



Guide sur l'utilisation de l'intelligence artificielle générative

Publié : le 2024-02-12

© Sa Majesté le Roi du chef du Canada,
représenté par la présidente du Conseil du Trésor, 2024

Publié par le Secrétariat du Conseil du Trésor du Canada
90 rue Elgin, Ottawa, Ontario, K1A 0R5, Canada

No de catalogue BT48-37/2024F-PDF
ISBN: 978-0-660-70253-7

Ce document est disponible sur le site Web du gouvernement du Canada à l'adresse www.canada.ca

Ce document est disponible en médias substituts sur demande.

Nota : Pour ne pas alourdir le texte français, le masculin est utilisé
pour désigner tant les hommes que les femmes.

Also available in English under the title: Guide on the use of generative AI

Guide sur l'utilisation de l'intelligence artificielle générative

Table des matières

- [Vue d'ensemble](#)
- [Qu'est-ce que l'IA générative?](#)
- [Défis et préoccupations](#)
- [Approche recommandée](#)
- [Considérations politiques et pratiques exemplaires](#)
- [Utilisation de ce guide et soutien supplémentaire disponible](#)
- [Foire aux questions](#)

Vue d'ensemble

Les outils d'intelligence artificielle (IA) générative offrent de nombreux avantages potentiels aux institutions du gouvernement du Canada (GC). Les institutions fédérales devraient explorer les utilisations potentielles des outils d'IA générative afin d'appuyer et d'améliorer leurs opérations.

Toutefois, parce que ces outils sont en évolution, ils ne devraient pas être utilisés dans tous les cas. Les institutions fédérales doivent faire preuve de prudence, et évaluer les risques avant de commencer à les utiliser.

L'utilisation de ces outils devrait être limitée aux situations où les risques peuvent être gérés de façon efficace.

Le présent document, sous forme préliminaire, a pour but d'aider les institutions fédérales à utiliser les outils d'IA générative. Cela inclut les cas où ces outils sont déployés par les institutions fédérales. Il donne une vue d'ensemble de l'IA générative, identifie les défis et les préoccupations liés à son utilisation, propose des principes pour l'utiliser de manière responsable, et suggère des considérations politiques et des pratiques exemplaires.

Le présent Guide vise également à sensibiliser les institutions fédérales à l'IA générative et à favoriser la coordination entre elles. Il souligne l'importance de mobiliser les principaux intervenants avant de déployer des outils d'IA générative pour un usage public et avant d'utiliser ces outils à des fins telles que la prestation de services. Ces intervenants comprennent des conseillers juridiques, des experts en matière de protection de la vie privée et de sécurité, ainsi que le bureau de la dirigeante principale de l'information du Secrétariat du Conseil du Trésor du Canada (SCT).

Le présent Guide complète et appuie la conformité à plusieurs lois et politiques fédérales qui existent déjà, y compris dans les domaines de la vie privée, de la sécurité, de la propriété intellectuelle, et des droits de la personne. Il reconnaît le besoin d'une itération afin de suivre le rythme des modifications réglementaires et technologiques.

Qu'est-ce que l'IA générative?

La Directive sur la prise de décisions automatisée définit l'IA comme la technologie de l'information qui exécute des tâches pour lesquelles il faut habituellement faire appel à l'intelligence biologique, comme comprendre le langage parlé, apprendre des comportements ou résoudre des problèmes.

L'IA générative est un type d'IA qui produit du contenu comme du texte, de l'audio, du code, des vidéos et des images. ¹ Ce contenu est produit sur la base des informations saisies par l'utilisateur, qui consistent en des messages-guides (généralement de courts textes didactiques).

Parmi les exemples d'outils d'IA générative, on peut citer les robots conversationnels comme ChatGPT et Bing Chat, GitHub Copilot, qui produit du code à partir de messages-guides textuels, et DALL-E, Midjourney et Stable Diffusion, qui produisent des images à partir de messages-guides textuels ou d'images. En outre, les modèles d'IA générative peuvent être mis au point, ou des modèles personnalisés peuvent être entraînés et déployés pour répondre aux besoins d'une organisation. ²

De nombreux modèles d'IA générative ont été entraînés sur de grands volumes de données, notamment des données publiques accessibles à partir d'Internet. Sur la base des données d'apprentissage, ces modèles génèrent un contenu statistiquement probable en réponse à un message-guide, ³ par exemple, en prédisant le prochain mot d'une phrase. Des techniques comme la supervision humaine et l'apprentissage par renforcement peuvent également être appliquées pour améliorer les résultats, ³ et les utilisateurs peuvent fournir un retour d'information ou modifier leur message-guide pour affiner la réponse. L'IA générative peut donc produire du contenu qui semble avoir été produit par un humain.

Elle peut être utilisée pour effectuer ou appuyer diverses tâches, notamment :

- rédiger et modifier des documents et des courriers électroniques;
- exécuter des tâches de codage comme le débogage et la création de modèles et de solutions communes;
- résumer l'information;
- lancer des idées;
- rechercher, traduire et apprendre;
- aider les clients (p. ex., répondre à leurs questions, rechercher la cause d'un problème).

