



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE**

Développement d'un outil éthique exploratoire pour guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA pour les entreprises

Rapport de la démarche de co-construction avec les acteurs du milieu

Mandat réalisé par un groupe de travail de l'axe éthique, gouvernance et
démocratie de l'OBVIA pour le Forum IA Québec

Québec, Canada

Mars 2022

L'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA)

L'OBVIA est un réseau de recherche ouvert qui contribue aux réflexions entourant l'utilisation responsable de l'IA et du numérique en favorisant les discussions et le maillage entre les communautés de pratique, la société civile et les décideurs publics. Soutenu financièrement par les Fonds de recherche du Québec (FRQ), il fédère les expertises de plus de 260 chercheur-es couvrant les secteurs des sciences humaines et sociales, des sciences et génie et de la santé. Il bénéficie également de l'appui de plus de 125 partenaires issus de centres de recherche et d'organisations publiques et privées du Québec, du Canada et de l'étranger.

Son objectif est de mettre en lumière les enjeux cruciaux soulevés par les développements et l'utilisation de l'IA et du numérique et d'identifier des solutions plausibles aux problèmes et opportunités qu'ils soulèvent.

L'OBVIA réalise des travaux de recherche et des mandats d'intervention en partenariat en vue de développer de nouvelles connaissances, de les diffuser et d'aider les organisations québécoises à maximiser les impacts positifs du numérique et de l'IA dans nos sociétés. Ces travaux sont priorisés en fonction des orientations scientifiques de l'OBVIA et visent à répondre à la fois aux objectifs des partenaires et des équipes de recherche impliquées. Ils peuvent prendre plusieurs formes, comme des études sur le terrain, des synthèses et analyses de la littérature, la conception d'ateliers ainsi que le soutien dans l'élaboration, l'adoption et la mise en œuvre de bonnes pratiques en matière d'enjeux touchants à l'IA et au numérique.

**Observatoire international sur les
impacts sociétaux de l'intelligence artificielle et du numérique**

Pavillon Charles-De Koninck, local 2489

1030, avenue des Sciences-Humaines

Université Laval

Québec (Québec) G1V 0A6

observatoireia@ulaval.ca

observatoire-ia.ulaval.ca

ISBN : 978-2-925138-21-1

Crédits

Chercheuses responsables :

- Lyse Langlois, professeure titulaire au Département des relations industrielles de l'Université Laval et directrice générale de l'OBVIA
- Allison Marchildon, professeure titulaire au Département de philosophie et d'éthique appliquée de l'Université de Sherbrooke et coresponsable de l'axe éthique, gouvernance et démocratie de l'OBVIA

Auteurs

- Aude Marie Marcoux, candidate au doctorat à UQAM et boursière à l'OBVIA
- Justin Lawarée, conseiller en affaires publiques à l'OBVIA
- Lyse Langlois, professeure titulaire au Département des relations industrielles de l'Université Laval et directrice générale de l'OBVIA

Soutien à la recherche

- Michelle Albert-Rochette, étudiante à la maîtrise en Droit à l'Université Laval
- Émilie Guiraud, candidate au doctorat en Droit à l'Université Laval

Coordination du projet

- Nicolas Martin, responsable des mandats et des partenariats à l'OBVIA

Remerciements

Ce rapport a été réalisé dans le cadre d'un mandat avec le Forum IA Québec. L'équipe de recherche tient à remercier particulièrement Anne Boily, directrice performance de l'écosystème au Forum IA Québec et Dany Plourde, conseiller aux partenariats industriels chez IVADO, pour leur contribution dans la sélection et le recrutement des entreprises et des participants à l'étude.

Table des matières

Contexte.....	1
Mandat et objectifs.....	1
État des connaissances et justification de la question à l'étude	2
Méthodologie.....	3
1 - Revue des outils formalisés dans la littérature	3
2- Atelier de co-construction	7
Recrutement des participants	7
Rencontre délibérative : déroulement	10
3- Analyse et codage des contenus de l'atelier	13
4- Analyse et synthèse des principes et définitions qui guideront l'outil préliminaire	14
Résultats de la co-construction	14
Évaluation des sept outils	14
Table 1 : Outil ALTAI.....	14
Table 2: Responsible AI Design Assitant	15
Table 3: The BOX.....	15
Table 4: Data Ethics Decision Aid (DEDA)	16
Table 5: Everyday Ethics for AI.....	17
Table 6: Grille de réflexivité	17
Table 7: Data Ethics Canvas	17
Préférences des participants à l'étude	18
Outil intégrateur préliminaire proposé	20
Les principes et définitions retenus.....	20
Les caractéristiques.....	24
Conclusion.....	31
Références	32
Annexes.....	34
Annexe 1 : documents de référence - Sites et publications recensant les outils et lignes directrices pour le développement éthique de l'IA.....	34
Sites et publications recensant les outils et lignes directrices pour le développement éthique de l'IA.....	34
Annexe 2 : Vingt-trois modèles retenus	36

Annexe 3 : Sept outils sélectionnés	39
Annexe 4 : modèles d’invitation	40
Annexe 5 : formulaire de consentement.....	44

Table des tableaux

Tableau 1- Les sept outils d'auto-évaluation du risque sélectionnés.....	5
Tableau 2 - Caractéristiques des entreprises productrices et utilisatrices d’IA	8
Tableau 3 - Caractéristiques des participants à l’atelier de co-construction	9
Tableau 4 - Rencontre délibérative - Bloc 5, table 1	19
Tableau 5 - Rencontre délibérative - Bloc 5, table 2	20
Tableau 6 - La présence des principes directeurs au sein des outils éthiques analysés ..	22
Tableau 7 - Les sept principes consensuels et leurs définitions proposées	23
Tableau 8 - Outil intégrateur à partir des principes retenus	26

Contexte

Afin de prévenir les risques éthiques liés aux avancées en intelligence artificielle (IA), plusieurs initiatives se sont penchées sur le développement de documents normatifs visant à définir des principes éthiques directeurs. À présent, les efforts se tournent vers les mécanismes de mise en œuvre desdits principes. En effet, plusieurs dénoncent un écart difficile à surmonter entre les principes et les pratiques. Par ailleurs, aucun outil intégrateur n'a été développé au Québec. Cependant, les outils déjà développés constituent un excellent point de départ pour déterminer la forme que devrait prendre un outil intégrateur en vue d'une innovation responsable en IA au Québec. Considérant que l'innovation numérique responsable relève d'une responsabilité partagée, le développement d'un tel outil s'avère essentiel et ne peut se faire sans le concours des acteurs du milieu.

Mandat et objectifs

L'un des objectifs de l'OBVIA est de produire des connaissances publiques sur les modalités de conception et d'utilisation responsable de l'IA dans la société. En ce sens, l'OBVIA mène actuellement un chantier visant le développement d'outils éthiques pour différentes catégories d'utilisateurs, dont les entreprises. Ainsi, le présent projet consistant à produire un outil éthique intégrateur préliminaire applicable en contexte d'entreprise, s'inscrit dans un projet plus large qu'est le chantier en éthique de l'IA de l'OBVIA.

Dans le cadre de ce chantier, le Forum IA Québec a soutenu l'OBVIA dans la réalisation du présent projet visant à valider la pertinence d'outils d'autoévaluation du risque auprès d'entreprises québécoises productrices et utilisatrices d'IA. En effet, le Forum IA Québec a pour mission de faire de l'IA un levier de développement économique et social pour le Québec. Il développe une série de projets variés dont l'objectif est notamment de favoriser l'appropriation responsable de l'IA par l'ensemble des organisations, en particulier les petites et moyennes entreprises (PME). Pour mener à bien cet objectif, le Forum IA Québec a manifesté à l'OBVIA son désir de soutenir la mise en place d'un outil éthique visant à guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA pour les entreprises québécoises. Le but de cette première phase de validation consiste à produire un outil éthique intégrateur applicable en contexte d'entreprise.

Plus précisément, l'objectif de ce volet du projet s'avère de mettre en évidence les besoins de la communauté des entreprises productrices et utilisatrices de technologies relevant de l'IA et du numérique en matière d'éthique. Deux sous-objectifs s'en dégagent:

1. Identifier les préférences des acteurs québécois concernés relativement au développement d'un outil éthique en évaluant la pertinence de sept outils d'autoévaluation du risque avec ou sans score;
2. Identifier la forme que pourrait prendre un outil éthique québécois en créant collectivement un outil exploratoire.

Cette première phase de validation viendra nourrir le développement d'un outil intégrateur préliminaire et nécessairement évolutif visant à guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA pour l'ensemble de la communauté québécoise. Dans cette visée, l'outil développé sera diffusé publiquement et gratuitement sur la vitrine québécoise de l'IA du Forum et le site de l'OBVIA.

État des connaissances et justification de la question à l'étude

La numérisation croissante de la société, notamment accélérée par les récentes avancées majeures en intelligence artificielle (IA) (Cardon et al., 2018), a ouvert la voie à de prometteuses perspectives dans tous les secteurs - amenant certains auteurs à décrire l'avènement d'une quatrième révolution industrielle (Schwab, 2016). Si ces développements ont fait l'objet d'un engouement majeur, ils s'accompagnent également de nombreux risques et enjeux éthiques potentiels. Ces risques sont notamment liés à la protection de la vie privée et de la confidentialité des données nécessaires à l'apprentissage des algorithmes (Stahl & Wright, 2018); à la transparence des systèmes d'IA considérant l'opacité des réseaux de neurones (Selbst & Barocas, 2018); au respect du consentement des utilisateurs (Jones et al., 2018) ou encore à l'atténuation des discriminations et des biais de l'analyse des algorithmes (Campolo et al., 2017).

Afin de prévenir ces risques, de nombreuses initiatives à travers le monde se sont penchées sur le développement de documents normatifs visant à définir des principes éthiques directeurs pour assurer une innovation responsable en IA (84 documents ayant été recensé en 2019 - Jobin et al., 2019). Alors que le travail sur l'identification des enjeux et des principes de l'éthique de l'IA a été conséquent, les efforts se tournent à présent sur les mécanismes de mise en œuvre des chartes éthiques.

Plusieurs études empiriques vont dans le sens de ce défi et montrent la nécessité d'outiller la communauté de l'IA – à commencer par les scientifiques de données et les ingénieurs du domaine. Des chercheurs ont mis en évidence que si ces acteurs considèrent les dimensions éthiques de leur travail comme importantes, ils n'aiment pas travailler sur ces sujets et ont du mal à les adapter à leur environnement organisationnel (Spiekermann et al., 2018).

Pour répondre de cette ambiguïté et faciliter l'appropriation des principes par les différents acteurs de la communauté de l'IA, plusieurs se sont alors tournés vers le développement de modèles de mise en œuvre des principes, qu'il s'agisse d'acteurs gouvernementaux et paragonnementaux (Office of the Director of National Intelligence, 2020; Groupe d'experts indépendants de haut niveau sur l'IA, 2020);

d'entreprises privées (IBM, 2019; Microsoft, 2018), d'OSBL (AIEI, 2020; Data Ethics, 2021) ou encore de membres de la communauté académique (Floridi et al., 2020; Ryan et Stahl, 2020). Parmi ces modèles, on retrouve plusieurs outils basés sur l'auto-évaluation des risques par les acteurs de la communauté de l'IA eux-mêmes, comme les décideurs, développeurs, chercheurs, etc. (Open Data Institute, 2019; AI Ethics Lab, 2020), lesquels offrent ainsi un premier niveau de mise en œuvre pour guider la pratique des acteurs concernés.

La formalisation d'un outil éthique comme instrument d'orientation

Qu'entendons-nous par formalisation d'une charte ou d'un outil éthique en organisation ou au sein d'un secteur? Formaliser un outil éthique, c'est dans un premier temps en reconnaître l'importance et ensuite instituer une législation interne en vue d'orienter les conduites en matière d'un développement responsable de l'IA. Son utilité doit être réelle, les modalités de son application bien justifiées et sa finalité doit être clairement comprise de tous les acteurs, le tout appuyé par la direction.

Ainsi, les outils déjà développés dont certains sont formalisés au sein des organisations constituent un excellent point de départ visant à savoir lequel de ces outils retient les faveurs des entreprises participantes. Déjà au Québec, la Déclaration de Montréal (2018) a fait l'objet d'un travail de formalisation en organisation, pensons au secteur de la santé (grille de réflexivité du CHUM (Passalacqua et al., 2021), à une validation auprès d'un secteur privé (grille de réflexivité de l'OBVIA, (Marchildon, Jacob et Marcoux, 2021) et au travail normatif qui a été formalisé pour l'Autorité des marchés financiers (Dilhac et al., 2021).

Considérant que l'innovation numérique responsable relève d'une responsabilité partagée, le développement d'un tel outil s'avère essentiel, et ne peut se faire sans le concours des acteurs utilisateurs et par l'intelligence collective.

Méthodologie

Le développement de l'outil exploratoire est basé sur quatre étapes méthodologiques menées par l'équipe scientifique de l'OBVIA, à savoir: 1) une revue des outils formalisés dans la littérature; 2) un atelier de co-construction avec les acteurs du milieu; 3) une analyse des contenus de l'atelier et des préférences des participants et 4) une analyse et une synthèse des principes fondateurs de l'outil exploratoire et leur définition.

1 - Revue des outils formalisés dans la littérature

Une revue des outils éthiques qui ont été formalisés en organisation a été effectuée par l'OBVIA à l'été 2021, dans le cadre du chantier portant sur les outils éthiques dans le domaine de l'IA. Cette revue conduite sur *Google Scholar* a identifié quatorze sites et publications recensant les outils et lignes directrices pour le développement éthique de l'IA (Annexe 1).

