelques questions d'ordre général concernant la technologie et l'intelligence artificielle

- Quel est votre intérêt pour les nouvelles technologies?
 - Utilisez-vous des objets intelligents ou connectés (montre, cellulaire, maison connectée, etc.)?
- Comment décririez-vous votre niveau d'aisance/confort face aux nouvelles technologies?
- À quel niveau êtes-vous à l'aise de partager vos données personnelles avec des applications, des objets connectés ou des sites Web? (**Ex. Données de consommation** de types préférences d'achats, achats passés, recherches effectuées; **Données de santé** de type dossier de santé, fréquence cardiaque, poids, habitudes de vies); **Données sociales** (amis, famille, photos, etc.);

- Quels types de données personnelles acceptez-vous de partager sans problème?
 - Pourquoi?
- Quels types de données personnelles vous n'accepteriez pas, ou vous seriez réticent, à partager?
 - Pourquoi?
- Sentez-vous que vous avez le contrôle sur vos données personnelles dans l'univers du numérique?
- Si je vous dis « intelligence artificielle », quel est le premier mot qui vous vient à l'esprit?
- Pourriez-vous me dire ce qu'est l'intelligence artificielle dans vos mots? Quelle serait votre définition de l'Intelligence Artificielle?
- Quelle est votre opinion générale des technologies d'intelligence artificielle?
 - Quels en sont les bénéfices?
 - Quelles seraient vos mises en garde face à l'IA, quels risques est-ce qu'elle soulève?

Concernant votre expérience précise de la technologie proposée par EmoScienS :

- Quelles furent vos premières impressions à l'idée de tester cette technologie, avant même d'en débiter l'utilisation? Sentiments et perceptions que ça a soulevés en vous.
- Lorsque vous avez commencé à utiliser EmoScienS avez-vous pris le temps de lire l'information reliée au consentement sur la plateforme? Pourquoi?
- Lorsque vous utilisez différents sites, téléchargez de nouvelles applications, prenez-vous le temps de lire les termes et conditions? Pourquoi?
- Pouvez-vous me parler de votre expérience générale liée à l'utilisation de la technologie d'EmoScienS?
 - Quels sont les aspects que vous avez appréciés de cette technologie?
 - Quels sont les aspects que vous avez moins aimés?
- Que pourriez-vous me dire face à la protection de votre intimité et de votre vie privée en lien avec l'utilisation d'EmoScienS?

- De quelle façon votre compréhension et vos impressions face à la technologie ont-elles évolué pendant et après l'utilisation?
- Quelles sont vos impressions entourant la « Facilité d'utilisation » de la technologie
- Qu'avez-vous pensé de vos « tableaux de bord »?
- Est-ce que vous avez trouvé que le compte rendu sur votre expérience émotionnelle était relativement juste? Pourriez-vous développer?
- Que pensez-vous qu'une utilisation prolongée d'EmoScienS signifierait face à votre intelligence émotionnelle et votre sensibilisation à vos émotions?
- Trouvez-vous qu'EmoScienS a amélioré votre bien-être, votre qualité de vie? (Sur une plus longue période d'utilisation, entrevoyez-vous cette possibilité?) Avez-vous ressenti des bénéfices à l'utilisation reliés à l'utilisation d'EmoScienS? Lesquels?
- Entrevoyez-vous certains bénéfices reliés à l'utilisation de la technologie dans les organisations? Si oui, lesquels?
- Dans un contexte où vous utiliseriez EmoScienS au sein d'une/de votre organisation, que penseriez-vous de partager vos données avec votre patron à titre d'outil de gestion?
 - S'il/elle pouvait avoir le pouls de votre expérience émotionnelle au travail, de façon anonyme au sein de votre équipe ou de façon individuelle, cela pourrait lui permettre d'ajuster son mode de gestion pour viser à avoir une équipe plus stable et heureuse? Qu'en pensez-vous?
- Avez-vous des craintes ou entrevoyez-vous des dangers/effets néfastes reliés à l'utilisation de la technologie dans son état actuel? Pouvez-vous élaborer?
 - Considérez-vous que ces dangers/effets néfastes soient applicables à l'ensemble des potentiels utilisateurs? Pourquoi?
- Quels enjeux concernant la technologie que propose EmoScienS dans son état actuel devraient être adressés pour favoriser son adoption chez les consommateurs ou dans une organisation?

