Guide de mise en œuvre pour les gestionnaires de systèmes d'intelligence artificielle

De : Innovation, Sciences et Développement économique Canada

Soutien à la mise en œuvre du code de conduite volontaire visant les systèmes avancés d'intelligence artificielle générative

Table des matières

Introduction

Ce guide a pour but d'aider les gestionnaires de systèmes d'intelligence artificielle (IA) à mettre en œuvre le Code de conduite volontaire visant un développement et une gestion responsables des systèmes d'IA générative avancés (le Code).

Les gestionnaires de systèmes d'IA exploitent un système d'IA pour fournir un produit ou un service aux utilisateurs. Ils peuvent le faire à des fins commerciales internes, par exemple pour trier les CV en vue de recruter des employés, ou pour des clients, par exemple pour fournir un service ou un produit à d'autres entreprises ou particuliers. La gestion d'un système d'IA peut inclure des activités telles que la mise en service d'un système, le contrôle de son fonctionnement, le contrôle de l'accès et la surveillance de son fonctionnement.

Le code établit six principes (sécurité, responsabilité, transparence, justice et équité, surveillance humaine, et validité et fiabilité) et dix-huit mesures qui peuvent être mises en œuvre par les développeurs et les gestionnaires de systèmes d'IA. Ces mesures sont conformes aux initiatives internationales visant à promouvoir une IA responsable, telles que les Principes directeurs internationaux du processus Hiroshima du G7 pour les organisations développant des systèmes d'IA avancés.

De par leur conception, les mesures et principes du Code sont énoncés de manière générale pour permettre une mise en œuvre flexible et pratique en fonction des différents profils d'entreprise et cas d'utilisation. Le présent guide vise à fournir des conseils et des suggestions plus détaillés, conformes au code, afin d'aider les gestionnaires à exploiter leurs systèmes d'IA de manière responsable. Il ne s'agit pas d'une liste de contrôle ni d'un ensemble de mesures rigides. Les organisations sont plutôt encouragées à utiliser ces informations pour éclairer leur approche de l'IA responsable, adaptée à leurs activités commerciales et à leurs cas d'utilisation, et proportionnelle au profil de risque de leurs activités.

Bien que le code ait été conçu à l'origine pour guider les développeurs et les gestionnaires de systèmes d'IA générative avancés, bon nombre des mesures qu'il contient décrivent de manière plus générale les pratiques responsables en matière d'IA. Ce guide peut être utilisé de manière appropriée pour informer la gouvernance d'un large éventail de systèmes d'IA, y compris ceux qui ne sont pas génératifs.

Les suggestions établies dans ce guide sont destinées à compléter les exigences d'autres politiques et lois au Canada, telles que les lois canadiennes sur la protection de la vie privée, la concurrence, la protection des consommateurs et le droit d'auteur, qui s'appliquent déjà au développement commercial et à la gestion des systèmes d'IA. En outre, d'autres politiques peuvent s'appliquer dans des secteurs particuliers, telles que les règles relatives au développement et à l'exploitation des systèmes d'IA en tant que dispositifs médicaux. Ce guide n'a pas pour but d'expliquer ces politiques, et les gestionnaires sont encouragés à se familiariser avec leurs obligations existantes en vertu de la loi au Canada.

Le cycle de vie des systèmes d'IA

Les systèmes d'IA sont des systèmes technologiques complexes composés de différents éléments, tels qu'un ou plusieurs modèles, et pouvant inclure une interface utilisateur. Leur création et leur fonctionnement impliquent généralement différents acteurs, depuis les entités participant à la collecte et à la préparation des données jusqu'à la gestion des systèmes d'IA après leur déploiement, en passant par le développement de modèles et le développement et la validation de systèmes. D'autres facteurs peuvent avoir une incidence sur la chaîne de valeur de l'IA et sur les différents rôles au sein de celle-ci, notamment la manière dont le système est déployé, par exemple au moyen d'une interface de programmation d'applications (API) ou d'autres méthodes.[Note de bas de page 1]

Le développement et le fonctionnement des systèmes d'IA sont complexes et itératifs, et différentes organisations peuvent constater que leurs activités ou opérations ne correspondent pas exactement à ce guide. Les gestionnaires sont encouragés à appliquer ce guide en fonction de leurs activités.

