GUIDE DES BONNES PRATIQUES D'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE

APPLICABLE AUX OUTILS D'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE EXTERNES

MINISTÈRE DE LA CYBERSÉCURITÉ ET DU NUMÉRIQUE

OCTOBRE 2024

Votre gouvernement

Québec

Table des matières

Liste des acronymes

CAI :	Commission d’accès à l’information
CQEN :	Centre québécois d’excellence numérique
IA :	Intelligence artificielle
MCN :	Ministère de la Cybersécurité et du Numérique
OP :	Organismes publics
PI :	Propriété intellectuelle

Public cible

Ce guide est d'abord destiné aux instances de gouvernance des organismes publics (OP) de l'administration publique québécoise dans le cadre de la gestion et du soutien du personnel dans l'utilisation d'outils numériques d'intelligence artificielle (IA) générative.

Le guide s'adresse également aux employés de l'administration publique québécoise susceptibles d'utiliser, dans le cadre de leurs activités professionnelles, des solutions d'IA générative.

Ce guide porte sur les outils d'IA générative qui sont librement accessibles à tous et offerts par un service externe, et ce, sans avoir été développés ou acquis par une organisation.

Objectifs

Les objectifs de ce guide sont de :

1. Guider les OP vers une utilisation responsable et adéquate de cette technologie;
2. Sensibiliser les employés de l'administration publique aux possibilités et aux risques de l'utilisation d'outils d'IA générative.

Les moyens pour y parvenir sont de :

1. Préconiser la prudence lors de l'utilisation de l'IA générative en général;
2. Proposer des consignes mettant l'accent sur l'acquisition des réflexes prudents.

Ce guide est un document qui sera mis à jour de manière itérative. Le guide propose aux OP et à leurs employés des éléments à prendre en considération dans l'utilisation et dans l'encadrement de l'utilisation des outils d'IA générative, plutôt que d'imposer des interdictions ou des limitations quant à l'utilisation des outils. Il s'agit de donner aux OP et aux employés les moyens de faire des choix responsables et de développer de bonnes habitudes applicables dans différents contextes, ce qui est essentiel pour utiliser correctement cette technologie en constante évolution. Au même titre que l'utilisation de tout autre outil numérique, comme les réseaux sociaux, l'utilisation des outils d'IA générative doit être responsable.

Cadre légal et réglementaire

Les bonnes pratiques de ce guide doivent toujours être prises en compte parallèlement aux lois, règlements, directives, politiques et orientations concernant l'utilisation de la technologie et des services numériques, entre autres, et sans être exhaustif :

- *Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement* (RLRQ, c. G-1.03);
- *Loi concernant le cadre juridique des technologies de l'information* (RLRQ, c. C-1.1);
- *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* (RLRQ, c. A-2.1);
- *Politique gouvernementale de cybersécurité;*
- *Directive gouvernementale sur la sécurité de l'information* (décret numéro 1514-2021 du 8 décembre 2021);
- *Exigences en matière de ressources informationnelles au regard de l'utilisation de l'intelligence artificielle par les organismes publics* (arrêté numéro 2024-01 du ministre de la Cybersécurité et du Numérique en date du 28 février 2024);
- *Énoncé de principes pour une utilisation responsable de l'intelligence artificielle par les organismes publics* (arrêté numéro 2024-02 du ministre de la Cybersécurité et du Numérique en date du 27 juin 2024).

Stratégie d'intégration de l'IA dans l'administration publique

Afin de profiter pleinement du potentiel de l'IA comme outil d'optimisation et de modernisation de ses processus, le gouvernement du Québec a lancé officiellement en juin 2021 la *Stratégie d'intégration de l'intelligence artificielle dans l'administration publique 2021-2026*.

Cette Stratégie vise à positionner l'administration publique comme acteur exemplaire de l'IA en prenant notamment appui sur le leadership du Québec dans ce domaine. Le présent guide s'appuie particulièrement sur les trois axes de la Stratégie :

- Des services publics renouvelés et optimisés par l'IA;
- Une administration publique outillée et proactive à l'égard des changements engendrés par l'IA;
- Une action gouvernementale en IA fondée sur des pratiques responsables.

De plus, ce guide concourt à la réalisation de deux objectifs de cette Stratégie :

- L'objectif 2 : Soutenir les organismes publics afin qu'ils tirent profit de l'IA dans le déploiement des services publics;
- L'objectif 8 : Encadrer la conception et l'utilisation de l'IA par des balises éthiques et des pratiques de sécurité adaptées à l'IA.