Premièrement, alors que de nombreuses lignes directrices éthiques ont été mises en évidence dans la littérature, les vingt-trois modèles retenus (Annexe 2) dans le cadre de cette revue ont été sélectionnés sur la base des deux critères d'inclusion suivants : 1) identification et définition des principes éthiques, 2) formulation de recommandations pratiques relatives à leur mise en œuvre de façon à développer des systèmes d'IA (SIA) de manière responsable. Les données relatives aux vingt-trois modèles ont été extraites en fonction des quatre critères suivants : 1) leur crédibilité et pertinence, 2) leur source (par exemple, modèles gouvernementaux, modèles privés, modèles d'associations et d'organismes sans but lucratif, modèles académiques et autres modèles), 3) leur format (par exemple, PDF, site internet ou affiche) et 4) leur contenu (par exemple, score, recommandations, outil de sensibilisation et/ou à visée éducationnelle).

Deuxièmement, à la lumière de cette revue, nous avons sélectionné sept outils d'auto-évaluation du risque (Annexe 3). Les critères qui ont permis de retenir ces outils sont : 1) la pertinence de l'outil pour le milieu, 2) leur facilité d'usage, 3) la formalisation dans les milieux de pratiques, 4) la diversité du secteur (par exemple, gouvernemental, paragouvernemental, privé, public organismes sans but lucratif) de même que 5) l'approche. En matière d'approche, certains de ces sept outils permettent d'arriver à une forme de « score » pour évaluer la dimension éthique de la démarche de développement ou d'utilisation de technologies relevant de l'IA et du numérique, alors que d'autres se veulent davantage des outils de sensibilisation et de réflexion (Tableau 1)¹.

¹ Il est à noter que depuis la sélection des sept outils qui ont été évalués par les participants dans le cadre de l'atelier de co-construction, le AI Global Institute est devenu le Responsible Artificial Intelligence Institut, leur outil « Responsible AI Design Assistant » est devenu le « RAI Design Assistant » et les principes sur lesquels il se base ont été revus. Le présent rapport se base sur la première version de l'outil, celui qui a été évalué par les participants au projet.

Tableau 1- Les sept outils d'auto-évaluation du risque sélectionnés

Outil	Source	Format	Contenu	Approche
ALTAI - The Assessment List on Trustworthy Artificial Intelligence	(Groupe d'experts indépendants de haut niveau sur l'IA, Commission européenne, 2020)	Questionnaire interactif semi-automatisé, en ligne.	Les questions réfèrent à sept principes éthiques: 1) <i>Human Agency and Oversight</i> ; 2) <i>Technical Robustness and Safety</i> ; 3) <i>Privacy and Data Governance</i> ; 4) <i>Transparency</i> ; 5) <i>Diversity</i> , 6) <i>Non-discrimination and Fairness</i> ; 7) <i>Environmental and Societal well-being</i> ; 9) <i>Accountability</i> .	Avec score et recommandations
Responsible AI Design Assistant	(AI Global, 2020)	Questionnaire interactif semi-automatisé, en ligne.	Les questions se réfèrent à des risques éthiques liés à cinq dimensions : 1) <i>Accountability</i> , 2) <i>Explainability and Interpretability</i> , 3) <i>Data Quality</i> , 4) <i>Bias and Fairness</i> , 5) <i>Robustness</i> . L'outil peut être utilisé à toutes les phases d'un projet et est accompagné d'un guide de l'utilisateur.	Avec score et recommandations
TOOL: the Box	(AI Ethics Lab, 2020)	Questionnaire en ligne doté d'une échelle de mesure attribuant un score.	L'outil se divise en trois principes éthiques généraux : 1) <i>Autonomy</i> , 2) <i>Harm-Benefit</i> , 3) <i>Justice</i> . Ces principes se déclinent en plusieurs sous-principes, lesquels sont accompagnés de questions permettant d'évaluer leur mise en œuvre.	Auto-évaluation avec score (None, Low, Medium, High) pour chaque principe.
Everyday Ethics for Artificial Intelligence	(IBM, 2019)	PDF	Le document se concentre sur cinq domaines d'intérêt éthique : 1) <i>Accountability</i> , 2) <i>Value Alignment</i> , 3) <i>Explainability</i> , 4) <i>Fairness</i> et 5) <i>User Data Rights</i> .	Outil de sensibilisation et éducatif

			Chaque domaine est abordé au moyen de trois sections : recommandations pratiques, actions à considérer et questions à examiner.	
Data Ethics Decision Aid (DEDA)	(Utrecht Data School et Université d'Utrecht, 2020)	PDF sous forme d'affiche	Sous forme de spirale, il met en évidence différentes étapes à compléter successivement. Le centre de la spirale marque la fin du parcours. Le modèle se compose de deux principales sections : 1) <i>Data related considerations</i> (comprenant les sous sections : <i>Collecting, Using et Storing</i>), 2) <i>General considerations</i> .	Outil de sensibilisation et éducatif
Grille de réflexivité portant sur les enjeux éthiques des systèmes d'intelligence artificielle (SIA)	(OBVIA, 2021)	PDF	Une grille de réflexivité avec réponses ouvertes construite à partir des dix principes composant la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle: 1) Bien-être, 2) Autonomie, 3) Intimité et vie privée, 4) Solidarité, 5) Participation démocratique, 6) Équité, 7) Diversité, 8) Prudence, 9) Responsabilité, 10) Développement soutenable.	Outil de sensibilisation et éducatif
The Data Ethics Canvas	(Open Data Institute, 2019)	PDF	Le modèle prenant la forme d'une toile divisée en 15 sections permettant d'identifier et d'aborder les questions éthiques d'un projet.	Outil de sensibilisation et éducatif

2- Atelier de co-construction

La présentation des sept outils sélectionnés au terme de la première étape de travail a permis de lancer les discussions et d'offrir une base commune en vue de la co-construction d'un outil exploratoire et intégrateur. Le processus de co-construction a été l'objet d'un atelier de délibération réalisé le 9 décembre 2021. En effet, la pratique de la délibération est particulièrement pertinente lorsqu'il s'agit de saisir la complexité de pratiques professionnelles données et des défis en découlant (Lambert, 2016, p.2). Dans cette perspective délibérative, le processus de co-construction vise à faire interagir différents groupes d'acteurs (par exemple, producteur ou utilisateur d'intelligence artificielle) afin d'arriver de faire émerger de nouvelles propositions issues notamment de l'expérience sociale des utilisateurs (Foudriat, 2019; Déclaration de Montréal IA Responsable, 2018).

Recrutement des participants

Le recrutement des participants a été facilité avec l'aide de Forum IA Québec, un OBNL qui détient une position privilégiée vis-à-vis de la mobilisation des entreprises productrices et utilisatrices de technologies relevant de l'IA et du numérique au Québec, de même qu'avec le concours d'IVADO. Le recrutement des participants repose sur une double technique d'échantillonnage non-probabiliste : 1) par choix raisonné pour la sélection des entreprises et 2) par quotas pour la sélection des participants. Dans une optique exploratoire, une telle technique permet d'étudier des problématiques peu étudiées, sans se concentrer sur la représentativité de l'ensemble de la population (Dufour et Larivière, 2012).

La constitution de l'échantillon a débuté par la sélection des entreprises (n=12). Pour ce faire, deux critères de sélection ont été utilisés : 1) le rapport de l'organisation à l'IA (c'est-à-dire, productrice ou utilisatrice de technologies relevant de l'IA et du numérique), ainsi que 2) la taille des entreprises et leur secteur d'activité (Tableau 2). Par la suite, l'échantillon a été construit pour garantir une présence égale des développeurs, gestionnaires et analystes.

En pratique, l'atelier de co-construction a réuni 32 participants provenant de 6 entreprises utilisatrices d'IA et de 6 entreprises utilisatrices d'IA (Tableau 3). Deux modèles d'invitation de participation à ce projet ont été envoyés à Forum IA Québec (Annexe 4).

Tableau 2 - Caractéristiques des entreprises productrices et utilisatrices d'IA

Productrices d'IA		
Entreprises	Taille	Secteur(s) d'activité
1	Startup	Technologies et services de l'information (Ressources humaines)
2	Petite	Technologies et services de l'information (Aide aux entreprises, consultant en IA)
3	Startup	Technologies et services de l'information (Aide aux entreprises, consultant en IA)
4	Startup	Technologies / développement logiciel (Éducation, santé et services sociaux, gestion de projet)
5	Startup	Développement logiciel IA (algorithmes de recommandation)
6	Startup	Développement logiciel IA - compréhension du langage, analyse de documents (Assurances, services financiers, milieu juridique)

Utilisatrices d'IA		
Entreprises	Taille	Secteur(s) d'activité
7	Grande	Finance, assurance (Services financiers particuliers/entreprises, gestion de patrimoine, assurance de personnes, assurance de dommages)
8	Grande	Assurance (Assurance de dommages pour particuliers et entreprises, assurance spécialisée)
9	Grande	Recherche appliquée pour XXX qui elle se spécialise en TI / consultant en TI (pour les secteurs suivants : aéronautique, cybersécurité, défense, espace, identité et sécurité numériques, transport, "Big Data", IoT)

10	Grande	Assurances, finances (Assurance collective/individuelle/dommages, investissement, épargne)
11	Grande	Aérospatiale (Conception, fabrication et entretien de moteurs avions et hélico.)
12	Moyenne	Recherche (Fondamentale et appliquée en IA pour le secteur des services financiers.)

La participation à ce projet s'est faite sur une base volontaire. Un formulaire de consentement détaillant notamment que ce projet a été approuvé par le Comité d'éthique de la recherche de l'Université Laval (N° d'approbation 2021-324 / 03-11-2021), de même que les mesures de confidentialité, de droit de retrait et de participation volontaire ont été signés par chacun des participants (Annexe 5). De plus, lors de la rencontre, chaque participant avait la responsabilité de respecter les opinions, les expériences et les choix des autres participants et à s'engager dans la réflexion sur les outils éthiques présentés. Les participants ont la responsabilité de respecter la confidentialité des échanges qui ont eu lieu, de même que l'identification des autres participants.

Tableau 3 - Caractéristiques des participants à l'atelier de co-construction

Titre	Identité de genre	Catégorie d'âge
CEO et Co-fondateur	M	41-55
Scientifique de données	M	21-40
Analyste de projet	M	21-40
Gestionnaire senior, développement des affaires	M	21-40
Senior Machine Learning Research Team Lead	F	21-40
Gestionnaire projet	F	41-55
Analyste	M	21-40
Analyste	M	21-40
Chief AI Scientist	M	21-40
Directeur/Manager	M	21-40

Directeur de compte, Client	M	41-55
PDG	F	41-55
Scientifique de données sénior	M	21-40
Directeur IA	M	21-40
Chercheur	M	21-40
Scientifique de données	F	21-40
Machine learning researcher	M	21-40
Conseiller principal (analyste)	M	21-40
AI Delivery Manager	F	21-40
Directeur, innovation, partenariat et développement de la pratique d'analytique avancée	M	41-55
Analyste de projet	F	21-40
PDG	M	41-55
Gestionnaire	M	21-40
Director, Agile Delivery	F	21-40
Directeur de la recherche	M	21-40
Technical Manager of Product Development	M	41-55
Responsable de l'IA	M	21-40
Conseiller en science de la donnée	M	41-55
Chief Technical Officer	M	21-40
Chercheur	F	56-70
Gestionnaire principal	M	41-55
Director, Applied Machine Learning	M	41-55

Rencontre délibérative : déroulement

La rencontre délibérative a duré trois heures et s'est tenue à distance via la plateforme Zoom. Les conversations en séances plénières ont été audio-enregistrées afin de recueillir le contenu des échanges, d'en assurer l'intégrité lors de la transcription et de procéder à

l'analyse de ce contenu. Les conversations en petits groupes ont quant à elles été recueillies par l'équipe de recherche. L'atelier a suivi le plan de déroulement suivant, en six blocs.

Bloc 1 : travail préparatoire

Questionnaire préparatoire

Un questionnaire préparatoire a été envoyé à chaque participant afin d'arrimer la préparation/l'adaptation de la formation en éthique de l'IA prévue en début d'atelier au calibre des participants et également amorcer un début de réflexion/sensibilisation. Les données collectées par le biais du questionnaire ont été anonymisées: il n'est pas possible de mettre en lien le participant et son questionnaire.

Test individuel d'un des sept outils

Comme second travail préparatoire à l'atelier, l'équipe de recherche a également demandé aux participants de tester préalablement au moins un des sept outils. Ces derniers devaient répondre à des questions en vue d'évaluer l'outil assigné et en retenir les principaux avantages et inconvénients en vue de les partager au groupe lors de l'atelier. Chacun des participants s'est vu attribuer un outil à tester de manière aléatoire lors de son inscription. Plus précisément, l'équipe de recherche s'est assurée qu'un même outil soit testé à la fois par un participant issu d'une entreprise productrice d'IA et par un participant provenant d'une entreprise utilisatrice d'IA, et a veillé ainsi à ce que les sept outils aient été testés par les participants d'une même table.

Bloc 2 : Introduction en plénière (35 min)

Ce bloc d'introduction en plénière a débuté par une allocution de Lyse Langlois, directrice de l'OBVIA, suivie d'un rappel des objectifs de l'atelier de co-construction et de l'approbation éthique chapeautant ce projet, de même que d'une explication du déroulement de cette rencontre délibérative.

