

Recommandations éthiques de l'intelligence artificielle en éducation

MINISTÈRE DE L'ÉDUCATION

2026



Le présent document a été produit par le Centre d'expertise de l'IA en éducation (CEIAE) du ministère de l'Éducation. Il découle d'un travail de collaboration entre les membres du comité d'encadrement éthique du CEIAE composé de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (Obvia), de la GRICS, du Mila et de spécialistes en droit, en éthique de l'IA et en éducation. L'ensemble des travaux de rédaction a été coordonné par l'Obvia.

Autrice

- **Viviane Vallerand**, coordonnatrice et auxiliaire de recherche de l'axe Éducation et capacitation pour l'Obvia

Conseillères et conseillers scientifiques

- **Antoine Lefebvre-Brossard**, scientifique des données à la GRICS
- **Charles Canavaggio** (contribution et révision technique), scientifique des données à la GRICS
- **Simon Collin** (contribution et révision éthique), titulaire de la Chaire de recherche du Canada sur l'équité numérique en éducation, professeur à la Faculté des sciences de l'éducation de l'UQAM et membre-chercheur de l'axe Éducation et capacitation de l'Obvia
- **Emma Damitio**, scientifique des données à la GRICS
- **Rose Landry**, gestionnaire de projet, Gouvernance de l'IA, Affaires juridiques et gouvernance de l'IA à Mila
- **Lyse Langlois**, directrice générale de l'Obvia et professeure titulaire au Département des relations industrielles à l'Université Laval
- **Nadia Naffi**, chercheuse responsable de l'axe Éducation et capacitation de l'Obvia, titulaire de la Chaire de leadership en enseignement (CLE) sur les pratiques pédagogiques innovantes en contexte numérique – Banque Nationale et professeure adjointe au Département d'études sur l'enseignement et l'apprentissage à l'Université Laval
- **Ledy Rivas Zannou** (contribution et révision légale), coordonnateur scientifique à la Chaire L. R. Wilson en droit des technologies de l'information et du commerce électronique
- **Nesrine Zemirli**, directrice principale de l'innovation axée sur les données et l'intelligence artificielle à la GRICS

Groupes responsables du cadre éthique

- Ministère de l'Éducation
- GRICS, entreprise de technologie de l'information spécialisée en éducation
- Mila – Institut québécois d'intelligence artificielle
- Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (Obvia)

Coordination et rédaction

Direction de la gouvernance et des relations avec les partenaires
Direction générale du soutien stratégique et de la gouvernance du numérique
Sous-ministériat de l'innovation et de la performance

Pour information

Renseignements généraux
Ministère de l'Éducation
1035, rue De La Chevrotière, 27^e étage
Québec (Québec) G1R 5A5
Téléphone : 418 643-7095
Ligne sans frais : 1 866 747-6626

© Gouvernement du Québec, 2026
Ministère de l'Éducation

ISBN 978-2-550-84717-5 (PDF)

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2026

26-105-02_w4

Mise en garde

Certains des éléments juridiques du document proviennent de la jurisprudence et de publications de la Commission d'accès à l'information (CAI), de la documentation gouvernementale sur la [protection des renseignements personnels](#) ainsi que de travaux internes de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (Obvia), rédigés en collaboration avec Forum AI. Ces travaux portent sur l'interprétation de la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (ci-après « *Loi sur l'accès* »), selon les sept principes éthiques de l'Obvia. **Le contenu de ce document NE constitue PAS un avis juridique.**

Ce document s'appuie sur l'Énoncé de principes pour une utilisation responsable de l'intelligence artificielle par les organismes publics, publié par le ministère de la Cybersécurité et du Numérique en 2024. Il sera mis à jour en fonction des nouvelles versions de cet énoncé afin de demeurer aligné sur les orientations gouvernementales.

Table des matières

Introduction	1
Contexte.....	1
Centre d'expertise de l'IA en éducation	1
Objectif et méthodologie	2
Éthique de l'IA	3
Principes pour une utilisation responsable de l'IA par les organismes publics	4
Structure du document	6
1 Planification du système d'IA	7
1.1 Recours à l'IA	7
1.2 Responsabilités en IA	8
1.3 Participation des parties prenantes	8
1.4 Objectifs du système d'IA	12
1.5 Type d'IA	13
1.5.1 IA simple	13
1.5.2 IA analytique et prédictive	13
1.5.3 L'IA générative	14
1.6 Gestion des risques	16
2 Gestion éthique des données	21
2.1 Mesures pour les renseignements personnels.....	22
2.2 Mesures pour les données confidentielles	24
2.3 Mesures pour respecter le droit d'auteur.....	25
2.4 Mesures pour l'ensemble des données	26
3 Développement du système d'IA.....	28
3.1 Fournisseur externe	28
3.2 Méthodes de débiaisage et d'équité.....	30
3.3 Méthodes d'explicabilité	31
3.4 Design du système d'IA	33
3.5 Mesures pour le coût énergétique des algorithmes.....	34

4	Mise en œuvre.....	35
4.1	Mesures de pré-déploiement	36
4.2	Essai pilote.....	36
4.3	Formation au développement et à l'utilisation éthiques de l'IA	37
4.4	Accessibilité.....	38
4.5	Milieu inclusif et sécuritaire	39
4.6	Campagne d'information et de sensibilisation.....	40
4.7	Transfert des connaissances	41
5	Maintenance et amélioration	42
5.1	Mécanismes de suivi	42
5.2	Veille du système d'IA	42
5.3	Mises à jour régulières	43
	Conclusion	44
	Références.....	45

Introduction

Contexte

Face aux avancées récentes de l'intelligence artificielle (IA) en éducation, plusieurs applications voient le jour, ouvrant la porte à de nouvelles possibilités d'enseignement et d'apprentissage. Parmi elles, l'IA pourrait améliorer les services offerts aux élèves, faciliter les interventions précoces et ciblées, soutenir la prise de décision de gestionnaires scolaires et assister les membres du personnel scolaire et administratif dans leurs tâches quotidiennes en réduisant leur charge de travail, entre autres. L'IA est vouée à transformer les manières de concevoir et d'offrir les services publics, et l'éducation n'en fera pas exception. Toutefois, elle comporte certains risques notables. Ainsi, elle peut porter atteinte à la vie privée des élèves et des parents, renforcer des biais discriminatoires à leur égard, et même réduire l'agentivité (pouvoir d'agir) des acteurs du milieu de l'éducation (Conseil supérieur de l'éducation, 2024).

L'IA ne remplace pas l'être humain. La présence et l'intelligence humaines dans les processus d'automatisation des systèmes d'IA s'avèrent essentielles pour créer la confiance des usagères et usagers à l'égard de ces outils et favoriser leur acceptabilité sociale. L'IA transforme la manière de travailler, en permettant aux personnes d'investir leur temps dans des tâches à plus forte valeur ajoutée. Par exemple, bien qu'un agent conversationnel puisse assister les personnes dans des tâches à faible valeur ajoutée, il est nécessaire de garder l'humain au cœur du processus afin d'assurer la fiabilité, l'empathie et l'adaptabilité du service, tout en garantissant une adoption durable de l'outil et en respectant le pouvoir d'agir de l'être humain.

Devant ces préoccupations, nous proposons des recommandations éthiques pour le développement et l'utilisation d'outils d'IA en éducation, dans la foulée des travaux entamés par le Centre d'expertise de l'IA en éducation (CEIAE).

Centre d'expertise de l'IA en éducation

Le CEIAE est une initiative dirigée par le ministère de l'Éducation (Ministère) qui regroupe plusieurs partenaires principaux : l'Institut de valorisation des données (IVADO), le Mila, la GRICS, l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (Obvia), le Groupe de recherche interétablissement sur l'intégration pédagogique des technologies de l'information et de la communication (GRIPTIIC), le Centre de transfert pour la réussite éducative du Québec (CTREQ) ainsi que des chercheuses et chercheurs en éducation. Globalement, le CEIAE a pour mission d'établir une vision commune de l'IA en éducation au Québec et de concrétiser celle-ci sur le terrain. Précisément, le CEIAE a pour mandat d'accompagner le Ministère et le réseau scolaire public pour une utilisation judicieuse de l'IA en éducation au Québec.

Plusieurs initiatives ont été achevées ou sont en cours de réalisation. Elles comprennent notamment :

- la conception d'une solution d'IA en matière de prévention du décrochage scolaire;
- l'offre de recommandations éthiques d'IA en éducation;
- la sensibilisation et la formation à l'IA du personnel du Ministère;
- l'établissement de guides pour une utilisation éthique de l'IA par le personnel administratif et scolaire;
- la sensibilisation à l'IA et l'information destinée aux parents et au personnel scolaire;
- l'actualisation du cadre de référence et de ses outils en y intégrant la littératie de l'IA.

Le CEIAE se divise en deux regroupements :

- un comité de suivi, où des représentantes et représentants du Ministère et de chaque partenaire peuvent discuter et coordonner les activités; ce comité s'assure de la coordination des mandats que se donne le CEIAE;
- des groupes de travail *ad hoc* ayant des objectifs précis, qui réunissent au besoin certains partenaires concernés. L'un de ces groupes s'est penché sur les recommandations éthiques de l'IA en éducation pour coconstruire diverses solutions qui soutiennent le développement et l'utilisation éthiques de l'IA, dont le présent document.

Objectif et méthodologie

L'objectif du présent document est d'accompagner les gestionnaires, les directions et le personnel administratif des centres de services scolaires (CSS) et des commissions scolaires (CS) ainsi que les établissements scolaires voulant concevoir et utiliser de façon éthique un système d'IA en éducation au Québec. À cette fin, il présente un ensemble de considérations éthiques et de recommandations qui permettront de maximiser les bénéfices liés à un tel système et de minimiser les risques pouvant en découler, tout en assurant le respect des droits de la personne.

Pour ce faire, le document se base sur un corpus de textes axés sur l'éthique de l'IA :

- [Orientations pour l'intelligence artificielle générative dans l'éducation et la recherche](#) (UNESCO, 2024)
- [Intelligence artificielle générative en enseignement supérieur : enjeux pédagogiques et éthiques](#) (Conseil supérieur de l'éducation [CSE] et Commission de l'éthique en science et en technologie [CEST], 2024)
- [Guide des bonnes pratiques en intelligence artificielle : sept principes pour une utilisation responsable des données](#) (Gautrais *et al.*, 2023)
- [L'intelligence artificielle en éducation : un aperçu des possibilités et des enjeux](#) (CSE, 2020)

- [Trustworthy artificial intelligence \(AI\) in education: Promises and challenges](#) (OCDE, 2020)
- [Orientations stratégiques sur l'IA destinée aux enfants](#) (UNICEF, 2021)
- [Consensus de Beijing sur l'intelligence artificielle et l'éducation](#) (UNESCO, 2019)

Le document tient compte également des orientations gouvernementales, telles que les suivantes :

- [Cadre de référence de la compétence numérique](#) (Ministère, 2026)
- [Principes pour une utilisation responsable de l'IA par les organismes publics](#) (ministère de la Cybersécurité et du Numérique [MCN], 2024)
- [Stratégie gouvernementale de cybersécurité et du numérique](#) (Gouvernement du Québec, 2024)
- [Évaluation de la mise en œuvre du Plan d'action numérique en éducation et en enseignement supérieur](#) (Ministère, 2022)
- [Politique de la réussite éducative](#) (Ministère, 2016)

En vue de faciliter l'opérationnalisation des recommandations éthiques de l'IA établies dans ce document, une trousse d'outils a été élaborée pour accompagner les développeurs et les usagers et usagères des systèmes d'IA. Cette trousse d'outils prévoit, notamment :

- une feuille de route interactive qui facilite l'exploration des différentes phases et principes de l'IA;
- un guide sur l'utilisation pédagogique, éthique et légale de l'intelligence artificielle destiné au personnel enseignant;
- une fiche diagnostique éthique pour favoriser l'utilisation responsable d'un système d'IA.

Éthique de l'IA

Puisque ce document vise à développer une culture d'éthique de l'IA, il importe de préciser ce qui est entendu par « éthique » et « éthique de l'IA ». L'éthique est une discipline philosophique qui permet d'établir des principes ou des valeurs communes à respecter afin de préserver un vivre-ensemble harmonieux (Obvia, 2023). Par exemple, un principe commun auquel nous adhérons afin de préserver une relation de confiance avec les autres est de ne pas leur faire de mal (Dilhac, 2023). L'éthique peut être considérée comme « prescriptive », en ce sens qu'elle prescrit des principes à respecter (Dilhac, 2023). Dans cet ordre d'idées, elle n'est donc pas obligatoire comme le sont les règles de droit, puisque le respect de principes éthiques dans une diversité de contextes nécessite des réflexions collectives propres à l'action humaine et sur la façon dont celle-ci peut préserver ces principes (Langlois, 2008). C'est pourquoi il importe de rappeler que l'éthique, par sa nature réflexive, ne consiste pas à « cocher des éléments sur une liste ». Elle cherche plutôt à préserver des principes par la réflexion, les considérations et les actions qui sont mises en œuvre dans un contexte particulier (Obvia, 2023).

Étant donné que les applications de l'IA ont des impacts sociétaux préoccupants, le domaine de l'« éthique de l'IA » a peu à peu émergé (Dilhac, 2023). Ainsi, ce qui est considéré comme un « enjeu éthique » (Dilhac, 2023) émerge lorsqu'un ou plusieurs principes éthiques sont ébranlés par l'application de l'IA dans un contexte particulier. Un exemple de principe pouvant être mis en péril par un système d'IA est le risque de perte de pouvoir d'agir des acteurs scolaires, qui pourraient utiliser inadéquatement l'intelligence artificielle générative. Pour prévenir de tels enjeux éthiques, il devient nécessaire d'adopter une éthique de l'IA permettant de respecter des principes éthiques particuliers.

Principes pour une utilisation responsable de l'IA par les organismes publics

Ce document se base sur les dix principes pour une utilisation responsable de l'IA par les organismes publics du MCN. Il présente les principales lignes directrices à suivre pour le développement et l'utilisation éthiques de l'IA dans le secteur public au Québec. Ce cadre s'inspire de l'ensemble de ces principes, adaptés aux particularités du réseau scolaire québécois.

1. Principe du respect des personnes et de la règle de droit

L'utilisation responsable de systèmes d'IA doit se faire dans le respect de la primauté du droit, des droits et libertés de la personne, des droits d'auteur, de la loi ainsi que des valeurs de l'administration publique québécoise. Plus particulièrement, les organismes publics doivent veiller à ce que les données d'apprentissage ou les autres données d'entrée utilisées par les systèmes d'IA soient légalement collectées, utilisées et divulguées, en tenant compte des droits et des obligations applicables en matière de protection de la vie privée.

2. Principe de l'inclusion et de l'équité

L'utilisation responsable de systèmes d'IA doit viser à répondre aux besoins de la population québécoise à l'égard des services publics, tout en faisant la promotion de la diversité et de l'inclusion. Tout système d'IA doit minimiser les risques et les inconvénients pour la population et éviter de causer une fracture numérique. Les membres du personnel des organismes publics doivent pouvoir bénéficier de l'accompagnement nécessaire par la mise en place de mécanismes et d'outils, notamment lorsque des métiers sont appelés à être transformés grâce aux avancées technologiques.