Enfin, le guide appuie les autres mesures prévues à la Stratégie et est complémentaire à ces dernières. Parmi celles-ci, notons celles de doter l'administration publique d'un cadre de développement et d'utilisation responsables de l'IA et d'adopter un cadre de gouvernance de l'usage de l'IA dans l'administration publique. Des références vers ces documents seront indiquées lors des mises à jour subséquentes de ce guide.

Contexte

À la fin de l'année 2022 et au cours de l'année subséquente, la possibilité d'utiliser des outils d'IA générative dans un contexte professionnel au sein de l'administration publique québécoise s'est étendue avec une offre d'outils plus performants. L'IA générative est donc apparue comme un moteur de changement majeur. À une époque où les services publics visent une plus grande efficacité et une adaptation aux besoins individuels, l'IA générative peut redéfinir la manière dont le secteur public est capable d'accomplir ses missions et de communiquer avec les citoyens. Dans ce contexte, la détermination et l'adoption des outils d'IA générative sont encouragées considérant l'aide que ces outils sont susceptibles d'apporter aux employés, sans toutefois les remplacer. Tout en gardant à l'esprit les aspects positifs qui peuvent être générés par l'utilisation de ces outils, les utilisateurs doivent également prendre garde aux risques que cela peut soulever.

Qu'est-ce que l'IA générative?

La définition présentée dans cette section est tirée du *Guide d'introduction en intelligence artificielle*[1] conçu par le Centre québécois d'excellence numérique (CQEN) au ministère de la Cybersécurité et du Numérique (MCN). Elle permet de comprendre ce à quoi l'on réfère dans le cas de l'IA générative ainsi que les cas d'application.

L'IA générative peut être définie comme une classe ou des types d'algorithmes issus de diverses branches, méthodes et techniques de l'IA telles que l'apprentissage machine, l'apprentissage profond, le traitement automatique du langage ou la vision par ordinateur, ayant des capacités de générer des données synthétiques, en texte, images, vidéos, musique, etc. (Gartner, 2023).

La catégorisation de ce concept est assez ardue, particulièrement en l'absence d'un domaine d'application précis. L'IA générative peut être considérée comme une catégorie de l'IA. Cependant, l'implémentation des algorithmes requiert diverses approches et branches de l'IA. Les algorithmes catégorisés comme génératifs existent depuis

[1] À paraître.

des dizaines d'années comme la machine Boltzmann, le classifieur naïf de Bayes, le modèle de Markov caché et les réseaux antagonistes génératifs ou GAN (Goodfellow et coll., 2016) ainsi que d'autres.

L'IA générative a récemment été mise de l'avant grâce aux avancées technologiques en matière de gigantesque puissance de calcul et aux percées de la recherche, notamment en apprentissage automatique, profond et par renforcement. Actuellement, ces systèmes et applications comptent parmi les plus populaires et suscitent à la fois de l'engouement, mais aussi des débats sur son utilisation. Parmi les exemples bien connus du grand public, on peut citer le modèle ChatGPT, le copilote de GitHub ou les hypertrucages.

Cas d'utilisation

L'IA générative offre des applications variées dans la prestation et la gestion des services publics. Sans remplacer la réflexion et le jugement humains, l'IA générative peut apporter du soutien dans divers domaines d'activités.

Voici quelques exemples :

Idéation/remue-méninge : l'IA générative peut servir de catalyseur pour stimuler la créativité et générer de nouvelles idées.

Création de contenu : l'IA générative peut être utilisée pour créer de nouveaux contenus, par exemple :

1. **Rédaction de brouillons pour des documents** : aider à la rédaction de brouillons de présentations, de rapports et de notes.
2. **Rédaction de courriels** : aider à la rédaction de correspondance courante tout en maintenant une communication claire et professionnelle.
3. **Création d'actualités, contenus pour la publicité et le marketing** : générer des contenus visant à informer le public des initiatives gouvernementales, des changements de politique et des services disponibles, en personnalisant le contenu pour différents groupes démographiques et régions par exemple.

Création de persona : l'IA générative peut aider à développer des personas pour les campagnes de communication publique et proposer des messages ciblés pour les différents segments de la population.

Création de contenus multimédias : l'IA générative peut créer de nouveaux sons ou de nouvelles images, voix, vidéos ou musique.

Programmation de code informatique : l'IA générative peut appuyer les travaux de développement de systèmes informatiques, notamment pour le débogage, l'explication de code et la conversion à un autre langage de programmation.

Production de listes de contrôle : l'IA générative peut générer des listes de contrôle à partir de courriels ou de comptes-rendus de réunions pour assurer le suivi des actions et des décisions.