Ensuite, les chercheuses ont procédé à une brève introduction/formation à l'éthique de l'IA. Cette formation a été conçue sur-mesure, de façon à être notamment adaptée aux connaissances et au degré de sensibilité à l'éthique de l'IA des participants et enjeux éthiques rapportés par ceux-ci. En effet, grâce à une lecture diligente des questionnaires préparatoires complétés par tous les participants, l'équipe de recherche s'est assurée de leur offrir un contenu pertinent et adapté à leur réalité. De plus, cette formation a été conçue de façon à poursuivre le processus de sensibilisation à l'éthique de l'IA et ainsi nourrir les échanges et réflexions de l'atelier de co-construction qui a suivi.

Enfin, l'équipe de recherche a présenté les sept outils sélectionnés sur la base desquels les participants ont délibéré. En effet, dans le cadre du travail préparatoire, les participants ont été appelés à tester, préalablement à l'atelier, l'un des sept outils et à en retenir les principaux avantages et inconvénients en vue de les partager au groupe. Comme ces sept outils avaient pour visée de servir de déclencheurs aux discussions et de permettre d'offrir une base en vue de la co-construction de l'outil exploratoire et intégrateur, cette présentation bien que succincte des sept outils s'est avérée essentielle question d'offrir un préambule de ces derniers, avant la suite de l'atelier.

Bloc 3 : Évaluation en sept tables de co-construction (30 min)

Durant ce bloc de l'activité, sept tables de co-constructions ont chacune réuni des participants (développeur, gestionnaire, analyste de projet) provenant d'entreprises productrices et/ou utilisatrices d'IA ayant testé un même outil. Ce choix méthodologique de rassembler en une même table des participants provenant des deux types d'entreprises et ayant des rôles distincts a visé à assurer la multidisciplinarité des discussions et échanges. Une feuille de route a été préparée et ensuite remise aux participants afin d'y consigner leurs échanges. De plus, l'équipe de recherche a circulé entre les tables, virtuellement, afin de faciliter les échanges au besoin et de s'assurer que l'évaluation tienne compte à la fois d'éléments de forme (par exemple, format de l'outil, temps d'utilisation, accessibilité) et de fond (par exemple, besoins en termes de compréhension des principes éthiques, dimensions essentielles de la réalité pratique des participants, etc.). À cet effet, une liste de questions pour relancer la discussion a préalablement été développée.

Concrètement, une fois réunis en sept tables, les participants ont d'abord été appelés à se présenter et à désigner un responsable de consigner les échanges dans la feuille de route. S'en est suivi un premier tour de table de présentation des évaluations individuelles des outils testés, où chaque participant a présenté les points positifs et négatifs retenus, ainsi que leurs premières impressions concernant l'utilisation de l'outil qu'ils ont testé.

Un deuxième tour de table a été amorcé afin d'échanger sur les points positifs ou négatifs additionnels et aux réactions des autres participants. Finalement, cette partie a été conclue par un retour constructif sur les apprentissages et par une synthèse collective.

Bloc 4 : Retour en plénière (25 min)

Lors de ce retour en plénière, un représentant de chaque table a présenté à l'ensemble des participants à l'atelier l'évaluation de l'outil testé par les participants de sa table.

Bloc 5 : Développement d'un outil exploratoire et intégrateur en deux tables de co-construction

Durant ce cinquième bloc de l'atelier, deux tables de co-construction ont chacune regroupé des participants (développeur, gestionnaire, analyste de projet) provenant d'entreprises productrices et/ou utilisatrices. Chaque groupe de participants a eu pour objectif d'évaluer les outils existants présélectionnés et de collectivement proposer un design d'outil exploratoire et intégrateur destiné à la communauté québécoise sur la base de cette évaluation. Une feuille de route a été remise aux participants afin d'y consigner les échanges par un responsable désigné par table. De plus, l'équipe de recherche a circulé entre les tables, virtuellement, afin de faciliter les échanges et de prendre des notes.

Cette deuxième étape de l'approche délibérative avait pour objectif d'amener les participants à réfléchir à un outil idéal pouvant être facilement applicable dans leur organisation.

Lors d'un premier tour de table, trois grandes questions ont alors été soumises aux participants:

- Dites-nous si vous pensez utiliser votre outil ? Pourquoi?
- Quel serait l'outil idéal (que vous savez pertinemment que vous allez l'utiliser)?
- Vendez votre outil idéal! Quels sont vos arguments (forme, contenu)?

Ensuite, lors d'une deuxième tour de table, les participants ont été appelés à construire collectivement leur « outil idéal », pour enfin procéder à une synthèse collective.

Bloc 6 : Conclusion en plénière (20 min)

Lors de la conclusion en plénière de l'atelier de co-construction, un ou des responsables de chacune des deux tables ont présenté la forme et le contenu que devrait présenter un outil exploratoire et intégrateur en éthique de l'IA, en fonction des échanges et réflexions nourris par les participants de chaque table. Cette conclusion a été animée par les chercheuses, de façon à inciter les participants à non seulement faire le pont entre les deux tables, mais à poursuivre la délibération constructive.

Ce bloc final de l'atelier s'est conclu par les chercheuses en rappelant les objectifs et la poursuite du projet.

3- Analyse et codage des contenus de l'atelier

Chacun des sept outils sélectionnés dans le cadre de ce projet se base sur divers principes directeurs en matière d'éthique de l'IA (par exemple, transparence, équité, etc.), à partir desquels découlent l'orientation et le contenu des questions.

Lors du bloc d'évaluation en sept tables de co-constructions, les participants ont d'abord présenté leurs évaluations individuelles des outils testés, à savoir les points positifs et négatifs retenus, ainsi que leurs premières impressions concernant l'utilisation de l'outil qu'ils ont testé. Ensuite, les participants ont échangé au sujet de points positifs ou négatifs additionnels et ont conclu ce bloc par une synthèse collective.

Puis, durant le cinquième bloc de l'atelier de délibération, deux tables de co-construction ont chacune regroupé des participants (développeur, gestionnaire, analyste de projet) provenant d'entreprises productrices et/ou utilisatrices. Au cours de cette deuxième étape de l'approche délibérative, les participants devaient réfléchir à un outil idéal pouvant être facilement applicable dans leur organisation.

Les conversations en séances plénières ont été audio-enregistrées afin de recueillir le contenu des échanges. Les conversations ont ensuite été retranscrites sous la forme de verbatim afin de procéder à l'analyse de contenu. Les conversations en petits groupes ayant quant à elles été recueillies sous forme de compte-rendu par l'équipe de recherche, de même que les deux feuilles de route complétées par les participants, ont été regroupées, comparées et synthétisées.

Les données collectées lors de cette activité ont fait l'objet d'une analyse thématique (Braun & Clarke, 2006) inductive des préférences des participants en termes d'outil

éthique (chaque « thème » correspondant à des catégories d'avantages ou d'inconvénients soulevés par les participants).

4- Analyse et synthèse des principes et définitions qui guideront l'outil préliminaire

Une analyse thématique comparée des principes (et leur définition) guidant chacun des outils à l'étude a été réalisée, en y ajoutant deux recensions exhaustives des principes guidant les chartes, déclarations et outils existants (Morley and al., 2020; Fjeldans al., 2020) de façon à en dégager des principes et définitions communs et consensuels. Une triangulation des données a permis de dégager les grands principes qui faisaient consensus et d'élaborer des définitions intégratives sur lesquels se base l'outil intégrateur préliminaire proposé.

Résultats de la co-construction

Parmi les sept outils présentés, aucun n'a fait l'objet d'un consensus. Il résulte des échanges entre les participants à l'étude qu'un outil universel (*one-size-fits-all*) en matière d'éthique de l'IA s'avère irréaliste et utopique, étant donné la diversité de ses utilisateurs potentiels (par exemple, fonctions organisationnelles, secteurs d'activités ou tailles des entreprises). Malgré ce constat, tous les participants au projet s'entendent sur l'utilité et la nécessité d'avoir un tel outil intégrateur pour mieux encadrer les développements et systèmes algorithmiques.

Évaluation des sept outils

Les résultats présentés dans cette section découlent des évaluations individuelles et des échanges des participants lors du bloc 3 de l'atelier de délibération, en sept tables de co-construction.

Table 1 : Outil ALTAI

Points positifs : les membres du groupe sont tous d'accord sur le fait que l'interface est facile d'utilisation et claire. Ils ont aimé la simplicité de l'outil sous la forme de questions/réponses. C'est un bon outil pour un premier pas. Ils ont aussi trouvé que l'outil était complet dans le sens où on retrouve tous les thèmes de l'éthique de l'IA. L'outil dispose donc d'une bonne structure. De plus, chaque mot est détaillé avec une explication. Également, ils ont jugé que la multidisciplinarité de l'équipe conceptuelle était un très bon point.

Points négatifs : même si on retrouve des recommandations à la fin, les participants ont dit qu'il manquait d'accompagnement, car trop général. En effet, l'utilisateur n'a aucune indication sur ce qu'il doit faire une fois le questionnaire rempli. Les questions auxquelles l'utilisateur doit répondre par un simple oui ou non ne correspondent pas à une démarche sérieuse d'analyse. Plusieurs ont affirmé que la question n'était pas toujours claire et compréhensible. Il y a un manque de guidance sur ce point, car la réponse va dépendre

de la personne qui y répond. Un soutien et accompagnement quant aux questions seraient utiles. Les questions sont souvent sommaires, peu développées.

L'équipe a aussi répertorié un problème au niveau technique quant à la compatibilité avec des navigateurs internet.

Enfin, au niveau du format, le design du produit n'a pas plu et a été jugé « lourd ou difficile ». L'outil semble mieux convenir à une personne travaillant dans un domaine administratif.

Table 2: Responsible AI Design Assitant

Points positifs : le format de l'outil a été bien apprécié. L'équipe a trouvé que la mise en place d'un score à la fin était pertinente. Le contenu leur a aussi semblé complet. Plusieurs étaient d'accord sur le fait que cet outil éveille la conscience de l'utilisateur sur des questions et des standards importants. Cela aide à formaliser les questions autour de certains contextes.

Points négatifs : pour commencer, tout est séquentiel et les utilisateurs ont regretté le fait de ne pouvoir passer d'une question à une autre ou d'un thème à un autre. Ils ont trouvé cela irritant, surtout quand on est spécialisé dans un domaine.

L'équipe a aussi tenu à souligner que le score manque de transparence. Il leur a été impossible de savoir pourquoi le score était plus haut ou plus bas. L'outil lui-même manque donc de transparence.

Enfin, plusieurs personnes ont trouvé que le temps passé à compléter le document était trop long. L'équipe s'est questionnée sur l'utilité de mettre en place un arbre décisionnel plutôt que des questions. En effet, cela pourrait aider à gagner du temps. Le gain de temps a été jugé problématique, car toutes les entreprises ou organisations ne disposent pas de ressources dédiées à l'éthique de l'IA ou aux bonnes pratiques. La question du temps est très importante.

En résumé, les utilisateurs ont regretté que l'outil ne soit pas séquencé et qu'il ne soit pas présenté sous la forme d'un arbre de décision.

Les membres de l'équipe retiennent qu'ils n'ont pas réellement su ce qu'il fallait faire pour s'améliorer après avoir rempli le questionnaire. Ils n'ont pas été aiguillés sur les actions critiques à prendre dans le cadre du projet. L'outil porte donc un flou artistique.

Table 3: The BOX

Points positifs : les membres de l'équipe ont été d'accord sur le fait que le système est relativement simple, car il s'agit d'un questionnaire en ligne doté d'une échelle de mesure permettant de s'auto-attribuer un score. Les utilisateurs eux-mêmes peuvent tenter d'évaluer le projet sur une échelle minimale et/ou maximale.

Toujours concernant les points forts, l'équipe trouve que l'outil est bon tant qu'on ne réfléchit pas trop aux questions éthiques. En effet, il est rapide et très simple à utiliser, surtout dans un cadre, une structure où cela coûterait trop cher de s'en poser.

Points négatifs : pour commencer, il ne propose aucune forme d'accompagnement et pas de tables d'analyse. Il est donc difficile de savoir si on fait bien les choses, si le projet va dans le bon sens.

De plus, l'outil ne comporte pas de définition, alors même que certaines questions ne sont pas très claires.

On ne sait pas non plus à « quels humains » cet outil s'adresse, ni quels sont les degrés d'expertises visés. De ce fait et en sachant qu'aucun exemple n'est présenté, chacun pourrait arriver à un score différent. C'est donc très subjectif.

En définitive, l'outil s'avère difficile à utiliser et le cadre n'est pas adapté à tout le monde.

Un dernier point mérite aussi d'être soulevé : l'outil donne l'impression à l'utilisateur d'avoir fait le tour des questions éthiques en IA alors qu'on devrait, au contraire, rester vigilants. Ceci constitue un réel danger.

Table 4: Data Ethics Decision Aid (DEDA)

Points positifs : concernant le fond, les utilisateurs sont d'accord sur le fait que le contenu de l'outil est intéressant et qu'il serait d'ailleurs utile en début de projet. On peut l'illustrer avec un sprint où des projets de différents horizons pourraient tous répondre aux questions au début du processus. Ou bien encore lorsqu'un commanditaire souhaite connaître les différents aspects du projet.

Points négatifs : les questions réflexives sont trop vagues. Il faudrait au contraire que les questions soient plus adaptées au cycle de vie du projet. Concernant la forme du produit, l'équipe revient sur le fait qu'il s'agit d'un outil réflexif composé d'une série de questions. Plusieurs utilisateurs ont retenu qu'il était impossible à lire / utiliser et que le format PDF nécessite de tout retaper.