Défis et préoccupations

Avant que les institutions fédérales ne commencent à utiliser des outils d'IA générative, elles doivent évaluer et atténuer certains risques éthiques, juridiques et autres. Par exemple, ces outils peuvent générer du contenu inexact, amplifier les biais et porter atteinte à la propriété intellectuelle, la vie privée et d'autres lois. En outre, certains outils peuvent ne pas répondre aux exigences fédérales en matière de protection de la vie privée et de sécurité. Lorsque les institutions utilisent ces outils, elles doivent protéger les renseignements personnels et les données sensibles. De plus, comme ces outils génèrent un contenu qui peut donner l'impression d'avoir été produit par un être humain, les gens peuvent ne pas savoir avec exactitude s'ils interagissent avec une personne ou un outil. L'utilisation de ces outils peut également exercer une incidence sur les compétences et le jugement des fonctionnaires et peut engendrer des coûts environnementaux.

Les outils d'IA générative s'appuient sur des modèles qui posent divers défis, notamment une transparence et une explicabilité limitées. Ils s'appuient également sur des données d'apprentissage qui sont difficiles à obtenir et à évaluer. Ces défis découlent en partie de la taille importante des modèles, des volumes élevés de données d'apprentissage et de la nature exclusive de nombreux outils. En outre, les résultats des modèles sont limités par les messages-guides et les données d'apprentissage, qui peuvent manquer un contexte qui n'est pas accessible au public à partir d'Internet. Les données d'apprentissage peuvent également être obsolètes; par exemple, ChatGPT est mis au point selon des données jusqu'en 2021, ce qui limite sa capacité à fournir des informations sur les événements postérieurs. ⁴ ⁵ De plus, ces outils présentent des limites qui réduisent leur utilité à certaines fins; par exemple, ils ont tendance à être peu performants pour les tâches liées aux émotions. ⁶ ⁷

L'IA générative peut également présenter des risques pour l'intégrité et la sécurité des institutions fédérales, compte tenu de son utilisation malveillante potentielle par des auteurs de menace. Les institutions fédérales doivent être conscientes de ces risques et veiller à ce que les mesures d'atténuation nécessaires soient mises en place conformément aux orientations du Centre canadien pour la cybersécurité sur l'IA générative.

Approche recommandée

Les institutions fédérales devraient étudier comment elles pourraient utiliser des outils d'IA générative afin d'appuyer leurs opérations et améliorer les résultats pour la population canadienne. Toutefois, compte tenu des défis et des préoccupations liés à ces outils, les institutions devraient évaluer et atténuer les risques et limiter leur utilisation aux activités pour lesquelles elles peuvent gérer efficacement les risques. Compte tenu de l'adoption croissante de ces technologies dans différents secteurs et par le public, leur utilisation par le gouvernement permettra de suivre l'évolution du paysage numérique.

Les institutions fédérales devraient évaluer les outils en fonction de leur capacité à aider les employés, et non à les remplacer. Lorsqu'ils décident d'utiliser ou non des outils d'IA générative, les fonctionnaires doivent se référer au guide de prise de décisions éthiques à la section 6, Les valeurs au quotidien : Guide de discussion du Code de valeurs et d'éthique du secteur public.

Pour maintenir la confiance du public et garantir une utilisation responsable des outils d'IA générative, les institutions fédérales devraient s'harmoniser avec les principes « PRETES » :

- **Pertinente** : Veiller à ce que l'utilisation d'outils d'IA générative réponde aux besoins des utilisateurs et de l'organisation et contribue à améliorer les résultats pour la population canadienne; déterminer les

outils appropriés pour la tâche; les outils d'IA ne sont pas le meilleur choix dans toutes les situations.

- **Responsable** : Assumer la responsabilité du contenu généré par ces outils, notamment veiller à ce que les informations soient factuelles, légales, éthiques et conformes aux conditions d'utilisation.
- **Équitable** : Veiller à ce que le contenu de ces outils ne comporte pas ou n'amplifie pas les biais et qu'il respecte les droits de la personne, l'accessibilité et les obligations en matière d'équité procédurale et de fond.
- **Transparente** : Identifier le contenu produit à l'aide de l'IA générative; informer les utilisateurs qu'ils interagissent avec un outil d'IA; documenter les décisions et être en mesure de fournir des explications si des outils sont utilisés afin d'appuyer la prise de décisions.
- **Éclairée** : Apprendre à connaître les points forts, les limites et l'utilisation responsable des outils; apprendre à créer des messages-guides efficaces et à relever les faiblesses potentielles des résultats.
- **Sécurisée** : Garantir que l'infrastructure et les outils sont adaptés à la classification de sécurité des renseignements et que la vie privée et les renseignements personnels sont protégés.

Pour obtenir de l'aide afin de déterminer l'utilisation appropriée de ces outils, les fonctionnaires devraient s'adresser aux partenaires appropriés tels que les services juridiques de leur institution, les experts en matière de protection de la vie privée et de sécurité, le bureau principal de l'information, le bureau principal des données et les spécialistes de la diversité et de l'inclusion. En outre, le Centre canadien pour la cybersécurité, Statistique Canada et le Bureau de la dirigeante principale de l'information du SCT sont également disponibles pour aider les institutions fédérales à utiliser ces outils de manière responsable.

Considérations politiques et pratiques

exemplaires

La Directive sur la prise de décisions automatisée s'applique-t-elle?