Formation : l'IA générative peut être utilisée de différentes manières dans le contexte de formation des employés, par exemple :

1. **Parcours de formation** : proposer des parcours personnalisés pour les employés en ciblant les compétences à bonifier et en recommandant des ressources éducatives pertinentes.
2. **Génération de contenu de formation** : contribuer à créer du contenu de formation, tel que des manuels, des guides ou des supports de cours.
3. **Conception de tests de connaissances** : aider à créer des évaluations et des tests pour mesurer les compétences et les connaissances.

Analyse de données : l'IA générative peut traiter et analyser des données, notamment pour aider à la prise de décision, en cernant les tendances, les modèles et les prévisions pour la planification et l'allocation des ressources.

Traitement de contenu : l'IA générative peut être utilisée pour faciliter le traitement de contenu déjà existant :

1. **Vulgarisation de concepts complexes** : transformer des informations techniques ou des données complexes en explications simples et en visuels clairs pour faciliter la compréhension.
2. **Correction de texte** : effectuer la relecture et la correction de documents officiels, en s'assurant que le texte est exempt de fautes d'orthographe, de grammaire et de ponctuation.
3. **Normalisation de texte** : aider à uniformiser le ton et le style des communications et s'assurer qu'elles répondent aux normes de l'administration publique et qu'elles sont cohérentes sur tous les canaux lors de communications officielles.
4. **Conversion de contenu** : faciliter la réutilisation de contenu d'une plateforme à une autre, par exemple en adaptant des documents textuels pour une présentation en ligne ou en transformant des données en infographie.
5. **Traduction de contenu** : traduire rapidement du contenu.
6. **Synthèse de textes** : résumer de longs documents en points clés.

Dans tous les cas d'utilisation, il est proscrit de :

- saisir de l'information sensible[2], confidentielle[3] ou personnelle[4];
- demander d'utiliser ou de s'inspirer de matériel protégé par des droits d'auteur;
- demander à l'IA d'enfreindre ou d'aider à enfreindre une loi, une directive ou un règlement.

---

[2] Selon le Grand dictionnaire terminologique, une information sensible est une « information confidentielle dont la divulgation, l'altération, la perte ou la destruction sont susceptibles de porter préjudice à la personne ou à l'organisation qu'elle concerne. » Office québécois de la langue française (2020), page consultée le 30 novembre 2023, à l'adresse information sensible | GDT (gouv.qc.ca).

[3] Selon le Grand dictionnaire terminologique, une donnée confidentielle est une « information qui ne doit être communiquée ou rendue accessible qu'aux personnes et aux entités autorisées. » Office québécois de la langue française (2019), page consultée le 30 novembre, à l'adresse : information confidentielle | GDT (gouv.qc.ca).

[4] Selon la Commission de l'accès à l'information, un renseignement personnel « est un renseignement qui permet d'identifier une personne physique, **directement ou indirectement**. Les renseignements personnels sont confidentiels. Leur confidentialité découle du droit à la vie privée, permettant à toute personne d'exercer un contrôle sur l'utilisation et la circulation de ses renseignements. »
Commission de l'accès à l'information (2023), page consultée le 30 novembre 2023, à l'adresse Qu'est-ce qu'un renseignement personnel? – Commission d'accès à l'information du Québec (gouv.qc.ca).

L’organisation et l’employé demeurent imputables de l’utilisation des résultats générés par les outils d’IA.

Risques

Les outils qui utilisent l’IA ont le potentiel de libérer du temps pour se concentrer sur des tâches à plus forte valeur ajoutée. Cependant, il est important que le personnel de l’administration publique soit pleinement sensibilisé aux risques et adopte de bonnes pratiques d’utilisation responsable de ces technologies. Bien que cette section ne puisse remplacer une analyse de risques en bonne et due forme et qu’il puisse exister d’autres risques, voici trois grandes catégories de risques à garder en tête lors de l’utilisation de ces outils :

Catégorie 1 : Risques liés à l’utilisation d’outils d’IA générative par le personnel de l’administration publique.

Catégorie 2 : Risques de l’IA générative liés à la cybercriminalité

2.1 Risques d’utilisation de l’IA générative par les cybercriminels.

2.2 Risques d’exploitation des vulnérabilités d’un modèle d’IA générative par les cybercriminels.

Catégorie 1 : Risques liés à l'utilisation d'outils d'IA générative par le personnel de l'administration publique

Confidentialité de l'information sensible[5] et protection des renseignements personnels[6]

Les systèmes d'IA générative sont souvent munis de composantes de collecte des données servant à des fins d'analyse et même de réutilisation pour améliorer l'apprentissage et les performances. Par exemple, une personne peut involontairement, fournir des informations confidentielles, sensibles ou personnelles pour qu'elles soient traitées par l'outil comme données d'entrée. Les données saisies pourraient être conservées dans le système, pour des fins d'entraînement, et ensuite divulguées en réponse à la demande d'un autre utilisateur qui ne devrait pas normalement y avoir accès.