Concernant justement les questions réflexives, les utilisateurs pensent que certaines d'entre elles devraient être illustrées à l'aide d'exemples plus concrets.

L'équipe a regretté également que cette série de questions séparées en catégorie ne mène à la fin à aucune synthèse. Après cette belle période de réflexion, l'utilisateur est laissé à lui-même et n'est pas aiguillé sur ce qu'il convient de faire après.

Dans ces mêmes questions, pour celles qui semblaient avoir le plus de valeur, il était difficile d'y répondre. Bien que le fait de répondre aux questions puisse aider l'utilisateur, lorsqu'il se rend compte qu'il y a des questions auxquelles il n'arrive pas à répondre, son premier réflexe sera de se dire qu'il manque peut-être quelque chose au projet.

Un participant rajoute qu'il a trouvé que l'outil n'était pas assez concret. On a du mal à se projeter sur l'impact qu'aura le projet sur ces questions. Les questions sont trop larges et jamais mises au regard des risques identifiés. Une solution pourrait être de superposer

les risques qu'on connaît déjà depuis longtemps sur ces questions qui n'ont pas de fils directeurs. Il s'agirait d'un moyen de rendre cela plus concret.

Table 5: Everyday Ethics for AI

Points positifs : l'équipe était d'accord sur les cinq principes de base énoncés dans l'outil. L'apparence de l'outil est un peu « léchée », ce qui donne l'impression que toutes les ressources y sont passées (contenant au lieu du contenu).

Points négatifs : cet outil ne nous permet pas de savoir à quel public cible il s'adresse. Le document est lourd et bavard sans dire grand-chose. Bien qu'il utilise plusieurs termes, on ne creuse pas les sujets assez en profondeur. Il en résulte que cet outil n'est pas simple d'utilisation pour quelqu'un qui commence et il n'est pas non plus assez poussé pour quelqu'un de plus expérimenté. En définitive, cela n'aide pas du tout.

De plus, sur certains points, l'outil vient mélanger l'égalité à l'éthique.

Aucune personne de l'équipe n'a vu comment mettre cet outil en pratique.

En somme, cet outil est davantage un argument marketing pour IBM pour dire qu'on se préoccupe de l'éthique en apparence.

Table 6: Grille de réflexivité

Points positifs : l'équipe a trouvé que cet outil était intéressant. Il amène une large réflexion au niveau de l'éthique de l'IA par les nombreuses questions posées.

Points négatifs : les utilisateurs ont regretté que le sondage soit trop qualitatif. Ils auraient aimé voir le résultat du sondage avec peut-être des recommandations et ce qu'il y aurait à faire, à mettre en place, par la suite.

L'équipe a aussi trouvé que le sondage était long à remplir. Pour autant, ils n'ont pas trouvé que cela était une mauvaise chose, car plus on lit et plus on réfléchit. Ce genre de sondage pourrait donc être utilisé sous la forme d'un barème avec tout un tas de critères à remplir pour des projets de création de produit et de développement de systèmes d'IA.

De façon plus générale, les utilisateurs se sont sentis plus outillés au sujet de l'éthique grâce à tous les principes référencés dans le sondage.

Un participant rajoute qu'il aurait aimé pouvoir utiliser l'outil de façon interactive en ligne. De cette manière, l'outil serait plus facile d'accès et cela pourrait aider de futurs utilisateurs. Il rajoute que pour chaque question, il serait pertinent d'avoir des exemples.

Table 7: Data Ethics Canvas

Points positifs : l'équipe a trouvé que de façon générale, il s'agissait d'un bon outil réflexif pour internaliser l'aspect éthique. Les questions font réfléchir l'utilisateur à l'éthique. De plus, le format « one pager » est une belle présentation. Par ailleurs, le temps requis pour utiliser l'outil a été jugé adéquat et conforme à sa visée réflexive.

Points négatifs : il est difficile d'évaluer un projet, car l'outil est trop général. Les utilisateurs aimeraient pouvoir se servir de quelque chose de plus précis et détaillé.

Il manque également un lexique ; les termes sont restés non définis. Ceci semble essentiel, particulièrement lorsque l'on aborde des sujets importants comme l'explicabilité. Le site web n'était sur ce point d'aucune utilité et les utilisateurs l'ont trouvé mal fait.

Un autre regret partagé par les membres de l'équipe est qu'il n'y a pas de recommandations ni de synthèse.

En définitive, cet outil s'avère plus utile pour la réflexion que pour évaluer un projet.

Un participant ajoute que cet outil peut aider pour voir si le projet reste éthique. En revanche, il est nécessaire de faire un autre exercice supplémentaire sur le plan de la réflexion, car trop général.

Préférences des participants à l'étude

Durant le cinquième bloc de l'atelier, deux tables de co-construction ont chacune regroupé des participants (développeur, gestionnaire, analyste de projet) provenant d'entreprises productrices et/ou utilisatrices.

Au cours de cette deuxième étape de l'approche délibérative, les participants avaient pour objectif de réfléchir à un outil idéal pouvant être facilement applicable dans leur organisation. Les résultats présentés dans les tableaux 4 et 5 reflètent la synthèse collective des deux tables.

Les différentes propositions d'outils exploratoires et intégrateurs proposés par les participants à l'atelier de co-construction ont servi de base à la création d'une première version d'un outil exploratoire intégrant les préférences mises en évidence, présentée dans la section suivante.

Tableau 4 - Rencontre délibérative - Bloc 5, table 1

Thème important	Explication
Collaboration et interdisciplinarité	<ul style="list-style-type: none"> • Aspect collaboratif de l'outil très important; • Différentes personnes avec différentes spécialités doivent pouvoir répondre aux questions.
Maturité	<ul style="list-style-type: none"> • L'outil doit être ajusté à la maturité de l'organisation en matière d'IA. • Ex : mode embryonnaire VS produit final.
Recommandations	<ul style="list-style-type: none"> • Des recommandations doivent être proposées à la suite de l'utilisation de l'outil (actions à prendre, aspects à considérer). • Il doit y avoir des exemples pour les niveaux de risque;
Conscientisation	<ul style="list-style-type: none"> • L'aspect « sensibilisation » est important pour conscientiser l'équipe aux divers éléments.
Score	<ul style="list-style-type: none"> • Ce que le score représente sur le plan éthique doit être clarifié; • Des exemples pour les niveaux de gradation de risque doivent être fournis
Ancré dans les recherches en éthique de l'IA	<ul style="list-style-type: none"> • Aspect évolutif très important; • L'outil doit évoluer selon les avancées en recherche. Les aspects à considérer vont varier dans le temps.
Adaptabilité	<ul style="list-style-type: none"> • L'outil doit pouvoir s'adresser à des personnes ayant différentes connaissances sur l'éthique.
Transparence	<ul style="list-style-type: none"> • L'outil et le score doivent être transparents / explicables; • Si une section est sautée, l'outil devrait le mentionner (transparence des « raccourcis »)
Outil en « deux phases »	<ul style="list-style-type: none"> • Au départ, c'est la conscientisation qui est importante. • Ensuite, à mesure qu'une entreprise devient plus sérieuse dans le projet, l'outil pourrait être intégré dans le cycle de développement de produit. Dépend des incitatifs.
Contextualisation	<ul style="list-style-type: none"> • Contextualiser l'information produite / les réponses (parce que des personnes différentes vont donner des réponses différentes.)
Arbre décisionnel VS score par section	<ul style="list-style-type: none"> • Arbre décisionnel important pour certains; • Score détaillé par section important pour certains (forme d'audit interne).

Pour la table 1, aucun consensus n'a été dégagé parmi les participants.

Tableau 5 - Rencontre délibérative - Bloc 5, table 2

Thème important	Explication
Format	<ul style="list-style-type: none"> Format interactif, intuitif (par exemple, pas juste un PDF à dérouler).
Clarté / lexique	<ul style="list-style-type: none"> Lexique qui permet de définir certains mots pour que tout le monde puisse profiter de l'outil.
Spécifique	<ul style="list-style-type: none"> Questionnement : est-ce possible d'avoir un outil exhaustif en termes d'éthique de l'IA? Il vaudrait peut-être mieux que l'outil soit exhaustif pour un problème spécifique (ex : transparence, biais).
Audience	<ul style="list-style-type: none"> L'audience devrait être identifiée dès le début (équipe de projet, <i>data scientists</i>).
Score	<ul style="list-style-type: none"> Préférable d'avoir un score pour pouvoir mesurer le risque et le communiquer assez clairement.
Pistes de réflexion	<ul style="list-style-type: none"> L'outil, à défaut de proposer des recommandations qui puissent s'appliquer à tout le monde, devrait au moins proposer des pistes de réflexion. « Avoir une idée » des recommandations pour remédier aux problèmes identifiés.
Intégration aux processus opérationnels	<ul style="list-style-type: none"> L'outil devrait être utilisé comme garde-fou à chaque étape du développement de projet; Il devrait donc poser les bonnes questions aux bons moments et faire intervenir tous les acteurs du projet dans leurs fonctions spécifiques (par exemple, gestionnaire se pose une question x, le <i>data scientist</i> se pose une question y.)
Éducatif	<ul style="list-style-type: none"> L'outil devrait être éducatif par rapport à l'IA et ainsi mettre en lumière et éduquer sur les problématiques spécifiques à l'IA.

Pour la table 2, aucun consensus n'a été dégagé parmi les participants

Outil intégrateur préliminaire proposé

Les principes et définitions retenus

Chacun des sept outils sélectionnés dans le cadre de ce projet se base sur divers principes directeurs en matière d'éthique de l'IA, à partir desquels découlent l'orientation et le contenu des questions. Ainsi, les principes et leur définition retenus composant l'outil intégrateur préliminaire proposé se basent sur les cycles trois et quatre de notre méthodologie, soit :

- Les données collectées lors de l'activité de co-construction ayant fait l'objet d'une analyse thématique inductive des préférences des participants quant aux principes qui devraient guider l'outil intégrateur exploratoire; (Braun & Clarke, 2006)

- L'analyse thématique comparée des principes (et leur définition) guidant chacun des outils à l'étude réalisée, en y ajoutant trois recensions exhaustives des principes guidant les chartes, déclarations et outils existants (Floridi and Cows, 2019; Morley and al., 2020; Fjeld and al., 2020) de façon à en dégager des principes et définitions communs et consensuels.

Une triangulation des données a permis aux chercheurs de dégager les grands principes qui faisaient consensus et qui étaient en lien avec les avancées en éthique de l'IA (Tableau 6). De plus, des définitions ont été proposées tel que demandé par les participants (Tableau 7) et sur lesquelles se base l'outil intégrateur préliminaire proposé.

Il est à noter qu'il ressort de cette analyse que l'un des sept outils soumis à l'évaluation des participants se base sur les mêmes sept principes retenus par les chercheurs, soit l'outil ALTAI. L'analyse thématique inductive des préférences des participants, de même que l'analyse comparée des principes guidant les six autres outils, incluant l'ajout de deux recensions exhaustives des principes guidant les chartes, déclarations et outils existants (Floridi and Cows, 2019; Morley and al., 2020; Fjeld and al., 2020) permettent en effet de conclure que les principes éthiques guidant l'outil ALTAI font consensus. Toutefois, tel que démontré dans la section « Résultats de l'atelier de co-construction » de ce rapport, les participants ont jugé que l'outil ALTAI comporte plusieurs points négatifs, et ce, tant sur le plan de la forme que du contenu. Ces résultats révèlent la pertinence de la création du nouvel outil intégrateur préliminaire proposé.

Tableau 6 - La présence des principes directeurs au sein des outils éthiques analysés

	Outil ALTAI	Responsible AI Design Assitant	The BOX	Data ethics decision aid	Everyday Ethics for AI (IBM)	Grille de réflexivité (OBVIA)	Data Ethics Canvas	Academics ²	Bernam Klein Center
Reddition de compte et responsabilité	*	*	-	*	*	*	-	*	*
Explicabilité et transparence	*	*	-	*	*	-	-	*	*
Vie privée	*	*	-	*	-	*	-	*	*
Robustesse et sécurité	*	*	-	*	*	*	*	*	*
Autonomie	*	-	*	-	-	*	*	*	*
Équité, non-discrimination et inclusion	*	*	*	*	*	*	-	*	*
Bien-être et développement soutenable et environnemental	*	-	*	-	*	*	*	*	*
Légende	*	= Présence du principe éthique			-	= Absence du principe éthique			

² Les références académiques utilisées pour compléter le tableau relatif à la présence ou à l'absence des principes éthiques sont : 1) MORLEY, JESSICA & FLORIDI, LUCIANO & KINSEY, LIBBY & ELHALAL, ANAT. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and engineering ethics; 2) FLORIDI, L., & COWLS, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review.