La Directive sur la prise de décisions automatisée s'applique aux systèmes automatisés, notamment ceux qui reposent sur l'IA, utilisés pour influencer ou prendre des décisions administratives. Comme les autres systèmes d'IA, les systèmes d'IA générative ont des capacités qui permettent de faire des évaluations ou des déterminations sur les clients dans le cadre de la prestation de services. Par exemple, un système d'IA générative pourrait être utilisé pour résumer les données d'un client ou pour déterminer s'il est admissible à un service. ⁸ Ces utilisations administratives peuvent influencer sur la manière dont un agent perçoit et décide d'un cas, ce qui entraîne des conséquences sur les droits, les intérêts ou les privilèges du client. La Directive s'applique donc à l'utilisation de systèmes d'IA générative pour prendre des décisions administratives ou les éclairer.

Toutefois, l'IA générative n'est peut-être pas adaptée à la prise de décisions administrative à ce stade. La conception et le fonctionnement des modèles d'IA générative peuvent limiter la capacité des institutions fédérales à garantir la transparence, la responsabilité et l'équité des décisions prises par les systèmes d'IA générative ou éclairées par leurs résultats. De même, les conditions d'utilisation des produits d'IA générative de nombreuses grandes entreprises technologiques interdisent d'utiliser ces produits pour prendre des décisions à forte incidence. Par exemple, OpenAI demande aux utilisateurs de ne pas utiliser ChatGPT dans les décisions relatives au crédit, à l'emploi, aux établissements d'enseignement ou aux services d'assistance publique, à l'application de la loi et à la justice pénale, ainsi qu'à l'immigration et à l'asile. ⁹ De même, Google interdit aux utilisateurs de son produit d'IA générative de prendre des « décisions automatisées dans des domaines qui affectent les droits ou le bien-être matériels ou des individus ». ¹⁰ Ces limites soulignent l'importance de se conformer à l'exigence de la Directive de consulter des services juridiques pendant la

phase de conception d'un projet d'automatisation. La consultation permet aux institutions fédérales de comprendre les risques juridiques liés aux utilisations administratives des systèmes d'IA générative, tant pour eux-mêmes que pour leurs clients.

Toutes les utilisations de l'IA générative ne sont pas soumises à la Directive. Par exemple, l'utilisation d'outils d'IA générative dans le cadre d'une recherche ou d'un remue-méninge, d'une planification ou d'une correspondance courante n'entre pas dans la portée de la Directive. Toutefois, ces utilisations non administratives restent soumises aux lois et politiques qui régissent les institutions fédérales.

Considérations relatives à la protection de la vie privée

Comme pour tout système en ligne, les renseignements personnels ne doivent pas être intégrés à un outil ou à un service d'IA générative, à moins qu'un contrat n'ait été conclu avec le fournisseur et qu'il n'indique comment les renseignements seront utilisés et protégés. Avant d'utiliser un outil d'IA générative, les institutions fédérales doivent également veiller à ce que la collecte et l'utilisation des renseignements personnels, notamment les renseignements utilisés pour perfectionner l'outil, respectent leurs obligations en matière de la protection de la vie privée.

Tous les renseignements personnels que les institutions fédérales ont utilisés, créés ou obtenus au moyen de l'IA générative et communiqués à cette fin sont soumis aux exigences de la Loi sur la protection des renseignements personnels et des instruments politiques connexes. Cela signifie ce qui suit :

- les renseignements personnels ne peuvent être recueillis que s'ils sont directement liés au programme ou à l'activité;
- ils ne peuvent être utilisés qu'aux fins pour lesquelles ils ont été recueillis ou pour les usages compatibles avec ces fins;
- leur communication autorisée par la loi est limitée;

- les institutions doivent être transparentes sur la manière dont elles traitent et protègent les renseignements personnels qu'elles recueillent une fois ces renseignements sont sous le contrôle du gouvernement.

Les risques pour la protection de la vie privée varient en fonction de la manière dont l'outil d'IA recueille et traite les informations sur les personnes et, éventuellement, prend des décisions à leur sujet. Un outil d'IA pourrait, par exemple, décider si une personne est admissible à un service, déterminer le niveau de prestation auquel elle a droit ou traiter des données d'enquête afin de faire la lumière sur l'orientation politique.

La *Loi sur la protection des renseignements personnels* exige qu'une institution gouvernementale prenne toutes les mesures raisonnables afin de veiller à ce que les renseignements personnels utilisés à des fins administratives par l'institution soient aussi exacts, à jour et complets que possible. Lorsqu'elles utilisent un système d'IA générative pour prendre ou mettre au clair certaines décisions concernant des personnes, les institutions fédérales doivent avoir la certitude que les renseignements personnels recueillis, créés ou utilisés par le système sont exacts. C'est pourquoi la collecte directe auprès de la personne est nécessaire dans la plupart des cas. La collecte directe permet également à la personne d'être informée de la collecte et de la manière dont ses renseignements seront utilisés et gérés.

Si les résultats d'un outil d'IA générative entraînent la création de nouveaux renseignements personnels, ces derniers doivent également être gérés conformément aux exigences en matière de protection des renseignements personnels. Par exemple, un résumé d'une demande de service ou de prestation produit par une IA générative pourrait constituer de nouveaux renseignements personnels. Les utilisateurs devraient valider tous les renseignements personnels créés par un outil d'IA générative pour s'assurer qu'ils sont à jour, exacts et complets. De plus, ils doivent veiller à ce que les nouveaux renseignements personnels ne soient pas divulgués à des fins incompatibles avec celles pour lesquelles ils ont été recueillis. Dans

l'exemple ci-dessus, le fait de partager les nouveaux renseignements sur la personne avec un autre programme pour une prestation non connexe peut être inapproprié et constituer une atteinte à la vie privée.