Biais

Les modèles de langage sont entraînés sur des données existantes pouvant contenir des biais implicites ou explicites, ce qui peut amener une production de texte ou d'autres actions biaisées. Ces biais peuvent provenir de différentes sources, telles que les préjugés sociétaux, les données déséquilibrées ou les erreurs humaines lors de l'étiquetage des données. Le biais dans l'IA se réfère à des tendances ou à des préjugés non intentionnels qui peuvent être présents dans les résultats produits par l'IA et créer un préjudice à un citoyen ou au personnel de l'administration publique, par exemple.

Fiabilité

L'IA générative peut parfois produire des erreurs ou des « hallucinations », c'est-à-dire inventer des faits dans sa réponse étant donné qu'elle est basée sur des modèles statistiques et ne comprend pas le sens du texte. En effet, l'IA générative peut créer de fausses informations qui semblent plausibles à première vue, d'où l'importance de toujours vérifier les faits, même pour des questions simples. Des décisions, prises à partir d'informations erronées, peuvent entraîner des conséquences indésirables pour l'organisme ou créer un préjudice à un citoyen ou au personnel de l'administration publique, par exemple.

---

[5] Voir la note 2.
[6] Voir la note 4.

Dépendance technologique

L'adoption de l'IA dans l'administration publique, au sein des entreprises ainsi que dans notre quotidien peut entraîner une dépendance excessive à ces systèmes. Au sein du monde professionnel, les personnes ayant recours à ces outils peuvent devenir moins enclines à effectuer des tâches analytiques ou décisionnelles sans l'assistance de l'IA, ce qui peut diminuer leur capacité à fonctionner indépendamment de la technologie.

Respect de la législation

L'IA générative propose des informations et des contenus qui pourraient soutenir une prise de décisions ayant un impact réel sur les clientèles des OP (individus, communautés ou entreprises). Si les orientations, les actions ou les décisions adoptées ne respectent pas les lois et la réglementation en vigueur, elles pourraient générer des préjudices et même possiblement des sanctions. Aussi, l'IA générative n'est pas nécessairement alimentée par les textes de loi en vigueur au Québec et au Canada et elle peut proposer des actions ou du contenu qui contreviendraient à des lois ou règlements en vigueur. Enfin, nourrir l'outil avec des renseignements personnels pourrait être considéré comme une infraction à la législation entourant la protection des renseignements personnels.

Respect de la propriété intellectuelle

La propriété intellectuelle (PI) est : « le résultat d'un travail de création de l'esprit qui fait l'objet d'un droit ».[7] On reconnaît notamment : le droit d'auteur, le brevet, le modèle industriel, la marque de commerce, l'appellation d'origine et le secret industriel.

Le risque lié à la PI concerne les incertitudes et les complications juridiques quant à la propriété et aux droits sur les contenus créés par des outils d'IA générative. En effet, ces systèmes peuvent imiter, reproduire ou s'inspirer de travaux protégés par différents types de droits de PI lors de la génération de contenu, soulevant des questions de violation potentielle de ces droits.

Génération de code non sécurisé

La génération de code non sécurisé désigne le risque que les outils basés sur l'intelligence artificielle produisent du code informatique qui ne respecte pas le cadre de développement sécuritaire et qui peut comporter des vulnérabilités ou du code

[7] OQLF propriété intellectuelle | GDT (gouv.qc.ca)

malveillant. Ces vulnérabilités peuvent permettre l'accès non autorisé à des données, des intrusions dans des systèmes, des fuites d'informations sensibles, etc.

Catégorie 2 : Risques de l'IA générative liés à la cybercriminalité

2.1 Risques d'utilisation de l'IA générative par les cybercriminels

L'IA est devenue un moteur d'innovation incontournable pour les organisations, y compris les gouvernements. Cependant, le développement des modèles d'IA présente des défis uniques en matière de sécurité.

L'IA peut être utilisée pour amplifier les campagnes de mésinformation et de désinformation

Les modèles de langage peuvent générer automatiquement du contenu trompeur, tel que de fausses nouvelles, des théories du complot ou des hypertrucages, qui peuvent influencer l'opinion publique et saper la confiance dans les institutions. La diffusion massive de ces contenus générés par l'IA peut rendre plus difficile la distinction entre les informations véridiques et les manipulations.

