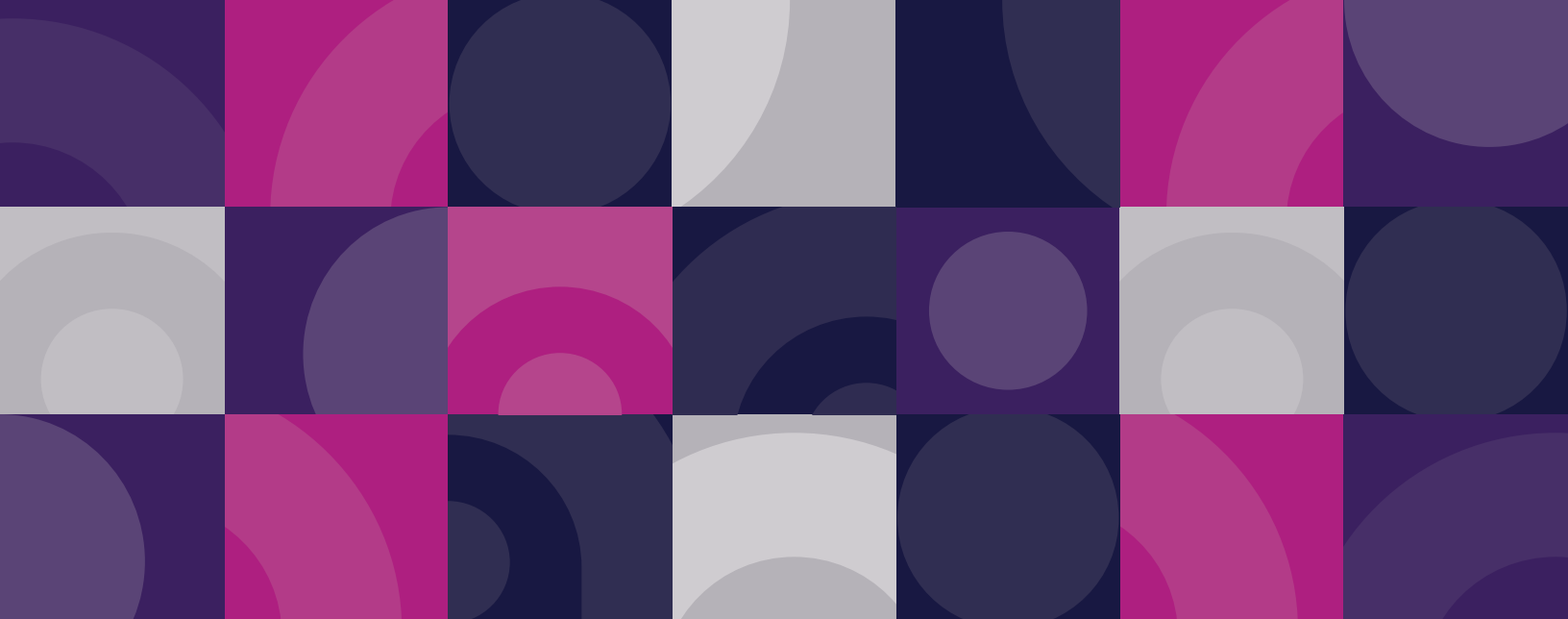




TRAVAUX EXPLORATOIRES POUR LE DÉVELOPPEMENT D'UN OUTIL D'ÉVALUATION DES IMPACTS SOCIÉTAUX DE L'IA ET DU NUMÉRIQUE

Québec, Canada
Septembre 2022



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



Forum
IA Québec

Ce document a été réalisé par l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) soutenu par les Fonds de recherche du Québec (FRQ), dans le cadre d'une collaboration avec Forum IA Québec.