Une gouvernance sûre et responsable de l'IA commence par le développement, qui comprend des étapes préliminaires telles que l'idéation, la planification et la conception. Les développeurs d'IA sont des organisations ou des entités qui conçoivent et développent des systèmes d'IA et/ou des composants de systèmes d'IA tels que des modèles. Le développeur d'un système d'IA peut être une entité distincte du gestionnaire du système, ou une entité peut être à la fois développeur et gestionnaire. En outre, il peut y avoir plusieurs développeurs pour un seul système ou pour des composants de système tout au long de la chaîne de valeur de l'IA.

Généralement, le développement d'un système d'IA est entrepris avant le déploiement, en commençant par l'énoncé du problème, la collecte et la préparation des données avant de passer à la conception du modèle et du système, au développement du modèle, au réglage fin et aux tests. Une fois le système prêt à être déployé, le développeur peut le mettre à la disposition des gestionnaires en aval, qui gèrent les opérations du système en tant que produit ou service après le déploiement, ou le développeur peut gérer lui-même les opérations du système, pour fournir un produit ou un service à leurs utilisateurs ou clients.

Les gestionnaires jouent un rôle de gouvernance crucial dans le cycle de vie des systèmes d'IA en raison de leur place dans la chaine de valeur de l'IA. Bien que les gestionnaires ne puissent pas atténuer tous les risques – par exemple, ils ne sont pas en mesure de traiter toutes les questions liées au modèle ou à l'utilisation du système par les utilisateurs finaux – les gestionnaires de systèmes d'IA sont bien placés pour traiter les risques découlant de la conception et des choix opérationnels au niveau du système, en raison de leur proximité avec le contexte d'utilisation. Par exemple, les gestionnaires peuvent entreprendre une série d'activités, telles que garantir la transparence de la conception et des opérations du système, fournir une expérience utilisateur accessible, gérer les risques de cybersécurité, identifier et traiter les dérives du modèle, et identifier et signaler les incidents graves. Les gestionnaires sont aussi généralement bien placés pour remarquer et effectuer d'autres ajustements qui pourraient être nécessaires pour continuer à exploiter le système de manière sûre et transparente en tant que produit ou service (ou en tant que partie d'un produit ou d'un service). Pour mieux comprendre leurs rôles et responsabilités, et pour s'assurer qu'ils disposent des informations nécessaires pour gérer efficacement leur système, il est important que les gestionnaires travaillent avec les autres entités de la chaîne de valeur de l'IA.

Orientations pour les gestionnaires de systèmes d'IA

Le code recommande une série de mesures que les gestionnaires peuvent mettre en œuvre pour s'assurer que les systèmes d'IA sont exploités de manière sûre et responsable. Les sections suivantes proposent des bonnes pratiques pour soutenir la mise en œuvre de ces mesures, en tant que point de référence pour les organisations qui cherchent à mettre en place des pratiques de gouvernance responsable de l'IA. Bien que le code s'articule autour de six principes généraux, les mesures à prendre par les gestionnaires ne sont recommandées que dans le cadre de cinq de ces principes.

estion responsable de l'IA, les organisations doivent poser des bases solides pour la gouvernance de l'IA. Cela commence par l'établissement d'une vision claire de la manière

dont une organisation a l'intention d'utiliser l'IA, à quelles fins et dans quel contexte. Les organisations doivent également examiner leurs structures et pratiques organisationnelles existantes afin de déterminer la meilleure façon d'y ancrer leur gouvernance de l'IA.

Une autre étape préliminaire importante pour les gestionnaires consiste à faire preuve de la diligence requise lors du développement ou de l'acquisition d'un système d'IA, conformément aux mesures contenues dans le Code, avant le déploiement du système.