L'IA peut également être exploitée pour des attaques d'hameçonnage plus sophistiquées

Les modèles de langage peuvent être utilisés pour créer des courriels personnalisés et convaincants, ciblant des individus spécifiques avec des informations obtenues à partir de leurs données personnelles. Ces attaques de piratage psychologique améliorées par l'IA augmentent les risques de compromission des identifiants et des informations sensibles.

L'IA peut jouer un rôle dans la propagation de rançongiciels

Les attaquants peuvent utiliser des modèles d'IA pour identifier les cibles potentielles, optimiser leurs méthodes d'infection et automatiser certaines parties de leurs attaques. De plus, l'IA peut être utilisée pour générer des variantes de rançongiciels plus difficiles à détecter par les solutions de sécurité traditionnelles, augmentant ainsi la probabilité de succès des attaques.

L'IA générative renforce les menaces d'ingénierie sociale

Les attaques d''ingénierie sociale se produisent lorsque des auteurs de menace se servent d'un lien social et de manipulation pour pousser ou inciter des utilisateurs à divulguer quelque chose qui va à l'encontre des intérêts d'une organisation, comme le fait

de fournir des détails sensibles, des mots de passe ou de l'information financière.[8] L'IA générative amplifie ces risques en permettant la création d'hypertrucages audio et vidéo très réalistes, ainsi que de faux profils sur les réseaux sociaux, rendant ces tentatives de manipulation plus crédibles et difficiles à détecter.

2.2 Risques d'exploitation des vulnérabilités d'un modèle d'IA générative par les cybercriminels

Les modèles d'IA générative, bien que puissants et innovants, ne sont pas à l'abri des cyberattaques. Les cybercriminels peuvent exploiter activement les failles de sécurité de ces systèmes, compromettant ainsi leur intégrité, leur confidentialité et leur disponibilité. Parmi les principales menaces pesant sur les modèles d'IA générative, trois types d'attaques se distinguent particulièrement : les attaques par empoisonnement, l'extraction de données et les violations par abus.

Les attaques par empoisonnement

Les attaques par empoisonnement consistent à introduire des données trompeuses ou malveillantes dans les ensembles de données utilisés pour entraîner les systèmes d'IA. Cette manipulation peut amener l'IA à produire des résultats erronés, biaisés ou même dangereux, compromettant ainsi son intégrité et sa fiabilité.

L'extraction de données

Des individus malintentionnés peuvent essayer d'accéder à des informations confidentielles ou sensibles stockées dans les systèmes d'IA. Des données personnelles, des informations classifiées ou même des secrets d'État pourraient tomber dans les mains de cybercriminels.

Les violations par abus

Des personnes mal intentionnées détournent les capacités des systèmes d'IA pour générer du contenu trompeur, offensant, biaisé ou préjudiciable. Notamment, cela peut inclure la création de fausses nouvelles, de propagande ou de contenus discriminatoires, portant atteinte à la confiance du public et à la réputation des institutions.

---

[8] Piratage psychologique – ITSAP.00.166 – Centre canadien pour la cybersécurité

Perte de confiance envers les institutions publiques et le gouvernement du Québec

Si certains des risques énumérés précédemment et associés à cette technologie devaient se concrétiser, cela pourrait affecter négativement la perception publique du gouvernement. En particulier, cela risque de diminuer la confiance des citoyens envers les actions et les décisions gouvernementales, notamment si ces risques touchent la sécurité, la confidentialité des données ou l'éthique.

Recommandations à l'échelle organisationnelle

Pour une utilisation responsable et adéquate de l'intelligence artificielle, les OP de l'administration publique québécoise doivent suivre l'Énoncé de principes pour une utilisation responsable de l'intelligence artificielle par les organismes publics (arrêté 2024-02).

