GUIDE SUR L'UTILISATION RESPONSABLE DE L'INTELLIGENCE ARTIFICIELLE AGENTIQUE

APPLICABLE AUX OUTILS D'INTELLIGENCE ARTIFICIELLE AGENTIQUE PUBLICS

MINISTÈRE DE LA CYBERSÉCURITÉ
ET DU NUMÉRIQUE

Bâtir l'avenir

Québec

RÉDACTION
La Direction du centre d'expertise en intelligence artificielle du Sous-ministériat adjoint aux partenariats et à l'intelligence artificielle du ministère de la Cybersécurité et du Numérique

Si vous éprouvez des difficultés techniques ou si vous souhaitez obtenir une version adaptée du document, veuillez communiquer avec la Direction des communications :

Direction des communications
Ministère de la Cybersécurité et du Numérique
425, rue Jacques-Parizeau, 2e étage
Québec (Québec) G1R 2G3
Courriel : information@mcn.gouv.qc.ca

TABLE DES MATIÈRES

1. OBJET ET PORTÉE DU GUIDE

Ce guide vise à aider les organismes publics à évaluer, à autoriser, à encadrer, à surveiller et à faire évoluer les usages de l'intelligence artificielle (IA) agentique de façon responsable, sécuritaire et proportionnée. Il ne remplace pas les cadres existants sur l'utilisation responsable de l'IA générative, la protection des renseignements personnels, la sécurité de l'information, les ressources informationnelles et la gouvernance des données. Il propose plutôt une application pratique de ces cadres aux systèmes capables d'agir, de consulter des outils, de suivre des étapes et d'intervenir dans un processus exécuté au moyen d'un système d'information.

L'idée centrale est simple : plus un système d'IA peut agir, accéder à des outils, interagir avec des systèmes ou enchaîner des étapes de manière autonome, plus les exigences en matière de gouvernance, de sécurité, de supervision humaine, de traçabilité et de contrôle des coûts doivent être élevées.

L'objectif n'est pas de freiner l'expérimentation, mais de proposer un cadre simple permettant de distinguer les usages acceptables, ceux qui exigent des validations renforcées et ceux qui devraient être reportés. Le guide met donc l'accent sur les conditions minimales d'acceptabilité : finalité claire, valeur démontrable, responsabilités attribuées, accès limités, supervision humaine, journalisation, sécurité, protection des renseignements personnels, contrôle des coûts et capacité de retour au mode manuel.

L'encadrement proposé vise à permettre une expérimentation responsable de l'IA agentique afin que les organismes publics puissent développer progressivement leur capacité organisationnelle à utiliser ces systèmes lorsque leur valeur est démontrable, tout en maintenant l'imputabilité publique, la sécurité, la protection des renseignements personnels, le contrôle des coûts et la continuité des services.

Les attentes présentées dans ce guide doivent être appliquées selon une logique de proportionnalité, en tenant compte du niveau d'autonomie de l'agent d'IA, des données utilisées, des systèmes touchés, des actions possibles, des coûts anticipés et des effets potentiels sur les personnes ou les services publics.

> Dans le présent guide, l'expression « IA agentique » désigne une catégorie de systèmes d'intelligence artificielle capables de poursuivre un objectif, d'orchestrer des étapes, de consulter des sources ou d'utiliser des outils dans un périmètre défini. L'expression « agent d'IA » désigne une instance concrète d'un tel système, utilisée dans un cas d'usage, un processus ou un environnement déterminé.

2. COMPRENDRE L'IA AGENTIQUE

L'IA agentique doit être comprise comme une capacité d'action numérique. Contrairement à un simple outil de génération de contenu, elle peut combiner des instructions, des sources d'information, des outils, des connecteurs, des systèmes et des règles opérationnelles afin de soutenir l'exécution d'une tâche ou d'un processus.