Les gestionnaires qui acquièrent des systèmes d'IA doivent mettre en place des processus d'approvisionnement rigoureux, car ces décisions déterminent les systèmes qui seront déployés. Les mesures que les gestionnaires peuvent envisager lors de l'acquisition de systèmes d'IA comprennent :

- Développer des critères d'évaluation standardisés qui évaluent les capacités techniques, les considérations éthiques et l'alignement avec les valeurs organisationnelles.
- Exiger des fournisseurs qu'ils fournissent une documentation complète sur le développement du système, les méthodologies de test et les limites connues.
- Créer des comités d'approvisionnement interfonctionnels comprenant des parties prenantes techniques, juridiques, éthiques et commerciales.
- Exiger des fournisseurs la transparence de l'architecture des modèles, des sources de données d'apprentissage et des mesures de performance.
- Mettre en œuvre des processus formels de diligence raisonnable pour la sélection des fournisseurs pour évaluer leurs antécédents en matière de pratiques responsables en matière d'IA.
- Répondre aux préoccupations en matière d'équité et de justice en intégrant des mesures de performance pertinentes dans les critères d'évaluation des fournisseurs, y compris les résultats des tests de biais sur diverses populations.
- Exiger des fournisseurs qu'ils démontrent leur conformité aux cadres réglementaires et aux normes industrielles en vigueur.

Une fois ces éléments fondamentaux en place, les organisations peuvent mettre en œuvre plus efficacement les meilleures pratiques décrites dans les sections suivantes en ce qui concerne la gestion des systèmes d'IA.

Meilleures pratiques en matière de sécurité

L'IA est une famille polyvalente de technologie qui sont utiles à de nombreuses fins différentes, notamment pour l'intégration dans différents types de produits et services. Cela signifie que le profil des organisations qui gèrent des systèmes d'IA variera considérablement, des petites et moyennes entreprises aux grandes entreprises, dans différents secteurs et industries, et que les systèmes gérés par les organisations varieront également considérablement, en fonction des capacités du système et de son contexte d'utilisation. Il est

s prennent le temps d'identifier les risques qui peuvent survenir dans

leur contexte opérationnel. La compréhension de ces risques est essentielle à la bonne gestion d'un système d'IA. Les risques qu'un système d'IA peut poser comprennent manque de fiabilité des résultats, le partage d'informations exclusives, le dysfonctionnement du système, l'échec pour les groupes vulnérables ou historiquement marginalisés, les vulnérabilités aux actes malveillants ou aux utilisations abusives par les utilisateurs ou la création d'effets d'entraînement pouvant avoir un impact sur la société dans son ensemble. Les risques pour la sécurité associés aux systèmes d'IA diffèrent en fonction de leur contexte d'utilisation et peuvent être plus ou moins graves selon ce contexte et la manière dont les différents acteurs gèrent ces risques.

Afin d'atténuer les risques et de promouvoir une utilisation sûre des systèmes d'IA, le code recommande aux gestionnaires des systèmes d'IA de :

Effectuer une évaluation complète des répercussions négatives potentielles raisonnablement prévisibles, notamment des risques associés à une utilisation inappropriée ou malveillante du système.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Identifier et évaluer les risques pouvant survenir en raison du fonctionnement du système, y compris les risques découlant : i) de l'utilisation prévue du système; ii) des utilisations non intentionnelles mais raisonnablement prévisibles, mauvaises utilisations ou utilisations malveillantes; iii) d'autres risques opérationnels, et classer ces risques en fonction de leur probabilité, des personnes ou des éléments susceptibles d'être affectés et de la gravité des impacts, y compris leur ampleur et leur portée. Cela doit être régulièrement revu et mis à jour.
- Prendre en compte un éventail de risques potentiels, notamment les risques de biais, de protection des données et de confidentialité, les risques découlant de l'utilisation du système à des fins de désinformation ou à d'autres fins malveillantes, les risques liés à la cybersécurité, à la conformité et à la réputation.
- Impliquer diverses parties prenantes internes (notamment les ressources humaines, les technologies de l'information, les services juridiques, la conformité, les produits, le service clientèle et les unités commerciales) dans les processus d'évaluation des risques afin de garantir la prise en compte de multiples perspectives organisationnelles.
- Identifier et évaluer comment les droits fondamentaux peuvent être affectés par le fonctionnement du système d'IA.
- Identifier et évaluer comment les groupes vulnérables (tels que les enfants, les personnes âgées ou les groupes historiquement marginalisés) peuvent être affectés par le fonctionnement du système d'IA.
- Élaborer des scénarios d'impact détaillés pour différents groupes d'utilisateurs et cas d'utilisation.
- Organiser des ateliers structurés avec diverses parties prenantes afin d'identifier les impacts potentiels, y compris les effets potentiels de deuxième et troisième ordre du déploiement du système.