3. Principe de la fiabilité et de la robustesse

Des mesures doivent être prises pour vérifier la fiabilité et la robustesse des systèmes d'IA utilisés ou à être utilisés. Des mesures correctives et des moyens de contrôle doivent également être mis en place pour s'assurer que ces systèmes fonctionnent de manière stable et constante, même en présence de perturbations ou de scénarios inédits. La qualité des données est un élément clé pour viser la fiabilité et la robustesse d'un système d'IA. Notamment, les données doivent être exactes et exemptes de biais pouvant comporter des risques, causer des préjudices ou renforcer diverses formes de discrimination.

4. Principe de la sécurité

L'utilisation responsable de systèmes d'IA doit se faire dans le respect des obligations relatives à la sécurité de l'information. Des mesures de sécurité doivent être mises en place afin de limiter les risques encourus et de protéger adéquatement l'information concernée.

5. Principe de l'efficience, de l'efficacité et de la pertinence

L'utilisation responsable de systèmes d'IA doit permettre d'offrir à la population et aux entreprises des services publics simplifiés, intégrés et de qualité. L'utilisation de tels systèmes doit également viser une gestion optimale des ressources informationnelles et des services publics.

6. Principe de la durabilité

L'utilisation responsable de systèmes d'IA doit s'inscrire dans la recherche d'un développement durable.

7. Principe de la transparence

L'utilisation responsable de systèmes d'IA doit se faire dans le respect du principe de transparence. Les organismes publics informent clairement la population et les entreprises de la nature et de la portée de tels systèmes ainsi que du moment de leur utilisation, de manière à promouvoir la confiance du public dans ceux-ci.

8. Principe de l'explicabilité

L'utilisation responsable de systèmes d'IA implique de fournir à la population et aux entreprises l'explication claire et sans ambiguïté des décisions, des prédictions ou des actions les concernant. L'explication doit permettre de comprendre les interactions et ses conséquences au regard d'une décision ou d'un résultat.

9. Principe de la responsabilité

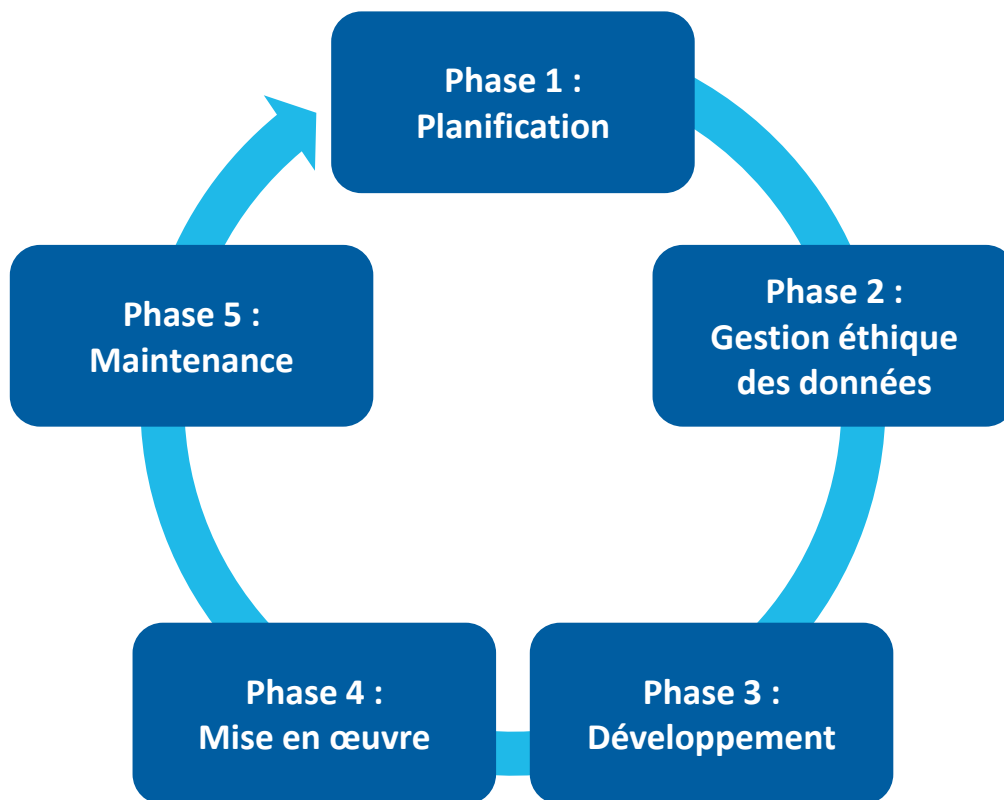
L'utilisation de systèmes d'IA entraîne une responsabilité, notamment quant à leur bon fonctionnement. Elle implique de mettre en place des mesures de contrôle et une gouvernance adéquate, y compris la supervision ou la validation humaine.

10. Principe de la compétence

Les membres du personnel des organismes publics doivent être sensibilisés à l'utilisation, aux bonnes pratiques et aux enjeux pouvant survenir tout au long du cycle de vie des systèmes d'IA dans le cadre de l'exercice de leurs fonctions, en plus de favoriser le développement de leur compétence numérique. Il est important que les équipes qui se consacrent à la conception et au développement de solutions liées à de tels systèmes disposent de l'expertise de pointe pour permettre à l'administration publique d'offrir des services publics simplifiés, intégrés et de qualité (MCN, 2024, p. 3-12).

Structure du document

Ce document se divise en cinq phases propres au cycle de vie de l'IA : la planification, la gestion éthique des données, le développement, la mise en œuvre et la maintenance. À noter que la phase de développement s'applique seulement dans le cas où le système d'IA est conçu par l'organisation.



Pour chacune de ces phases, des recommandations éthiques sont formulées, auxquelles s'ajoutent des questions réflexives et des principes éthiques de l'IA parmi les dix qui ont été présentés. Des exemples et des contre-exemples sont également proposés tout au long du document pour illustrer l'importance du respect de chaque phase et le rôle qu'elles jouent dans la mise en œuvre d'une réelle éthique de l'IA.

Il ne s'agit pas ici de suivre à la lettre chacune des actions proposées aveuglément, mais plutôt d'entamer une réflexion au sein de votre organisation afin de faire en sorte que le développement et l'utilisation de l'IA et du numérique puissent rester en tout temps au service de l'humain – en l'occurrence les élèves, le corps enseignant, le personnel administratif, les parents et le public.



1 Planification du système d'IA

Avant de recourir à un système d'IA au sein de votre organisme, il faut d'abord planifier plusieurs aspects pour assurer que son développement et son utilisation se font de façon éthique. Tout au long de la planification du système d'IA, les personnes qui en sont responsables doivent veiller à ce que son développement et sa mise en œuvre se fassent dans l'optique de promouvoir avant tout l'intérêt public. Voici les six principales actions à prendre en considération :

- Recours à l'IA
- Responsabilités en IA
- Participation des parties prenantes
- Objectifs du système d'IA
- Type d'IA
- Gestion des risques

1.1 Recours à l'IA

Questions réflexives

- À quel besoin l'IA peut-elle répondre?
- Le recours à l'IA est-il le meilleur moyen de répondre à ce besoin?
- Existe-t-il d'autres options plus pertinentes?



Avant même d'amorcer les travaux de planification du système d'IA, la première décision que devra prendre votre organisation est de déterminer si le recours à celle-ci est véritablement nécessaire pour répondre au besoin ciblé (5. Principe de l'efficience, de l'efficacité et de la pertinence). En effet, le discours actuel laisse parfois sous-entendre que l'IA peut faire des prédictions ou des recommandations de façon plus efficace que les systèmes numériques traditionnels. Or, ce n'est pas toujours le cas (ONU-Habitat et Mila, 2022).

Pour déterminer si votre organisation devrait recourir à l'IA, il est recommandé de préciser le besoin pouvant être pallié par cette technologie, de déterminer comment celui-ci est actuellement comblé, de réfléchir aux solutions susceptibles de mieux y répondre et de vérifier s'il existe d'autres options pour remplacer l'IA (5. Principe de l'efficience, de l'efficacité et de la pertinence). En ce sens, rappelons que l'IA est une technologie énergivore (Luccioni *et al.*, 2023). Il est donc recommandé de privilégier des solutions de rechange, numériques ou non, si cela est possible (6. Principe de la durabilité).

1.2 Responsabilités en IA

Questions réflexives

- À qui incombe la responsabilité de superviser les projets d'IA au sein de votre organisation?
- Cette personne a-t-elle les compétences et les savoirs nécessaires pour assumer cette responsabilité?
- Quels sont les rôles et les responsabilités des autres membres du personnel impliqués dans le projet de système d'IA?



Lorsqu'il est prévu de recourir à l'IA de manière générale par votre organisation, il vaut la peine de choisir un ou des responsables de l'IA parmi vos membres du personnel (9. Principe de la responsabilité) (Gautrais *et al.*, 2023). Dès la planification de tout système d'IA, cette ou ces personnes doivent être consultées et doivent approuver la solution envisagée. Une attention particulière par cette ou ces personnes responsables doit être apportée sur le plan de l'évaluation des aspects éthiques du projet. Il est recommandé de s'assurer que la ou les personnes ont les qualifications nécessaires et une bonne connaissance de base en sciences des données et en IA. De plus, le choix de plusieurs responsables pourrait permettre une meilleure répartition des pouvoirs qui découlent de cette responsabilité. Ces responsables doivent avoir les compétences nécessaires en IA et en éducation pour assumer pleinement les responsabilités s'y rattachant. Dans le cas où ces compétences sont déficientes, votre organisation doit s'assurer que ces responsables suivent les formations adéquates pour développer et maîtriser les compétences nécessaires (10. Principe de la compétence).

1.3 Participation des parties prenantes

Questions réflexives

- Qui sont les parties prenantes qui utiliseront le système d'IA et lesquelles seront concernées par celui-ci?
- Comment pouvez-vous impliquer ces parties prenantes dans le développement et l'utilisation du système d'IA?



La recherche est sans équivoque : pour favoriser une éthique de l'IA au sein de votre organisme, il est essentiel de faire participer les parties prenantes lors du développement ou de l'utilisation d'un système d'IA (UNICEF, 2021; UNESCO, 2019; CSE, 2020; OCDE, 2020). Cette participation est une façon de s'assurer que le développement et la mise en œuvre du système d'IA répondent à l'intérêt du public. Leur participation peut se faire à des moments programmés, mais peut aussi être réalisée lors de zones grises ou de conflits entre acteurs pour orienter le dénouement du problème vers l'intérêt public (5. Principe de l'efficacité, de l'efficacé et de la pertinence).

Les parties prenantes sont des individus, des groupes d'individus, des organismes ou des établissements concernés par le développement et l'utilisation d'un système d'IA. En éducation, elles regroupent principalement l'équipe de développement du système d'IA, les acteurs éducatifs (élèves, parents, regroupements de parents, personnel enseignant, syndicats, personnel scolaire, personnel administratif et directions d'établissements scolaires et de CSS ou de CS), le gouvernement, le public, des membres de la fonction publique québécoise, la communauté de recherche et de spécialistes (ex. : des psychologues et des médecins spécialistes en santé mentale de l'adolescence). Concernant ces deux dernières parties prenantes, il peut être intéressant d'entrevoir le projet en IA comme une occasion de profiter de l'expertise d'une équipe de chercheuses et chercheurs en éducation, en éthique ou en informatique, souvent en quête de terrains pour approfondir leurs recherches. Il en va de même pour les spécialistes visés par le développement du système d'IA. Ils peuvent bien sûr provenir du milieu de l'éducation, mais il peut être pertinent d'envisager la participation de professionnelles et professionnels provenant d'autres domaines, dans une optique de collaboration interdisciplinaire.

À noter que la démarche pour faire participer les parties prenantes doit se faire avec prudence. Il est plus aisé de les faire participer afin d'explorer leurs besoins et les moyens d'y répondre par l'IA. Dans d'autres cas, lorsque le développement du système d'IA est plus avancé, particulièrement pour la prise de décisions, les parties prenantes impliquées pourraient recevoir des informations erronées qui n'ont pas encore été validées.

Cette participation peut soutenir le développement et l'utilisation de l'IA de plusieurs façons. Elle peut notamment aider à définir l'objectif du système d'IA, à anticiper les risques (4. Principe de la sécurité) et à déterminer le potentiel de l'IA lié au bien-être des parties prenantes (5. Principe de l'efficacité, de l'efficacé et de la pertinence) ainsi qu'au respect de leurs droits et de leur autonomie (1. Principe du respect des personnes et de la règle de droit). Enfin, une telle initiative favorise également la transparence du processus du développement ou d'utilisation du système d'IA (7. Principe de la transparence), tout en renforçant la confiance des parties prenantes, ce qui facilitera à long terme son adoption. Sans ce processus de participation et de négociation, le risque est qu'une équipe administrative ait recours à un système d'IA et que les bénéficiaires, comme le personnel enseignant ou les élèves, n'en voient pas la pertinence.

Pour susciter leur participation, l'organisme britannique en innovation sociale Nesta (Berditchevskaia *et al.*, 2022) recommande l'adoption de l'un des quatre modèles participatifs décrits dans le tableau 1, ci-après, selon le rôle des parties prenantes. Pour chaque modèle, une ou plusieurs méthodes peuvent être sélectionnées pour favoriser cette participation.

Tableau 1 : Quatre modèles participatifs des parties prenantes (Berditchevskaia *et al.*, 2022)

Modèle	Rôles des parties prenantes	Méthodes	Avantages	Désavantages
1. Consultation	Comprendre la perspective du public par rapport à un système d'IA ou à l'IA	<ul style="list-style-type: none"> ▪ Sondage ▪ Entrevue ▪ Groupe de discussion 	Peu coûteux en temps et en ressources	Impact éthique faible sur le développement et l'utilisation du système d'IA
2. Contribution	<ul style="list-style-type: none"> ▪ Collecter ou catégoriser des données par des parties prenantes ciblées ▪ Valider un système d'IA 	Approvisionnement par la foule (« <i>crowdsourcing</i> »)	Augmentation de la diversité de données ou des perspectives des parties prenantes	Aucune interaction encouragée entre les parties prenantes
3. Collaboration	<ul style="list-style-type: none"> ▪ Tester le système d'IA et les méthodes utilisées pour le développement d'un système d'IA ▪ Partager les meilleures pratiques 	<ul style="list-style-type: none"> ▪ Ateliers ▪ Espaces collaboratifs ▪ Outils partagés ▪ Rencontres régulières en ligne ou en personne 	<ul style="list-style-type: none"> ▪ Participation sur une plus longue période ▪ Obtention de perspectives plus approfondies 	Nécessité de former préalablement les parties prenantes sur certains aspects en IA
4. Cocréation	<ul style="list-style-type: none"> ▪ Connaître les besoins, valeurs et priorités des parties prenantes – surtout les bénéficiaires et les spécialistes – par rapport au système d'IA ▪ Définir l'objectif du système d'IA ▪ Améliorer le design, l'interface ou la mise en œuvre du système d'IA 	<ul style="list-style-type: none"> ▪ Rencontres régulières entre l'équipe de développement ou de mise en œuvre du système d'IA et un petit groupe de participantes et participants ▪ Communauté de pratique 	Modèle de participation le plus complet et ayant le plus grand impact éthique sur le développement et l'utilisation du système d'IA	Modèle de participation le plus onéreux en temps et en ressources

Pour chacun de ces modèles, il faut prévoir par la suite des mécanismes qui permettent d'inclure la perspective des parties prenantes et une communication des moyens utilisés pour répondre aux points qu'elles auront soulevés, une fois le système d'IA conçu (voir la phase 4, « Mise en œuvre du système d'IA », p. 56) afin de leur fournir une reddition de comptes (7. Principe de la transparence).