Les institutions fédérales doivent également veiller à ce que tous les renseignements personnels qu'elles recueillent et utilisent puissent être mis à la disposition de la personne concernée, et que celle-ci puisse y accéder et y apporter des corrections sur demande. Les institutions fédérales doivent conserver pendant au moins deux ans les renseignements personnels utilisés pour prendre une décision concernant une personne. La personne dispose ainsi d'un délai suffisant pour exercer son droit d'accès et de correction des renseignements. Les institutions fédérales ne doivent pas conserver les renseignements personnels plus longtemps que nécessaire. Plus elles conservent longtemps des renseignements personnels, plus la probabilité d'une atteinte à leur protection est élevée.

La dépersonnalisation et l'utilisation de données synthétiques peuvent aider les institutions à réduire l'incidence et la probabilité d'atteinte à la vie privée lors du perfectionnement, de l'utilisation et de l'évaluation des résultats des outils d'IA générative. L'[Avis de mise en œuvre de la protection des renseignements personnels 2023-01 : Dépersonnalisation](#) contient de plus amples informations sur ces techniques de préservation de la vie privée. D'autres mesures de protection comme les contrôles administratifs, les droits d'accès et l'audit, sont également importantes pour réduire le risque de divulgation involontaire ou d'accès non autorisé, de repersonnalisation et d'inférence et, d'une manière générale, pour préserver la vie privée des personnes.

Avant d'envisager l'acquisition, l'utilisation ou le déploiement d'outils d'IA générative, les responsables de la protection de la vie privée des institutions fédérales doivent déterminer si une évaluation des facteurs relatifs à la vie privée est nécessaire.

Lorsque les institutions fédérales mettent en place des solutions informatiques qui utilisent l'IA générative, elles doivent veiller à ce qu'elles respectent les exigences en matière de protection de la vie privée. [Le Guide sur les pratiques relatives à la vie privée numérique](#) contient plus de renseignements sur ces exigences et sur la manière d'intégrer des conseils en matière de protection de la vie privée aux solutions informatiques qui utilisent l'IA générative.

Enjeux potentiels et pratiques exemplaires

La section suivante donne un aperçu de plusieurs domaines de risque et présente les pratiques exemplaires pour une utilisation responsable de l'IA générative dans les institutions fédérales. Outre les pratiques exemplaires identifiées pour tous les utilisateurs de l'IA générative au sein du gouvernement fédéral, des pratiques exemplaires propres aux institutions fédérales qui développent ou déploient ces outils sont également identifiées pour s'assurer que les risques sont correctement évalués et atténués, et pour faire la distinction entre les responsabilités des utilisateurs et des développeurs.

Protection des renseignements

Enjeu : Certains outils d'IA générative ne répondent pas aux exigences du gouvernement en matière de sécurité de l'information.

La protection des renseignements personnels, classifiés, protégés et exclusifs, est essentielle lors de l'utilisation de systèmes d'IA générative. Les fournisseurs de certains outils d'IA générative peuvent inspecter les données d'entrée ou utiliser ces données pour perfectionner leurs modèles, ce qui pourrait entraîner des atteintes à la vie privée et à la sécurité. Des risques peuvent également découler du stockage des données d'entrée sur des serveurs non contrôlés par le gouvernement du Canada (GC), où les données peuvent être conservées plus longtemps que nécessaire, rendues accessibles, diffusées ultérieurement ou vulnérables à une violation des

données. ¹¹ Certains outils, publics ou non, peuvent ne pas répondre aux exigences en matière de protection de la vie privée et de sécurité établies par la loi et la politique fédérales.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Ne pas saisir de renseignements sensibles ou personnels dans des outils qui ne sont pas gérés par le GC.
- Ne pas soumettre de requêtes au moyen d'outils non gérés par le GC qui pourraient compromettre la confiance du public si ces demandes étaient divulguées. Veuillez consulter l'annexe B de la Directive sur les services et le numérique pour obtenir des exemples d'utilisations inacceptables.
- Comprendre comment un système utilise les données d'entrée (p. ex., si elles sont utilisées comme données d'apprentissage ou si elles sont accessibles aux fournisseurs).
- Demander aux services juridiques et au (à la) dirigeant(e) principal(e) de la sécurité (DPS) de votre institution d'examiner les conditions d'utilisation, la politique sur la protection de la vie privée et les autres documents juridiques d'un fournisseur avant d'utiliser un système pour traiter des informations sensibles ou de nature exclusive.
- Utiliser une infrastructure et des outils adaptés à la classification de sécurité des renseignements, conformément à la Directive sur la gestion de la sécurité.
- Consulter le (la) DPS de votre institution avant d'utiliser, d'acquérir ou de déployer une IA générative pour obtenir des renseignements protégés ou d'autres renseignements sensibles.
- Tenir compte des exigences relatives à la résidence des informations et des données dans la Directive sur les services et le numérique et l'orientation correspondante dans la Ligne directrice sur les services et le numérique

- Utiliser si possible la fonction de « non-participation » afin de veiller à ce que les messages-guides ne sont pas utilisés pour entraîner ou développer davantage un système d'IA.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Procéder à des essais réguliers avant et pendant l'exploitation du système afin de s'assurer qu'il répond aux principaux objectifs de performance.
- Planifier des audits indépendants pour évaluer les systèmes d'IA générative par rapport aux cadres de risque et d'incidence.