- Mettre en œuvre une analyse régulière de l'horizon pour détecter les risques émergents et les vecteurs de menace, tels que les attaques malveillantes.
- Effectuer des tests pour identifier les vulnérabilités du système d'IA et de son environnement de déploiement,, y compris des tests adversatifs et des tests réguliers pour détecter les dysfonctionnements.

Meilleures pratiques en matière de responsabilité

En plus de passer du temps à identifier les risques qui peuvent survenir dans leur contexte opérationnel, il est tout aussi important que les organisations mettent en place des politiques et des procédures pour faire face à ces risques. Cela implique de socialiser ces informations avec leurs employés, qui peuvent être chargés de maintenir le système, d'identifier les incidents et d'y répondre, de s'adresser aux utilisateurs finaux et de surveiller son fonctionnement. En établissant des pratiques, des politiques et des procédures pour gérer et traiter les risques, les organisations et les employés comprendront leurs responsabilités et pourront réagir rapidement et de manière appropriée aux incidents et aux problèmes lorsqu'ils se produisent.

Il est essentiel pour une gouvernance responsable de l'IA de s'assurer que l'organisation, y compris ses employés et les autres personnes avec lesquelles elle collabore, comprend ses responsabilités et sait comment agir en cas de problème. Une bonne connaissance de l'IA à tous les niveaux de l'organisation permet une meilleure gestion des risques.

Pour établir ces normes, il est important de mettre en place et de maintenir un cadre de gestion des risques. Le code recommande aux gestionnaires des systèmes d'IA de :

Mettre en œuvre un cadre complet de gestion des risques adapté à la nature et au profil de risque des activités. Ce cadre comprend la mise en place de politiques, de procédures et de formations pour veiller à ce que les employés connaissent bien leurs responsabilités et les pratiques de gestion des risques de l'organisation.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Élaborer et tenir à jour un cadre de gestion des risques qui explique comment les risques identifiés sont atténués (par le gestionnaire ou par d'autres acteurs de la chaîne de valeur), qui détient les pouvoirs de décision et quels sont les délais de réaction attendus pour faire face aux risques.
- Établir une politique indiquant quand désactiver ou cesser les opérations des systèmes, ainsi qu'une procédure de mise hors service des systèmes de manière à atténuer les risques.
- Établir des politiques pour le personnel, y compris des formations, afin de faire connaître les attentes de l'organisation, les procédures et les autorités en cas d'incident. Cette formation doit être régulièrement mise à jour pour refléter la nature évolutive des risques et des meilleures pratiques en

- Offrir des formations spécifiques aux différents rôles et des possibilités de perfectionnement. Il peut s'agir, par exemple, d'une formation générale pour tous les employés sur l'utilisation responsable des outils d'IA générique, et d'une formation spécialisée pour les équipes techniques sur le développement, le déploiement et la maintenance de l'IA.
- Mettre en place un contrôle des versions pour le système d'IA et ses composants, et établir un processus formel de gestion des changements pour suivre et évaluer l'impact des mises à jour et des modifications.
- Tenir un répertoire centralisé de toute la documentation relative au système d'IA, y compris les évaluations des risques, les rapports d'incidents, les modifications du système, le retour d'information des utilisateurs et les mesures de performance, avec une période de conservation appropriée.
- Fournir des conseils clairs aux utilisateurs, y compris des politiques d'utilisation acceptable qui décrivent l'utilisation appropriée du système, les activités interdites, les responsabilités des utilisateurs et les conséquences potentielles d'une mauvaise utilisation. Ces directives doivent être facilement accessibles, rédigées dans un langage simple et mises à jour régulièrement.

Les cadres d'évaluation et de gestion des risques de l'organisation devront être régulièrement revus et mis à jour afin d'intégrer les nouvelles informations et de s'assurer qu'ils continuent à répondre aux besoins de l'organisation.

Afin de promouvoir une culture de la responsabilité tout au long de la chaîne de valeur de l'IA et dans l'ensemble de l'industrie, le code recommande également aux gestionnaires des systèmes d'IA de :

Transmettre l'information et les pratiques exemplaires visant la gestion des risques aux entreprises qui jouent des rôles complémentaires dans l'écosystème.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Publier les résultats dépersonnalisés de l'évaluation des risques et les stratégies d'atténuation;
- Collaborer avec d'autres organisations pour mettre au point des outils normalisés d'évaluation des risques;
- Contribuer aux forums et groupes de travail du secteur sur la gestion des risques liés à l'IA.