Cette capacité peut prendre différentes formes. Un agent d'IA peut, selon les limites établies par l'organisation, rechercher de l'information dans des sources autorisées, comparer des données, préparer une recommandation, proposer une action, remplir une étape d'un processus ou interagir avec un système. Son autonomie doit toutefois être comprise comme une capacité encadrée, et non comme une autorisation générale d'agir sans supervision humaine.

La distinction avec l'IA générative est importante : l'IA générative est principalement utilisée pour produire du contenu en réponse à une demande, alors que l'IA agentique peut tenter d'exécuter une tâche en orchestrant des actions, des outils et des systèmes pour soutenir une tâche composée de plusieurs étapes. Ses effets ne se limitent donc pas à la qualité d'une réponse produite : ils peuvent aussi concerner les accès, les actions, les coûts, la continuité des opérations, la sécurité et l'imputabilité publique.

3. PRINCIPES DIRECTEURS

Voici les principes directeurs qui encadrent le recours à l'IA agentique. Ils visent à assurer un usage proportionné, sécurisé et responsable de ces systèmes, en fonction de leur niveau d'autonomie, de leur portée et des effets qu'ils peuvent produire.

- L'IA agentique ne doit être utilisée que lorsqu'une solution moins autonome ne permet pas d'atteindre adéquatement l'objectif poursuivi;
- Le cas d'usage doit être clairement défini, circonscrit, mesurable et justifié par une valeur ajoutée démontrable;
- Les accès, données, outils et actions autorisés doivent être limités au strict nécessaire, selon le principe de minimisation;
- Les actions sensibles doivent demeurer assujetties à une validation humaine préalable;
- Les journaux doivent permettre de retracer et comprendre ce que l'agent a consulté, proposé, exécuté et généré comme coûts;

- Un mécanisme d’arrêt, de suspension et de retour au mode manuel doit être prévu dès la conception;

- L’organisme public demeure responsable des effets, erreurs, décisions préparatoires et actions associées à l’agent.

Ces principes directeurs doivent être appliqués selon une logique de proportionnalité. Ainsi, un agent utilisé uniquement pour préparer une synthèse interne n’exige pas le même niveau d’encadrement qu’un agent capable de modifier des données officielles, d’interagir avec plusieurs systèmes ou de soutenir une décision ayant un effet sur une personne. Toutefois, même les usages à faible risque doivent reposer sur des limites claires et sur une validation humaine des résultats avant toute utilisation officielle.

En contexte public, les principes pour une utilisation responsable de l’IA, notamment les principes de la responsabilité, de la transparence, de l’explicabilité, de fiabilité et de la robustesse, de la sécurité, de l’inclusion et d’équité, de la compétence et de l’autonomie et de la supervision humaine, doivent demeurer applicables tout au long du cycle de vie de l’agent. L’organisme doit pouvoir expliquer pourquoi l’agent est utilisé, ce qu’il peut faire, ce qu’il ne peut pas faire, quelles données il consulte, quelles actions il peut proposer ou exécuter, et à quels moments une personne doit intervenir. Ces principes doivent être traduits en exigences concrètes selon le niveau d’autonomie, les données utilisées, les systèmes touchés et les effets possibles sur les personnes.

4. QUAND UTILISER OU ÉVITER UN AGENT D’IA

Le recours à un agent d’IA doit être une décision d’affaires et non seulement une décision technologique. L’organisation devrait d’abord déterminer si l’autonomie de l’agent est réellement nécessaire, si les bénéfices attendus compensent les risques et si les conditions de contrôle sont réunies. Cette analyse permet d’éviter d’utiliser un agent lorsque le besoin pourrait être satisfait par une solution plus simple, plus transparente ou moins risquée.