Tableau 7 - Les sept principes consensuels et leurs définitions proposées

Principes retenus	Définitions
Responsabilité et Reddition de compte	Le développement et l'utilisation des SIA doivent se faire de façon responsable, ce qui signifie qu'ils doivent prendre en considération leurs impacts sur les différents acteurs/parties prenantes, prendre les mesures adéquates pour les minimiser et rendre compte de ces mesures de façon transparente.
Explicabilité et transparence	<p>Le développement et l'utilisation des SIA doivent se faire de façon transparente, ce qui comprend la mise en place de mécanismes permettant la traçabilité de leurs systèmes ainsi que la communication ouverte à cet égard, incluant les limites de leurs SIA.</p> <p>Les concepteurs et utilisateurs de SIA doivent en parallèle s'assurer que les fonctionnalités de leurs systèmes satisfont au critère d'explicabilité, c'est-à-dire qu'elles soient intelligibles et qu'elles soient en mesure de fournir des explications à savoir où, quand, comment et pourquoi ils sont utilisés, de même que d'expliquer les résultats obtenus grâce aux SIA.</p>
Vie privée	<p>La vie privée doit être protégée de l'intrusion de SIA et de systèmes d'acquisition et d'archivage des données personnelles (SAAD).</p> <p>Le développement et l'utilisation des SIA doivent respecter le droit fondamental au respect à la vie privée qui se révèle particulièrement sensible aux incidences des SIA. Des mécanismes permettant la prévention d'atteinte au respect de la vie privée requièrent notamment une gouvernance appropriée des données qui porte sur la qualité et l'intégrité des données utilisées, leur pertinence et leur proportionnalité par rapport au domaine dans lequel les SIA sont déployés, leurs protocoles d'accès et la capacité à traiter les données d'une manière qui protège la vie privée.</p>
Robustesse et sécurité	Le développement et l'utilisation des SIA doivent s'assurer de la robustesse technique et de la sécurité des systèmes. Pour ce faire, des mécanismes doivent être mis en place afin de s'assurer que les SIA soient fiables, qu'ils fonctionnent et soient utilisés avec justesse et en fonction de leur objectif premier, et qu'ils ne risquent pas d'être compromis par des tiers. Cela signifie que leurs SIA soient mis au point selon une approche de prévention des risques, c'est-à-dire, de manière à ce que ces systèmes se comportent de manière fiable, conformément aux attentes, tout en réduisant le plus possible les atteintes involontaires et inattendues sur les personnes. Ce principe s'applique également aux modifications potentielles de l'environnement dans lesquels sont exploités les SIA, ou de leur cadre d'application, ainsi qu'à la présence d'autres agents (humains et artificiels) pouvant avoir des interactions antagonistes avec le système.

Autonomie (agentivité et respect des valeurs humaines)	<p>Le développement et l'utilisation des SIA doivent soutenir l'autonomie et la prise de décisions humaines et non s'y substituer. Cela suppose notamment qu'un contrôle humain doive être mis de l'avant lors de tout processus impliquant un SIA et ce, afin d'éviter la déresponsabilisation des agents humains face aux décisions en découlant.</p> <p>De façon plus large, les SIA doivent être à la fois protecteurs d'une société démocratique, prospère et équitable, tout en étant au service de l'utilisateur, et ce, en protégeant les droits fondamentaux.</p>
Équité, non-discrimination et inclusion	<p>Le développement et l'utilisation des SIA doivent s'assurer de favoriser l'équité, la non-discrimination et l'inclusion tout au long du cycle de vie d'un SIA.</p> <p>Cela implique la prise en compte et la participation démocratique de l'ensemble des parties prenantes concernées tout au long du processus de développement et de mise en œuvre d'un SIA de veiller à l'égalité de traitement et d'accès au moyen de processus inclusif, ainsi qu'à l'égalité de traitement.</p>
Bien-être et développement soutenable et environnemental	<p>Le développement et l'utilisation des SIA doivent considérer la société au sens large, les êtres sensibles et l'environnement comme des parties prenantes, et ce, tout au long du cycle de vie de la SIA.</p> <p>La durabilité et la responsabilité écologique des SIA doivent être encouragées, et il convient de promouvoir la recherche de solutions d'IA ancrées dans les objectifs de développement durable afin d'y intégrer une préoccupation pour les générations futures et ne pas aller à l'encontre du bien-être humain des êtres sensibles et de l'environnement naturel.</p>

Les caractéristiques

Les caractéristiques de l'outil intégrateur préliminaire doivent s'appuyer sur des éléments jugés importants par les participants à ce projet, et ce, tant sur le plan de la conception technique que sur le plan pédagogique et scientifique.

Les participants se sont entendus à majorité sur la forme que doit prendre l'outil. Cet outil doit être intuitif, interactif et hébergé sur un site web. Le format PDF n'est pas une option viable lorsqu'on prévoit une utilisation régulière. L'outil doit être simple d'utilisation, même pour des utilisateurs possédant moins ou pas de compétences techniques, facile à comprendre.

L'aspect évolutif de l'outil s'est révélé une autre importante caractéristique. En ce sens, une veille est suggérée pour suivre les meilleures pratiques en éthique de l'IA de même que la mise en place d'une boucle de rétroaction « feed-back loop » auprès des

participants/utilisateurs de l'outil/groupes de parties prenantes. Le but ici est de le rendre toujours plus adéquat en fonction des expériences. Le domaine de l'éthique de l'IA évolue rapidement et il est important d'en suivre les avancées.

Les participants ont aussi demandé que l'outil soit flexible et n'enferme l'utilisateur dans une marche à suivre de type séquentiel. Les participants ont précisé l'importance d'aller directement aux principes concernés. La dimension collaborative lors de l'utilisation de l'outil s'avère cruciale pour soutenir la réflexion éthique. Différentes spécialités et différents profils sont souvent mis à profit pour un projet en particulier. L'outil devrait être adaptable pour plusieurs acteurs et modulé en fonction des différentes étapes de développement d'un SIA.

Les participants souhaitent obtenir un outil basé sur un arbre décisionnel en fonction du niveau de risque (mitigation des risques de faible à élevé) et ce, pour chacun des principes. Un arbre décisionnel (questions influencées par les réponses aux questions précédentes) a été jugé utile pour la gestion du temps, mais aussi pour mieux cerner l'impact de sa réponse, ce qui contribue à une forme de sensibilisation éthique. De plus, les participants souhaitent que la gradation des risques soit expliquée clairement³ et accompagnée d'exemples concrets.

Les participants ont demandé à ce que l'outil inclue des scores en fonction des réponses issues de l'arbre décisionnel et de la mitigation des risques. Des recommandations adaptées au score obtenu et aux principes concernés devraient être offertes offrant par exemple, des suggestions sur les actions à prendre, sur les aspects à considérer, et ce, en fonction du score obtenu. En effet, si l'outil permet d'obtenir un score détaillé par section, cela permet à l'utilisateur d'avoir accès à un rapport complet sur les forces/faiblesses pour chaque domaine de même qu'aider à la détection des « zones » à améliorer en termes de mitigation du risque. Une telle formule permet aussi à une forme d'audit interne qui génère des solutions et des recommandations basées sur les standards de bonnes pratiques, mais aussi sur les normes ISO et légales. Cette caractéristique renforce la pertinence d'un outil dynamique, interactif et évolutif.

Dernière caractéristique demandée par les participants concerne le besoin d'avoir un ancrage pédagogique, éducatif en matière de sensibilisation à l'éthique de l'IA. Ainsi, les participants souhaitent que l'outil présente des questions claires et accessibles, accompagnées d'un lexique et d'exemples pour chacun des principes.

³ Un participant a mentionné le fait d'être souvent face à une sorte de paradoxe lorsque l'on souhaite s'améliorer sur le plan de l'explicabilité et par le fait même d'avoir un outil qui n'explique pas clairement comment sont établis les scores.

Enfin, nous notons que l’outil ne pourra certes répondre aux besoins de tous, qu’il s’avère probablement impossible de créer un outil exhaustif et parfait. Toutefois, les quelques caractéristiques mises de l’avant lors de ce mandat correspondent aux besoins exprimés des participants. Ils recommandent aussi d’avoir un document d’accompagnement expliquant la démarche, le tout accompagné de références et lectures complémentaires.

À la suite des principes retenus (Tableau 7), tout en prenant en considération les caractéristiques souhaitées, il est proposé de concevoir des sous-principes et des définitions qui illustrent l’architecture de cet outil intégrateur (Tableau 8).

Tableau 8 – Sous-principes et unités d’analyse de l’outil intégrateur à partir des principes retenus⁴

Reddition de compte		
Sous-principe	Unité d'analyse	Définition
<i>Responsabilité</i>		Les droits et devoirs des employés au regard du système d'intelligence artificielle sont formalisés et explicitement décrits dans un document interne.
	Définition des responsabilités	L'entreprise formalise dans un document accessible à l'ensemble des employés les tâches en lien avec le système d'intelligence artificielle et la répartition claire de celles-ci entre les employés.
	Responsabilité sociale	L'entreprise veille également à prendre en compte les retombées sociales et environnementales engendrées par la mise en œuvre du système d'intelligence artificielle.
<i>Évaluation</i>		L'entreprise met en œuvre les procédures requises pour suivre, mesurer et s'assurer de la légalité des décisions et des composants du système d'intelligence artificielle.
	Impact	L'entreprise développe ou a développé une méthodologie d'évaluation des impacts et de l'incidence du système d'intelligence artificielle qu'elle utilise.
	Conformité	L'entreprise évalue la conformité du système d'intelligence artificielle et des décisions rendues au regard du cadre réglementaire en vigueur.
	Audit	L'entreprise ou une partie prenante extérieure peut évaluer les données, leur fiabilité ainsi que les résultats du système d'intelligence artificielle.

⁴ Il est suggéré de mettre le tableau 8 sous forme d’arbre décisionnel.

Indépendance		Le processus de décision, son suivi et son évaluation sont confiés à un organisme ou une équipe d'employés n'ayant aucun lien direct avec les
	Protection des employés	Le statut de lanceur d'alerte est créé afin de protéger tout employé dénonçant des pratiques ou des décisions abusives en lien avec le système d'intelligence artificielle.
	Équipe de suivi	Une équipe indépendante est dédiée à l'évaluation, au suivi ou à l'audit des systèmes d'intelligence artificielle utilisés ou développés au sein de l'entreprise
Robustesse et sécurité		
Décision humaine		La décision finale revient exclusivement à un être humain et non au système d'intelligence artificielle.
	Opérateur	Un membre ou une équipe de l'entreprise est responsable de valider les décisions rendues par les systèmes d'intelligence artificielle.
	Correction	En cas d'erreur, un membre ou une équipe de l'entreprise est en mesure de corriger la décision et, le cas échéant, de réviser l'algorithme.
Données sensibles		Les données individuelles dont le dévoilement peut causer un préjudice majeur doivent faire l'objet d'une sécurité maximale.
	Protection	Les données sensibles et personnelles sont stockées à l'écart des données non-sensibles et font l'objet d'un degré de sécurité plus élevé.
	Inventaire	L'ensemble des données sensibles collectées sur les usagers est listé dans un document crypté.
Infrastructures		Les données sont entreposées dans des conditions de sécurité optimale.
	Décentralisation	L'entreposage des données sensibles est divisé entre plusieurs centres d'entreposage.
	Expertises	L'entreprise, que cela soit en interne ou en externe, dispose des compétences nécessaires en matière de cybersécurité.
Prévention des risques		Tout préjudice potentiel ou réel occasionné par le système d'intelligence artificielle doit être évité.
	Cycle de vie	Les préjudices doivent être identifiés et évalués en fonction des étapes du cycle de vie du système d'intelligence artificielle durant laquelle ils se produisent.
	Test	Le système d'intelligence artificielle a conclu avec succès un ou plusieurs essais dans des conditions similaires à son utilisation effective.
	Certification	Le système d'intelligence artificielle ou les employés ont été certifiés dans un domaine pertinent tels que la cybersécurité, la prévention des risques ou encore la gestion de données sensibles.
Explicabilité		

Explicabilité		Les décisions rendues par le système d'intelligence artificielle peuvent être explicitées aux usagers.
	Facteurs	Le membre ou l'équipe de l'entreprise est en mesure les facteurs à la base de la décision rendue.
	Résultat	La décision prise par le système d'intelligence artificielle peut-être expliquée dans des termes adaptés aux usagers.
Transparence		Le processus menant à la décision prise par le système d'intelligence artificielle peut être explicité par les employés chargés de l'utiliser.
	Traçabilité	Toute modification apportée au système d'intelligence artificielle est consignée dans un document dédié.
	Consultation	Les données sur la base desquelles la décision est prise peuvent être consultables par les usagers.
Interprétabilité		L'interprétabilité se réfère à la connaissance de la logique interne du système d'intelligence artificielle.
	Algorithme	La composition et l'évolution de l'algorithme à l'origine de la prise de décision peuvent être explicitées par les employés chargés de l'utiliser.
	Évaluation	Une méthode d'évaluation de l'interprétabilité est élaborée et mise en œuvre dans l'entreprise.
Protection de la vie privée		
Anonymisation des données privées		Le processus de transformation des données personnelles en vue de minimiser la ré-identification des usagers.
	Risque de ré-identification	Une évaluation du risque de ré-identification (par exemple, via le calcul d'un score) est conduite.
	Méthode de généralisation	Une ou plusieurs méthodes de généralisation pertinentes sont mobilisées pour l'ensemble des données afin de diminuer le risque de ré-identification des usagers.
Consentement des usagers		Le consentement explicite, simple, informé et spécifique est demandé aux usagers.
	Accès au formulaire de consentement	Le formulaire de consentement est consultable et accessible en tout temps par les usagers.
	Contenu du formulaire de consentement	Le formulaire de consentement contient les informations nécessaires pour éclairer de façon pertinente la décision des usagers.
Données		L'entreprise limite la collecte des données privées et emploie des méthodes de collecte pertinentes au regard des objectifs poursuivis
	Objectifs de la collecte	L'entreprise informe les usagers des données privées collectées et des objectifs justifiant cette collecte.
	Méthode de collecte	L'entreprise identifie et explicite les mécanismes par lesquels les données privées sont collectées.
	Données nécessaires	La collecte des données est strictement limitée aux données privées nécessaires pour le fonctionnement du SIA