Dans tout projet, pour favoriser la participation des parties prenantes, il est essentiel de cibler celles qui risquent d'être moins impliquées, telles que les parents, le personnel administratif, scolaire et enseignant ainsi que les élèves (UNICEF, 2021). Même si celles-ci ne sont pas nécessairement des usagères ou usagers du système d'IA, elles ont le droit de s'exprimer et d'être entendues à propos d'un système d'IA qui les concerne (2. Principe de l'inclusion et de l'équité). Une fois que les parties prenantes susceptibles d'être moins impliquées ont été identifiées, des moyens réalistes de les faire participer peuvent être envisagés. En plus des moyens mentionnés précédemment, il est possible d'impliquer des spécialistes des droits de ces groupes (ex. : spécialistes des droits des enfants, syndicats, chercheuses et chercheurs en éducation, représentantes et représentants de groupes de parents).

Parmi ces parties prenantes, il est également essentiel de penser à inclure des représentantes et représentants de celles qui appartiennent à des groupes marginalisés ou sous-représentés, tels que les communautés autochtones ou des groupes ethniques sous-représentés, si nécessaire. La participation des parties prenantes marginalisées permet d'assurer que le système d'IA est plus équitable, tout en favorisant une plus grande diversité de points de vue pour l'améliorer (2. Principe de l'inclusion et de l'équité) (UNICEF, 2019).

Le développement de l'agent conversationnel en santé mentale des adolescentes et adolescents – prénommé Milli – par le Centre hospitalier universitaire d'Helsinki en Finlande, en 2019, représente un exemple notable d'une participation réussie des parties prenantes pertinentes se rapprochant du modèle de « collaboration ». Des parties prenantes expertes de l'objectif du système d'IA, soit la santé mentale, ont pu participer au développement du système d'IA, comme des psychologues, des spécialistes en santé mentale et du personnel infirmier, en concertation avec l'équipe de développement. Un cours de développement a aussi été donné à l'Université Aalto, durant lequel des étudiantes et étudiants ont agi en tant que « spécialistes de l'expérience ». Grâce à la participation de cette partie prenante, l'agent conversationnel a été révisé pour qu'il apparaisse comme un agent virtuel, afin de réduire le risque que ses interventions soient perçues comme celles d'un être humain (UNICEF, 2021).

Ressource :

- [Participatory AI for humanitarian innovation](#)

1.4 Objectifs du système d'IA

Questions réflexives

- Quel est l'objectif ou quels sont les objectifs du système d'IA?
- Quelle est la plus-value visée par l'atteinte de cet objectif par le système d'IA?



Bien que l'IA puisse contribuer à l'amélioration de l'éducation, son utilisation n'est pas une garantie en soi, car elle ne peut résoudre de façon technique des problèmes de nature sociale et éducative : elle ne peut que soutenir une intervention de cette nature. C'est pourquoi, pour permettre de maximiser le potentiel de cette technologie, il est recommandé de préciser l'objectif du système d'IA (IEAIE, 2020; Gouvernement du Québec, 2023). Cet objectif devrait toujours être directement lié à la promotion de l'intérêt du public. Si ce n'est pas le cas, il est préférable de le réévaluer (5. Principe de l'efficacité, de l'efficacité et de la pertinence). Le fait d'établir dès le début l'objectif du système d'IA permet ainsi de mieux orienter les décisions quant au choix des méthodes d'IA et du design pédagogique du système, tout en prévenant des enjeux légaux et pédagogiques (UNICEF, 2021).

Une fois l'objectif établi, l'organisme doit analyser si celui-ci respecte différents aspects. D'abord, l'objectif ciblé doit être une priorité au sein de l'organisme et apporter une plus-value dans le milieu scolaire. Ensuite, il devra respecter les orientations éducatives de l'organisme (ex. : plan d'engagement vers la réussite), les politiques en éducation, les conventions collectives des membres du personnel ainsi que les lois (ex. : *Loi sur l'instruction publique*, *Loi sur l'accès*). Par la suite, cet objectif devra être au diapason des méthodes pédagogiques probantes validées par la recherche scientifique en éducation (UNESCO, 2019). Il doit également garantir le bien-être de l'ensemble des parties prenantes, y compris les élèves, leurs parents et le personnel scolaire, tout en respectant leurs droits et leur autonomie (5. Principe de l'efficacité, de l'efficacité et de la pertinence). Par ailleurs, il doit aussi s'accorder avec les définitions d'équité, de diversité et d'inclusion (EDI) de votre organisme (2. Principe de l'inclusion et de l'équité). Enfin, dans une optique de coopération mondiale, il est aussi recommandé de justifier en quoi l'objectif s'insère dans un ou quelques objectifs du Plan d'action de développement durable 2023-2028 du ministère de l'Éducation (2023), qui s'appuie sur les 17 objectifs de développement durable de l'ONU (2015).

L'objectif du système d'IA peut être ensuite précisé et justifié dans un dossier d'opportunité (ou dossier d'affaires), afin de démontrer que « le recours à l'intelligence artificielle est essentiel à la résolution du problème ou à l'opportunité identifiée, ou encore que cette technologie permet d'améliorer un processus » (MCN, 2024, p. 7).

Ressource :

- [Plan d'action de développement durable 2023-2028 du ministère de l'Éducation](#)

1.5 Type d'IA

Questions réflexives

- Quel type d'IA sera utilisé?
- Quels sont les risques associés à ce type d'IA?
- Quelles mesures de mitigation pourront être mises en place pour minimiser ces risques?



Il existe plusieurs types d'IA. Toutefois, leur incidence de risque n'est pas la même. Selon l'objectif que le système d'IA cherche à atteindre, il est important de déterminer quel type d'IA sera utilisé et de mettre en place les mesures de mitigation nécessaires. Parmi les types d'IA pouvant être utilisés en éducation, ceux qui seront explorés ici sont l'IA simple, l'IA analytique et prédictive et l'IA générative.

1.5.1 IA simple

L'IA simple regroupe des méthodes d'IA comme l'arbre décisionnel, qui offre des résultats parfois aussi robustes que des algorithmes d'IA plus complexes, tels que les réseaux de neurones artificiels (*neural network*). Or, pourquoi est-il préférable de recourir à des algorithmes d'IA utilisant peu de données? D'une part, l'utilisation de peu de variables et d'observations permet de diminuer les risques associés à la vie privée (1. Principe du respect des personnes et de la règle de droit), tout en garantissant une robustesse du système d'IA (3. Principe de la fiabilité et de la robustesse). En effet, un système d'IA peut conserver une grande robustesse, tout en restant relativement simple (Khosravi *et al.*, 2022). D'autre part, les algorithmes utilisant peu de données permettent de diminuer le nombre d'observations utilisées pour modéliser un phénomène, ce qui diminue l'empreinte carbone (6. Principe de la durabilité). Le recours à des algorithmes simples a aussi comme avantage d'être parfaitement interprétable, contrairement aux algorithmes d'IA plus complexes qui nécessitent l'intégration de méthodes d'explicabilité (8. Principe de l'explicabilité). En revanche, le choix du système d'IA ne peut se limiter au seul critère de simplicité. D'autres critères de performance (précision, sensibilité, exactitude, etc.) doivent être mis à contribution pour le choix du système à retenir.

1.5.2 IA analytique et prédictive

Il existe aussi des systèmes d'IA qui peuvent analyser des données – IA analytique – et d'autres qui peuvent faire des prédictions à partir de données – IA prédictive. Également connus sous le nom d'« analyse des données » ou d'« analytique des données », ces types d'IA sont principalement utilisés par les gestionnaires pour les soutenir dans la prise de décision. Toutefois, malgré ce potentiel, ces types d'IA ne devraient pas automatiser une telle prise de décision, c'est-à-dire qu'un humain devrait toujours être dans la boucle du fonctionnement de tout système d'IA analytique ou prédictif. Ce genre de mesure permet ainsi d'éviter que des humains ne soient remplacés (10. Principe de la compétence) et de

minimiser les risques de discrimination envers certains groupes marginalisés (1. Principe du respect des personnes et de la règle de droit).

Certains de ces systèmes d'IA utilisent des méthodes d'IA qui nécessitent une grande quantité de données et un fonctionnement complexe, ce qui rend difficile pour les usagères et usagers d'en comprendre le mécanisme, et peut donc nuire au recours de ces systèmes pour la prise de décision (Khosravi *et al.*, 2022). Ce phénomène, aussi appelé « boîte noire », nécessite, d'une part, la valorisation de systèmes d'IA simples ou, du moins, qui intègrent des méthodes d'explicabilité, c'est-à-dire des méthodes qui rendent compréhensibles les données et les résultats utilisés par le système d'IA aux usagères et usagers (8. Principe de l'explicabilité). D'autre part, il importe que les usagères et usagers soient sensibilisés à ce risque et outillés pour le minimiser.

Un autre risque à prendre en considération quant à ces types d'IA concerne directement les gestionnaires et le personnel scolaire, qui pourraient juger les résultats obtenus par le système d'IA comme entièrement neutres et objectifs. Selon Collin et Hennetier (2024), une telle perspective pourrait nuire directement au jugement de ces membres du personnel, qui pourraient le déléguer entièrement au système d'IA analytique ou prédictif. L'équipe de recherche fait les recommandations suivantes pour prévenir ce genre de risque éthique et ainsi préserver le jugement professionnel du personnel scolaire (10. Principe de la compétence) :

- Remplacer l'imputabilité de performance par une imputabilité « intelligente » (Mockler et Stacey, 2020) et privilégier avant tout une finalité d'amélioration continue orientée vers l'équité scolaire;
- Établir des objectifs scolaires et pédagogiques signifiants, puis déterminer les données scolaires pertinentes en conséquence, plutôt que de définir des objectifs faciles à mesurer et des données scolaires faciles à collecter;
- Diversifier les données scolaires prises en compte (formelles/informelles, qualitatives/quantitatives, scolaires/scientifiques) afin de trianguler la prise de décision;
- Privilégier l'interprétation des données à l'échelle locale, qui est plus informative pour guider les interventions scolaires et pédagogiques (Collin et Hennetier, 2024, p. 61).

Ainsi, les résultats de ces systèmes d'IA devraient être considérés comme une source complémentaire à plusieurs autres pour nourrir les réflexions collectives dans la prise de décision.

1.5.3 L'IA générative

L'IA générative peut être définie comme un « système informatique qui utilise des modèles d'intelligence artificielle probabilistes pour générer automatiquement des contenus variés (texte, image, voix, musique, vidéo, présentation, etc.) en réponse à une requête faite par une personne utilisatrice » (Gouvernement du Québec, 2024, p. 1).

Avec la prolifération des systèmes d'IA générative depuis 2022, plusieurs activités potentielles sont associées à ce type d'IA, telles que la personnalisation du parcours d'apprentissage des élèves, ou encore l'assistance du personnel scolaire dans la réalisation de certaines tâches. Il reste que ce type d'IA pose plusieurs risques à prendre en compte.

Parmi ces risques, on note le fait que l'IA générative est considérée comme l'un des types d'IA les plus polluants. Une seule requête sur un système d'IA générative utilisant des données issues d'Internet représente l'équivalent de la recharge d'un téléphone intelligent (Luccioni *et al.*, 2024). En ce sens, il est recommandé de recourir, lorsque cela est possible, à un système d'IA générative utilisant des données plus limitées que les systèmes entraînés à l'aide de données d'Internet. De plus, une culture de « sobriété numérique » se définit comme « une réflexion critique et créative » face aux enjeux environnementaux du numérique et encourage « la priorisation des usages face aux limites planétaires » (Obvia, 2024). Cette culture peut se développer lorsque l'on sensibilise les usagères et usagers à ces enjeux et que l'on met en place des moyens qui leur permettent de limiter le recours à ces technologies en valorisant d'autres solutions.

Un autre aspect important à prendre en considération avec l'IA générative est que les réponses qu'elle peut donner ne sont pas toujours exactes ni de bonne qualité, bien que leur contenu semble les présenter comme si elles l'étaient (CSE et CEST, 2024). Ce phénomène, connu sous le nom d'« hallucination d'IA » ou de « fabulation », doit donc être pris en compte par les usagères et usagers. Ainsi, il importe de valoriser leur jugement critique et de les inviter à valider les contenus générés à l'aide de sources externes.

Ces systèmes d'IA peuvent avoir des effets préjudiciables à l'égard de certains groupes (UNESCO, 2024). En effet, parce qu'ils utilisent des données parfois biaisées, ils peuvent générer des contenus qui le sont également. Les usagères et usagers doivent être sensibilisés à cet enjeu et modifier leur requête ou les contenus générés pour éviter les préjudices qui peuvent en découler. Il arrive aussi qu'une requête concerne un groupe sous-représenté par les données utilisées par le système d'IA. Dans de tels cas, votre organisation peut valoriser le recours à des systèmes d'IA générative utilisant des données plus représentatives de ces groupes. Par exemple, les systèmes d'IA générative qui utilisent des données tirées d'Internet produisent des contenus de meilleure qualité en anglais qu'en français. Ici, il peut être pertinent de se tourner vers des systèmes d'IA entraînés principalement à partir de données en français, de façon à obtenir de meilleurs résultats.

À l'heure actuelle, les effets de ces systèmes sur l'apprentissage et le développement de compétences ne sont pas encore clairs. Il convient donc de rester vigilant quant aux risques qu'ils peuvent représenter sur l'« agentivité » des actrices et acteurs scolaires qui les utilisent. Il est recommandé de ne pas imposer ces systèmes aux membres du personnel de votre organisation, mais plutôt de sensibiliser ces derniers à cet enjeu, afin qu'ils s'assurent que leurs usages permettent de conserver leur agentivité. **Une attention particulière doit être apportée à la sensibilisation des usagères et usagers au fait qu'ils ne doivent pas partager de données personnelles, confidentielles ou protégées par les droits d'auteur dans ces systèmes (1. Principe du respect des personnes et de la règle de droit).**

Dans le cas où votre organisation envisage de recourir à l'IA générative avec des personnes mineures, une population considérée comme vulnérable, il est recommandé de préciser l'âge minimal pour l'utilisation autonome du système d'IA générative que vous prévoyez adopter dans votre organisation (UNESCO, 2024).

Ressources :

- [Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative](#) (MCN, 2024)
- [L'utilisation pédagogique, éthique et légale de l'intelligence artificielle générative – Guide destiné au personnel enseignant](#) (Ministère, 2024)

1.6 Gestion des risques

Questions réflexives

- Qui seront les membres du personnel de votre organisation responsables d'établir les risques possibles du système d'IA?
- Quels sont les risques associés au système d'IA?
- Quelles sont les mesures de mitigation mises en place pour prévenir les risques du système d'IA?
- Est-ce que le système d'IA que vous comptez utiliser pose des risques trop élevés par rapport aux bénéfices?