Équipe de recherche

- Lyse Langlois
- Justin Lawarée
- Aude Marie Marcoux

Avec contributions de

- Richard Khoury
- Nathalie de Marcellis-Warin
- Allison Marchildon
- Christophe Mondain
- Denis Poussart
- Bryn Williams-Jones

Soutien à la recherche :

- Michelle Albert-Rochette
- Émilie Guiraud

Coordination du projet

- Nicolas Martin

ISBN 978-2-925138-19-8

Remerciements

L'équipe de recherche tient à remercier les participants des entreprises productrices et utilisatrices d'IA qui ont généreusement participé à l'atelier de co-construction visant à valider la pertinence d'outils éthiques ainsi qu'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 participants à l'étude.

MISE EN CONTEXTE

Dans le contexte des récentes évolutions en matière de recherche et développement sur l'intelligence artificielle (IA), plusieurs initiatives de cadres normatifs ont vu le jour ces dernières années, visant à définir des principes éthiques directeurs afin de mieux évaluer les risques éthiques liés à ces systèmes. Alors que l'on constate une démocratisation du développement et du déploiement des systèmes d'intelligence artificielle (SIA) dans nos organisations, les efforts se tournent à présent vers la formalisation de mécanismes de mise en œuvre de ces principes. La demande pour des outils d'opérationnalisation des principes éthiques de l'IA est forte et plusieurs organisations cherchent à formaliser de tels dispositifs éthiques en vue de maintenir de bonnes pratiques. C'est dans ce contexte que l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) développe une série de travaux portant sur le déploiement d'outils éthiques appliqués à l'IA.

Afin de mettre un outil intégrateur à la disposition des organisations québécoises, le Forum IA Québec a confié à l'OBVIA le mandat de mener un projet de recherche exploratoire visant à tester des chartes éthiques et voir laquelle pourrait correspondre aux besoins des entreprises participantes. Les chartes éthiques, sélectionnées à partir d'une revue de littérature des outils éthiques qui ont été formalisés en organisation, ont constitué un point de départ aux discussions et échanges sur la conception d'un outil éthique idéal. Ce projet a eu lieu durant l'automne 2021 et l'hiver 2022 et un atelier de co-construction a servi à valider la pertinence des outils sélectionnés auprès d'entreprises québécoises productrices et utilisatrices d'IA. Parmi les outils présentés, aucun des modèles n'a retenu l'attention du groupe. La plupart ont souligné des sections ou des aspects qui faisaient l'unanimité sans trouver l'outil adapté à leurs besoins. L'analyse des données collectées lors de l'activité ainsi qu'une analyse thématique comparée de principes consensuels dans la littérature et dans les outils à l'étude ont permis de dégager 7 principes et 25 sous-principes ainsi qu'une série de caractéristiques spécifiques importantes à intégrer dans l'outil intégrateur.



Dans le présent document, nous présentons la première partie de cet outil comprenant les principes, les définitions qui ont été validées par un groupe d'experts, de même qu'une première série de questionnements éthiques et d'évaluation des risques.

Une deuxième phase est en cours visant à finaliser le modèle global. Ces travaux serviront à intégrer les divers besoins mis en évidence lors de l'atelier de co-construction avec les entreprises, incluant des pistes d'action et recommandations en lien avec la législation ainsi que des liens vers des ressources pertinentes. Cette prochaine phase mènera à la mise à disposition pour l'écosystème québécois d'une première version de l'outil, qui pourra être utilisé, validé et commenté par la communauté des entreprises utilisatrices et productrices d'IA.

STRUCTURE DE L'OUTIL PRÉLIMINAIRE POUR ÉVALUER LES IMPACTS SOCIÉTAUX DE L'IA ET DU NUMÉRIQUE

DÉFINITION DES PRINCIPES ÉTHIQUES

ROBUSTESSE ET SÉCURITÉ

La robustesse technique est la capacité du SIA à bien fonctionner malgré les changements dans l'environnement dans lequel il opère. La sécurité est la capacité du système d'intelligence artificielle (SIA) à bien fonctionner malgré la présence d'autres agents (humains et artificiels) pouvant avoir des interactions antagonistes (volontaires ou accidentelles) avec lui. Le développement et l'utilisation des SIA doivent s'assurer de la robustesse technique et de la sécurité des systèmes. Pour ce faire, des politiques et 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 dans une variété de conditions, et qu'ils ne risquent pas d'être compromis accidentellement ou délibérément par des tiers. Cela signifie que leurs SIA soient mis au point selon une approche de prévention des risques en matière de sécurité, de façon à envisager le degré de vulnérabilité du SIA, de même qu'élaborer des stratégies d'atténuation des risques identifiés.