- Quel(s) condition(s) ou critère(s) vous a fait accepter d'utiliser la technologie d'EmoScienS?
- Quelles sont les conditions que vous jugez acceptables pour initier l'utilisation d'une nouvelle technologie d'intelligence artificielle? Quels critères motivent votre acceptation ou refus en lien avec le site, l'application ou l'objet?
- Qu'est-ce qui démontre un comportement organisationnel éthiquement responsable à vos yeux?
- Que pensez-vous des initiatives et mesures mises en place pour encadrer l'intelligence artificielle? Sont-elles suffisantes à vos yeux?
 - Connaissez-vous la Déclaration de Montréal?
 - Connaissez-vous le projet de loi 64?
 - Connaissez-vous le RGPD (Règlement général pour la protection des données)
- Comment faire selon vous afin de s'assurer que les entreprises développent des technologies éthiques et responsables en intelligence artificielle?

Conclusion

Avant de conclure l'entrevue, avez-vous des commentaires ou des questions, concernant l'entrevue ou le projet en général?

Annexe B : Guide d’entrevue révisé

Guide d’entrevue semi-dirigée

Mars 2021

Bonjour,

Suite à la période d’utilisation de la technologie d’EmoScienS qui vient de se terminer, vous avez été invité à réaliser une entrevue dans le cadre du projet de recherche pour en apprendre davantage sur votre expérience. Je tiens à vous remercier de votre participation à cette entrevue qui devrait durer d’environ 45 minutes. Je vais vous poser différentes questions pour nous permettre de mieux comprendre votre appréciation de la technologie EmoScienS et de l’intelligence artificielle en général.

Je tiens à reconfirmer que vous êtes à l’aise avec l’enregistrement de cette entrevue pour permettre la transcription dans le cadre de l’analyse des données. Sentez-vous aussi à l’aise d’arrêter la rencontre en tout temps ou de ne pas répondre à certaines questions.

Questions pour établir le profil du participant pour démontrer de l’hétérogénéité au sein de notre échantillon :

Votre rôle actuel?

Âge?

Éducation complétée?

Nous aimerions entamer l’entrevue avec quelques questions d’ordre général concernant les technologies et l’intelligence artificielle

Commencer avec les questions sur les technologies de façon générale, et ensuite sur l’IA

- Quel est votre intérêt pour les nouvelles technologies?
Utilisez-vous des objets intelligents ou connectés (montre, cellulaire, maison connectée, etc.)?
- Comment décririez-vous votre niveau d’aisance/confort face aux nouvelles technologies?

- Considérez-vous que vous êtes parmi les premiers à adopter une nouvelle technologie ou attendez-vous que la technologie soit largement utilisée avant d’y emboîter le pas?
- À quel niveau êtes-vous à l’aise de partager vos données personnelles avec des applications, des objets connectés ou des sites Web?
(Ex. Données de consommation de types préférences d’achats, achats passés, recherches effectuées; Données de santé de type dossier de santé, fréquence cardiaque, poids, habitudes de vies); Données sociales (amis, famille, photos, etc.);
 - Quels types de données personnelles acceptez-vous de partager sans problème? Pourquoi?
 - Quels types de données personnelles êtes-vous réticent, à partager? Pourquoi?
- Sentez-vous que vous avez le contrôle sur vos données personnelles dans l’univers du numérique?
- Si je vous dis “intelligence artificielle”, quel est le premier mot qui vous vient à l’esprit?
- Pourriez-vous me décrire ce qu’est l’intelligence artificielle dans vos mots? Quelle en serait votre définition?
- Quelle est votre opinion générale face à l’intelligence artificielle?
 - Quels en sont les bénéfices?
 - Quelles seraient vos mises en garde face à l’IA, quels risques est-ce qu’elle soulève?