Avant d’autoriser un agent d’IA, l’organisation doit vérifier si le besoin peut être satisfait par une amélioration de processus, une automatisation classique, un formulaire structuré, un assistant d’IA sans capacité d’action ou un outil déjà encadré. Le recours à l’IA agentique devrait être réservé aux situations où le processus comporte plusieurs étapes, mobilise plusieurs sources, nécessite une adaptation au contexte ou produit une valeur ajoutée mesurable justifiant le niveau d’autonomie envisagé.

Les premiers pilotes devraient être internes, réversibles, limités dans le temps, peu sensibles, supervisés, journalisés, mesurables et assortis d’un plafond budgétaire. Les usages touchant des droits, des décisions administratives, des communications externes,

des transactions financières ou des modifications de données officielles devraient être évités en première phase, sauf encadrement renforcé et validation spécialisée.

Un cas d'usage est généralement plus approprié lorsqu'il est interne, répétitif, documenté, contrôlable, mesurable et réversible. Il devient plus risqué lorsqu'il touche des droits, des prestations de services, des communications externes, des renseignements personnels sensibles, des décisions administratives, des engagements financiers ou des systèmes critiques. Dans ces situations, l'agent ne devrait jamais agir seul : il devrait soutenir une personne responsable, fournir une analyse préparatoire ou proposer une action soumise à validation.

La qualification devrait également vérifier les conditions d'exclusion. L'usage devrait être écarté si le processus est mal défini, si les règles d'affaires sont instables, si les données sont de qualité insuffisante, si les accès ne peuvent pas être limités, si les coûts ne peuvent pas être suivis, si la supervision humaine n'est pas réaliste ou si l'organisation ne peut pas reprendre rapidement le traitement en mode manuel.

Point de vigilance — Un agent d'IA ne devrait pas être retenu par défaut : il doit être justifié par un besoin réel d'autonomie, une valeur mesurable et des conditions de contrôle suffisantes.

5. PARCOURS D'AUTORISATION EN TROIS ÉTAPES

L'autorisation d'un agent d'IA devrait suivre un parcours simple en trois étapes. La première consiste à confirmer que le cas d'usage justifie réellement une capacité agentique. La deuxième vise à vérifier si les conditions d'un pilote contrôlé sont réunies. La troisième permet de déterminer si le passage en production est acceptable, contrôlé et soutenable dans le temps.

- **Étape 1 – Filtrer le cas d'usage** : vérifier la présence d'un processus multiétapes, de données dynamiques, d'entrées non structurées, de chemins variables, de volumes élevés, d'indicateurs mesurables et d'un besoin réel d'autonomie.

- **Étape 2 – Vérifier la préparation du pilote** : confirmer les objectifs, indicateurs, données, accès, architectures, coûts, garde-fous, supervision, journalisation, continuité, essais, critères d'arrêt et responsabilités.

- **Étape 3 – Vérifier la préparation à la production** : valider les résultats du pilote, obtenir l'approbation requise, activer la surveillance, consigner l'agent au registre approprié, tester le retour manuel, former l'équipe et planifier les revues périodiques.

Ce parcours évite de traiter tous les projets de la même façon. Un usage expérimental, interne et réversible peut être autorisé avec un encadrement proportionné, tandis qu'un agent destiné à modifier des données, à interagir avec plusieurs systèmes ou à soutenir une décision sensible doit faire l'objet de validations renforcées.

6. CHOISIR LA BONNE SOLUTION

Le choix d'un agent d'IA ne devrait pas être automatique. Avant de retenir une solution agentique, l'organisation devrait comparer le besoin avec des options plus simples, comme une automatisation classique ou un assistant d'IA. Le tableau suivant propose une grille de comparaison synthétique.