Bonnes pratiques en matière de surveillance humaine

En raison de leur place dans la chaîne de valeur de l'IA, les gestionnaires sont les mieux placés pour s'assurer que les systèmes ne fonctionnent pas de manière totalement autonome et qu'une personne est présente pour surveiller, mettre à jour et maintenir les opérations du système. Ce rôle peut également permettre d'identifier et de traiter rapidement les incidents lorsqu'ils surviennent, garantissant ainsi une expérience utilisateur fluide et atténuant le risque qu'un petit incident ne devienne grave.

 recommande aux gestionnaires des systèmes d'IA de :

Surveiller le fonctionnement du système pour s'assurer qu'il n'est pas utilisé à des fins nuisibles ou qu'il n'a pas des répercussions néfastes après qu'on l'ait rendu accessible, y compris par l'intermédiaire de canaux de rétroaction tiers, et informer le développeur et/ou mettre en œuvre des contrôles d'utilisation au besoin pour atténuer les biais.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Mettre en place des procédures de suivi et d'évaluation des systèmes d'IA déployés.
- Développer des systèmes de détection automatisés pour les utilisations potentiellement préjudiciables.
- Surveiller les performances du système d'IA en fonction de différents groupes démographiques ou d'autres catégories pertinentes.
- Surveiller le comportement des utilisateurs à l'égard du système et leur permettre de donner leur avis sur leur expérience du système.
- Recueillir et analyser les commentaires des utilisateurs, les rapports d'incidents et d'autres données pertinentes.
- Procéder à des évaluations régulières des performances du modèle afin de détecter et de corriger les dérives du modèle.
- Créer plusieurs canaux de retour d'information pour les utilisateurs et les parties concernées.
- Mettre en place des procédures d'examen régulier des incidents signalés.
- Mettre en œuvre des mécanismes permettant de traiter et d'atténuer les utilisations ou les impacts préjudiciables.
- Maintenir des équipes de réponse aux incidents avec des procédures d'escalade claires.
- Établir des protocoles et des canaux de communication pour informer les développeurs des problèmes identifiés ou des préoccupations en matière de performance, y compris le partage des données de surveillance pertinentes.

Bonnes pratiques en matière de transparence

La place qu'occupent les gestionnaires dans la chaîne de valeur de l'IA signifie qu'ils sont bien placés pour assurer la transparence du système auprès des utilisateurs. Des pratiques de transparence solides peuvent promouvoir la confiance, améliorer la satisfaction des utilisateurs, atténuer les risques de mauvaise utilisation et de dysfonctionnement, et garantir que le système continue à fonctionner comme prévu. Pour améliorer la transparence, le code recommande aux gestionnaires de systèmes d'IA de :

Veiller à ce que les systèmes qui pourraient être confondus avec des êtres humains soient clairement et visiblement identifiés comme des systèmes d'IA.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Élaborer et mettre en œuvre des protocoles normalisés d'identification de l'IA pour tous les types d'interaction (par exemple, les chatbots, les courriels, le téléphone), y compris des avis de divulgation systématiques.
- Fournir aux utilisateurs des informations gratuites et accessibles sur la nature et les capacités des systèmes d'IA, y compris des informations sur la manière dont ils sont développés, exploités et entretenus.
- Examiner si les choix de l'interface utilisateur, par exemple l'utilisation de pronoms personnels, l'auto-attribution d'états mentaux ou d'émotions par les chatbots en contact avec l'utilisateur, sont nécessaires et appropriés pour le cas d'utilisation.
- Mettre en place des processus pour documenter le fait que le contenu a été généré par un système d'IA, par exemple en ajoutant des étiquettes normalisées aux extrants de l'IA lorsqu'ils sont stockés ou distribués.