Biais

Enjeu : Le contenu généré peut amplifier les biais ou d'autres idées nuisibles qui dominent dans les données d'apprentissage.

Les outils d'IA générative peuvent produire des contenus discriminatoires ou non représentatifs, ou qui comportent des biais ou des stéréotypes (p. ex., des biais liés à des facteurs identitaires multiples et croisés comme le sexe, la race et l'appartenance ethnique). ¹² ¹³ ¹⁴ De nombreux modèles d'IA générative sont formés sur de grandes quantités de données provenant d'Internet, qui est souvent la source de ces biais. Par exemple, les données d'apprentissage sont susceptibles de refléter les biais historiques prédominants et peuvent ne pas inclure des perspectives qui sont moins répandues dans les données ou qui sont apparues depuis que le modèle a été mis au point. ¹² D'autres sources peuvent contribuer à un contenu biaisé, notamment le filtrage des données, qui peut amplifier les biais de l'ensemble d'entraînement initial, ¹⁵ le cadre des messages-guides ¹⁶ et les biais du modèle. L'utilisation généralisée de ces technologies pourrait amplifier ou renforcer ces biais et les points de vue dominants, et conduire à une moindre grande diversité des idées, des perspectives et du langage, ¹² ¹⁷ ainsi que des préjudices potentiels.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Examiner le contenu généré afin de veiller à ce qu'il soit conforme aux engagements, aux valeurs et à l'éthique du GC et qu'il respecte les obligations légales, notamment l'évaluation des biais ou des associations stéréotypées.
- Formuler des messages-guides pour produire un contenu qui offre des perspectives holistiques et limite les biais.
- S'efforcer de comprendre les données utilisées pour entraîner l'outil, par exemple leur origine, leur contenu et la manière dont elles ont été sélectionnées et préparées.
- Apprendre ce que sont les biais, la diversité, l'inclusion, l'antiracisme, les valeurs et l'éthique, afin d'améliorer votre capacité à relever les contenus biaisés ou discriminatoires.
- Notifier les destinataires lorsque le contenu a été produit par l'IA générative.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Prendre en compte les biais potentiels et les approches d'atténuation dès la phase de planification et de conception, notamment en réalisant une analyse comparative entre les sexes Plus (ACS Plus) afin de comprendre comment votre déploiement d'outils d'IA générative pourrait avoir une incidence sur différents groupes de population.
- Consulter des experts en matière d'ACS Plus et de diversité et d'inclusion au sein de votre organisation afin de déterminer les incidences de l'utilisation d'outils d'IA générative sur différents groupes de population et d'élaborer des mesures pour remédier à ces incidences.
- Tester les biais dans les données, le modèle et les résultats avant de déployer un système, et le faire de manière continue.

Qualité

Enjeu : Le contenu généré peut être inexact, incohérent ou incomplet.

Les technologies d'IA générative peuvent produire un contenu qui semble bien élaboré, crédible et raisonnable, mais qui est en fait inexact, absurde ou incohérent par rapport aux données sources. ¹⁸ ¹⁹ Ce contenu est parfois appelé « hallucination ». En outre, le contenu généré par les outils d'IA peut ne pas fournir une vision holistique d'un problème. Au lieu de cela, il peut porter sur les perspectives prédominantes des données d'apprentissage. ¹² Le contenu peut également ne plus être à jour en fonction de la période couverte par les données d'apprentissage et du fait que le système puisse ne pas avoir accès directement à des données récentes. La qualité des outils et des résultats dans les différentes langues doit également être prise en compte pour garantir le respect des exigences en matière de langues officielles.

Les risques associés à un contenu inexact varient en fonction du contexte et doivent être évalués. Par exemple, l'utilisation d'outils d'IA générative pour apprendre sur un sujet peut produire des renseignements inexacts ou des sources non existantes, ²⁰ qui, s'ils étaient utilisés dans la prise de décisions, pourraient conduire à un traitement injuste de personnes ou à une politique malavisée. En outre, l'utilisation d'outils d'IA générative pour les communications destinées au public pourrait mener le gouvernement à communiquer des renseignements inexacts, ce qui contribuerait à la désinformation et à l'érosion de la confiance du public.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Indiquer clairement que vous avez utilisé l'IA générative pour élaborer du contenu.
- Ne pas considérer le contenu généré comme faisant autorité. Vérifier l'exactitude des faits et du contexte, par exemple en les comparant à des renseignements provenant de sources fiables.
- Examiner les renseignements personnels créés à l'aide de l'IA générative pour s'assurer qu'ils sont exacts, à jour et complets.

- Évaluer l'incidence des résultats inexacts. Ne pas utiliser l'IA générative lorsque l'exactitude des faits ou l'intégrité des données est nécessaire.
- S'efforcer de comprendre la qualité et la source des données d'apprentissage.
- Penser à votre capacité à relever les contenus inexacts avant d'utiliser l'IA générative. Ne pas utiliser l'IA générative si vous ne pouvez pas confirmer la qualité du contenu.
- Apprendre à créer des messages-guides efficaces et à fournir un retour d'information pour affiner les résultats afin de limiter la production de contenu inexact.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Veiller à ce que la qualité des outils et des résultats réponde aux exigences en matière de langues officielles avant de les déployer.
- Aviser les utilisateurs qu'ils interagissent avec l'IA générative.
- Utiliser des filigranes pour aider les utilisateurs à déterminer le contenu généré par l'IA.
- Lorsque le contenu est généré par l'IA, fournir des liens vers des sources faisant autorité et d'encourager les utilisateurs à vérifier le contenu à l'aide des liens fournis.
- Fournir des renseignements sur la source des données d'apprentissage et sur la manière dont les modèles ont été élaborés.