Critère	**Automatisation classique**	**Assistant d'IA**	**Agent d'IA**
Nature du besoin	Étapes fixes, répétitives et prévisibles.	Aide à la rédaction, à la recherche, à la synthèse ou à l'analyse.	Processus variable, multiétapes, fondé sur plusieurs sources ou outils.
Rôle de l'humain	Conçoit les règles et traite les exceptions.	Formule la demande, vérifie et adapte le résultat.	Fixe les objectifs, les limites, les validations et surveille l'exécution.
Gouvernance requise	Faible à modérée, selon le processus.	Modérée, centrée sur les données, les résultats et les usages.	Élevée : supervision, journalisation, garde-fous, arrêt, coûts et incidents.
Risque si mal utilisé	Blocage ou erreur si les règles changent.	Réponse inexacte ou non vérifiée.	Action incorrecte, coût imprévu, incident de sécurité ou effet en cascade.

7. RISQUES PROPRES À L'IA AGENTIQUE

L'IA agentique reprend les risques associés à l'IA générative, notamment l'inexactitude, les biais, les hallucinations, l'usage non autorisé de données, les enjeux de confidentialité et les limites d'explicabilité. Elle ajoute toutefois des risques particuliers liés à sa capacité d'agir, d'enchaîner des étapes, de consulter des outils et d'interagir avec des systèmes d'information.

- Actions non prévues ou mal exécutées, surtout lorsque l'agent interprète incorrectement une instruction ou un contexte.
- Interopérations inattendues avec des systèmes, applications, connecteurs ou sources de données.
- Propagation d'erreurs dans un processus composé de plusieurs étapes.
- Injection d'instructions, contournement de règles ou manipulation du comportement de l'agent par une source externe.
- Exfiltration ou exposition de renseignements personnels, confidentiels ou sensibles.
- Élévation de privilèges, usage abusif des accès ou actions au-delà du périmètre autorisé.
- Dérive fonctionnelle lorsque l'agent est utilisé pour des tâches différentes de celles autorisées.
- Dépendance opérationnelle excessive ou perte de capacité à reprendre rapidement le traitement en mode manuel.
- Hausse imprévue des coûts lorsque l'agent multiplie les appels d'outils, les requêtes, les étapes ou le volume de contexte traité.

Point de vigilance — Les risques des agents d'IA ne concernent pas seulement la qualité des réponses : ils concernent aussi les actions, les accès, les coûts, les incidents et les effets en cascade.

8. EXIGENCES MINIMALES AVANT UN PILOTE

Avant d'autoriser un projet pilote en IA agentique, un ensemble minimal d'exigences doit être satisfait afin d'assurer un déploiement encadré, proportionné et conforme aux responsabilités de l'organisme public.

- Désigner un propriétaire d'affaires, un responsable technologique et des personnes responsables de la supervision.
- Décrire l'objectif, la portée, la durée, les utilisateurs visés et les indicateurs de succès.
- Documenter le niveau d'autonomie, les actions permises, les actions interdites et les limites d'usage.
- Identifier les données utilisées, leur classification, leur provenance, leur qualité et leurs droits d'usage.
- Évaluer la présence de renseignements personnels et réaliser une EFVP lorsque requise.
- Limiter les accès, les outils, les connecteurs et les systèmes selon le principe du moindre privilège.
- Prévoir les points de validation humaine, les seuils d'escalade et les critères de suspension.
- Vérifier les risques de cybersécurité, notamment l'injection d'instructions, l'exfiltration, l'usage abusif, l'élévation de privilèges et la dérive fonctionnelle.
- Définir les journaux requis, leur conservation et leur protection contre toute modification par l'agent.
- Établir un mécanisme d'arrêt, une procédure d'incident, un retour manuel et un plafond de consommation.

L'autorisation du pilote devrait être appuyée par une documentation légère, proportionnée à la nature du cas d'usage et au niveau de risque, afin de soutenir une décision éclairée. Cette documentation peut notamment préciser la valeur attendue, les scénarios d'utilisation, les personnes autorisées, les actions permises ou interdites, les données accessibles, les environnements utilisés, les dépendances technologiques, les validations requises, les seuils d'alerte et les critères d'arrêt. Les risques identifiés doivent être traduits en mesures concrètes avant le lancement du pilote : restrictions d'accès, limites d'action, validation humaine, essais contrôlés, surveillance des journaux, seuils de

suspension, formation des utilisateurs et suivi budgétaire. Le pilote doit permettre de confirmer que ces mesures fonctionnent dans des conditions réelles, sans exposer l'organisation à des impacts irréversibles.