Conservation		L'entreposage des données est limité dans le temps.
	Période de validité	Une date de fin d'entreposage des données collectées est établie au terme de laquelle les données sont détruites.
	Méthode de destruction des données	L'entreprise utilise des méthodes de destruction efficaces des données et protégeant les usagers de toutes récupérations possibles.
Droits individuels		Les usagers ont la possibilité d'avoir accès à leurs données personnelles, à les réviser ou au droit à l'oubli
	Droit à l'oubli	Les usagers ont droit à demander que leurs données personnelles soient supprimées des ensembles de données utilisés par l'entreprise.
	Révision	Les usagers ont droit à demander que leurs données personnelles soient révisées des ensembles de données utilisés par l'entreprise.
	Accès des participants aux données	Les usagers ont droit à consulter leurs données personnelles utilisées par l'entreprise.
Entreposage sécuritaire		Les données collectées sont protégées des accès non autorisés, des pertes, des manipulations ou destruction.
	Politique d'accès	Une liste des employés ayant accès aux ensembles de données est établie au sein de l'entreprise.
	Clé de sécurité	L'accès aux données sensibles se fait uniquement au moyen de l'utilisation d'une clé de sécurité efficace.
	Prévention des dommages	Les usagers ne peuvent subir de préjudices causés par des accès non autorisés, des pertes, des manipulations ou destruction.
Non-Discrimination		
Prévention de biais		
	Algorithme	L'algorithme a été conçu par une équipe de personnes aux parcours, compétences, genres ou origines différents.
	Données	Les données font l'objet d'une analyse de la présence de biais tels que les biais de mesure, d'agrégation, de déploiement ou encore d'évaluation.
Qualité des données		
	Entraînement	La taille des ensembles de données utilisés pour effectuer les tests du système d'intelligence artificielle est représentative de la population et de la diversité des situations rencontrées.
	Source	Les données proviennent d'organisations et de méthodes de collecte transparente et ayant fait l'objet d'un consentement libre et éclairé des usagers

<i>Inclusion</i>		
	Participation des usagers	Des usagers ou d'autres parties prenantes externes à l'entreprise ont collaboré au développement et à la mise en œuvre du système d'intelligence artificielle.
	Publics marginalisés	L'utilisation du système d'intelligence artificielle ne renforce pas l'exclusion des publics fragilisés ou marginalisés.
	Rédaction	L'interface d'utilisation du système d'intelligence artificielle emploie l'écriture inclusive.
<i>Équité</i>		
	Équité procédurale	Les usagers ont accès au même service et aux mêmes procédures.

Comme mentionné précédemment, les tableaux 7 et 8 représentent le contenu de l'outil intégrateur qui pourra être raffiné en fonction des caractéristiques souhaitées par les participants. Cette section, soit les caractéristiques demandées ne font pas partie du présent mandat et nécessite une phase complémentaire pour atteindre les besoins des utilisateurs.

Conclusion

Conformément à sa mission de faire de l'IA un levier de développement économique et social pour le Québec, Forum IA Québec développe et soutient des projets visant à favoriser l'appropriation responsable de l'IA par l'ensemble des organisations, en particulier les petites et moyennes entreprises (PME). En ce sens, Forum IA Québec souhaite outiller les entreprises québécoises par la mise en place d'un outil éthique visant à guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA. Les participants à ce projet de recherche-action ont d'ailleurs démontré que cette initiative répond à un réel besoin. Ils ont en effet manifesté une sensibilité éthique forte à l'égard des préoccupations avancées dans les différents outils testés, de même que des attentes très précises sur la conception technique et scientifique de l'outil à développer. Ces attentes démontrent le fort engagement du groupe à se doter d'un outil qui pourra les guider vers les bonnes pratiques en matière d'éthique de l'IA et qui correspond à leurs besoins.

La démarche co-construction qui a été initiée avec les participants a permis de produire le contenu et la matrice structurelle de l'outil désiré qu'on retrouve dans les tableaux 7 et 8. Une phase supplémentaire sera nécessaire pour rencontrer les exigences demandées compte tenu du fait qu'aucun des sept outils sélectionnés et soumis à l'évaluation des participants n'a obtenu l'aval du groupe. Cette seconde phase permettra de produire concrètement un outil intégrateur en matière d'éthique de l'IA dont la forme et le contenu répondent aux préférences de la communauté québécoise d'entreprises productrices et utilisatrices de technologies relevant de l'IA. Il s'avère qu'un outil bien conçu et bien ancré dans la pratique aura plus de chance d'influencer significativement les actions en matière de développement et d'utilisation responsables des SIA.

Références

Passalacqua, A., Blanckaert, C. Bédard, K., Bédard, S., Beffert, U., Blouin, C., De Patureaux, R., Petitgand, C., Reeves, S. et Malas, K. (2021). Guide des principes d'innovation et d'IA responsables en santé, CHUM, CEPPP, CHU Ste-Justine, Collège John Abott, CRCHUM, EmoScienS, Forum IA Québec, Institut du Nouveau monde, Imagia, IUCPQ, IVADO, Observatoire québécois des inégalités, Polytechnique Montréal, Prudence AI, Université de Montréal, Université libre de Bruxelles.

Marchildon, Jacob et Marcoux. (2021). Grille de réflexivité portant sur les enjeux éthiques des systèmes d'intelligence artificielle (SIA). <https://observatoire-ia.ulaval.ca/grille-de-reflexivite-portant-sur-les-enjeux-ethiques-des-systemes-dintelligence-artificielle-sia/>

Dilhac et al. (2021). L'intelligence artificielle en finance. Recommandations pour une utilisation responsable. https://lautorite.qc.ca/fileadmin/lautorite/grand_public/publications/professionnel/s/rapport-intelligence-artificielle-finance-fr.pdf

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1).

AI EI. "AI Ethics Impact Group: From Principles to Practice – VDE", (2020), En ligne : <https://www.ai-ethics-impact.org/en>

AI Ethics Lab. "Tool: The Box", (2020), En ligne : Toolbox: Dynamics of AI Principles <http://aiethicslab.com/the-box/>

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 Report*. https://ainowinstitute.org/AI_Now_2017_Report.pdf

Cardon, D., Cointet, J.-P., & Mazières, A. (2018). La revanche des neurones. *Rezeaux*, n° 211(5), 173-220.

Déclaration de Montréal IA Responsable. (2018). *Rapport de la Déclaration de Montréal pour un Développement Responsable de l'Intelligence Artificielle* (p. 321). https://docs.wixstatic.com/ugd/ebc3a3_d806f109c4104c91a2e719a7bef77ce6.pdf

Data Ethics. "Data Ethics Impact Assessment", (2021), En ligne : <https://dataethics.eu/wpcontent/uploads/dataethics-impact-assessment-2021.pdf>

Floridi, Luciano et al. "How to Design AI for Social Good: Seven Essential Factors" (2020) 26:3 *SciEng Ethics* 1771–1796, En ligne : <https://link.springer.com/article/10.1007/s11948-020-00213-5>

Groupe d'experts indépendants de haut niveau sur l'IA. "ALTAI - The Assessment List on Trustworthy Artificial Intelligence", (2020), En ligne : <https://futurium.ec.europa.eu/en/europeanai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>

Hagendorff, T. (2020). The Ethics of AI Ethics : An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99-120. <https://doi.org/10.1007/s11023-020-09517-8>

IBM. "Everyday Ethics for Artificial Intelligence", (2019), En ligne : <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence : The global landscape of ethics guidelines. *arXiv:1906.11668 [cs]*. <http://arxiv.org/abs/1906.11668>

Jones, M. L., Kaufman, E., & Edenberg, E. (2018). AI and the Ethics of Automating Consent. *IEEE Security Privacy*, 16(3), 64-72. <https://doi.org/10.1109/MSP.2018.2701155>

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 729-733. <https://doi.org/10.1145/3236024.3264833>

Microsoft. "Responsible bots: 10 guidelines for developers of conversational AI" (2018), En ligne : <https://www.microsoft.com/en-us/research/publication/responsible-bots/>

Mittelstadt, B. (2019). *AI Ethics – Too Principled to Fail?* (SSRN Scholarly Paper ID 3391293). Social Science Research Network. <https://papers.ssrn.com/abstract=3391293>

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From What to How. An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. *arXiv:1905.06876 [cs]*. <http://arxiv.org/abs/1905.06876>

Office of the Director of National Intelligence. "Artificial Intelligence Ethics Framework for the Intelligence Community", (2020), En ligne : https://www.dni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf

Open Data Institute. "The Data Ethics Canvas", (2019), En ligne : <https://theodi.org/article/dataethics-canvas/>

Ryan, Mark et Bernd Carsten Stahl. "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications" (2020), *Journal of Information, Communication and Ethics in Society*, En ligne : <https://doi.org/10.1108/JICES-12-2019-0138>

Schwab, K. (2016). *La quatrième révolution industrielle* (Dunod). World Economic Forum.

Selbst, A. D., & Barocas, S. (2018). *The Intuitive Appeal of Explainable Machines* (SSRN Scholarly Paper ID 3126971). Social Science Research Network. <https://papers.ssrn.com/abstract=3126971>

Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) Ethics : Ethics of AI and Ethical AI. *Journal of Database Management (JDM)*, 31(2), 74-87. <https://doi.org/10.4018/JDM.2020040105>

Spiekermann, S., Korunovska, J., & Langheinrich, M. (2018). Inside the Organization : Why Privacy and Security Engineering Is a Challenge for Engineers[40pt]. *Proceedings of the IEEE*, 1-16. <https://doi.org/10.1109/JPROC.2018.2866769>

Stahl, B. C., & Wright, D. (2018). Ethics and Privacy in AI and Big Data : Implementing Responsible Research and Innovation. *IEEE Security Privacy*, 16(3), 26-33. <https://doi.org/10.1109/MSP.2018.2701164>

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The Role and Limits of Principles in AI Ethics : Towards a Focus on Tensions. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 195-200. <https://doi.org/10.1145/3306618.3314289>

Annexes

Annexe 1 : documents de référence - Sites et publications recensant les outils et lignes directrices pour le développement éthique de l'IA

Sites et publications recensant les outils et lignes directrices pour le développement éthique de l'IA ⁵			
Liens URL	Mise à jour	Auteur	Notes
AI Ethicist	2021	Merve Hickok	Sections : AI Principles AI Governance National Strategies AI Organizations Ethics Cases Research on AI Biais, Fairness
AI Ethics Guidelines Global Inventory	2021-04	AlgorithmWatch (Allemagne, Berlin)	Cartographie des cadres de l'IA opérationnalisant des principes éthiques.
AI Principles Map	2021	AI Ethics Lab (É-U, Cambridge)	Cartographie des cadres de l'IA opérationnalisant des principes éthiques.
AI Ethics Typology (2019)	- (Article)	Morley, Floridi et al. (R-U, Oxford)	Typologie discutée dans From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principle into Practices (2019)
AI Global Bubble	2018	N/D	Cartographie des cadres de l'IA opérationnalisant des principes éthiques. Version interactive Données
Dote Everyone	2021	DoteEveryone *Site désormais géré par le Ada Lovelace Institute (R-U, Londres)	"This directory focuses on initiatives which help make more tech responsible/ethical, rather than individually ethical products or services. So: standards, training courses, advocacy, guidelines, campaigns, regulations, networks, tools, etc."
Ethical frameworks, tool kits, principles, and oaths	2020-10	Kathy Baxter (É-U, San Francisco)	Frameworks, Tools & Toolkits, Checklists, Principles, Oaths, Manifestoes, and Codes of Conducts, Policy Papers, White Papers, Statements, Reports, Other Resources

⁵ Mise à jour du tableau : 2021-06-15.

Ethics in AI research papers and articles	2021-05		Peer-reviewed, Government, NGO and expert publications, Popular Press
Ethics Standards repository	2021	Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS) (R-U)	"The Global AI Standards Repository is the world's first centralized, transparent notification system that captures AI and Autonomous and Intelligent Systems standards and standards in progress."
Linking-ai-principles	2020	Yi Zeng – Institute of Automation, Chinese Academy of Sciences (Chine, Beijing)	"Linking Artificial Intelligence Principles (LAIP) is an initiative and platform for integrating, synthesizing, analyzing, and promoting global Artificial Intelligence Principles and their social and technical practices World Wide, from different research institutes, non-profit organizations, non-governmental organizations, companies, etc. The efforts aim at understanding in which degree do these different AI Principles proposals share common values, differ and complete each other."
An Updated Round Up of Ethical Principles of Robotics and AI	2019	Alan Winfield	"This blogpost is an updated round up of the various sets of ethical principles of robotics and AI that have been proposed to date, ordered by date of first publication."
Ethics in Context	2021-06	Centre for Ethics (Canada, Université de Toronto)	The Oxford Handbook of Ethics of AI: Online Supplement The Oxford Handbook of Ethics of AI: An Annotated Bibliography
Stanford HAI: AI Index Report 2021	N/A <i>(Rapport)</i>	Human-Centered Artificial Intelligence (É-U, Université de Stanford)	Rapport Données
Implementing AI Principles: Frameworks, Processes, and Tools (2021)	N/A <i>(Article)</i>	Boza et Evgeniou (Hongrie, Budapest)	Examen des principes, cadres et outils éthiques actuels et identification des lacunes potentielles.