Autonomie des fonctionnaires

Enjeu : Une confiance excessive dans l'IA peut nuire grandement au jugement, étouffer la créativité et éroder les capacités de l'effectif.

Une confiance excessive dans les outils d'IA générative peut nuire à l'autonomie et au jugement. Par exemple, certains utilisateurs peuvent être enclins à accepter sans esprit critique les recommandations du système ou d'autres résultats, qui peuvent être inexacts. ²¹ ²² Une confiance excessive dans le système peut être le signe d'un biais d'automatisation, c'est-à-dire

d'une tendance à favoriser les résultats générés par des systèmes automatisés, même en présence de renseignements contradictoires provenant de sources non automatisées. ²¹ De même, le biais de confirmation peut contribuer à une confiance excessive, ²¹ car les résultats des systèmes d'IA générative peuvent renforcer les idées préconçues des utilisateurs, en particulier lorsque les messages-guides sont rédigés d'une manière qui reflète les hypothèses et les croyances des utilisateurs. ²³ Une confiance excessive dans les systèmes d'IA peut entraîner une baisse de l'esprit critique et limiter la diversité de la pensée, étouffant ainsi la créativité et l'innovation et aboutissant à des analyses partielles ou incomplètes. Une confiance excessive dans l'IA peut entraver la capacité des employés à acquérir et à entretenir les compétences dont ils ont besoin pour accomplir les tâches confiées aux systèmes d'IA générative. Cela pourrait renforcer la dépendance du gouvernement à l'égard de l'IA et potentiellement éroder les capacités de l'effectif.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Déterminer si vous **devez** utiliser l'IA générative pour répondre aux besoins des utilisateurs et de l'organisation.
- Tenir compte des capacités et des limites de l'IA générative lors de l'attribution des tâches et de l'examen des résultats du système.
- Perfectionner vos connaissances en matière d'IA afin de pouvoir évaluer de manière critique ces outils et leurs résultats.
- Utiliser les outils d'IA générative pour vous aider et non pour remplacer votre jugement. Ne pas externaliser une compétence que vous ne comprenez pas ou que vous ne possédez pas.
- Former votre propre opinion avant de chercher des idées ou des recommandations à l'aide des outils d'IA.
- Apprendre à rédiger des messages-guides susceptibles de déboucher sur un contenu qui offre une perspective holistique et qui limite les biais.

- Toujours examiner le contenu généré par l'IA, même si le système semble fournir des réponses précises.

Risques juridiques

Enjeu : L'IA générative présente des risques pour les droits de la personne, la vie privée, la protection de la propriété intellectuelle et l'équité procédurale.

L'utilisation par le gouvernement de systèmes d'IA générative présente des risques pour les droits et obligations juridiques des institutions fédérales et de leurs clients. Ces risques découlent des données utilisées pour perfectionner des modèles d'IA, de la manière dont les systèmes traitent les données d'entrée et de la qualité des résultats du système.

L'utilisation par des fournisseurs ou des institutions fédérales de documents protégés par le droit d'auteur comme des articles, des livres, du code, des peintures ou de la musique, pour perfectionner des modèles d'IA peut porter atteinte aux droits de propriété intellectuelle. L'utilisation ou la reproduction des résultats générés par ces modèles pourrait également porter atteinte à de tels droits, s'ils contiennent des renseignements qui sont identiques ou substantiellement semblables à des travaux dont les droits d'auteur sont déjà protégés. En outre, la propriété du contenu créé par ou avec l'aide de l'IA générative est incertaine. Le droit à la vie privée pourrait également être menacé parce que les données utilisées pour perfectionner les modèles d'IA générative pourraient inclure des renseignements personnels recueillis ou utilisés de manière illégale, notamment des renseignements personnels obtenus à partir de sources en ligne accessibles au public.

Des risques pourraient également découler de l'opacité des modèles d'IA générative et de leur capacité à produire des résultats inexacts, biaisés ou incohérents. Cette opacité rend difficiles la traçabilité et la compréhension de la manière dont le système d'IA produit des résultats, ce qui peut nuire à l'équité procédurale dans les cas où une institution fédérale est tenue de fournir à ses clients les raisons de ses décisions administratives comme les

décisions de refus de prestations. La qualité des résultats de l'IA peut également avoir une incidence sur les droits juridiques des personnes. Par exemple, des résultats biaisés pourraient conduire à une discrimination dans les services, ce qui pourrait constituer une violation des droits de la personne.