S'il est essentiel de maintenir la transparence en ce qui concerne les cas où un produit ou un service utilisant l'IA pourrait être confondu avec un être humain, il est tout aussi important de veiller à ce que les utilisateurs finaux sachent quand l'IA est utilisée pour façonner leur expérience d'un produit ou d'un service basé sur l'IA, et comment le système d'IA contribue à leur expérience. Cela favorise le choix de l'utilisateur et lui permet de mieux comprendre quand il est en contact avec un système d'IA et ce qu'il fait. Il est également recommandé de faire preuve de transparence en ce qui concerne les capacités, les risques et les limites du système, ainsi que les attentes de l'opérateur quant à la manière dont les utilisateurs peuvent utiliser le système, et ce qui est considéré par l'entreprise comme une utilisation abusive.

Bonnes pratiques pour la validité et la fiabilité

Les performances d'un système d'IA sont valides lorsqu'il fonctionne comme prévu pour les utilisations pour lesquelles il a été conçu. Un système est fiable lorsqu'il fonctionne comme prévu dans de nombreux types de scénarios, y compris des scénarios divers ou inhabituels. La validité et la fiabilité réfèrent donc à la performance optimale et fiable du système dans de nombreuses conditions différentes.

Pour garantir que les systèmes d'IA fonctionnent de manière optimale et fiable, les gestionnaires doivent envisager de tester les performances de leur système par rapport à des données réelles diverses et dans des conditions défavorables ou difficiles, de procéder à de nouveaux tests après des mises à jour importantes, d'identifier et de documenter les limites du système et, en particulier dans les applications à enjeux élevés où les erreurs pourraient avoir des conséquences importantes, de vérifier les extrants critiques.

Bien que les mesures décrites précédemment soient recommandées aux gestionnaires des systèmes d'IA accessibles au public et non accessibles au public, le code recommande en outre aux gestionnaires des systèmes d'IA accessibles au public de prendre des mesures supplémentaires pour protéger la validité et la

n prenant les mesures suivantes :

Effectuer une évaluation des risques en matière de cybersécurité et mettre en œuvre des mesures adaptées pour atténuer les risques, notamment en ce qui a trait à l'empoisonnement des données.

Les gestionnaires peuvent envisager les étapes suivantes pour mettre en œuvre cette mesure :

- Mettre en œuvre des protocoles complets de tests de sécurité.
- Créer des outils et des procédures automatisés d'analyse de la sécurité.
- Mettre en place des procédures d'audit de sécurité régulières.
- Élaborer des plans d'intervention en cas de violation de la sécurité.
- Maintenir des systèmes de surveillance de la sécurité pour la détection précoce des menaces.
- Adopter les meilleures pratiques générales en matière de cybersécurité.

Recueil de ressources pertinentes pour les gestionnaires de systèmes d'IA

Reconnaissant le rôle important que les gestionnaires des systèmes d'IA jouent dans la gouvernance de l'IA, ce guide établit un point de départ pour les organisations qui cherchent à obtenir des informations sur l'IA responsable.

Il existe de nombreuses ressources pour aider les entreprises d'IA à réussir leur gouvernance de l'IA. La liste ci-dessous est un sous-ensemble de ressources qui peuvent présenter un intérêt particulier pour les gestionnaires de systèmes d'IA qui cherchent à mettre en œuvre les mesures prévues par le code.

Organisation	Document	Date	Brève description
Organisation internationale de normalisation (ISO)	ISO/IEC 42001:2023 – Système de management de l'IA	2023	La norme ISO 42001 établit une norme de gestion des risques pour les entreprises qui développent, fournissent ou utilise un système d'IA.
Institut des normes de gouvernance numérique	CAN/DGSI 101 – Conception et utilisation éthiques de l'intelligence artificielle par les petites et moyennes organisations	2025	Fournit un cadre permettant aux petites et moyennes organisations d'évaluer et de gérer les risques liés aux systèmes d'IA et de s'aligner sur les orientations internationales et canadiennes en matière d'IA sûre et responsable.