Les risques liés aux systèmes d'IA sont compris comme un préjudice potentiel ou réel occasionné par le système d'IA qui doit être évité. Il est recommandé que votre organisme adopte une stratégie de gouvernance et de gestion des risques de l'IA (Gautrais *et al.*, 2023). Dans cette gouvernance, vous devez prévoir des procédures en cas d'urgence ou d'échec du système d'IA. Cette gestion des risques peut être réalisée par un comité de membres qualifiés du personnel de votre organisation.

Pour ce faire, il est recommandé de structurer cette gestion des risques en quatre étapes : gouverner, cartographier, mesurer et gérer les risques (ensemble des principes) (National Institute of Standards and Technology, 2023). Pour la première étape, gouverner les risques, il s'agit d'instaurer une culture de gestion des risques intégrant les trois autres étapes (cartographier, mesurer et gérer les risques) à l'ensemble de votre organisation pour tous les systèmes d'IA que vous comptez développer ou utiliser. La seconde, cartographier les risques, concerne directement le système d'IA à développer ou à utiliser. Ici, il est plutôt question de bien déterminer tous les risques de l'IA et de les prioriser en prenant en compte le contexte dans lequel elle sera mise en œuvre, son fonctionnement, ses objectifs et son potentiel. En fonction des risques décelés, des mesures de mitigation doivent être prévues. Pour la troisième étape,

mesurer les risques, les mesures d'évaluation des risques sont définies et planifiées pendant tout le cycle de vie du système d'IA. Pour la dernière étape, gérer les risques, il s'agit de s'assurer que, face aux risques cartographiés et mesurés, les mesures nécessaires sont mises en place pour les mitiger, pour réagir rapidement à tout risque détecté et pour apporter les correctifs nécessaires au système d'IA, pour prévenir une récurrence.

Il faut toutefois signaler que, dans certains cas, l'évaluation du système d'IA peut difficilement se faire. Dans leur recherche sur les modèles prédictifs du décrochage scolaire aux États-Unis, Perdomo *et al.* (2023) soulignent que certains modèles prédictifs sont inefficaces s'ils se basent sur des aspects environnementaux (ex. : le niveau de défavorisation de l'établissement scolaire) et que, même s'ils sont efficaces, il faut du temps avant de voir s'ils posent un risque pour les élèves, puisque certaines variables nécessitent un délai avant de déterminer si le système fonctionne en les utilisant. Ce phénomène est d'ailleurs appelé « paradoxe de Simpson » (Berger et Caravenna, 2021). À noter que, dans le cas de la solution d'IA pour prévenir le décrochage scolaire, des mesures de mitigation ont été mises en place pour prévenir ce risque d'inefficacité.

Face à ces risques, il est attendu que l'organisme prendra les mesures raisonnables pour éviter qu'un incident de la sorte ne survienne. L'ensemble des mesures d'atténuation pour prévenir le ou les risques décelés devrait être documenté (Gautrais *et al.*, 2023).

À noter que certains systèmes d'IA posent un risque trop élevé par rapport à leurs bénéfices pour que leur utilisation soit recommandée, que ce soit par les données principalement sensibles qu'ils traitent ou par les effets que leurs décisions ont sur le bien-être et la sécurité des usagères et usagers (1. Principe du respect des personnes et de la règle de droit), surtout lorsque cela implique des mineurs, une population considérée comme vulnérable (2. Principe de l'inclusion et de l'équité) (Langlois *et al.*, 2022).

L'Union européenne (UE) définit comme risque inacceptable en IA « tous les systèmes d'IA considérés comme une menace évidente pour la sécurité, les moyens de subsistance et les droits des personnes [...] » (2023). Par exemple, il peut s'agir de jouets utilisant l'assistance vocale susceptibles d'encourager des comportements dangereux ou un système de crédit social. L'exemple d'un tel système en Chine est révélateur des risques encourus par ce type de technologie (UE, 2023). En effet, il a été constaté que ce système de crédit social a depuis exacerbé un état de surveillance, réduit considérablement l'agentivité de ses citoyennes et citoyens et bafoué plusieurs de leurs droits (Creemers, 2018). Il est aussi recommandé d'éviter de recourir à des systèmes d'IA de reconnaissance faciale en éducation, comme celui utilisé par une école suédoise en 2018 pour la prise de présence des élèves, étant donné les risques élevés en matière de protection de la vie privée (1. Principe du respect des personnes et de la règle de droit) et de la sécurité des données (4. Principe de la sécurité) que de tels systèmes posent (Kun, 2022). Ces systèmes d'IA ne sont pas recommandés, puisqu'ils posent des risques trop élevés par rapport aux bénéfices.

Si vous comptez utiliser un autre système d'IA que ceux présentés précédemment, voici une liste non exhaustive d'exemples de risques par catégories :

1. Risques liés aux données

- a) L'utilisation de données personnelles et la façon dont elles sont stockées en matière d'infrastructures augmentent les risques d'incidents de confidentialité (CAI, 2023).
- b) La variabilité entre les CSS ou CS des données utilisées et des méthodes de collecte peut nuire à la qualité générale des données.
- c) Il est possible qu'il y ait une quantité insuffisante de données disponibles sur des groupes sous-représentés (ex. : élèves d'origines autochtones) (Ministère, 2023).

2. Risques liés au système d'IA et aux méthodes utilisées

- a) Les systèmes d'IA peuvent donner l'impression aux usagères et usagers que les décisions prises par le système d'IA sont d'origine humaine (UNICEF, 2021).
- b) Il existe un risque que le système d'IA fasse des erreurs de classification (Collin *et al.*, 2022).
- c) Un système d'IA ne peut garantir à 100 % un résultat adéquat. Il y a donc un risque d'engendrer de faux positifs ou de faux négatifs, particulièrement auprès d'élèves appartenant à des groupes marginalisés (UNICEF, 2021). Il est ainsi nécessaire de réduire les faux positifs et les faux négatifs selon les conséquences de chaque type d'erreur.
- d) La grande quantité de données ou d'algorithmes qu'utilise le système d'IA peut rendre difficilement compréhensibles ses résultats auprès des usagères et usagers, ce qui peut mener à une mauvaise interprétation de ceux-ci, et même à une méfiance à leur égard (Khosravi *et al.*, 2022).
- e) Certaines méthodes d'IA posent des risques élevés sur le plan des droits et libertés de certaines parties prenantes (ex. : géolocalisation, reconnaissance faciale, décision fondée exclusivement sur un traitement automatisé) (voir la p. 40) (UE, 2023).

3. Risques liés aux parties prenantes

- a) La pénurie de main-d'œuvre peut complexifier le développement et l'utilisation de systèmes d'IA par le milieu scolaire.
- b) Le manque de compétences sur le plan de l'intelligence numérique et de la cybersécurité peut rendre difficile le développement de projets éthiques en IA.

- c) Les systèmes d'IA peuvent entraver les droits des enfants (droits de s'exprimer et de penser librement; à la vie privée; à l'information; de jouer, de relaxer et de participer pleinement à la vie culturelle et artistique; de ne pas être exploités économiquement ou de toute manière qui nuit à leur bien-être [Commission des droits de la personne et des droits de la jeunesse du Québec, 1989]).
- d) Les systèmes d'IA peuvent empiéter sur le jugement professionnel, notamment pédagogique, du personnel enseignant.
- e) Les systèmes d'IA peuvent renforcer des injustices vécues par certaines ou certains élèves, parents ou personnel scolaire appartenant à des groupes marginalisés (UNICEF, 2021).
- f) Les systèmes d'IA peuvent créer de la dépendance chez les usagères et usagers, particulièrement auprès de populations vulnérables telles que les élèves (UNICEF, 2021).
- g) Le manque de littératie numérique des parties prenantes (personnel scolaire, parents, élèves, direction des établissements scolaires et des CSS ou CS, membres administratifs, public, etc.) peut les amener soit à faire confiance aveuglément aux prédictions et aux recommandations du système d'IA, soit à s'en méfier déraisonnablement.
- h) Le manque de participation de certaines parties prenantes au développement du système d'IA peut nuire à son acceptabilité sociale (Arnold et Sclater, 2017).
- i) Plusieurs parties prenantes ont des préoccupations liées à l'IA et donc s'en méfient, que ce soit concernant l'usage de « vieilles » données pour prendre des décisions, l'usage de données à des fins commerciales ou la mise en place d'un état de surveillance diminuant leur autonomie et leur vie privée (OCDE, 2020).

4. Risques liés à la qualité des services éducatifs

- a) Le recours à des données biaisées, « le design de systèmes d'IA, l'implantation de méthodes principalement instructionnistes qui mettent l'accent sur le transfert de connaissances sans prendre en considération le contexte et les facteurs sociaux peuvent amplifier les biais propres à des approches d'enseignement et d'apprentissage inadéquates » [traduction libre] (UNESCO, 2019).
- b) Même si un système d'IA devrait permettre de cibler, sur le plan scolaire, les principaux élèves ou groupes d'élèves pour lesquels il faut davantage s'investir – ou pour lesquels il faut modifier le type d'encadrement afin de leur permettre d'atteindre les objectifs –, le système d'IA peut pousser le personnel scolaire, lorsque ce dernier n'a pas été adéquatement formé, à prédire la performance future de leurs élèves et à ne se concentrer que sur ceux qui réussissent déjà bien (effet Pygmalion) (Beattie *et al.*, 2017).
- c) Les systèmes d'IA peuvent diminuer le temps que les élèves passent à interagir les uns ou les uns avec les autres (UNESCO, 2019).

- d) Les systèmes d'IA peuvent « diminuer les occasions pour les élèves de cultiver leur débrouillardise, leur efficacité, leur capacité à s'autoréguler, leur pensée critique et leurs indépendances de pensée » [traduction libre] (UNESCO, 2019).
- e) Les systèmes d'IA peuvent promouvoir un enseignement standardisé où les élèves sont perçus comme de simples unités métriques (Arnold et Sclater, 2017).
- f) Les systèmes d'IA peuvent encourager l'adoption de comportements dysfonctionnels, tels que l'exclusion d'élèves à risque élevé de décrochage par un établissement scolaire pour préserver son statut (OCDE, 2022), ou la modification de données pour augmenter le nombre d'élèves à risque dans un établissement scolaire en vue d'obtenir plus de financement (Khosravi *et al.*, 2022).
- g) Les systèmes d'IA peuvent appuyer des idées pédagogiques pauvres (Collin et Marceau, 2021).
- h) Les systèmes d'IA peuvent créer une distance psychologique entre ceux-ci et l'utilisateur ou l'usager, par exemple le personnel enseignant, et la personne dont les données sont utilisées, par exemple l'élève (Crawford *et al.*, 2023).

5. Risques sociétaux

- a) Les systèmes d'IA peuvent avoir des effets néfastes sur l'environnement, notamment concernant la production de gaz à effet de serre, la production de déchets électroniques et l'extraction de ressources naturelles (Université de Montréal, 2018).

Ressources :

- [Artificial Intelligence Risk Management Framework](#) (National Institute of Standards and Technology, 2023)
- [Artificial Intelligence and Cities: Risks, Applications and Governance](#) (ONU-Habitat et Mila, 2022)



2 Gestion éthique des données

Questions réflexives

- Quels types de données seront utilisés?
- Quels sont les risques associés à ces types de données?
- Quelles mesures de mitigation pourront être mises en place pour minimiser ces risques?



L'IA requiert l'usage de données pour se perfectionner. C'est pourquoi il est nécessaire de mettre en place une gestion éthique de celles-ci. Cette phase regroupe un ensemble de mesures adoptées pour assurer que les données sont utilisées de façon appropriée tout au long de leur cycle de vie, afin de protéger le public. Une donnée peut être définie comme une « information sous une forme conventionnelle destinée à en permettre le stockage, la transmission ou le traitement par des moyens automatiques » (Office québécois de la langue française, 2022). Par exemple, des données peuvent consister en un pixel sur une image, en un nombre dans un tableau statistique ou en un mot dans un texte.

Les mesures à prendre pour respecter les principes éthiques dépendent principalement des types de données utilisées par le système d'IA. Ainsi, les mesures seront ici explorées selon les types de données suivants :

- Mesures pour les renseignements personnels
- Mesures pour les données confidentielles
- Mesures pour respecter le droit d'auteur
- Mesures pour l'ensemble des données

2.1 Mesures pour les renseignements personnels

Questions réflexives

- Qu'est-ce qui justifie le recours à des renseignements personnels ou sensibles par le système d'IA?
- Quelles mesures de mitigation pourront être mises en place pour assurer leur sécurité?
- Si vous travaillez pour un organisme public, est-ce que le système d'IA respecte la *Loi sur l'accès*?
- S'il s'agit plutôt d'une école privée non subventionnée, est-ce que le système d'IA respecte la *Loi sur la protection des renseignements personnels dans le secteur privé*?



La *Loi sur l'accès*, modernisée en septembre 2022, détermine quels sont les renseignements qui sont personnels et confidentiels et la façon dont les organismes publics doivent les protéger en adoptant un ensemble de mesures. Pour leur part, les écoles privées non subventionnées sont assujetties à la *Loi sur la protection des renseignements personnels dans le secteur privé*, comme toutes les entreprises privées. À noter que, dans le domaine de l'IA, les renseignements personnels sont mieux connus sous le nom de « données personnelles », tandis que, pour les renseignements sensibles, il est question de « données sensibles ».

Dans le cas de la *Loi sur l'accès* et de tout organisme public, le gouvernement du Québec (2025) définit un renseignement personnel ainsi :

Un renseignement est personnel lorsqu'il concerne une personne physique et qu'il en permet, directement ou indirectement, l'identification.

Voici ce qui le caractérise : il doit faire connaître quelque chose à quelqu'un; il doit avoir un rapport avec une personne physique; il doit être susceptible de distinguer cette personne par rapport à une autre ou de reconnaître sa nature. (Voir la présentation des concepts clés liés aux renseignements personnels.)

« Un nom, un âge, une position GPS précise, une adresse IP, un identifiant publicitaire ou une adresse courriel, mais aussi [des] informations concernant les intérêts, les émotions, les attitudes, les habitudes, l'identité, les activités, les relations, etc. » (CAI, 2022, p. 7). Tout renseignement personnel est confidentiel, et un consentement de la personne concernée¹ peut être requis selon les différentes utilisations envisagées. Leur utilisation dans un système d'IA nécessite l'adoption de mesures

¹ Personne qui doit donner son consentement pour qu'une donnée personnelle ou sensible soit utilisée par un système d'IA. Pour le mineur de moins de 14 ans, c'est au titulaire de l'autorité parentale ou au tuteur de donner ce consentement. Pour celui de 14 ans ou plus, le consentement est donné par le mineur, le titulaire de l'autorité parentale ou le tuteur.

supplémentaires pour assurer leur sécurité. Toutefois, si le système d'IA utilise un renseignement personnel qui a déjà été recueilli pour les mêmes fins, le consentement pourrait ne pas être nécessaire. Par exemple, les résultats scolaires d'élèves sont des renseignements personnels qui sont collectés pour suivre leur avancement dans leur parcours éducatif et veiller à leur offrir les services et l'accompagnement requis pour leur réussite. Donc, si un système d'IA utilisant les mêmes renseignements déjà détenus par l'organisme scolaire est conçu pour mieux cibler les interventions à faire pour soutenir la réussite de l'élève, l'obtention du consentement pourrait ne pas être requise. Chaque usage doit donc être validé auprès de la personne responsable de la protection des renseignements personnels de votre organisation. Celle-ci pourra alors déterminer quelles conditions doivent être respectées (obtenir le consentement ou informer les personnes concernées qu'un système d'IA est utilisé).