VIE PRIVÉE

La vie privée doit être protégée de l'intrusion de SIA et de systèmes d'acquisition et d'archivage des données personnelles (SAAD). 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.

EXPLICABILITÉ ET TRANSPARENCE

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. 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.

**ÉQUITÉ,
NON-DISCRIMINATION
ET INCLUSION**

Le développement et l'utilisation des SIA doivent s'assurer de favoriser l'équité, la non-discrimination et l'inclusion, et ce, tout au long du cycle de vie d'un SIA. Cela implique la prise en compte et la participation de l'ensemble des parties prenantes concernées depuis le processus de développement d'un SIA jusqu'à sa mise en œuvre, dans le but de veiller à l'égalité de traitement et d'accès à chaque SIA. De façon plus large, les SIA doivent ainsi être non seulement au service de l'utilisateur, mais doivent aussi être protecteurs d'une société démocratique et équitable.

**BIEN-ÊTRE ET
DÉVELOPPEMENT
SOUTENABLE ET
ENVIRONNEMENTAL**

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 chaque SIA. 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.

**AUTONOMIE
(AGENTIVITÉ ET
RESPECT DES DROITS
HUMAINS)**

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'assurer que les décisions qui en découlent soient justes et respectent les droits des personnes touchées, ainsi que d'éviter la déresponsabilisation des agents humains face à ces décisions.

**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.

ROBUSTESSE ET SÉCURITÉ

Avez-vous établi une procédure visant à offrir la possibilité aux individus de corriger ou supprimer certaines informations (renseignement personnel et sensible) et de désindexer des contenus (géoblocage) ?

Êtes-vous en mesure d'identifier les risques sur la vie privée des personnes concernées ?

Données sensibles

Y a-t-il un risque d'identification des données sensibles et personnelles contenues dans votre ensemble de données ?

Infrastructures

Y a-t-il des risques pour que la sécurité des données soit compromise ?

Prévention des risques

Y a-t-il des risques pour que l'utilisation du système d'intelligence artificielle engendre des préjudices potentiels ou réels ?

VIE PRIVÉE

Collectez-vous des informations personnelles à l'insu des personnes ? (reconnaissance faciale, géolocalisation, empreintes, etc.)

Avez-vous mis en place des mesures visant à protéger la vie privée ?

Est-ce que ces mesures offrent toutes les garanties de protection de la vie privée en matière juridique ?

Connaissez-vous la provenance de votre base de données ?

Anonymisation
des données
privées

Y a-t-il des risques pour la réidentification des clients à partir des données collectées ou analysées ?

Conservation

Y a-t-il des risques que les données soient conservées au-delà de la durée de vie du système d'intelligence artificielle ?

Consentement
des usagers

Y a-t-il des risques que les clients ne puissent pas donner un consentement explicite de façon simple, informée et spécifique ?

Droits
individuels

Y a-t-il des risques que les clients n'aient pas la possibilité d'accéder à leurs données personnelles, de les réviser ou de les supprimer ?

Données

Y a-t-il des risques que des données non pertinentes pour le fonctionnement du système d'intelligence artificielle soient collectées ?

Entreposage
sécuritaire

Y a-t-il des risques que l'organisation perde, manipule ou détruise les données ?

EXPLICABILITÉ ET TRANSPARENCE

Avez-vous mis en place une procédure écrite pouvant aider la personne à bien comprendre les termes techniques et les éléments de langage plus complexe ?

Est-ce que les utilisateurs du cyberspace sont informés du contenu de leurs droits ?