Plus précisément pour les outils d'IA générative, les OP devraient adopter une posture de prudence et préconiser l'application des principes spécifiques suivants :

- **Protection** : veiller à la sécurité des informations sensibles[9] ou stratégiques et à la protection des renseignements personnels.[10] L'organisation demeure imputable de la protection des informations qu'elle détient;
- **Responsabilité** : assumer la responsabilité du contenu généré par l'IA, en veillant à ce qu'il soit factuel, légal et éthique. Il faut s'assurer que ces résultats sont cohérents, valides et pertinents dans le cadre du mandat concerné. L'organisation demeure imputable des résultats qu'elle produit et de l'information qu'elle diffuse;
- **Utilité** : s'assurer que les outils d'IA générative sont utilisés de manière efficace et pertinente pour répondre aux besoins de l'organisation ou des citoyens;
- **Diligence** : être diligent et proactive dans la gestion des risques et des incidents liés à l'IA;
- **Neutralité** : veiller à ce que les résultats générés par les outils d'IA générative soient validés et reformulés au besoin afin d'être exempts de biais et qu'ils soient utilisés de manière équitable;
- **Traçabilité** : évaluer la pertinence de suivre et de documenter l'utilisation de l'IA générative, offrant ainsi une forme de transparence dans le suivi et la justification des actions et décisions;
- **Sensibilisation et gestion du changement** : informer et sensibiliser sur le fonctionnement, les forces, les limites et l'utilisation responsable des outils d'IA générative;
- **Santé et bien-être** : s'assurer que les employés ont accès à des ressources en cas d'impact négatif par rapport à leur santé ou à leur bien-être, selon les ressources disponibles au sein de l'organisation.

Le contexte de chaque OP est différent. Les principes énumérés ci-haut peuvent être bonifiés pour s'adapter à leur contexte particulier. Les organisations ont généralement différents mécanismes entourant la sécurité et l'accès à l'information et la protection des

[9] Voir la note 2 à la page 9.
[10] Voir la note 4 à la page 9.

renseignements personnels et ceux-ci pourraient être interpellés en lien avec l'utilisation de tels outils ou d'incidents liés à leur utilisation. Il pourrait être pertinent de désigner une personne responsable pour répondre aux questions ou gérer les incidents liés à ces outils.

Les bonnes pratiques individuelles

Ces pratiques prudentes sont des mesures préventives et des comportements à intégrer dans la routine quotidienne afin de mitiger les risques associés à l'utilisation d'outils d'IA générative. Voici des réflexes prudents à adopter.

1. Ne jamais divulguer d'informations personnelles, confidentielles ou sensibles.

La protection des données est primordiale. Il ne faut jamais soumettre des informations personnelles[11], confidentielles[12], sensibles[13] ou qui font l'objet de propriété intellectuelle de l'organisme public à des systèmes d'IA générative. Le personnel demeure imputable de l'utilisation qu'il fait des données de son organisation. Pour des informations supplémentaires, les unités responsables de la protection des renseignements personnels et de la sécurité informatique de votre organisation peuvent être interpellées.

2. Reconnaître que l'outil, bien que puissant, peut contenir des erreurs ou des imprécisions.

Bien que l'IA générative soit un outil performant, elle n'est pas exempte d'erreurs. Les réponses générées ne doivent pas être considérées comme des informations définitives, mais plutôt comme une base à la réflexion et aux vérifications supplémentaires. Il incombe donc à la personne utilisatrice de revoir, de réviser et de vérifier les informations générées. Pour y parvenir, vous pouvez entre autres :

- vérifier la source (si elle est disponible);
- vérifier l'auteur;
- croiser et diversifier les sources pour corroborer les réponses de l'IA générative;
- utiliser seulement les outils d'IA approuvés par votre organisation;
- solliciter l'expertise humaine;
- utiliser des outils de vérification des faits : [14]
    - Recherche des faits sur HabiloMédias,
    - Vérification des faits AFP.

Le recours à un outil d'IA générative devrait prendre en considération le contexte dans lequel il est utilisé. Par exemple, les conséquences de son utilisation lorsqu'on fait un

[11] Voir la note 4, à la page 9.
[12] Voir la note 3, à la page 9.
[13] Voir la note 2, à la page 9.
[14] Désinformation en ligne – Canada.ca

remue-méninge ne sont différentes que lors d'une prise de décision critique, ce qui mérite des précautions supplémentaires.

3. Respecter les conditions d'utilisation d'un outil en ligne.

Il faut s'assurer de lire et de respecter les conditions d'utilisation. Cela peut inclure des restrictions sur le type de contenu généré, des limitations de temps d'utilisation ou d'autres règles spécifiques.

4. S'assurer de la conformité au cadre législatif et aux normes éthiques en vigueur.

Le matériel généré par les outils d'IA générative doit respecter les lois, règles et directives existantes. Le personnel de l'administration publique demeure imputable de respecter le cadre légal, les principes pour une utilisation responsable de l'IA et les normes éthiques gouvernementales[15] et de son organisation, s'il y a lieu.

5. S'assurer que le canal de communication réseau avec l'IA est chiffré et vérifier toujours l'URL dans la barre d'adresse.