Parmi l'ensemble des renseignements personnels, certains sont considérés comme « sensibles ». Le gouvernement du Québec les définit ainsi :

Un renseignement personnel est sensible lorsque par sa nature, notamment médicale, biométrique ou autrement intime, ou en raison du contexte de son utilisation ou de sa communication, il suscite un haut degré d'attente raisonnable en matière de vie privée.

Cette sensibilité doit donc être évaluée en fonction du degré d'attente raisonnable du renseignement personnel en matière de vie privée. En effet, lorsque le degré est élevé, cette information doit être considérée comme un renseignement personnel sensible.

En d'autres termes, il s'agit de tout renseignement qui peut affecter davantage la personne concernée, compte tenu : de sa source, du degré d'intimité auquel il est rattaché, des répercussions potentiellement liées à sa divulgation ou à son utilisation.

L'évaluation de la sensibilité d'un renseignement personnel dépend donc du contexte et de la nature des renseignements.

Un renseignement personnel inclus dans les catégories suivantes doit généralement être considéré comme sensible, soit les renseignements : financiers; génétiques ou biométriques; concernant la santé; concernant la vie sexuelle ou l'orientation sexuelle; concernant les convictions religieuses ou philosophiques; concernant les opinions politiques; concernant l'origine ethnique ou raciale. (Voir la présentation des concepts clés liés aux renseignements personnels.)

Une donnée sensible est définie comme une information portant sur une personne qui, si elle était divulguée, pourrait avoir un impact sur sa vie privée en raison de son caractère intime : numéro d'assurance sociale, date et lieu de naissance, numéro de téléphone, adresse postale, genre, groupe ethnique, renseignements relatifs à la santé, par exemple. Dans le domaine de l'éducation, les données sensibles peuvent être « l'identité de l'enfant et de ses parents, leurs coordonnées (adresses postales et de courriel, téléphone), leurs noms d'utilisateur pour accéder au portail et leurs alias » (Khuong, 2022, p. 4). Ainsi, en raison des impacts que ces données sensibles peuvent avoir sur la vie privée des personnes concernées si elles étaient divulguées, elles doivent faire l'objet d'une sécurité et d'un traitement adéquats.

Au Québec, c'est la Commission d'accès à l'information (CAI) qui est chargée de l'application de la *Loi sur l'accès* et de la *Loi sur la protection des renseignements personnels dans le secteur privé*. Elle veille à faire respecter le droit d'accès de toute citoyenne et de tout citoyen aux documents d'organismes publics et de voir ses renseignements personnels protégés (Gouvernement du Québec, 2023). Pour ce faire, la CAI peut émettre des avis sur des projets de développement de systèmes d'IA et offre des outils pour favoriser une application adéquate et conforme des lois. À noter qu'elle peut faire une demande péremptoire afin d'exiger la production de renseignements ou de documents pour vérifier si la loi dans votre organisme est bien appliquée.

Par conséquent, s'il est prévu que le système d'IA utilise ces types de données, il est recommandé de bien justifier son utilisation, et même, lorsque cela est possible, d'évaluer la possibilité de ne pas y recourir si une telle option ne modifie pas le fonctionnement du système. Il est aussi recommandé d'adopter les mesures prévues par la *Loi sur l'accès* (1. Principe du respect des personnes et de la règle de droit). Au besoin, il peut être pertinent de valider les mesures mises en place auprès d'une ou d'un juriste spécialisé dans la protection des renseignements personnels. L'ensemble de ces mesures et la validation du système doivent se faire auprès du responsable de la protection des renseignements personnels de votre organisation. Enfin, la création d'un tableau présentant l'ensemble des exigences et des actions prises en compte pourrait s'avérer utile pour démontrer la conformité de l'organisme public à cette loi (MCN, 2024).

Ressource :

- [Protection des renseignements personnels](#) (Gouvernement du Québec, 2024b)

2.2 Mesures pour les données confidentielles

Questions réflexives

- Qu'est-ce qui justifie le recours à des données confidentielles par le système d'IA?
- Quelles mesures de mitigation pourront être mises en place pour assurer leur sécurité?
- Est-ce que le système d'IA respecte la *Loi sur l'accès*?



Avant de débiter, il est à noter qu'en plus des renseignements personnels confidentiels spécifiés dans la *Loi sur l'accès* et la *Loi sur la protection des renseignements personnels dans le secteur privé*, il existe d'autres données confidentielles qui ne sont pas personnelles. Cette section concerne les données confidentielles autres que personnelles. Pour les données confidentielles également personnelles, se référer à la partie 2.1.

Une donnée confidentielle peut être définie comme une « information qui ne doit être communiquée ou rendue accessible qu’aux personnes et aux entités autorisées » (Office québécois de la langue française, 2024). Elle peut être de nature technique, financière ou administrative – il peut même s’agir d’un secret industriel – et ne peut être révélée, car cela compromettrait les processus décisionnels de l’organisme ou lui créerait un préjudice (ex. : estimation des coûts de travaux avant un appel d’offres ou la conclusion d’un contrat, réponses à un examen, critères ou pondération précis d’évaluations de projets ou de candidats à l’embauche, bilan financier d’une entreprise, etc.).

Tout comme pour les renseignements personnels, le recours aux données confidentielles (autres que personnelles) doit donc être justifié et bien encadré. Des mesures pour respecter la *Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement* ainsi que la *Loi concernant le cadre juridique des technologies de l’information* doivent aussi être mises en place et approuvées par la personne répondante de la sécurité de l’information de l’organisme. L’assistance d’une ou d’un juriste maîtrisant ces encadrements pourrait aussi être sollicitée (1. Principe du respect des personnes et de la règle de droit).

2.3 Mesures pour respecter le droit d’auteur

Questions réflexives

- Qu’est-ce qui justifie le recours à des œuvres protégées conformément à la *Loi sur le droit d’auteur* par le système d’IA?
- Est-ce que le système d’IA pourrait plutôt avoir recours à des œuvres libres de droits?
- Est-ce que le système d’IA respecte la *Loi sur le droit d’auteur*?



Dans le cas où il est prévu qu’un système d’IA ait recours à des œuvres protégées par la *Loi sur le droit d’auteur*, il est essentiel de respecter cette loi et, si cela ne modifie pas le fonctionnement du système, d’éviter d’y recourir et de se tourner plutôt vers des œuvres libres de droits (1. Principe du respect des personnes et de la règle de droit). La *Loi sur le droit d’auteur* vise à protéger les œuvres – artistiques, littéraires, musicales et dramatiques – qui ne sont pas libres de droits. Dans le cas d’une utilisation de telles œuvres, l’organisation doit s’assurer d’obtenir les droits d’utilisation ou d’exploitation. Elle peut être assistée, au besoin, par une ou un juriste se spécialisant dans ce champ.

Ressources :

- [Le droit d’auteur en bref](#) (Copibec, 2024a)
- [Où en sommes-nous avec l’intelligence artificielle?](#) (Copibec, 2024b)

2.4 Mesures pour l'ensemble des données

Questions réflexives

- Est-ce que les données sont entreposées de façon sécuritaire dans votre pays?
- Est-ce que les données sont biaisées?
- Est-ce que les données sont représentatives de l'ensemble des groupes concernés?
- Est-ce possible de prévenir le recours à des données biaisées ou non représentatives?



Peu importe les données auxquelles un système d'IA peut recourir, qu'elles soient personnelles, sensibles, confidentielles, qu'elles appartiennent à une autrice ou à un auteur, ou qu'elles n'appartiennent à aucune de ces catégories, des mesures doivent être mises en place pour mitiger les risques.

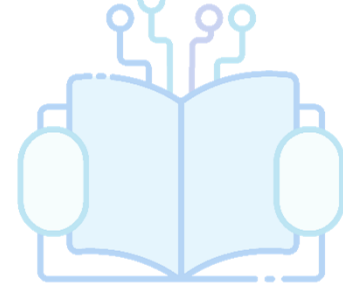
Entre autres, pour prévenir les bris de sécurité (4. Principe de la sécurité), une ligne de conduite sécuritaire est à prévoir afin de séparer « l'environnement du développement du modèle d'IA du reste de l'infrastructure informatique » (Gautrais *et al.*, 2023, p. 31). L'infonuagique est une technologie permettant le stockage et le traitement de données de façon délocalisée à partir de serveurs distants accessibles grâce à un appareil et d'une connexion Internet (Office québécois de la langue française, 2023). Il existe plusieurs modèles de service et de déploiement. Parmi eux, il est recommandé de choisir les modèles qui conviennent le mieux à votre organisation et à la nature des données utilisées. Le recours à la technologie infonuagique permet un entreposage plus sécuritaire que le recours à des entrepôts physiques de données et permet ainsi de diminuer les risques d'incident de confidentialité (4. Principe de la sécurité). Toutefois, malgré ses bienfaits, l'infonuagique comporte plusieurs risques, notamment le fait que les données peuvent être stockées dans des pays ayant moins de protection légale et peuvent être surveillées par d'autres gouvernements (Sivan et Zukarnain, 2024).

Par ailleurs, parce que certaines données peuvent comporter des biais, leur utilisation peut modifier les résultats du système. Un biais est un « préjugé en faveur ou contre une chose, une personne ou un groupe par rapport à un autre généralement d'une manière qui est considérée comme injuste » (Langlois, 2023). Dans le domaine de l'IA, une attention particulière devrait être portée à la prévention des biais pouvant renforcer les préjugés ou les inégalités. Ces biais peuvent être fondés sur :

[l]a race, la couleur, le sexe, l'identité ou l'expression de genre, la grossesse, l'orientation sexuelle, l'état civil, l'âge sauf dans la mesure prévue par la loi, la religion, les convictions politiques, la langue, l'origine ethnique ou nationale, la condition sociale, le handicap ou l'utilisation d'un moyen pour pallier ce handicap (art. 10 de la Charte des droits et libertés de la personne).

Parmi les raisons de ces possibles biais, on trouve des données qui sont erronées, biaisées ou pas assez représentatives, un manque de compréhension du contexte culturel, une utilisation inadéquate des résultats et un manque de contrôle par l'humain (UNICEF, 2021). Sans mesures préventives, ces biais peuvent toucher durablement le personnel scolaire et les élèves ainsi que limiter leurs possibilités et leur développement en les profilant (UNICEF, 2021). Dans de tels cas, il est recommandé d'éviter, lorsque cela est possible, de recourir à des données erronées, biaisées ou pas assez représentatives, ou de sensibiliser les usagères et usagers à ces possibles biais.

Un autre aspect à prendre en considération est la représentativité des données, c'est-à-dire le fait qu'elles représentent l'ensemble des groupes concernés (UNESCO, 2024). En effet, tout système d'IA a été entraîné à partir de données spécifiques. Or, il peut arriver que ces données ne permettent pas d'offrir un portrait global de l'ensemble des groupes concernés. Par exemple, un système d'IA pourrait être entraîné à partir de données d'une école située en ville ou en milieu favorisé. Si ce système est utilisé dans une école en milieu rural ou défavorisé, les résultats risquent de ne pas être fiables (3. Principe de la fiabilité et de la robustesse). Un réentraînement avec des données représentatives de ces milieux serait ici préférable.



3 Développement du système d'IA

Une fois les aspects plus généraux du système d'IA bien planifiés, il est possible que votre organisation envisage de concevoir son propre système d'IA. Dans un tel cas, plusieurs choix devront être faits afin d'assurer un développement du système d'IA de façon éthique. Il est à noter que cette section ne présente que les étapes ayant des implications éthiques validées par la recherche par lesquelles passent les équipes de développement dans la conception du système d'IA. Elle ne constitue pas une synthèse de l'ensemble des étapes de développement de l'IA. De plus, ces mesures éthiques sur le plan de la conception permettent, certes, de respecter plusieurs principes éthiques, mais ne sont pas suffisantes. Les mesures externes au système d'IA présentées dans les autres phases de ce document sont tout aussi importantes à prendre en considération pour éviter le recours à un solutionnisme technologique, c'est-à-dire l'utilisation exclusive de solutions technologiques pour prévenir des enjeux éthiques (Farnadi *et al.*, 2023). À noter que les équipes responsables de la conception du système d'IA doivent détenir non seulement les compétences techniques, mais aussi les compétences pédagogiques pour prendre en compte les particularités du contexte scolaire québécois (voir la phase 4, « Mise en œuvre du système d'IA »).

Voici les principaux éléments à prendre en considération :

- Fournisseur externe
- Méthodes de débiaisage et d'équité
- Méthodes d'explicabilité
- Design du système d'IA
- Mesures pour le coût énergétique des algorithmes

3.1 Fournisseur externe

Questions réflexives

- Prévoyez-vous faire affaire avec un fournisseur externe pour la conception de votre système d'IA?
- Est-ce possible de faire affaire avec un fournisseur externe local spécialisé en éducation ou dont les équipes sont diversifiées?
- Quelles questions pouvez-vous poser au fournisseur externe pour vérifier le respect des principes éthiques lors de la conception du système d'IA?



Dans le cas où il est prévu de concevoir un système d'IA, si vous n'envisagez pas de le faire à l'interne, il est recommandé de choisir, dans la mesure du possible, une entreprise régionale ou provinciale spécialisée dans le développement de solutions technologiques en éducation. Dans un premier temps, le choix d'une entreprise locale permet de recevoir un service mieux adapté au contexte particulier de votre région ou de votre province, d'encourager une économie locale et de décentraliser les pouvoirs des entreprises du numérique (UNICEF, 2021). Dans un deuxième temps, une entreprise spécialisée en éducation permet un meilleur respect des méthodes pédagogiques validées par la recherche, du contexte scolaire et des orientations éducatives, soutenant ainsi mieux le bien-être des acteurs éducatifs du milieu (5. Principe de l'efficacité, de l'efficacit  et de la pertinence). Dans un troisi me temps, pour pr venir les risques du syst me d'IA li s   l' quit ,   la diversit  et   l'inclusion, il est pr f rable de v rifier si la composition de l' quipe charg e du d veloppement du syst me d'IA est diversifi e et sp cialis e sur le plan de la pr vention de biais en IA (5. Principe de l'efficacit , de l'efficacit  et de la pertinence).