Annexe 2 : Vingt-trois modèles retenus

Modèles gouvernementaux

Artificial Intelligence Ethics Framework for the Intelligence Community

Office of the Director of National Intelligence. "Artificial Intelligence Ethics Framework for the Intelligence Community", (2020),

En ligne :

https://www.dni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf

Data Ethics Framework

Gouvernement du Royaume-Uni. "Data Ethics Framework", (2020), En ligne : GOV.UK

<<https://www.gov.uk/government/publications/data-ethics-framework>>

The Assessment List for Trustworthy Artificial Intelligence (ALTAIL)

Groupe d'experts indépendants de haut niveau sur l'IA. "ALTAI - The Assessment List on Trustworthy Artificial Intelligence", (2020), En ligne :

<<https://futurium.ec.europa.eu/en/europeanai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>>

Modèles d'entreprises privées

A Framework for Systematically Applying Humanistic Ethics when Using AI as a Design Material

Dent, Kyle, Richelle Dumond et Mike Kuniavsky. A Framework for Systematically Applying Humanistic Ethics when Using AI as a Design Material, SSRN, (Rochester, NY: Social Science Research Network, 2019), En ligne :

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3455518

Everyday Ethics for Artificial Intelligence

IBM. "Everyday Ethics for Artificial Intelligence", (2019),

En ligne : <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>

Guidelines for Human-AI Interaction

Amershi, Saleema et al. "Guidelines for Human-AI Interaction" (2019),

En ligne : <<https://www.microsoft.com/en-us/research/publication/guidelines-for-human-aiinteraction/>>.

Responsible AI by Design in Practice

Benjamins, Richard, Alberto Barbado et Daniel Sierra. "Responsible AI by Design in Practice" (2019) arXiv:1909.12838 [cs], En ligne : <<http://arxiv.org/abs/1909.12838>>.

Responsible bots: 10 guidelines for developers of conversational AI

Microsoft. "Responsible bots: 10 guidelines for developers of conversational AI" (2018), En ligne :

<<https://www.microsoft.com/en-us/research/publication/responsible-bots/>>

Modèles d'associations et d'OSBL

AI Ethics Label

AIEI. "AI Ethics Impact Group: From Principles to Practice – VDE", (2020),
En ligne : <<https://www.ai-ethics-impact.org/en>>

Data Ethics Impact Assessment

Data Ethics. "Data Ethics Impact Assessment", (2021), En ligne :
<https://dataethics.eu/wpcontent/uploads/dataethics-impact-assessment-2021.pdf>>

Éthique et numérique : un référentiel pratique pour les acteurs du numérique

Cigref. Éthique & numérique : un référentiel pratique pour les acteurs du numérique , (2018),
En ligne : Cigref <<https://www.cigref.fr/ethique-numerique-un-referentiel-pratique-pour-les-acteursdu-numerique>>.

Responsible AI Design Assistant

AI Global. "Responsible AI Design Assistant", (2020), En ligne :
<<https://oproma.github.io/raitrustindex/>>

Modèles académiques

An Ethical Framework for Good AI Society: Opportunities, Risks, Principles, and Recommendations

Floridi, Luciano et al. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations" (2018) 28:4 Minds & Machines 689–707

Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications

Ryan, Mark et Bernd Carsten Stahl. "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications" (2020), Journal of Information, Communication and Ethics in Society, En ligne : <<https://doi.org/10.1108/JICES-12-2019-0138>>

DEDA : Data Ethics Decision Aid

Utrecht Data School et Universit d'Utrecht. "Data Ethics Decision Aid (DEDA)", (2020),
En ligne : Utrecht Data School <<https://dataschool.nl/en/deda/>>.

FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity

Arnold, Matthew et al. "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity" (2019) arXiv:180807261 [cs], En ligne :
<<http://arxiv.org/abs/1808.07261>>

How to Design AI for Social Good: Seven Essential Factors

Floridi, Luciano et al. "How to Design AI for Social Good: Seven Essential Factors" (2020) 26:3 Sci Eng Ethics 1771–1796, En ligne : <https://link.springer.com/article/10.1007/s11948-020-00213-5>

Grille de réflexivité portant sur les enjeux éthiques des systèmes d'intelligence artificielle (SIA).
Marchildon, Jacob et Marcoux. (2021). *Grille de réflexivité portant sur les enjeux éthiques des systèmes d'intelligence artificielle (SIA)*. Récupéré de <https://observatoire-ia.ulaval.ca/grille-de-reflexivite-portant-sur-les-enjeux-ethiques-des-systemes-dintelligence-artificielle-sia/>

Autres modèles

AI Ethics Framework

Digital Catapult. "AI Ethics Framework", (2018), En ligne : Machine Intelligence Garage
<<https://www.migarage.ai/ethics/ethics-framework/>>

Data Ethics Canvas

Open Data Institute. "The Data Ethics Canvas", (2019), En ligne :
<<https://theodi.org/article/dataethics-canvas/>>

Ethical Application of Artificial Intelligence Framework

American Council for Technology – Industry Advisory Council. "Ethical Application of Artificial Intelligence Framework", (2020), En ligne: < <https://www.actiac.org/act-iac-white-paper-ethicalapplication-ai-framework>>.

Ethics and algorithms toolkit

Center for Government Excellence. "Ethics and algorithms toolkit", (2018)
En ligne : <<https://ethicstoolkit.ai/>>

The Box: Dynamics of AI principles

AI Ethics Lab. "Tool: The Box", (2020), En ligne : Toolbox: Dynamics of AI Principles
<<http://aiethicslab.com/the-box/>>

Annexe 3 : Sept outils sélectionnés

1. [ALTAI - The Assessment List on Trustworthy Artificial Intelligence](#) (Groupe d'experts indépendants de haut niveau sur l'IA, Commission européenne, 2020)
2. [Responsible AI Design Assistant](#) (AI Global, 2020)
3. [TOOL: the Box](#) (AI Ethics Lab, 2020)
4. [Everyday Ethics for Artificial Intelligence](#) (IBM, 2019)
5. [Data Ethics Decision Aid \(DEDA\)](#) (Utrecht Data School et Université d'Utrecht, 2020)
6. [Grille de réflexivité portant sur les enjeux éthiques des systèmes d'intelligence artificielle \(SIA\)](#) (OBVIA, 2021)
7. [The Data Ethics Canvas](#) (Open Data Institute, 2019)

Annexe 4 : modèles d'invitation

Courriel-type / Lettre de présentation, invitation à participer à un groupe de discussion - participants

Objet : Invitation à participer à un groupe de discussion pour un projet de recherche FORUM IA & OBVIA

Dans le cadre d'un projet de recherche mené par l'OBVIA (Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique), nous souhaiterions organiser un groupe de discussion avec 5 entreprises utilisatrices d'IA et 5 entreprises productrices d'IA installées au Québec.

Cette recherche est sous la direction de Lyse Langlois, professeure à la Faculté des sciences sociales (Département des relations industrielles) de l'Université Laval et directrice générale de l'OBVIA, avec la collaboration de Aude Marie Marcoux, coordonnatrice scientifique, et Nicolas Martin, responsable des partenariats à l'OBVIA.

Objectif du projet de recherche :

Ce premier volet s'inscrit dans un projet plus large et viendra nourrir le développement d'un outil intégrateur visant à guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA pour l'ensemble de la communauté québécoise.

L'objectif de ce volet est de mettre en évidence les besoins de la communauté des entreprises productrices et utilisatrices de technologies relevant de l'IA et du numérique en matière d'éthique. Deux sous-objectifs s'en dégagent :

1. Identifier les préférences des acteurs québécois concernés relativement au développement d'un outil éthique en évaluant la pertinence de 7 outils d'autoévaluation du risque avec ou sans score;
2. Identifier la forme que pourrait prendre un outil éthique québécois en créant collectivement des outils exploratoires.

Présentation de l'activité :

Cette activité de formation débutera par une introduction à l'éthique de l'IA, puis prendra la forme de 2 tables de coconstruction qui regrouperont trois participants (développeur, gestionnaire, analyste de projets) provenant de 5 entreprises productrices et trois participants (développeur, gestionnaire, analyste de projets) provenant de 5 entreprises utilisatrices d'IA, pour un total de 30 participants. Une compréhension de l'anglais écrit est requise, car certains outils à explorer sont en anglais.

Déroulement de l'activité :

Les groupes de discussion seront conduits en ligne par visioconférence, en utilisant la plateforme en ligne Zoom (possibilité de se connecter au salon de discussion par téléphone).

L'activité se déroulera sur Zoom, le jeudi 9 décembre, de 9h à 12h.

Avantages liés à votre participation et compensation :

Le fait de participer à cette recherche vous offre une occasion de réfléchir et de discuter en toute confidentialité, de la pertinence et de la faisabilité d'instaurer en entreprise un outil éthique à partir de sept outils qui ont fait leur preuve au point de vue des validations. Cela peut représenter

des avantages pour l'ensemble des entreprises participantes. En effet, c'est une occasion de contribuer à la réflexion éthique pour un développement responsable de l'IA, tant sur le plan de la faisabilité et la pertinence. De plus, il s'agit d'une occasion de bénéficier d'une formation en éthique de l'IA et des apprentissages collectifs générés au sein du groupe des participants. Une attestation de participation à une activité de formation sera d'ailleurs remise aux participants.

Veillez noter que vous ne recevrez pas de compensation financière pour votre participation à ce projet de recherche, mais votre participation s'avère endossée par votre employeur. Ainsi le temps investi dans l'exercice ne relève aucunement de votre temps personnel.

Note sur la confidentialité :

Soyez assuré de la confidentialité de votre participation. Les noms des personnes participantes ne seront pas cités et aucune information à leur sujet ne sera divulguée ni même les entreprises participantes. Toutes les données et les informations reçues seront utilisées uniquement à des fins de recherche. Si les chercheurs utilisent dans le rapport des données ou de l'information fournies pendant l'atelier en y associant le nom de votre entreprise, ou institution/organisme, ils devront s'assurer au préalable d'avoir eu votre consentement écrit leur accordant l'utilisation de ces informations. Dans le but de ne pas commettre d'erreurs ou de ne pas divulguer d'informations confidentielles, l'équipe vous transmettra la version de l'outil intégrateur avant publication.

L'OBVIA et le FORUM IA ont identifié votre entreprise comme collaboratrice des plus pertinentes à cet important projet et celle-ci a accepté de participer. Ainsi, nous espérons que vous accepterez de faire partie des participants volontaires. Le cas échéant, la coordonnatrice scientifique communiquera avec vous. Par ailleurs, lors de ces échanges, les mesures de confidentialité, de droit de retrait et de participation volontaire énoncées dans le formulaire de consentement en lien avec ce projet, vous seront énoncées.

Pour toutes informations complémentaires vous pouvez contacter Aude Marie Marcoux, coordonnatrice scientifique :

Audemarie.Marcoux@observatoire-ia.ulaval.ca

En vous remerciant par avance pour votre collaboration,

Cordialement,



Aude Marie Marcoux

Professionnelle de recherche / Doctorante

**Observatoire international sur les impacts sociétaux de
l'intelligence artificielle et du numérique (OBVIA)**

Courriel-type / Lettre de présentation, invitation à participer à un groupe de discussion - entreprises

Objet : Invitation à participer à un groupe de discussion pour un projet de recherche FORUM IA & OBVIA

Dans le cadre d'un projet de recherche mené par l'OBVIA (Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique), nous souhaiterions organiser un groupe de discussion avec 5 entreprises utilisatrices d'IA et 5 entreprises productrices d'IA installées au Québec.

Cette recherche est sous la direction de Lyse Langlois, professeure à la Faculté des sciences sociales (Département des relations industrielles) de l'Université Laval et directrice générale de l'OBVIA, avec la collaboration de Aude Marie Marcoux, coordonnatrice scientifique, et Nicolas Martin, responsable des partenariats à l'OBVIA.

Objectif du projet de recherche :

Ce premier volet s'inscrit dans un projet plus large et viendra nourrir le développement d'un outil intégrateur visant à guider la mise en œuvre du développement et de l'utilisation responsable des systèmes d'IA pour l'ensemble de la communauté québécoise.

L'objectif de ce volet est de mettre en évidence les besoins de la communauté des entreprises productrices et utilisatrices de technologies relevant de l'IA et du numérique en matière d'éthique. Deux sous-objectifs s'en dégagent :

1. Identifier les préférences des acteurs québécois concernés relativement au développement d'un outil éthique en évaluant la pertinence de 7 outils d'autoévaluation du risque avec ou sans score;
2. Identifier la forme que pourrait prendre un outil éthique québécois en créant collectivement des outils exploratoires.

Présentation de l'activité :

Cette activité de formation débutera par une introduction à l'éthique de l'IA, puis prendra la forme de 2 tables de coconstruction qui regrouperont trois participants (développeur, gestionnaire, analyste de projets) provenant de 5 entreprises productrices et trois participants (développeur, gestionnaire, analyste de projets) provenant de 5 entreprises utilisatrices d'IA, pour un total de 30 participants. Une compréhension de l'anglais écrit est requise, car certains outils à explorer sont en anglais.

Déroulement de l'activité :

Les groupes de discussion seront conduits en ligne par visioconférence, en utilisant la plateforme en ligne Zoom (possibilité de se connecter au salon de discussion par téléphone).

L'activité sera d'une durée comprise entre 2 et 3 heures.

La date finale de l'activité sera décidée en fonction des disponibilités des participants.

Voici une liste des dates ouvertes pour le déroulement des activités de coconstruction : 2 décembre ou 9 décembre, de 9h à 12h. Veuillez s'il-vous-plait transmettre vos disponibilités à l'équipe de recherche.