Concernant votre expérience précise de la technologie proposée par EmoScienS :

- Quelles furent vos premières impressions à l’idée de tester la technologie d’EmoScienS, avant même d’en débiter l’utilisation? Sentiments et perceptions que ça a soulevés en vous.
- De quelle façon votre compréhension et vos impressions face à la technologie ont-elles évolué pendant et après l’utilisation?
- Lorsque vous avez commencé à utiliser EmoScienS avez-vous pris le temps de lire l’information reliée au consentement sur la plateforme? Pourquoi?
- Lorsque vous utilisez différents sites, téléchargez de nouvelles applications, prenez-vous le temps de lire les termes et conditions? Pourquoi?

- Pouvez-vous me parler de votre expérience générale liée à l'utilisation de la technologie d'EmoScienS?
 - Quels sont les aspects que vous avez apprécié de cette technologie?
 - Quels sont les aspects que vous avez moins appréciés?
- Que pourriez-vous me dire face à la protection de votre intimité et votre vie privée en lien avec l'utilisation d'EmoScienS?
- Quelles sont vos impressions entourant la « facilité d'utilisation » d'EmoScienS?
- Qu'avez-vous pensé de vos « tableaux de bord »?
- Est-ce que vous avez trouvé que le compte rendu sur votre expérience émotionnelle était relativement juste? Pourriez-vous développer?
- Trouvez-vous qu'EmoScienS a amélioré votre bien-être, votre qualité de vie? (Sur une plus longue période d'utilisation, entrevoyez-vous cette possibilité? Pourquoi?)
- Avez-vous ressenti des bénéfices reliés à l'utilisation d'EmoScienS? Lesquels?
- Entrevoyez-vous certains bénéfices reliés à l'utilisation de la technologie dans les organisations? Si oui, lesquels?
- Dans un contexte où vous utiliseriez EmoScienS au sein d'une/de votre organisation, que penseriez-vous de partager vos données avec votre patron?
 - S'il/elle pouvait avoir le pouls de votre expérience émotionnelle au travail, de façon anonyme au sein de votre équipe ou de façon individuelle, cela pourrait lui permettre d'ajuster son mode de gestion pour viser à avoir une équipe plus stable et heureuse? Qu'en pensez-vous?
- Avez-vous des craintes ou entrevoyez-vous des dangers/effets néfastes reliés à l'utilisation de la technologie dans son état actuel? Pouvez-vous élaborer?
- Considérez-vous que ces dangers/effets néfastes soient applicables à l'ensemble des potentiels utilisateurs? Pourquoi?
- Quels enjeux concernant la technologie que propose EmoScienS dans son état actuel devraient être adressés pour favoriser son adoption chez les consommateurs, ou dans une organisation?

- Quelle(s) condition(s) ou critère(s) vous a fait accepter d'utiliser la technologie d'EmoScienS?
- Quelles sont les conditions que vous jugez acceptables pour initier l'utilisation d'une nouvelle technologie d'intelligence artificielle? Quels critères motivent votre acceptation ou refus en lien avec le site, l'application ou l'objet?
- Qu'est-ce qui démontre à vos yeux un comportement organisationnel éthiquement responsable qui inspire confiance en lien avec les nouvelles technologies d'intelligence artificielle?
- Que pensez-vous des initiatives et mesures mises en place pour encadrer l'intelligence artificielle? Sont-elles suffisantes à vos yeux?
- Comment faire selon vous afin d'assurer que les entreprises développent des technologies éthiques et responsables en IA?

Conclusion

Avant de conclure l'entrevue, avez-vous des commentaires ou des questions, concernant l'entrevue ou le projet en général?