Les fournisseurs ont plusieurs responsabilit s : garantir une responsabilit  humaine ainsi que des donn es et des mod les fiables, s rs et s curitaires; produire du contenu non discriminatoire et indiqu  lorsqu'il est g n r  par de l'IA g n rative; rendre les mod les d'IA explicables et transparents; mitiger les risques et indiquer les limites; mettre en place des m canismes de r clamation et de recours; contr ler et signaler lors d'une utilisation illicite (UNESCO, 2024, p. 22-23).   noter  galement que la conception d'un syst me d'IA demande des ajustements et ne peut  tre r alis e en un seul temps. Il est donc important de s'assurer qu'un service de soutien est garanti par le fournisseur d'IA apr s une p riode de deux ans apr s la conception du syst me d'IA. Durant cette p riode, il est toutefois important d'assurer un transfert de connaissances de fa on locale afin de rendre l'organisation ind pendante dans son utilisation du syst me d'IA.

Pour soutenir une collaboration transparente entre votre organisme et le fournisseur externe (7. Principe de la transparence), Gautrais *et al.* (2023) proposent une politique d'approvisionnement d'IA dont le niveau de partage d'information est suffisant pour faciliter des  valuations compl tes. Cette politique pourra  tre transmise au fournisseur pour faciliter le processus d' valuation entre les deux parties.

Questions   poser au fournisseur externe :

- Les donn es comportent-elles des biais et quels nouveaux pr judices ces derniers peuvent-ils entra ner?
- Est-ce que les donn es sont stock es localement? Si non, comment g rer ce d calage g ographique?
- Dans quelle mesure l'architecture de l'algorithme peut-elle  tre transparente?
- Le syst me est-il robuste? Est-il fiable?
- Est-ce que les m thodes ou le fonctionnement du syst me d'IA sont durables?
- Comment les donn es seront-elles conserv es? Quand et comment seront-elles d truites lorsqu'elles ne seront plus n cessaires?

3.2 Méthodes de débiaisage et d'équité

Questions réflexives

- Est-ce que les données du système d'IA comportent des biais?
- Est-il souhaitable d'intégrer des méthodes de débiaisage et d'équité pour minimiser les biais qui pourraient se trouver dans les résultats?



Dans le cas où des données utilisées par le système d'IA seraient biaisées, des mesures peuvent ainsi être prises en amont, notamment par la gestion des données (voir la phase 2, « Gestion éthique des données »), mais aussi par le recours à des méthodes de débiaisage en IA afin de garantir un traitement équitable des acteurs éducatifs (2. Principe de l'inclusion et de l'équité). Dans tous les cas, le choix des méthodes de débiaisage doit se faire avec prudence, puisque certaines d'entre elles peuvent amplifier des biais ou en minimiser d'autres, ce qui peut faire disparaître les cas plus rares et ainsi être discriminatoire envers certains groupes concernés. À noter qu'une ressource présentant les avantages et les désavantages des différentes méthodes de débiaisage pour faciliter la sélection des bonnes méthodes est en cours de développement.

Pour aider à déterminer les méthodes d'IA nécessaires pour prévenir les biais discriminatoires et soutenir une équité entre chacun des groupes concernés, il peut être pertinent que l'organisme établisse sa propre définition de « discrimination » et d'« équité » (Gautrais *et al.*, 2023). Pour établir ces définitions, il est préférable d'impliquer les parties prenantes et les personnes de ces parties qui risquent le plus de subir de la discrimination.

En revanche, il importe de rappeler que l'inclusion d'une variable discriminatoire dans le système d'IA ne rend pas nécessairement celui-ci discriminatoire. Au contraire, inclure cette variable peut même permettre de compenser le biais au moment de la prédiction. En d'autres mots, il peut être préférable, parfois, de conserver une variable discriminatoire plutôt que d'éviter de l'utiliser, pour mieux cerner des biais par le système et intervenir par la suite. Par exemple, dans le cas du système d'embauche d'Amazon, l'entreprise a choisi de supprimer l'indicateur relatif au sexe pour mesurer la discrimination, sans comprendre comment ce biais pouvait modifier d'autres variables, ce qui a rendu difficile la détection du fait que le modèle utilisait les bourses et activités destinées aux femmes comme variable importante pour ses choix (Dastin, 2022). La première étape nécessaire est de mesurer et d'identifier autant de sources de biais que possible, pour ensuite être capable de commencer la discussion pour les mitiger. Très souvent, il est impossible d'éliminer complètement les biais, puisque certains sont en opposition, notamment l'équité entre individus et l'équité entre groupes.

Un autre aspect important en éducation à prendre en considération par rapport à d'autres domaines d'application de l'IA est que les valeurs « aberrantes » (*outliers*) concernent de vraies personnes dont il faut tenir compte et qui sont potentiellement à risque (Chawla, 2023). Une valeur aberrante est toute erreur qui se trouve dans les données, comme une virgule mal placée dans un nombre. Les techniques usuelles de minimisation des données utilisées pour cibler des valeurs aberrantes pourraient créer de la discrimination par rapport à certaines personnes et sont donc à utiliser avec vigilance. Ces valeurs aberrantes pourraient présenter des défis dans l'analyse et la prédiction. Il serait donc important de mettre en place des procédures visant à les repérer et d'y apporter une gestion appropriée afin de pouvoir maintenir une intégrité des analyses statistiques, le but étant d'améliorer la fiabilité des conclusions à tirer.

Dans cette perspective, une ressource comme **Fairlearn** peut soutenir cette vigilance en proposant des outils permettant d'évaluer les écarts de performance d'un modèle entre différents groupes, d'identifier les risques de préjudices discriminatoires et d'explorer des méthodes d'atténuation lorsque certaines catégories d'apprenants sont désavantagées par les données ou par les prédictions produites.

Ressource :

- [Fairlearn](#)

3.3 Méthodes d'explicabilité

Questions réflexives

- Est-ce que le système d'IA utilise un grand nombre de données? Leur traitement est-il complexe au point où les usagères et usagers ne comprennent pas comment les résultats ont été obtenus?
- Est-il souhaitable d'intégrer des méthodes d'explicabilité pour rendre plus compréhensibles les données utilisées ou les résultats obtenus?



L'explicabilité permet de préciser pourquoi l'algorithme a pris telle ou telle décision. D'un point de vue éthique, en rendant un système d'IA « explicable », l'usagère ou l'utilisateur peut mieux comprendre comment le système a pu recommander une action ou prendre une décision, ou encore quelles données ont été utilisées et selon quelles proportions (8. Principe de l'explicabilité). L'interprétabilité se rapporte à la connaissance de la logique interne du système d'IA, tandis que l'explicabilité se concentre sur le « pourquoi ». L'interprétabilité permet de préciser le « comment » en identifiant les données, les variables et les moyens utilisés pour que l'algorithme prenne une décision. L'explicabilité et l'interprétabilité permettent ainsi à l'usagère ou à l'utilisateur de prendre des décisions mieux éclairées (équité) et de se sentir plus en confiance quant à son utilisation (Khosravi *et al.*, 2022). Ces informations doivent permettre à la personne concernée de contester une décision, si elle le juge nécessaire (Gautrais *et al.*, 2023).

Il existe aussi une différence entre les modèles « boîte noire » (*black box*), qui sont difficilement interprétables comme des modèles d'apprentissage automatique (*machine learning*) et qui nécessitent des méthodes d'explicabilité et d'interprétabilité, et les modèles de « boîte claire » (*glass box*), qui sont parfaitement interprétables. Elles sont donc préférables à des méthodes d'IA plus complexes dans la mesure du possible. Dans le cas où un modèle « boîte noire » est préféré parce qu'il est plus robuste, il devient donc nécessaire d'intégrer des méthodes d'explicabilité et d'interprétabilité. Cependant, il importe de noter que ce type de modèle ne donne qu'une approximation de la façon dont le modèle a réalisé ses analyses, ses prédictions ou ses décisions. En d'autres mots, il s'agira tout simplement de jouer aux équilibristes (un compromis entre la performance du modèle et l'interprétabilité ou l'explicabilité), car dans la pratique le choix d'une catégorie de modèles dépendra des objectifs spécifiques à atteindre.

Il est recommandé d'adopter une politique interne d'explicabilité incluant une définition du terme et précisant quand une explication de procédures techniques est pertinente (Gautrais *et al.*, 2023). À partir de cette politique, il importe d'inclure une ou plusieurs méthodes d'explicabilité et d'interprétabilité adaptées au système d'IA, au contexte, à son objectif et aux particularités des parties prenantes qui pourront accéder à ces explications et interprétations. De plus, il est préférable que la ou les méthodes choisies soient les plus simples, interprétables et explicables à l'aide de « techniques interactives et de visualisation » (Gautrais *et al.*, 2023, p. 34). Parmi les informations à communiquer, on trouve « le développement des fonctionnalités, des attributs et des modèles du système d'IA » (Gautrais *et al.*, 2023, p. 34). Pour orienter ce choix, Khosravi *et al.* (2022) ainsi que Biecek et Burzykowski (2020) présentent des cadres pour guider le choix de la ou des méthodes d'explicabilité et d'interprétabilité en fonction de ces divers facteurs. Dans tous les cas, aucune méthode ne garantit une explicabilité et une interprétabilité sans faille. Il vaut donc la peine de documenter les bénéfices et les risques du ou des modèles d'explicabilité et d'interprétabilité choisis (Khosravi *et al.*, 2022). Dans le cas où une équipe de développement externe à votre organisation conçoit le système d'IA, il lui est recommandé de demander de fournir les méthodes d'explicabilité et d'interprétabilité utilisées et de justifier ce choix (IEAIE, 2020).

En soutien avec les méthodes d'explicabilité, Gautrais *et al.* (2023) recommandent l'adoption d'une politique interne sur l'explicabilité incluant une définition de l'explicabilité ainsi que les moments opportuns pour expliquer les procédures et les résultats à la personne concernée.

Enfin, certains algorithmes simples parfaitement interprétables, comme la régression linéaire ou les arbres décisionnels, sont préférables à d'autres, plus complexes, même si ces algorithmes simples obtiennent de moins bons résultats que les algorithmes plus complexes.

3.4 Design du système d'IA

Questions réflexives

- Quelles sont les fonctionnalités du système d'IA?
- Comment sera son interface?
- Est-ce que son design est personnalisable pour chacune des usagères ou chacun des usagers?
- Est-ce que son design peut soutenir la capacité des usagères et usagers à prendre des décisions par eux-mêmes?
- Ce design permet-il de soutenir la pratique enseignante du personnel enseignant ou les apprentissages des élèves?
- Est-ce que des moyens complémentaires à ce design sont nécessaires pour soutenir l'enseignement et les apprentissages?



Tout ce qui touche le design du système, tel que les fonctionnalités, l'interface et la façon dont les résultats sont présentés, regroupe des aspects qui doivent être pris en considération pour assurer un développement éthique.

Entre autres, il peut arriver que le design du système d'IA puisse amener l'usagère ou l'utilisateur à se méfier des résultats ou à trop y faire confiance, ce qui dans les deux cas peut diminuer considérablement son autonomie à prendre ses propres décisions (10. Principe de la compétence). D'une part, certains usagers et usagères peuvent croire que les résultats sont d'origine humaine et y faire plus confiance. Pour éviter ce piège, il est important d'insister dans l'interface sur le fait que les résultats du système proviennent de l'IA. Même chose lorsque les résultats sont présentés à l'aide d'un personnage dans l'interface. Ce personnage doit avoir des attributs technologiques pour éviter toute confusion (UNICEF, 2021). D'autre part, d'autres pourraient croire à tort que, parce que les résultats sont d'origine technologique, et non humaine, les systèmes d'IA sont plus fiables puisque la technologie fait moins d'erreurs (Vincent-Lancrin et Van der Vlies, 2020). Or, au contraire, tout système d'IA se base sur des données d'origine humaine qui peuvent être biaisées. Pour prévenir une prise de décision humaine renforçant des biais négatifs (2. Principe de l'inclusion et de l'équité), un outil permettant une analyse critique de la décision par l'usagère ou l'utilisateur, comme une grille d'analyse comportant des questions réflexives (ex. : Est-ce que la décision semble discriminatoire envers des élèves? Est-ce que les informations relayées sont en conformité avec les politiques éducatives et des méthodes pédagogiques probantes?) est un moyen pour permettre un usage éclairé du système d'IA.

Tout système d'IA en éducation doit valoriser l'enseignement du personnel enseignant et les apprentissages des élèves (5. Principe de l'efficacité, de l'efficacité et de la pertinence). Pour ce faire, la présentation des résultats peut avoir un rôle majeur à cet effet, que ce soit en donnant un contexte au profil d'une élève identifiée comme étant à risque ou encore en fournissant des rétroactions adressées aux élèves au personnel enseignant pour qu'il puisse les présenter d'une façon plus adaptée au profil de l'élève et à partir de ses propres compétences professionnelles. Pour soutenir l'autonomie des usagers et usagers, il est aussi recommandé d'offrir une option de personnalisation de l'interface en fonction des besoins de ceux-ci et des particularités de leur environnement (5. Principe de l'efficacité, de l'efficacité et de la pertinence).

3.5 Mesures pour le coût énergétique des algorithmes

Questions réflexives

- Est-ce qu'un outil a été utilisé pour évaluer le coût énergétique et l'empreinte carbone du système d'IA?
- Comment sera son interface?
- Est-ce que ce coût énergétique et cette empreinte du système d'IA sont élevés?
- Est-ce que des mesures ont été mises en place pour réduire le coût énergétique et l'empreinte carbone du système d'IA?



Pour fonctionner, un système d'IA produit des algorithmes qui demandent une certaine quantité d'énergie (Luccioni *et al.*, 2023). De plus, l'IA fonctionne généralement à l'aide d'une grande quantité de données hébergées dans des centres de données physiques. Ce genre d'installation peut représenter un coût énergétique et une empreinte carbone importants. C'est pourquoi il est recommandé d'utiliser des mesures pour évaluer et réduire le coût énergétique et l'empreinte carbone causés par l'utilisation de données et d'algorithmes (6. Principe de la durabilité). Parmi ces mesures, il est possible d'utiliser des outils pour comparer l'empreinte carbone créée par différents modèles pour choisir celui ayant le moins grand impact environnemental ou de s'assurer d'utiliser des énergies renouvelables pour faire fonctionner le système d'IA. Un autre moyen recommandé pour diminuer l'empreinte carbone d'un système d'IA serait de le compresser (Rafat *et al.*, 2019). Dans le cas de la Solution d'IA pour prévenir le décrochage scolaire, l'empreinte carbone n'a pas été mesurée à l'heure actuelle, bien que des mesures aient été mises en place pour minimiser son impact environnemental et sensibiliser les usagers et usagers à l'importance de conserver une sobriété numérique lors de son utilisation.

Ressource :

- [Mitigating carbon footprint for knowledge distillation based deep learning model compression](#)



4 Mise en œuvre

Maintenant que le système d'IA est conçu et évalué, il est prêt à être utilisé. Toutefois, pour s'assurer que son utilisation se fait de façon éthique, que divers risques sont prévenus et que cette technologie est acceptée de manière générale, plusieurs mesures doivent être prises sur le plan de sa mise en œuvre. Il s'agit ici de veiller à ce que l'ensemble des parties prenantes impliquées directement ou indirectement dans le système d'IA aient les compétences nécessaires lors de sa mise en œuvre, c'est-à-dire qu'elles puissent le comprendre, l'utiliser de façon éthique et savoir quand réagir lorsque le système pose un risque (ensemble des principes). La formation au développement et à l'utilisation éthiques de l'IA ainsi que la campagne d'information et de sensibilisation représentent deux actions clés pour justement encapaciter l'ensemble de ces parties prenantes.