Êtes-vous en mesure d'expliquer dans les limites du possible et de la complexité des systèmes algorithmiques mis en place (apprentissage profond) le chemin de la décision prise par les SIA ?

Êtes-vous en mesure d'expliquer le SIA mis en place (à savoir où, quand, comment et pourquoi ils sont utilisés) de même que d'expliquer les résultats obtenus grâce aux SIA ?

Explicabilité

Y a-t-il des risques que l'apprentissage et l'évolution du système d'intelligence artificielle ne soient pas explicables ?

Transparence

Y a-t-il des risques que les personnes concernées ne soient pas en mesure de comprendre et d'expliquer la décision prise par le système d'intelligence artificielle ?

Interprétabilité

Y a-t-il des risques que l'apprentissage et l'évolution du système d'intelligence artificielle ne soient pas interprétables ?

ÉQUITÉ, NON-DISCRIMINATION ET INCLUSION

Est-ce qu'on connaît le contenu/provenance de la base de données sur lequel sera appliqué le SIA ?

Est-ce qu'on connaît les limites du SIA ? Les modèles ont-ils été testés dans divers contextes culturels ?

Est-ce que la base de données est diversifiée au plan populationnel (si cela s'avère nécessaire) : ex. étude sur les femmes blanches ayant eu un cancer du sein = non applicable) ?

Prévention de biais

Y a-t-il des risques que le système d'intelligence artificielle maintienne ou renforce des inégalités ou des préjugés ?

Qualité des données

Y a-t-il des risques que les données analysées par le système d'intelligence artificielle ne soient pas de qualité suffisante ?

Inclusion

Y a-t-il des risques que le système d'intelligence artificielle exclue une partie de la population ?

Équité

Y a-t-il des risques que certains clients ayant un profil et un revenu économique différents n'aient pas accès aux mêmes services et aux mêmes procédures de réparation ?

BIEN-ÊTRE ET DÉVELOPPEMENT SOUTENABLE ET ENVIRONNEMENTAL

Êtes-vous en mesure d'identifier si les exigences environnementales des Objectifs de développement durable de l'ONU (ODD) ont été rencontrées ?

Durabilité

Y a-t-il des risques que le système d'intelligence artificielle impacte négativement l'environnement ?

Responsabilité écologique

Y a-t-il des risques que l'organisation ne paie pas les réparations nécessaires pour remédier aux dommages causés ?

Bien-être des êtres sensibles

Y a-t-il des risques que l'organisation utilise les données récupérées au détriment de leurs clients et de la société ?

AUTONOMIE, AGENTIVITÉ ET RESPECT DES DROITS HUMAINS

Savez-vous si le SIA entraîne une dépendance psychologique ?

Savez-vous si des stratégies d'influence ont été mises en place dans la programmation du SIA ?

Autonomie

Y a-t-il des risques que le système d'intelligence artificielle affecte les décisions et le comportement des êtres humains ou de la société ?

Prise de décision humaine

Y a-t-il des risques que la décision finale ne soit pas systématiquement prise par un être humain ?

Respect des droits humains

Y a-t-il des risques que le système d'intelligence artificielle menace ou entrave un ou plusieurs droits humains ?

RESPONSABILITÉ ET REDDITION DE COMPTE

Est-ce que les responsabilités ont été bien définies lors de conception / l'utilisation d'un SIA ?

Êtes-vous en mesure de justifier chacune des décisions prises lors de l'utilisation d'un SIA ?

Responsabilité

Y a-t-il des risques que les droits et devoirs des employés en matière de gouvernance du système d'intelligence artificielle ne soient pas décrits dans un document interne ?

Évaluation

Y a-t-il des risques que les décisions et les utilisations du système d'intelligence artificielle contreviennent aux lois ?

Indépendance

Y a-t-il des risques que l'évaluation du système soit confiée à un organisme ou une équipe ayant des liens ou des intérêts directs avec l'équipe opérationnelle ?



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA 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