S'assurer, par exemple, d'utiliser le protocole de communication « https » entre un client Web et un serveur Web qui recourt au chiffrement afin de sécuriser les échanges et d'éviter qu'un tiers malveillant puisse y accéder.

Même si des sites Web ont des certificats valides, cela ne garantit pas toujours qu'ils soient dignes de confiance. Il faut donc toujours vérifier l'URL dans la barre d'adresse pour s'assurer que l'on se trouve sur le site prévu avant de saisir des informations.

[15] Secrétariat du Conseil du trésor (2009) « Éthique et valeurs : pour des choix éclairés », Gouvernement du Québec, consulté le 30 novembre 2023, à l'adresse : https://www.tresor.gouv.qc.ca/ressources-humaines/ethique-et-valeurs

6. Signaler toute information qui pourrait laisser croire à une fuite d'informations ou à la publication d'informations trompeuses, erronées ou diffamatoires au sujet de l'organisation.

Certaines applications d'IA générative conservent les données qui leur sont soumises pour les réutiliser (entraîner le modèle). Si en utilisant ces applications, les réponses obtenues laissent croire à une fuite d'informations ou à la publication d'informations trompeuses, erronées ou diffamatoires au sujet de l'organisation, il faut le signaler rapidement conformément à la procédure en vigueur au sein de cette dernière.

7. Être prudent dans le choix des outils et se tenir régulièrement informé des mises à jour, des améliorations et des éventuelles limitations de l'outil.

Avec la multiplication des offres d'IA sur le marché, il est essentiel de faire preuve de discernement dans le choix des outils que vous utilisez. Si les grands acteurs du domaine proposent des solutions fiables et reconnues, il existe une multitude d'IA dans tous les domaines et champs d'applications, dont la sécurité et la fiabilité peuvent varier considérablement. Il est crucial d'être vigilant face aux IA malveillantes ou développées sans les mesures de sécurité adéquates, car elles peuvent exposer votre organisation à des risques importants, tels que la fuite de données sensibles, la manipulation des résultats ou encore la compromission des systèmes.

Pour garantir une utilisation sécurisée et responsable de l'IA, il est fortement recommandé de privilégier les outils approuvés par votre organisation. Ces outils ont été évalués et validés pour répondre aux exigences de sécurité et de conformité. En cas de doute sur la fiabilité ou la sécurité d'une IA, il est impératif de consulter les autorités compétentes au sein de votre organisation, telles que le service informatique ou le responsable de la sécurité des systèmes d'information. Ils pourront vous guider dans le choix des outils les plus adaptés à vos besoins et vous aider à mettre en place les mesures de sécurité nécessaires pour protéger vos données et vos systèmes.

Enfin, étant donné la rapidité avec laquelle évolue la technologie, il est conseillé de se tenir régulièrement informé et de consulter les directives d'utilisation les plus récentes.

Qui contacter?

Pour des questions concernant le guide sur les bonnes pratiques d'utilisation des outils d'IA générative offerts par des services externes, n'hésitez pas à contacter la Direction de l'encadrement éthique de l'intelligence artificielle du MCN, à l'adresse suivante : encadrement.ia@mcn.gouv.qc.ca.

Références bibliographiques

A Legal and Ethical framework for gen AI. (n.d.). Retrieved from IDEO: https://www.ideo.com/journal/a-legal-and-ethical-framework-for-gen-ai

Coleman, T. (2023, 07 04). *How countries around the world are trying to regulate artificial intelligence*. Retrieved from The week Us: https://theweek.com/artificial-intelligence/1024605/ai-regulations-around-the-world

Commission de l 'accès à l'information . (2023). *Renseignement personnnel*. Retrieved from Qu'est-ce qu'un renseignement personnel?: https://www.cai.gouv.qc.ca/quest-ce-un-renseignement-personnel/#:~:text=C%27est%20un%20renseignement%20qui,la%20circulation%20de%20ses%20renseignements

Econ. (2023, 07 21). *How governments are looking to regulate AI*. Retrieved from Economist intelligence: https://www.eiu.com/n/how-governments-are-looking-to-regulate-ai/

European Parliament. (2023, 06 14). *EU AI Act: first regulation on artificial intelligence*. Retrieved from News European Parliament: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

François Candelon, R. C. (2023, 10). *Ai regulation is coming*. Retrieved from Harvard Business Review: https://hbr.org/2021/09/ai-regulation-is-coming

Gartner. (2023). *What is it, tools, models, applications and use cases*. Retrieved from Genertive AI: https://www.gartner.com/en/topics/generative-ai

Hauptfleisch, W. (2023, 06 02). *Where the world is on AI Regulation*. Retrieved from Medium: https://wolfhf.medium.com/where-the-world-is-on-ai-regulation-june-2023-d0ca0d31ce80

Ian Goodfellow, Y. B. (2016). *Deep Learning.* London, England: The MIT Press.