Lors de cette phase, il est possible que les données utilisées pour entraîner le système d'IA dévoilent des aspects des données personnelles.

Voici les mesures à adopter :

- Mesures de pré-déploiement
- Essai pilote
- Formation au développement et à l'utilisation éthiques de l'IA
- Accessibilité
- Milieu inclusif et sécuritaire
- Campagne d'information et de sensibilisation
- Transfert des connaissances

4.1 Mesures de pré-déploiement

Questions réflexives

- Est-ce qu'il est prévu que les résultats du système d'IA soient validés par un humain?
- Est-ce qu'un processus sera mis en place pour évaluer les erreurs et expliquer leurs origines, et pour qu'une ou un membre du personnel de votre organisation soit en mesure de les corriger?
- Est-ce qu'un système de signalement de problèmes liés à la discrimination ou à la mauvaise performance du système d'IA est prévu?
- Est-ce que la production d'un avis écrit de la part de la sécurité de l'information est prévue pour démontrer la sécurité du système d'IA?



Une fois le système d'IA prêt à être déployé, quelques mesures supplémentaires doivent être adoptées pour assurer, même à cette étape, une utilisation responsable. Parmi les principaux enjeux à garder en tête, on trouve l'importance de valider les résultats par un humain. De plus, pour éviter que des erreurs de faux négatifs ou de faux positifs ne surviennent, il importe, lors du déploiement, de s'appuyer sur un processus apte à repérer les erreurs et à expliquer leurs origines, et de désigner une personne responsable de corriger la décision rapidement dans un tel cas (9. Principe de la responsabilité). Pour prévenir des risques liés à la discrimination (2. Principe de l'inclusion et de l'équité), Gautrais *et al.* (2023) proposent aussi d'adopter « un système de signalement de problèmes liés à la discrimination ou à la mauvaise performance du système d'IA » (p. 45). Enfin, en matière de sécurité, l'organisation peut demander un avis écrit de la part de la sécurité de l'information pour son système d'IA « et en faisant la démonstration que les mesures de sécurité en place sont adéquates selon les risques encourus » (3. Principe de la fiabilité et de la robustesse) (MCN, 2024, p. 6).

4.2 Essai pilote

Questions réflexives

- Est-ce qu'un essai pilote du système d'IA est prévu avant qu'il soit mis en œuvre à plus grande échelle?
- Est-ce que des modifications doivent être apportées à la suite de l'essai pilote?



Malgré toutes les précautions prises pour prévenir les risques inhérents à un système d'IA et maximiser son potentiel, plusieurs aspects liés à son implantation peuvent être difficiles à anticiper (ensemble des principes). C'est pour cette raison qu'il peut être pertinent, selon le niveau de risque du système d'IA, de réaliser un essai pilote et d'apporter les modifications nécessaires pour faciliter sa mise en œuvre à plus grande échelle.

4.3 Formation au développement et à l'utilisation éthiques de l'IA

Questions réflexives

- Quels usagers et usagères devront être formés pour utiliser le système d'IA?
- Quels sujets devront être abordés durant leur formation?
- Quels sont les moyens les plus efficaces pour le faire?



Pour permettre un développement et une utilisation responsables de l'IA au sein de votre organisation, des formations devront être données auprès des groupes ciblés. Ces groupes sont les équipes de développement en IA engagées par l'organisation, les expertes et experts en science des données, les directions, le personnel administratif, le personnel scolaire et les élèves. Même si les élèves ne représentent pas un groupe habituel d'usagères et usagers du système d'IA, il importe de les former au numérique et à l'IA afin de leur donner les moyens d'en faire une utilisation éthique en tant que futurs citoyens et citoyennes, usagères et usagers, développeurs et développeuses ou dirigeants et dirigeantes (UNICEF, 2021).

La formation représente une pierre angulaire de l'implantation d'une culture de développement et d'utilisation éthiques de l'IA dans votre organisme. Elle devrait favoriser une utilisation de l'IA « effective, [réalisée] avec discernement et avec confiance », pour permettre au groupe ciblé d'être capable d'examiner les décisions du système d'IA et de surveiller les risques (IEAIE, 2020). La formation soutient également la littératie numérique des parties prenantes, laquelle leur permet de déconstruire des idées erronées sur l'IA, d'en comprendre les avantages et les risques, et d'adopter des stratégies pour mieux se protéger contre ces derniers (10. Principe de la compétence). Une littératie pédagogique pourra aussi être inculquée aux groupes moins familiers avec le milieu scolaire (équipes de développement et expertes et experts en science des données) afin de respecter le droit des enfants (2. Principe de l'inclusion et de l'équité) et l'autonomie professionnelle du personnel scolaire (5. Principe de l'efficacité, de l'efficacéité et de la pertinence). De manière générale, les usagères et usagers d'IA devraient être sensibilisés aux termes d'utilisation du système d'IA, être formés à son utilisation éthique et être renseignés quant aux moyens d'en contrôler et d'en signaler des applications inadéquates (UNESCO, 2024).

Pour éviter la duplication des formations sur le sujet, il vaut la peine, avant de les concevoir, de vérifier s’il en existe déjà qui répondent aux préoccupations et aux besoins ciblés de chaque partie prenante. Au Québec, il existe plusieurs organismes qui offrent des ressources informatives sur l’IA, des formations pour développer la compétence numérique, de l’accompagnement personnalisé auprès du personnel enseignant ainsi que des entreprises mettant au point des solutions technologiques en éducation. Une fois ces formations sélectionnées, il sera possible de les bonifier avec d’autres qui pourront être créées et présentées selon les besoins, les préoccupations et les préférences de chaque groupe ciblé, tout en considérant les ressources financières et humaines disponibles au sein de votre organisme.

Fait intéressant à souligner, chaque sujet abordé en formation s’appuie sur l’expertise de certaines parties prenantes (ex. : méthodes pédagogiques probantes par le personnel scolaire; fondements et utilisations de l’IA par les équipes de développement et les expertes et experts en science des données; droit des enfants par les élèves et leurs parents). La formation à un développement et à une utilisation éthique de l’IA représente donc une occasion d’impliquer chacune des parties prenantes dans l’élaboration des contenus de formation.

L’organisation peut ensuite démontrer « que les membres de son personnel ont suivi une formation adéquate les sensibilisant aux enjeux relativement à l’intégration de l’intelligence artificielle dans l’administration publique » (MCN, 2024, p. 12).

Ressources :

- [Matériel pour le personnel enseignant](#)
- [Idées de discussion pour les parents](#)
- [Conseils pour les jeunes du primaire](#)
- [Conseils pour les jeunes du secondaire](#)

4.4 Accessibilité

Question réflexive

- Est-ce que les usagères et usagers ont accès aux infrastructures et à la connexion Internet nécessaires pour utiliser le système d’IA?



Pour que chaque usagère ou usager de votre organisation puisse bénéficier des services découlant de l'utilisation de ce système d'IA, elle ou il doit avoir accès à l'équipement informatique nécessaire (2. Principe de l'inclusion et de l'équité). Du côté des établissements scolaires, un accès équitable à l'équipement informatique permet une utilisation adéquate du système d'IA (5. Principe de l'efficacité, de l'efficacit  et de la pertinence). Le cycle de vie de l' quipement informatique doit  galement  tre planifi  dans une optique de respect de l'environnement, que ce soit pour prolonger sa dur e de vie en le maintenant et en le r parant, ou en pr voyant de le recycler ad quatement en fin de vie (5. Principe de l'efficacit , de l'efficacit  et de la pertinence).

Pour que tous les  tablissements scolaires puissent acc der au syst me d'IA et puissent l'utiliser ad quatement, une connexion Internet de qualit  et s curitaire est requise (4. Principe de la s curit ). L'installation de logiciels antivirus et de tout autre logiciel visant la protection des donn es permet, quant   elle, de pr venir des incidents de s curit  (SRIDAIL, 2018).

4.5 Milieu inclusif et s curitaire

Questions r flexives

- Est-ce que le syst me d'IA augmente le risque d'imposer une culture de surveillance dans les  tablissements scolaires?
- Quelles mesures peuvent  tre mises en place pour valoriser l'autonomie des  l ves et celle du personnel scolaire?



Malgr  les bienfaits propres   l'IA, plusieurs chercheuses et chercheurs s'inqui tent que son usage instaure une culture de surveillance au sein des  tablissements scolaires, ressentie autant par les  l ves que par le personnel scolaire et administratif (IEAIE, 2021; UNICEF, 2021; UNESCO, 2019). Une telle culture pourrait nuire   l'autonomie professionnelle du personnel scolaire et entra ner chez l' l ve la crainte de commettre des erreurs, ce qui serait nuisible   l'apprentissage (2. Principe de l'inclusion et de l' quit ; 5. Principe de l'efficacit , de l'efficacit  et de la pertinence). Pour pr venir une telle situation, il est recommand  de soutenir l'autonomie professionnelle du personnel parall lement   l'implantation du syst me d'IA et de favoriser des espaces d'apprentissage o  le droit   l'erreur est garanti pour l' l ve (ex. :  valuation formative, tutorat, exercisation, ateliers).

4.6 Campagne d'information et de sensibilisation

Questions réflexives

- Qui sont les personnes concernées directement et indirectement par le système d'IA?
- Comment seront-elles informées du système d'IA?
- Quelles informations devront leur être transmises?
- Quels moyens ces personnes auront-elles pour communiquer avec votre organisation afin de se renseigner davantage sur le système d'IA?



Face aux risques de l'IA, les établissements ont aussi la responsabilité d'informer les parties prenantes touchées indirectement par l'IA en éducation, soit le public, les élèves, les parents et le personnel scolaire (7. Principe de la transparence). En effet, l'un des dangers de l'IA est que les personnes s'en méfient irrationnellement, ce qui les empêcherait de retirer des avantages de cette technologie pour soutenir le développement de l'élève et du personnel scolaire, ou qu'elles l'adoptent trop rapidement, ce qui peut mener à des risques liés à une mauvaise utilisation (Dignum, 2019). Pour contrer ces craintes et ces risques, une campagne d'information peut être réalisée par chaque CSS ou CS une fois le système d'IA prêt à être mis en œuvre. Une telle campagne permet aussi de sensibiliser ces groupes aux enjeux liés à la vie privée en IA, à leurs droits et aux moyens de les faire respecter (1. Principe du respect des personnes et de la règle de droit) (UNICEF, 2021).

Une campagne d'information se distingue d'une formation de différentes façons. D'une part, la formation concerne les parties prenantes directement touchées par le système d'IA et permet un développement de compétences en profondeur, de façon continue et active. D'autre part, la campagne d'information concerne les parties prenantes indirectement touchées par le système, présente surtout des informations générales à ce sujet et se fait de façon moins récurrente et active.

Les moyens de communication pour joindre chaque partie prenante pourront être sélectionnés, et les informations adaptées en fonction des ressources et du temps limités des parties prenantes et de leur niveau de littératie numérique (UNICEF, 2021). Par exemple, le public pourrait être joint par les médias, tandis que les élèves, les parents et le personnel scolaire peuvent être informés lors des rencontres ponctuelles dans les écoles. Tous pourraient aussi être informés au moyen d'outils en ligne. Un autre moyen pertinent pour communiquer avec les élèves et les parents serait de mettre en place un site Web à visiter, qui présenterait « les informations les concernant, incluant les détails des informations et les sources utilisées pour créer le profil » (Gautrais *et al.*, 2023, p. 25).

Un visuel peut aussi être prévu pour « confirmer aux usagers finaux que le service qu'ils reçoivent est généré par un système d'IA » (MCN, 2024, p. 9), de façon à conserver une transparence et une relation de confiance avec l'utilisateur ou l'usager (7. Principe de la transparence). À noter que, si ce service résulte

d'un processus décisionnel complètement automatisé, l'organisme a l'obligation d'en informer la clientèle et de lui présenter les recours offerts en cas de désaccord ou d'erreur (voir [Décision fondée exclusivement sur un traitement automatisé](#), Gouvernement du Québec, 2024).

4.7 Transfert des connaissances

Questions réflexives

- Prévoyez-vous partager votre expertise en développement de systèmes d'IA avec d'autres organisations?
- Quels moyens envisagez-vous pour y arriver?
- Quelles informations jugez-vous qu'il soit nécessaire de partager avec elles à ce sujet?



En raison des risques liés à l'IA, un travail en vase clos est à proscrire. Les organismes responsables du développement de l'IA et les organismes scolaires ont tous un rôle à jouer pour partager leurs expertises respectives et assurer la mise en place d'une IA au service de l'humain, au Québec comme à l'international. En revanche, pour partager cette expertise, il faut prendre en compte diverses limites, telles que la propriété intellectuelle, la protection des données personnelles et la cybersécurité. Ces limites peuvent être atténuées grâce à divers moyens, comme le fait de « partager des données en utilisant une fiducie de données » ou de « publier les métadonnées » du système d'IA (Gautrais *et al.*, 2023, p. 35).

Au Québec, les possibilités sont nombreuses, d'autant plus que le gouvernement cherche à adopter une culture d'ouverture et de partage en matière de données et d'IA (Gouvernement du Québec, 2019). Les jeux de données peuvent être partagés avec la plateforme Données Québec (avec bien sûr toutes les précautions propres à l'anonymisation des données). L'expertise développée par l'organisme peut être partagée dans le cadre de symposiums, de colloques et de congrès. À l'international, il est possible de documenter la démarche, de la publier et de la transmettre aux responsables de l'IA en éducation de l'UNICEF (2021), qui sont fort désireux de cartographier les initiatives de la sorte dans le monde.

Un tel partage, tant au Québec que dans le reste du monde, facilite une meilleure reddition de comptes (9. Principe de la responsabilité), favorise un échange des meilleures pratiques pour une IA robuste (3. Principe de la fiabilité et de la robustesse), équitable (2. Principe de l'inclusion et de l'équité) et respectant les droits de toutes les parties prenantes (1. Principe du respect des personnes et de la règle de droit). Le partage public de la démarche, tant sur le plan des bons coups que sur celui des moins réussis, permet également une plus grande transparence auprès des parties prenantes, ce qui soutient leur confiance à l'égard du système d'IA (7. Principe de la transparence).

Ressource :

- [Lignes directrices sur la diffusion de données ouvertes](#)



5 Maintenance et amélioration

Une fois le système d'IA déployé, il est recommandé d'assurer sa maintenance et, si requis, de lui apporter les améliorations nécessaires. Cette étape implique autant les équipes d'assistance technique et les analystes de données que la personne responsable de l'IA, qui doivent collaborer à assurer le bon fonctionnement du système d'IA jusqu'à la fin de son cycle de vie, par les étapes suivantes :

- Mécanismes de suivi
- Veille du système d'IA
- Mises à jour régulières

5.1 Mécanismes de suivi

Question réflexive

- Quels mécanismes de suivi sont prévus pour évaluer l'utilisation et les impacts réels du système d'IA au sein de votre organisation?