Annexe C : Grille de codage

Grille de codage		
Codes		Champs lexicaux
Variables	Sous-variables	
1. Intelligence artificielle		
1.1 Intérêt pour les nouvelles technologies (NT)		
	Intérêt moyen à fort	Le participant (pa) exprime un intérêt prononcé pour la NT
	Intérêt nul à modéré	Le pa est désintéressé de façon générale
1.2 Aisance face aux nouvelles technologies		
	Confortable	Le pa se dit à l'aise, confortable, une ressource pour son entourage
	Inconfortable	Le pa a besoin d'aide, doit être encadré, n'aime pas avoir à apprendre un nouveau système
1.3 Style de personnalité face à l'adoption de NT		
	Innovateur	Le pa est le premier à adopter les NT, aime découvrir les nouveautés
	Attente que la technologie ait fait ses preuves	Le pa préfère attendre qu'un grand nombre ait adopté la NT avant d'embroquer le pas
1.4 Littérature intelligence artificielle		
1.4.1 Un mot pour l'intelligence artificielle (IA)		Le pa utilisera un mot qui représente l'IA à ses yeux. Des mots comme « innovation technologique » ou « Surveillance » qui peuvent être lourds de sens
1.4.2 Définition pour l'intelligence artificielle		L'IA désigne l'ensemble des techniques qui permettent à une machine de simuler l'intelligence humaine, notamment pour apprendre, prédire, prendre des décisions et percevoir le monde environnant. Dans le cas d'un système informatique, l'intelligence artificielle est appliquée à des données numériques (Déclaration de Montréal).
1.5 Opinion générale face à l'intelligence artificielle		
	Bénéfices	Le pa partagera des bénéfices retrouvés dans la revue de littérature comme : Optimisation, productivité, découverte, rapidité, avancement.
	Mises en garde	Le pa partagera ses méfiances envers l'IA, les enjeux que ça soulève comme la sécurité, la perte d'emploi, la vie privée, transparence, la discrimination, l'explicabilité, la responsabilité.
1.6 Partage de données personnelles (DP)		

	Confortable	Le pa se dit tout à fait confortable à partager ses données, n'en voit pas les risques. Ou comprend les risques mais se sent en sécurité
	Inconfortables	Le pa est inconfortable à partager ses données, sentiment d'insécurité, sentiment de perte de contrôle, y voit des risques
1.7 Sentiment de contrôle sur les DP dans l'univers du numérique		
	Contrôle	Le pa sent qu'il a pleinement le contrôle sur ce qui advient de ses données personnelles dans la sphère du numérique.
	Aucun contrôle	Le pa sent qu'il n'a aucun contrôle sur la gestion de ses données personnelles dans la sphère du numérique.
2. EmoScienS		
2.1 Perception sur EmoScienS préutilisation		
Le pa partage ses premières impressions : Curiosité, craintes, méfiances, intérêt face à la technologie.		
2.2 Facilité d'utilisation d'EmoScienS		
	Facile	Le pa a trouvé que la technologie était facile, simple, pas compliquée, intuitive, claire, accessible.
	Difficile	Le pa a eu une expérience compliquée, a rencontré des embûches, a eu de la difficulté à s'y retrouver, a eu besoin de nombreuses assistances, pas intuitif.
2.3 Validité des résultats		
	Valide	Le pa a considéré que les résultats au sein de son tableau de bord étaient valides, relativement vrais, représentatifs de son état émotionnel.
	Invalide	Le pa a considéré que les résultats n'étaient pas représentatifs de son état émotionnel, pas valide, relativement faux.
2.4 Bénéfices perçus		
	Individuel	Prise de conscience face à son état émotionnel, augmenter son bien-être, permettre d'agir sur sa santé émotionnelle
	Organisationnel	Prendre le pouls du bien-être en entreprise
2.5 Confort à partager ses données avec son patron		
	Confortable	Le pa est à l'aise avec le partage de son profile émotionnel avec son patron, il n'a pas d'angoisse et est confortable.
	Inconfortable	Le pa ne serait pas à l'aise à partager son profile émotionnel avec son patron, inconfort, malaise.
2.6 Confort à partager ses données anonymisées		