IAPP. (2023). *Global AI Legislation tracker*. Retrieved from https://iapp.org/

Intelligence artificielle générative. (2023, 07). Retrieved from Centre canadien pour la cybersécurité: https://www.cyber.gc.ca/fr/orientation/lintelligence-artificielle-generative-itsap00041

Loi sur l'IA de l'UE : première réglementation de l'intelligence artificielle. (2023, 9 6). Retrieved from

https://www.europarl.europa.eu/news/fr/headlines/society/20230601STO93804/loi-sur-l-ia-de-l-ue-premiere-reglementation-de-l-intelligence-artificielle

OECD. (2023). *Policies, data and analysis for trustworthy artificial intelligence*. Retrieved from OECD: https://oecd.ai

Office québécois de la langue française. (2020). *Donnée sensible*. Retrieved from Grand dictionnaire terminologique.

Pieper, B. K.-U. (2023, 05). *AI regulation around the world*. Retrieved from TaylorWessing: https://www.taylorwessing.com/en/interface/2023/aiare-we-getting-the-balance-between-regulation-and-innovation-right/ai-regulation-around-the-world

Reuters. (2023, 08 22). *Governments race to regulate AI tools*. Retrieved from Reuters: https://www.reuters.com/technology/governments-race-regulate-ai-tools-2023-08-22

Sécrétariat du Conseil du trésor. (2009, ÉTHIQUE ET VALEURS : POUR DES CHOIX ÉCLAIRÉS). *Éthique et valeur: Pour des choix éclairés*. Retrieved from https://www.tresor.gouv.qc.ca/ressources-humaines/ethique-et-valeurs

Tawakley, T. (2023, 06). *Ai regulation arount the world*. Retrieved from Lewis Silkin: https://www.lewissilkin.com/en/insights/ai-regulation-around-the-world

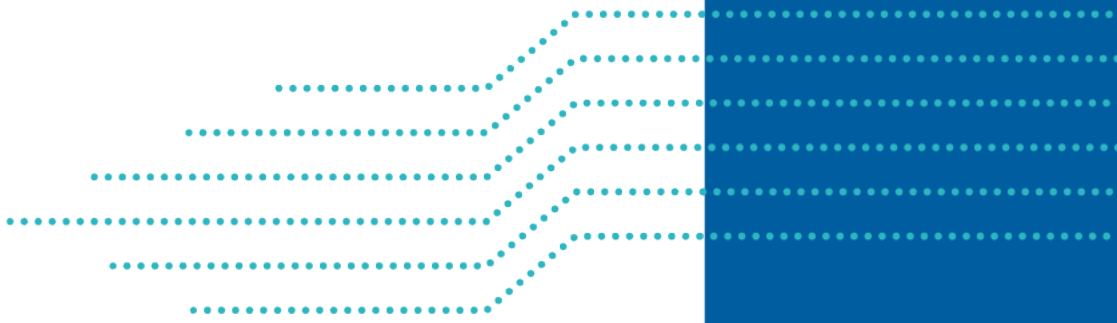
Wikipédia. (2023, 11 27). *Regulation of artificial intelligence*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Regulation_of_artificial_intelligence

Annexe 1

Résumé des bonnes pratiques individuelles

1. **Ne jamais divulguer d'informations personnelles, confidentielles ou sensibles.**
2. **Reconnaître que l'outil, bien que puissant, peut contenir des erreurs ou des imprécisions.**
3. **Respecter les conditions d'utilisation d'un outil en ligne.**
4. **S'assurer de la conformité au cadre législatif et aux normes éthiques en vigueur.**
5. **S'assurer que le canal de communication réseau avec l'IA est chiffré et vérifier toujours l'URL dans la barre d'adresse.**
6. **Signaler toute information qui pourrait laisser croire à une fuite d'informations ou à la publication d'informations trompeuses, erronées ou diffamatoires au sujet de l'organisation.**
7. **Se tenir régulièrement informé des mises à jour, des améliorations et des éventuelles limitations de l'outil.**

Ces bonnes pratiques devraient être appliquées pour toute solution informatique. Par exemple, si le correcteur automatique corrige une phrase et en change le sens, l'employé demeure responsable du texte, il en va de même pour les solutions d'IA générative.



Québec