Avantages liés à votre participation et compensation :

Le fait de participer à cette recherche offre à vos employés une occasion de réfléchir et de discuter en toute confidentialité, de la pertinence et de la faisabilité d'instaurer en entreprise un outil éthique à partir de sept outils qui ont fait leur preuve au point de vue des validations. Cela peut représenter des avantages pour l'ensemble des entreprises participantes. En effet, c'est une occasion de contribuer à la réflexion éthique pour un développement responsable de l'IA, tant sur le plan de la faisabilité et la pertinence. De plus, il s'agit pour vos employés d'une occasion de bénéficier d'une formation en éthique de l'IA et des apprentissages collectifs générés au sein du groupe des participants. Une attestation de participation à une activité de formation sera d'ailleurs remise à vos employés.

Veuillez noter que vos employés ne recevront pas de compensation financière pour leur participation à ce projet de recherche. Ainsi, vous devrez endosser leur temps investi dans l'exercice.

Note sur la confidentialité :

Soyez assuré de la confidentialité de la participation de vos employés. Les noms des personnes participantes ne seront pas cités et aucune information à leur sujet ne sera divulguée ni même les entreprises participantes. Toutes les données et les informations reçues seront utilisées uniquement à des fins de recherche. Si les chercheurs utilisent dans le rapport des données ou de l'information fournies pendant l'atelier en y associant le nom de votre entreprise, ou institution/organisme, ils devront s'assurer au préalable d'avoir eu votre consentement écrit leur accordant l'utilisation de ces informations. Dans le but de ne pas commettre d'erreurs ou de ne pas divulguer d'informations confidentielles, l'équipe vous transmettra la version de l'outil intégrateur avant publication.

L'OBVIA et le FORUM IA ont identifié votre entreprise comme collaboratrice des plus pertinentes à cet important projet. Ainsi, nous espérons une réponse positive de votre part. Le cas échéant, vous serez ensuite invité à transmettre l'invitation ci-jointe aux employés ciblés par le projet (un développeur, un gestionnaire et un analyste de projet). Une fois la liste de vos employés participants potentiels établie, la chercheuse principale et la coordonnatrice communiqueront avec les volontaires identifiés. Par ailleurs, lors de ces échanges, les mesures de confidentialité, de droit de retrait et de participation volontaire énoncées dans le formulaire de consentement en lien avec ce projet, seront énoncées aux participants.

Pour toutes informations complémentaires vous pouvez contacter Aude Marie Marcoux, coordonnatrice scientifique :

Audemarie.Marcoux@observatoire-ia.ulaval.ca

En vous remerciant par avance pour votre collaboration,

Cordialement,

Signature

Annexe 5 : formulaire de consentement

Formulaire de consentement

Titre de la recherche : Mise en œuvre des principes de l'éthique de l'IA : outiller les entreprises productrices et utilisatrices de technologies relevant de l'intelligence artificielle et du numérique au Québec

Présentation de la chercheuse principale et de la co-chercheure

Cette recherche est réalisée dans le cadre d'un mandat donné par FORUM IA QUEBEC et est sous la direction de Lyse Langlois, professeure à la Faculté des sciences sociales (Département des relations industrielles) de l'Université Laval et directrice générale de l'OBVIA, avec la collaboration de Aude Marie Marcoux, coordonnatrice scientifique, et Nicolas Martin, responsable des partenariats à l'OBVIA.

L'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) est un réseau de recherche ouvert qui fédère les expertises de 260 chercheuses et chercheurs de 18 établissements postsecondaires du Québec (universités et collèges) ainsi que de nombreux partenaires nationaux et internationaux financé par les Fonds de recherche du Québec.

Avant d'accepter de participer à ce projet de recherche, veuillez prendre le temps de lire et de comprendre les renseignements qui suivent. Ce document vous explique le but de ce projet de recherche, ses procédures, avantages, risques et inconvénients. Nous vous invitons à poser toutes les questions que vous jugerez utiles à la personne qui vous présente ce document.

Nature et objectifs du projet

L'objectif de ce volet est de mettre en évidence les besoins de la communauté des entreprises productrices et utilisatrices de technologies relevant de l'IA et du numérique en matière d'éthique. Deux sous-objectifs s'en dégagent :

1. Identifier les préférences des acteurs québécois concernés relativement au développement d'un outil éthique en évaluant la pertinence de 7 outils d'autoévaluation du risque avec ou sans score;
2. Identifier la forme que pourrait prendre un outil éthique québécois en créant collectivement des outils exploratoires

Déroulement du projet

Pour répondre à ces objectifs, une activité de coconstruction sera réalisée au courant du mois de novembre 2021. La coconstruction permet de mobiliser l'intelligence collective autour d'objectifs communs en vue de faire émerger de nouvelles propositions issues notamment de l'expérience sociale des utilisateurs (Déclaration de Montréal IA Responsable, 2018).

Cette activité d'une durée de 3 heures prendra la forme de 2 tables de coconstruction qui regrouperont trois participants volontaires (développeur, gestionnaire, analyste de projet) provenant de 5 entreprises productrices et trois participants volontaires (développeur,

gestionnaire, analyste de projet) provenant de 5 entreprises utilisatrices d'IA, pour un total de 30 participants. Plus précisément, une première table regroupera 7 participants provenant d'une entreprise productrice d'IA et 7 participants d'une entreprise utilisatrice d'IA. Une deuxième table rassemblera 8 participants provenant d'une entreprise productrice d'IA et 8 participants d'une entreprise utilisatrice d'IA. Ce choix méthodologique de rassembler les participants provenant des 2 types d'entreprise vise à assurer la multidisciplinarité des discussions et échanges.

La participation des participants se fait sur une base volontaire et les participants seront dégagés par leur employeur pour leur permettre de participer à cette activité.

Chaque groupe de participants aura pour objectif d'évaluer les outils existants présélectionnés et de collectivement proposer un design d'outil exploratoire et intégrateur destiné à la communauté québécoise sur la base de cette évaluation.

Comme point de départ à la coconstruction, sept outils d'auto-évaluation du risque ont été sélectionnés comme déclencheurs, sur la base de leur pertinence éthique, leur facilité d'usage (questionnaire en ligne) et du temps nécessaire pour les compléter. Certains outils permettent d'arriver à une forme de « score » pour évaluer la dimension éthique de la démarche de développement ou d'utilisation de technologies relevant de l'IA et du numérique, alors que d'autres se veulent davantage des outils de sensibilisation et de réflexion.

Chacun des participants se verra attribué un outil à tester de manière aléatoire lors de son inscription.

L'atelier de coconstruction, d'une durée de trois heures, se déroulera à distance, sur la plateforme Zoom.

Avantages, risques ou inconvénients possibles liés à votre participation et compensation

Le fait de participer à cette recherche vous offre une occasion de réfléchir et de discuter en toute confidentialité, de la pertinence et de la faisabilité d'instaurer en entreprise un outil éthique à partir de sept outils qui ont fait leur preuve au point de vue des validations. Cela peut représenter des avantages pour l'ensemble des entreprises participantes. En effet, c'est une occasion de contribuer à la réflexion éthique pour un développement responsable de l'IA tant sur la faisabilité et la pertinence. De plus, il s'agit d'une occasion de bénéficier d'une formation en éthique de l'IA et des apprentissages collectifs générés au sein du groupe des participants. Une attestation de participation à une activité de formation sera d'ailleurs remise aux participants.

Votre participation à ce projet s'avère endossée par votre employeur, ainsi le temps investi dans l'exercice ne relève aucunement de votre temps personnel. Celui-ci s'engage donc à assumer les possibles frais incombés aux participants (ex. : déplacement, stationnement, repas). Cela étant dit, la présente situation sanitaire nous porte à conclure que l'activité se déroulera sur la plateforme Zoom. De plus, ce faisant, les participants ne devraient aucunement être susceptibles de sentir qu'ils cumulent du retard dans leur travail alors qu'ils participent à ce projet.

Par ailleurs, question d'éviter l'éventualité que les participants puissent ressentir de l'inconfort en étant confrontés à des opinions divergentes ou en ayant à s'exprimer devant un groupe, la coordonnatrice de projet agira à titre de facilitatrice en atelier de dialogue pour les deux tables de coconstruction. De plus, dans le cadre du Bloc 1, la chercheuse principale et la coordonnatrice présenteront de façon informative et rassurante le fonctionnement d'une table de coconstruction.

Participation volontaire et droit de retrait

Vous êtes libre de participer à ce projet de recherche. Vous pouvez aussi mettre fin à votre participation sans conséquence négative ou préjudice et sans avoir à justifier votre décision. Si vous décidez de mettre fin à votre participation, il est important d'en prévenir la chercheuse dont les coordonnées sont incluses dans ce document. Tout le matériel permettant de vous identifier, incluant les enregistrements et les données que vous aurez fournies seront alors détruits, à moins que vous n'autorisiez le chercheur à les utiliser pour la recherche, malgré votre retrait. Le cas échéant, ils seront conservés selon les mesures décrites ci-après et qui seront appliquées pour tous les participants.

Confidentialité

Voici les mesures qui seront appliquées dans le cadre de la présente recherche :

Durant la recherche:

- votre nom et tous ceux cités durant l'entrevue seront remplacés par un nom fictif;
- tout le matériel de la recherche, incluant les formulaires de consentement et les enregistrements, sera conservé sur un espace OneDrive-Université Laval;
- les données en format numérique seront, pour leur part, conservées dans des fichiers encryptés dont l'accès sera protégé par l'utilisation d'un mot de passe et auquel seule la chercheuse et la coordonnatrice auront accès;

Lors de la diffusion des résultats :

- les noms des participants et de l'entreprise pour laquelle ils travaillent ne paraîtront dans aucun rapport;
- les résultats seront présentés sous forme globale de sorte que les résultats individuels des participants ne seront jamais communiqués;
- les résultats de la recherche seront publiés dans des revues scientifiques, et aucun participant ne pourra y être identifié ou reconnu;
- un court résumé des résultats de la recherche sera expédié aux participants qui en feront la demande en indiquant l'adresse où ils aimeraient recevoir le document, juste après l'espace prévu pour leur signature.

Après la fin de la recherche :

- tout le matériel et toutes les données seront utilisés dans le cadre exclusif de cette recherche et ils seront détruits au plus tard en juin 2026.

Le respect de la confidentialité des participants dépend de la collaboration des autres participants à cet égard.

- **Consentement à l'enregistrement audio de l'entrevue :**

- J'accepte que des ENREGISTREMENTS AUDIOS de ma personne puissent être utilisés par le chercheur pour l'analyse des données.

Si vous acceptez, toutes les VIDÉOS seront confidentielles, protégées par un mot de passe et conservées sous clef. Ces vidéos ne sont jamais publiées ou rediffusées publiquement d'une quelconque façon. Seuls les chercheurs impliqués dans le projet auront accès à ces vidéos.

- **Consentement à l'utilisation des données dans le cadre de la thèse:**

J'accepte que la coordonnatrice scientifique, aussi étudiante au 3e cycle, pourra réutiliser les données recueillies pour ce projet dans le cadre de sa thèse de doctorat et dans le cadre d'un article scientifique faisant état de cette intervention.

OUI

NON

Remerciements

Votre collaboration est précieuse pour nous permettre de réaliser cette étude. C'est pourquoi nous tenons à vous remercier pour le temps et l'attention que vous acceptez de consacrer à votre participation.

Signatures

Je soussigné(e) _____ consens librement à participer à la recherche intitulée : « *Mise en œuvre des principes de l'éthique de l'IA : outiller les entreprises productrices et utilisatrices de technologies relevant de l'intelligence artificielle et du numérique au Québec* ». J'ai pris connaissance du formulaire et j'ai compris le but, la nature, les avantages, les risques et les inconvénients du projet de recherche. Je suis satisfait(e) des explications, précisions et réponses que le chercheur m'a fournies, le cas échéant, quant à ma participation à ce projet.

—

Signature du participant, de la participante

Date

Un court résumé des résultats de la recherche sera expédié aux participants qui en feront la demande en indiquant l'adresse courriel où ils aimeraient recevoir le document, juste après l'espace prévu pour leur signature.

Adresse postale :

J'ai expliqué le but, la nature, les avantages, les risques et les inconvénients du projet de recherche au participant. J'ai répondu au meilleur de ma connaissance aux questions posées et j'ai vérifié la compréhension du participant.



_____ 2021 _____

Signature du chercheur ou de son représentant

_____22 novembre

Date

Renseignements supplémentaires

Si vous avez des questions sur la recherche, sur les implications de votre participation ou pour se retirer du projet, veuillez communiquer avec Lyse Langlois, professeure titulaire à l'Université Laval, responsable du volet éthique. Vous pouvez la contacter au numéro de téléphone suivant : (418)656-2131 poste 403483, ou à l'adresse courriel suivante : lyse.langlois@observatoire-ial.ca.

Programme d'aide aux personnels :

[PavillonAlphonse-Marie-Parent](#)

2255, rue de l'Université

Local 1661

Université Laval

Québec (Québec) G1V 0A7

Renseignements: 418 656-2131, poste 8008

Plaintes ou critiques

Toute plainte ou critique sur ce projet de recherche pourra être adressée au Bureau de l'Ombudsman de l'Université Laval :

Pavillon Alphonse-Desjardins, bureau 3320

2325, rue de l'Université

Université Laval

Québec (Québec) G1V 0A6

Renseignements - Secrétariat : (418) 656-3081

Ligne sans frais : 1-866-323-2271

Courriel : info@ombudsman.ulaval.ca

Copie du participant