	Confortable	Une fois les données anonymisées, le pa est confortable et à l'aise avec partager ses informations avec l'organisation et donc son patron.
	Inconfortable	Bien que les données soient anonymisées le pa n'est pas à l'aise avec partager son profile émotionnel et préfère garder l'information pour lui-même.
2.7 Lecture des conditions d'utilisation EmoScienS et pourquoi?		
	Oui	Le pa a pris le temps de lire les informations relatives aux termes et conditions avant d'accepter et d'utiliser la technologie
	Non	Le pa n'a pas pris le temps de lire les informations relatives aux termes et conditions avant l'accepter et d'utiliser la technologie
2.8 Lecture des conditions d'utilisation NT et pourquoi?		
	Oui	Le pa a prend généralement le temps de lire les informations relatives aux termes et conditions avant d'accepter et d'utiliser la technologie
	Non	Le pa n'a pas pris le temps de lire les informations relatives aux termes et conditions avant l'accepter et d'utiliser la technologie
3. EmoScienS et la Déclaration de Montréal		
	Principe de Bien-être	Perception du pa face à la capacité de la technologie d'accroître le bien-être chez les individus.
	Principe d'autonomie	Perception du pa face à la technologie en lien avec respect de l'autonomie des personnes et le contrôle des individus sur leur vie et leur environnement.
	Principe de protection de l'intimité et de la vie privée	Perception du pa face à la capacité de la technologie à protéger l'intimité et la vie privée des individus.
	Principe de solidarité	Perception du pa face à la capacité de la technologie à maintenir des liens de solidarité entre les individus.
	Principe de participation démocratique	Perception du pa face à la transparence et l'explicabilité de la technologie.
	Principe d'équité	Perception du pa face à la capacité de la technologie à contribuer à l'équité et à un sentiment de justice.

	Principe inclusion de la diversité	La perception du pa face à la capacité de la technologie de maintenir la diversité sociale et culturelle.
	Principe de prudence	Le pa perçoit que les mesures appropriées sont prises pour minimiser les risques entourant le développement et le déploiement de la technologie.
	Principe de responsabilité	La perception du pa face à la capacité de la technologie de maintenir la responsabilité décisionnelle des humains.
4. Initiatives influençant l'acceptabilité		
Conditions d'acceptabilité		
	EmoScienS	Le participant mentionne l'influence sociale, la réputation de la technologie ou de l'organisation, la transparence, le sentiment de sécurité face à la protection des données et de la vie privée, la fiabilité des résultats, la facilité d'utilisation, le bénéfice et l'utilité perçue en lien avec la technologie d'EmoScienS
	Nouvelles de technologies	Le pa mentionne l'influence sociale, la réputation de la technologie ou de l'organisation, la transparence, le sentiment de sécurité face à la protection des données et de la vie privée, la fiabilité des résultats, la facilité d'utilisation, le bénéfice et l'utilité perçue en lien avec une NT
5. Mesures organisationnelles qui augmentent la confiance des utilisateurs		Le pa partage sa perception face aux mesures qui lui donne confiance comme la transparence ou le sentiment de contrôle de ses données personnelles
6. Perceptions face aux initiatives pour encadrer l'IA		Le pa partage sa perception face aux mesures mises en place pour encadrer l'IA. Le pa semble connaître les initiatives en place, ou le pa semble peu informé.
	Suffisantes	Le pa considère que le cadre est suffisamment rigoureux et que les mesures sont appropriées
	Insuffisantes	Le pa considère que les initiatives ne sont pas suffisantes et que le cadre n'est pas assez rigoureux, plus devrait être fait.
7. Recommandations pour le développement responsable et éthique de l'IA		Liste des mesures proposées dans la revue de littérature dans la section 2.8 (RL)