Ces risques vont au-delà des scénarios de prise de décisions. Lorsque les institutions fédérales utilisent des outils d'IA générative pour aider le public à trouver des renseignements (comme c'est le cas, par exemple, avec l'utilisation des robots conversationnels sur les sites Web des ministères) ou pour rédiger des communications publiques, il existe un risque que ces outils génèrent un contenu inapproprié ou des renseignements erronés qui pourraient contribuer ou causer un préjudice pour lequel le gouvernement pourrait être tenu responsable.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Consulter les services juridiques de votre institution sur les risques juridiques liés au déploiement d'outils d'IA générative ou à leur utilisation dans la prestation de services. La consultation pourrait comprendre un examen des conditions d'utilisation, de la politique sur les droits d'auteur, de la politique de confidentialité et d'autres documents juridiques du fournisseur.
- Se conformer à la Directive sur la prise de décisions automatisée lors de l'utilisation de l'IA générative dans la prise de décisions administrative.
- Vérifier si les résultats du système sont identiques ou substantiellement semblables au matériel protégé par le droit d'auteur. Attribuer les droits là où il se doit, ou supprimer ce matériel afin de minimiser le risque d'atteinte aux droits de propriété intellectuelle.
- Consulter les fonctionnaires désignés pour l'octroi de licences et la gestion du droit d'auteur de la Couronne si vous prévoyez d'inclure des résultats dans des communications publiques, conformément aux Procédures sur l'édition.

- Évaluer la qualité des résultats pour détecter les inexactitudes factuelles, les biais ou les idées nuisibles qui peuvent être en conflit avec les valeurs du GC.
- Se tenir au courant des développements juridiques et politiques liés à la réglementation de l'IA.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Vérifier la légalité de la méthode utilisée pour obtenir les données nécessaires à l'entraînement des modèles d'IA, et veiller à ce que vous ayez l'autorisation d'utiliser les données à cette fin. Dans la mesure du possible, entraîner votre modèle en utilisant des données de source ouverte qui ne font l'objet d'aucune restriction quant à leur utilisation.
- Être transparent quant à l'utilisation de l'IA générative, notamment en informant les utilisateurs qu'ils interagissent avec un système plutôt qu'avec un être humain. Le cas échéant, inclure une clause de non-responsabilité afin de limiter les risques de responsabilité.
- Utiliser des filigranes pour aider les utilisateurs à déterminer le contenu généré.

Distinguer les êtres humains des machines

Enjeu : Les gens peuvent ne pas savoir qu'ils interagissent avec un système d'IA ou ils peuvent supposer à tort que l'IA est utilisée.

Les agents ou robots conversationnels qui utilisent l'IA générative peuvent produire des réponses qui ressemblent tellement à des réponses d'êtres humains qu'il peut être difficile d'en faire la distinction. ²⁴ En conséquence, les clients peuvent être induits en erreur et croire qu'ils interagissent avec un être humain. De même, les clients peuvent penser qu'un courriel qu'ils ont reçu a été écrit par une personne alors qu'il a été généré par un outil d'IA. Par ailleurs, les clients peuvent penser qu'ils interagissent avec un outil d'IA alors qu'ils ont en fait affaire à un être humain. La transparence sur la

question de savoir si un client interagit avec une personne ou un robot conversationnel est essentielle afin de veiller à ce que le client ne soit pas induit en erreur, et pour maintenir la confiance à l'égard du gouvernement.

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Communiquer clairement quand et comment le GC utilise l'IA dans ses interactions avec le public.
- Informer les utilisateurs lorsque les messages qui leur sont adressés sont générés par l'IA.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Envisager de proposer des moyens de communication non automatisés avec le GC.
- Utiliser des filigranes pour que les utilisateurs puissent déterminer le contenu généré par l'IA.
- Publier des renseignements sur le système comme une description en langage clair de son fonctionnement, des raisons de son utilisation et des mesures d'assurance qualité prises.

Incidences sur l'environnement

Enjeu : Le développement et l'utilisation de systèmes d'IA générative peuvent avoir des coûts environnementaux significatifs.

Le développement et l'utilisation de systèmes d'IA générative peuvent être une source importante d'émissions de gaz à effet de serre. Ces émissions proviennent du calcul utilisé pour perfectionner et exploiter les modèles d'IA générative et de la production et du transport des serveurs qui appuient les programmes d'IA. ²⁵ Bien que l'IA générative puisse contribuer à la lutte contre le changement climatique, son utilisation doit être mise en équilibre avec la nécessité de prendre des mesures rapides et radicales pour réduire les émissions mondiales de gaz à effet de serre et éviter des dommages irréversibles à l'environnement. ²⁶

Pratiques exemplaires pour tous les utilisateurs de l'IA générative dans les institutions fédérales

- Utiliser des outils d'IA générative hébergés dans des centres de données relatives à l'émission zéro.
- N'utiliser les outils d'IA générative que lorsqu'ils sont pertinents pour les objectifs du programme et les résultats escomptés.

Pratiques exemplaires additionnelles pour les institutions fédérales déployant un outil d'IA générative

- Vérifier si votre fournisseur d'IA s'est fixé des objectifs de réduction des gaz à effet de serre. ²⁷
- Réaliser une évaluation de l'incidence sur l'environnement dans le cadre de la proposition de développement ou d'acquisition d'outils d'IA générative. Veiller à ce que toute décision d'acquisition de ces outils soit prise conformément à la Politique d'achats écologiques.