Organisation	Document	Date	Brève description
Institut national des normes et de la technologie (NIST)	NIST AI Risk Management Framework	2023	Le NIST est une agence du ministère du commerce des États-Unis. Le RMF du NIST fournit des ressources pour comprendre les risques, les impacts et les mesures d'atténuation tout au long de la chaîne de valeur de l'IA.
Institut national des normes et de la technologie (NIST)	NIST AI 600-1 AI RMF Generative AI Profile	2024	Ce document est un profil intersectoriel et une ressource complémentaire au RMF pour l'IA générative.
Bureau européen de l'IA	Répertoire vivant pour favoriser l'apprentissage et l'échange sur l'alphabétisation en IA	Base de données vivante	Ce référentiel fournit des exemples de pratiques d'alphabétisation en matière d'IA en cours parmi les fournisseurs et les déployeurs de systèmes d'IA.
Bureau européen de l'IA	Code de pratique pour l'IA à usage général	2025 (ébauche)	Le code est un document d'orientation destiné aux fournisseurs de modèles d'IA à usage général, qui leur permet de démontrer qu'ils respectent la loi sur l'IA tout au long du cycle de vie des modèles. Bien que le code s'applique principalement aux développeurs, il est également pertinent pour les gestionnaires qui supervisent les systèmes d'IA à impact élevé. Il fournit des lignes directrices sur l'évaluation et l'atténuation des risques, ainsi que sur la gouvernance.
Organisation de coopération et de développement économiques (OCDE)	OECD AI Incidents Monitor (AIM)	Base de données vivante	Le Moniteur des incidents d'IA de l'OCDE (AIM) documente les incidents et les dangers liés à l'IA afin d'aider les décideurs politiques, les praticiens de l'IA et toutes les parties prenantes dans

Organisation	Document	Date	Brève description
			le monde entier à mieux comprendre les risques et les dangers des systèmes d'IA.
Organisation de coopération et de développement économiques (OCDE)	Catalogue of Tools & Metrics for Trustworthy AI	Base de données vivante	Ce catalogue facilite la recherche d'outils et de mesures en offrant un guichet unique pour les approches, les mécanismes et les pratiques utiles en matière d'IA digne de confiance.
Organisation de coopération et de développement économiques (OCDE)	Framework for the Classification of AI systems	2022	Un outil simple d'utilisation pour caractériser l'application d'un système d'IA déployé dans un contexte spécifique. Le cadre classe les systèmes et applications d'IA selon les dimensions suivantes : Personnes et planète, Contexte économique, Données et entrée, Modèle d'IA et Tâche et sortie. Chacune d'entre elles possède ses propres propriétés et attributs ou sous-dimensions pertinentes pour évaluer les considérations politiques de systèmes d'IA particuliers.
Massachusetts Institute of Technology (MIT)	MIT AI Risk Repository	Base de données vivante	Une base de données vivante et complète de plus de 1000 risques liés à l'IA, classés en fonction de leur cause et de leur domaine de risque.
AI Standards Hub	Standards Database	Base de données vivante	La base de données des normes est un catalogue consultable couvrant plus de 400 normes pertinentes en cours d'élaboration ou déjà publiées par un certain nombre d'organismes de normalisation de premier plan.
UK Department for Science, Innovation and	AI Management Essentials (AIME) tool	2024 (ébauche)	AIME est un outil d'auto-évaluation qui vise à aider les organisations à évaluer et à mettre en œuvre des systèmes et

Organisation	Document	Date	Brève description
Technology's (DSIT)			des processus de gestion responsables de l'IA. AIME peut être utilisé par toute organisation qui développe, fournit ou utilise des services qui utilisent des systèmes d'IA dans le cadre de ses activités commerciales. AIME est adapté à tous les secteurs et peut être utilisé par des organisations de différentes tailles. Cependant, il est principalement destiné aux petites et moyennes entreprises et aux start-ups qui rencontrent des obstacles lorsqu'elles naviguent dans le paysage en constante évolution des normes et des cadres de gestion de l'IA.

Notes de bas de page

Note de bas de page 1

L'OCDE définit un **système d'IA** comme : un système qui fonctionne grâce à une machine et capable d'influencer son environnement en produisant des résultats (tels que des prédictions, des recommandations ou des décisions) pour répondre à un ensemble donné d'objectifs. Il utilise les données et les intrants généré par la machine et/ou apportés par l'homme afin de (i) percevoir des environnements réels et/ou virtuels; (ii) produire une représentation abstraite de ces perceptions sous forme de modèles issus d'une analyse automatisée (ex. l'apprentissage automatisé) ou manuelle; et (iii) utiliser les déductions du modèle pour formuler différentes options de résultats. Les systèmes d'IA sont conçus pour fonctionner de façon plus ou moins autonome.

Retour à la référence de la note de bas de page 1