À la suite de la mise en œuvre du système d'IA, il faut veiller au suivi de son utilisation et à son adaptation à la lumière de ses impacts réels (ensemble des principes). Pour ce faire, votre organisation doit prévoir des mécanismes de suivi du système, comme une évaluation périodique, « des mécanismes dès l'implantation du système d'IA, notamment des processus d'évaluation périodique, des mécanismes de consultation et de rétroaction des utilisateurs, des bénéficiaires et de la population ainsi que des instances chargées d'assurer l'ajustement continu du système d'IA et des mesures de mitigation et de capacitation au sein de l'OP, [...] des mécanismes appropriés de répartition des responsabilités et de reddition de comptes quant à l'utilisation du système d'IA » (MCN, 2024, p. 15-16).

5.2 Veille du système d'IA

Questions réflexives

- Est-ce qu'une veille du système d'IA est prévue?
- Qui aura cette responsabilité?



Une veille de ce système doit être prévue (monitorage) sur une base régulière afin d'assurer sa robustesse (3. Principe de la fiabilité et de la robustesse) et d'éviter des risques associés à la sécurité des données (4. Principe de la sécurité). Tandis qu'un système informatique qui n'utilise pas d'IA continuera de fonctionner de la même façon une fois développé, un système d'IA peut changer en cours de fonctionnement selon les nouvelles données à partir desquelles il fait une prédiction ou une recommandation, ou prend une décision. Le monitoring du système permet d'assurer sa robustesse, c'est-à-dire qu'en cas de changements dans l'environnement qui se reflètent dans les données, des humains peuvent vérifier si les résultats obtenus sont les plus représentatifs du contexte.

5.3 Mises à jour régulières

Questions réflexives

- Est-ce que des mises à jour du système d'IA sont prévues? À quelle fréquence? Pour quelles raisons cela serait-il nécessaire?
- Qui aura cette responsabilité?



Tout au long du cycle de vie du système d'IA, ce dernier doit être souvent révisé, puis mis à jour, y compris les données nécessaires (3. Principe de la fiabilité et de la robustesse; 4. Principe de la sécurité). Il est aussi possible d'« effectuer les mises à jour avec de nouvelles données d'entrées obtenues par le système d'IA déployé » (Gautrais *et al.*, 2023, p. 37).



Conclusion

Ce document d'encadrement éthique a pour objectif de soutenir les acteurs du milieu scolaire vers un développement et une utilisation éthique de l'IA. À partir de la littérature sur l'éthique de l'IA en éducation (CSE et CEST, 2024; UNESCO, 2024; UNICEF, 2021; OCDE, 2020) et des orientations stratégiques gouvernementales, nous avons voulu adapter le tout aux particularités du système d'éducation québécois. Au total, cinq phases ont été formulées à partir des dix principes d'utilisation responsable de l'IA pour les organismes publics du Québec (MCN, 2024) qui ont structuré ce document : planification, gestion éthique des données, développement, mise en œuvre ainsi que maintenance et amélioration. Il importe de rappeler que l'IA est un domaine qui évolue à grande vitesse. D'autres lois et avancées en recherche sur l'éthique de l'IA peuvent s'ajouter.

Références

- Arnold, K. E. et Sclater, N. (2017). Student perceptions of their privacy in leaning analytics applications. *Proceedings of the seventh international learning analytics & knowledge conference*, 66-69. <https://doi.org/10.1145/3027385.3027392>
- Beattie, S., Woodley, C. et Souter, K. (2014). Creepy analytics and learner data rights. *Ascilite*. 421-425. <https://research.moodle.org/files/69-Beattie.pdf>
- Berditchevskaia, A., Maliaraki, E. et Stathoulopoulos, K. (2022). A descriptive analysis of collective intelligence publications since 2000, and the emerging influence of artificial intelligence. *Collective Intelligence*, 1(1), 26339137221107924. <https://doi.org/10.1177/26339137221107924>
- Berger, Q. et Caravenna, F. (2021, 3 novembre). Le paradoxe de Simpson illustré par des données de vaccination contre le Covid-19. *The Conversation*. <https://theconversation.com/le-paradoxe-de-simpson-illustre-par-des-donnees-de-vaccination-contre-le-covid-19-170159>
- Chawla, A. (2023, 8 novembre). Are you sure you are using the train, validation and test set correctly? *Daily Dose of Data Science*. <https://www.blog.dailydoseofds.com/p/are-you-sure-you-are-using-the-train>
- Collin, S. et Hennetier, C. (2024). Gouverner l'école à l'heure de l'IA : synthèse des connaissances sur la gouvernance scolaire basée sur les données. *Médiations & médiatisations*, (18), 43-66. <https://www.erudit.org/en/journals/media/2024-n18-media09698/1114742ar/>
- Collin, S., Lepage, A. et Nebel, L. (2022, 8 décembre). Enjeux éthiques et critiques des systèmes d'IA en éducation : entre tensions et risques [Présentation]. Webinaire de l'Obvia. <https://observatoire-ia.ulaval.ca/webinaire-enjeux-ethiques-et-critiques-des-systemes-dia-en-education-entre-tensions-et-risques/>
- Collin, S. et Marceau, E. (2021). L'IA en éducation : enjeux de justice. *Formation et profession*, 29(2), 1-4. <https://doi.org/10.4000/ethiquepublique.7619>
- Commission d'accès à l'information [CAI]. (2023). *Ministères et organismes publics*. <https://www.cai.gouv.qc.ca/organismes/>
- Commission d'accès à l'information. (2022). *Mieux protéger les renseignements personnels des jeunes à l'ère numérique : rapport de la Commission d'accès à l'information présenté au ministre responsable de l'Accès à l'information et de la Protection des renseignements personnels*. https://www.cai.gouv.qc.ca/uploads/pdfs/CAI_RAP_RP_Jeunes_Res.pdf

- Commission des droits de la personne et des droits de la jeunesse du Québec. (1989). *Convention relative aux droits de l'enfant*. <https://www.cdpedj.qc.ca/fr/vos-droits/lois-qui-protegent-vos-droits/convention-relative-aux-droits-de-lenfant>
- Conseil supérieur de l'éducation [CSE] et Commission de l'éthique en science et en technologie [CEST]. (2024). *Intelligence artificielle générative en enseignement supérieur : enjeux pédagogiques et éthiques*. <https://www.cse.gouv.qc.ca/wp-content/uploads/2024/04/50-0566-RP-IA-generative-enseignement-superieur-enjeux-ethiques.pdf>
- Conseil supérieur de l'éducation. (2020). *L'intelligence artificielle en éducation : un aperçu des possibilités et des enjeux*. <https://www.cse.gouv.qc.ca/publications/intelligence-artificielle-en-education-50-2113/>
- Copibec. (2024a). *Le droit d'auteur en bref*. <https://www.copibec.ca/fr/la-loi-en-bref>
- Copibec. (2024b). *Où en sommes-nous avec l'intelligence artificielle?* <https://www.copibec.ca/fr/nouvelle/593/ou-en-sommes-nous-avec-l-intelligence-artificielle->
- Crawford, J., Cowling, M. et Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*, 20(3).
- Creemers, R. (2018). China's Social Credit System: An evolving practice of control. Available at SSRN 3175792.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer.
- Dilhac, M.-C. (2021). *Penser l'intelligence artificielle responsable : un guide de délibération*. Obvia. <https://www.obvia.ca/sites/obvia.ca/files/ressources/202103-OBV-Pub-GuideDeliberationIA.pdf>
- Farnadi, F., De Lima Alves, A. L. et Salganik, R. (2023). Le secteur de l'IA du point de vue de l'éthique. *Angles morts de l'intelligence artificielle*. 31-56. <https://unesdoc.unesco.org/ark:/48223/pf0000384801>
- Gautrais, V., Tchiriaev, A. et Guiraud, E. (2023). *Guide des bonnes pratiques en intelligence artificielle : sept principes pour une utilisation responsable des données*. Obvia. <https://www.docdroid.com/UJM3Vpn/1-guide-de-bonnes-pratiques-en-ia-fr-pdf#page=2>
- Gouvernement du Québec. (2024). *Décision fondée exclusivement sur un traitement automatisé* <https://www.quebec.ca/gouvernement/travailler-gouvernement/normes-gouvernance-pratiques-internes/protection-des-renseignements-personnels/technologie-et-droit-a-la-protection-des-renseignements-personnels/decision-traitement-automatise>

- Gouvernement du Québec. (2024). *Protection des renseignements personnels*. <https://www.quebec.ca/gouvernement/travailler-gouvernement/travailler-fonction-publique/services-employes-etat/conformite/protection-des-renseignements-personnels>
- Gouvernement du Québec. (2019). *Stratégie de transformation numérique*. https://www.tresor.gouv.qc.ca/fileadmin/PDF/ressources_informationnelles/Strategie_TNG.pdf
- Institute for Ethical AI in Education. (2021). *The Ethical Framework for AI in Education*. <https://www.buckingham.ac.uk/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-The-Ethical-Framework-for-AI-in-Education.pdf>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J. et Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074.
- Khuong, R. (2022). *Enquête concernant le Centre de services scolaire du Val-des-Cerfs (anciennement Commission scolaire du Val-des-Cerfs)*. Commission de l'accès à l'information du Québec. https://www.cai.gouv.qc.ca/uploads/pdfs/decisions-en-surveillance/1020040-S_decision-cai-centre-services-scolaire-val-cerfs-2022-11-09.pdf?v=1764365041
- Langlois, L. (2023). *Les enjeux d'équité, de diversité et d'inclusion en IA et numérique*. <https://www.obvia.ca/equite-diversite-et-inclusion>
- Langlois, L., Lawarée, J. et Marcoux, A. M. (2022). *Travaux exploratoires pour le développement d'un outil d'évaluation des impacts sociétaux de l'IA et du numérique*. Obvia. <https://www.docdroid.com/wH3zTmH/travaux-exploratoires-pour-le-developpement-dun-outil-devaluation-des-impacts-societaux-de-lia-et-du-numerique-pdf>
- Langlois, L. (2008). Une éthique à deux vitesses : dangers et répercussions sur l'identité professionnelle. *Pyramides. Revue du Centre d'études et de recherches en administration publique*, (16/2), 33-52. <https://journals.openedition.org/pyramides/210>
- Luccioni, A. S., Jernite, Y. et Strubell, E. (2023). Power hungry processing: Watts driving the cost of AI deployment? *arXiv preprint arXiv:2311.16863*. <https://doi.org/10.48550/arXiv.2311.16863>
- Luccioni, S., Gamazaychikov, B., Hooker, S., Pierrard, R., Strubell, E., Jernite, Y. et Wu, C. J. (2024). Light bulbs have energy ratings—so why can't AI chatbots? *Nature*, 632(8026), 736-738. <https://doi.org/10.1038/d41586-024-02680-3>
- Ministère de la Cybersécurité et du Numérique [MCN]. (2024). *Cadre éthique de l'intelligence artificielle du gouvernement du Québec*. https://www.publicationsduquebec.gouv.qc.ca/fileadmin/gazette/pdf_encrypte/lois_reglements/2024F/83874.pdf

- Ministère de la Cybersécurité et du Numérique. (2024). *Guide des bonnes pratiques d'utilisation de l'intelligence artificielle générative*. https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/cybersecurite_numerique/Publications/Strategie_cybersecurite_numerique_2024-2028/GU_bonnes_pratiques_utilisation_IA_generative_VF.pdf
- Ministère de l'Éducation du Québec [Ministère]. (2026). *Cadre de référence de la compétence numérique* <https://cdn-contenu.quebec.ca/cdn-contenu/education/Numerique/Cadre-reference-competece-num.pdf>
- Ministère de l'Éducation du Québec. (2024). *L'utilisation pédagogique, éthique et légale de l'intelligence artificielle générative : guide destiné au personnel enseignant*. <https://cdn-contenu.quebec.ca/cdn-contenu/education/Numerique/Guide-utilisation-pedagogique-ethique-legale-IA-personnel-enseignant.pdf>
- Ministère de l'Éducation du Québec. (2023). *Indicateurs et statistiques*. <http://www.education.gouv.qc.ca/references/indicateurs-et-statistiques/indicateurs/>
- Ministère de l'Éducation du Québec. (2022). *Évaluation de la mise en œuvre du Plan d'action numérique en éducation et en enseignement supérieur*. <https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/education/publications-adm/education/evaluations-programmes/Plan-action-numerique-education-ens-sup-rapport-evaluation.pdf>
- Ministère de l'Éducation du Québec. (2017). *La politique de réussite éducative*. https://www.education.gouv.qc.ca/fileadmin/site_web/documents/ministere/brochure_consultations_16sept.pdf
- Muldoon, J., Cant, C., Graham, M. et Ustek Spilda, F. (2023). The poverty of ethical AI: Impact sourcing and AI supply chains. *AI & SOCIETY*, 1-15. <https://doi.org/10.1007/s00146-023-01824-9>
- National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework*. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>
- Obvia. (2023). *L'éthique au cœur de l'IA*. <https://www.obvia.ca/sites/obvia.ca/files/ressources/202310-OBV-Pub-EthiqueCoeurIA.pdf>
- Office québécois de la langue française. (2024). *Donnée confidentielle*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/2074793/information-confidentielle>
- Office québécois de la langue française. (2023). *Infonuagique*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26501384/infonuagique>
- Office québécois de la langue française. (2022). *Donnée*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8358482/donnee>

- ONU. (2015). *Objectifs de développement durable*. <https://www.un.org/sustainabledevelopment/fr/objectifs-de-developpement-durable/>
- ONU-Habitat et Mila. (2022). *AI and cities: Risks, applications and governance*. https://unhabitat.org/sites/default/files/2022/10/artificial_intelligence_and_cities_risks_applications_and_governance.pdf
- Organisation de coopération et de développement économiques. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. <https://doi.org/10.1787/19939019>
- Perdomo, J. C., Britton, T., Hardt, M., & Abebe, R. (2023). *Difficult lessons on social prediction from wisconsin public schools*. arXiv preprint arXiv:2304.06205.
- Santoni de Sio, F. et Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34(4), 1057-1084. <https://link.springer.com/article/10.1007/s13347-021-00450-x>
- Sivan, R. et Zukarnain, Z. A. (2021). Security and privacy in cloud-based e-health system. *Symmetry*, 13(5), 742. <https://doi.org/10.3390/sym13050742>
- UNESCO. (2019). *Consensus de Beijing sur l'IA et l'éducation*. https://unesdoc.unesco.org/ark:/48223/pf0000368303_fre
- UNICEF. (2021). *Orientations stratégiques sur l'IA destinée aux enfants 2.0*. https://www.unicef.org/innocenti/media/1346/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021_FR.pdf
- Union européenne. (2023). *Proposition de cadre réglementaire sur l'intelligence artificielle*. <https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai>
- Université de Montréal. (2018). *À propos de la Déclaration de Montréal IA responsable*. Déclaration de Montréal IA responsable. <https://declarationmontreal-iaresponsable.com/la-declaration/>
- Voarino, N. et Régis, C. (2023). Les dilemmes dans l'angle mort du développement responsable de l'IA en temps de pandémie. *Angles morts de la gouvernance de l'IA*. 95-116. <https://unesdoc.unesco.org/ark:/48223/pf0000384801>