Utilisation de ce guide et soutien supplémentaire disponible

Au fur et à mesure que les ministères développent leurs orientations sur l'utilisation de l'IA générative, ce document doit servir d'orientation générale sur laquelle s'appuyer. Pour obtenir de plus amples renseignements, notamment des conseils sur les utilisations spécifiques de l'IA générative, communiquez avec l'équipe du SCT chargée des données et de l'IA responsables (ai-ia@tbs-sct.gc.ca). Des ressources supplémentaires existent au sein du gouvernement fédéral, auxquelles les institutions peuvent accéder en contactant le Centre de la sécurité des télécommunications (notamment les orientations du Centre canadien pour la cybersécurité sur l'IA générative) et Statistique Canada. La communauté de pratique et le guide du SCT continueront d'évoluer au cours des prochaines années.

Foire aux questions

► Puis-je utiliser l'IA générative pour rédiger des courriels ou des notes d'information?

► Puis-je utiliser l'IA générative pour élaborer du contenu pour les communications publiques (p. ex., des publications Web, les médias sociaux)?

► Puis-je utiliser l'IA générative pour des tâches de programmation?

► Puis-je utiliser l'IA générative pour façonner les politiques?

► Puis-je utiliser l'IA générative pour automatiser les évaluations, les recommandations ou les décisions concernant les clients?

Bibliographie

- 1 McKinsey & Company, « [What is generative AI?](#) », 19 janvier 2023. [Accès le 8 mai 2023].
- 2 F. Candelon, A. Gupta, L. Kraye et L. Zhukov, « [The CEO's Guide to the Generative AI Revolution](#) », 7 mars 2023. [Accès le 11 mai 2023].
- 3 K. Martineau, « [What is generative AI?](#) », 20 avril 2023. [Accès le 8 mai 2023].
- 4 OpenAI, « [Documentation - Models Overview](#) », [Accès le 8 mai 2023].
- 5 OpenAI, « [What is ChatGPT?](#) », [Accès le 8 mai 2023].

- 6 J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieszczewski, P. Miłkowski, M. Oleksy, M. Piasecki, Ł. Radliński, K. Wojtasik, S. Woźniak et P. Kazienko, « ChatGPT: Jack of All Trades, Master of None », *SSRN preprint*, pp. 1-40, 28 février 2023.
- 7 K. Yang, S. Ji, T. Zhang, Q. Xie et S. Ananiadou, « On the Evaluations of ChatGPT and Emotion-enhanced Prompting for Mental Health Analysis », *arXiv preprint*, 2023.
- 8 Conseil de l'Union européenne, « ChatGPT in the Public Sector - overhyped or overlooked? », Union européenne, 2023.
- 9 OpenAI, « OpenAI usage policies », OpenAI, 2023. [Accès le 12 mai 2023].
- 10 Google, « Règlements sur les utilisations interdites de l'IA générative », Google, 2023. [Accès le 12 mai 2023].
- 11 OpenAI, « March 20 ChatGPT outage: Here's what happened », 24 mars 2023. [Accès le 10 mai 2023].
- 12 E. M. Bender, T. Gebru, A. McMillan-Major et S. Shmitchell, « On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 », *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610-623, 2021.
- 13 A. S. Luccioni, C. Akiki, M. Mitchell et Y. Jernite, « Stable Bias: Analyzing Societal Representations in Diffusion Models », *arXiv preprint*, vol. s.o, n° s.o., pp. 1-44, 2023.

- 14 W. Guo et A. Caliskan, « Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases », *AIES '21: Proceedings of the 2021 AAI/ACM Conference on AI, Ethics, and Society*, pp. 122-133, juillet 2021.
- 15 OpenAI, « DALL·E 2 pre-training mitigations », 28 juin 2022. [Accès le 4 mai 2023].
- 16 University of Waterloo Library, « ChatGPT and Generative Artificial Intelligence (AI): Potential for bias based on prompt », 27 avril 2023. [Accès le 5 mai 2023].
- 17 C. Bjork, « ChatGPT threatens language diversity. More needs to be done to protect our differences in the age of AI », 9 février 2023. [Accès le 5 mai 2023].
- 18 Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, W. Dai, A. Madotto et P. Fung, « Survey of Hallucination in Natural Language Generation », *arXiv preprint*, pp. 1-47, 7 novembre 2022.
- 19 OpenAI, « Introducing ChatGPT », 22 novembre 2022. [Accès le 1 mai 2023].
- 20 H. Alkaissi et S. I. McFarlane, « Artificial Hallucinations in ChatGPT: Implications in Scientific Writing », *Cureus*, vol. 15, n° 12, p. e35179, 2023.
- 21 S. Passi et M. Vorvoreanu, « Overreliance on AI: Literature review », juin 2022. [Accès le 3 mai 2023].
- 22 Z. Buçinca, M. B. Malaya et K. Z. Gajos, « To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making », *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, n° 1CSCW1, pp. 1-21, avril 2021.

- 23 M. Grawitch, « Confirmation Bias in the Era of Large AI », 1 mai 2023. [Accès le 11 mai 2023].
- 24 A. James, « ChatGPT has passed the Turing test and if you're freaked out, you're not alone », 29 mars 2023. [Accès le 4 mai 2023].
- 25 S. McLean, « The Environmental Impact of ChatGPT: A Call for Sustainable Practices In AI Development », 28 avril 2023. [Accès le 4 mai 2023].
- 26 IPCC, « Summary for Policymakers. Dans : Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty », Cambridge University Press, Cambridge, Royaume-Uni; New York, NY, É.-U., 2018.
- 27 Changements climatiques - Nations Unies, « Campagne Objectif zéro », [Accès le 5 mai 2023].
-

Date de modification :

2023-09-06