Point de vigilance — Aucun pilote ne devrait commencer sans propriétaire désigné, accès limités, supervision humaine, journalisation, critères d'arrêt et plafond de consommation.

9. SUPERVISION, TRAÇABILITÉ ET SÉCURITÉ

La supervision humaine doit être réelle, documentée et proportionnée au risque. Elle est obligatoire lorsque l'agent modifie une donnée officielle, déclenche une action dans un système, produit une recommandation ayant une incidence sur une personne ou une organisation, utilise des renseignements personnels ou confidentiels, communique avec un tiers, engage une dépense ou traite un cas ambigu, complexe ou litigieux.

Les journaux doivent permettre de reconstituer la demande initiale, les sources consultées, les données utilisées, les outils appelés, les étapes réalisées, les résultats produits, les actions proposées ou exécutées, les validations humaines, les erreurs, les incidents et les coûts générés. Ils doivent être conservés dans un environnement que l'agent ne peut pas modifier.

L'agent ne doit jamais pouvoir élargir lui-même ses permissions, modifier ses propres règles, effacer ses journaux, désactiver ses garde-fous ou contourner les validations humaines prévues.

La supervision peut prendre différentes formes selon le risque : validation avant exécution, surveillance pendant l'exécution ou revue après coup pour les actions internes, réversibles et de faible impact. Toutefois, les actions à incidence significative doivent être validées avant leur exécution. La supervision ne doit pas être symbolique : la personne responsable doit disposer de l'information nécessaire pour comprendre l'action proposée, l'accepter, la modifier, la refuser ou l'escalader.

Les mécanismes d'escalade doivent être définis à l'avance. Une escalade devrait être déclenchée lorsque les sources se contredisent, lorsque l'information est manquante ou incohérente, lorsque le cas est ambigu ou litigieux, lorsque le niveau d'incertitude est élevé, lorsque les coûts dépassent un seuil ou lorsqu'un comportement inattendu est observé. Cette approche permet de préserver le jugement humain dans les situations sensibles.

La capacité d'arrêt est une exigence centrale. L'organisation doit savoir qui peut suspendre l'agent, dans quelles circonstances, comment bloquer ses accès, comment préserver les journaux, comment aviser les personnes concernées, comment reprendre manuellement le processus et comment documenter l'incident. Un agent qui ne peut pas être interrompu rapidement ne devrait pas être autorisé.

Point de vigilance – La supervision doit permettre de comprendre, de valider, de corriger, de refuser ou de suspendre une action avant qu'elle ne produise un effet important.

10. GOUVERNANCE DES COÛTS

Les agents d'IA peuvent générer des coûts variables selon les modèles utilisés, le volume de contexte traité, le nombre d'outils appelés, la durée des tâches, la complexité des étapes et le niveau d'autonomie. Aucun pilote ne devrait être autorisé sans une compréhension du modèle de facturation, des principaux facteurs de consommation, des seuils d'alerte, des responsabilités de suivi, du plafond de consommation applicable et des mesures prévues en cas de dépassement.

La gouvernance des coûts doit être intégrée à la gouvernance du risque. Un agent peut consommer davantage lorsqu'il consulte un contexte volumineux, appelle plusieurs outils, répète des étapes, utilise un modèle plus coûteux, produit plusieurs livrables ou fonctionne sur une longue période. Cette consommation peut notamment être influencée par le nombre de jetons traités, les appels aux outils, les requêtes aux connecteurs, les traitements spécialisés, le stockage temporaire ou persistant, ainsi que les mécanismes d'orchestration utilisés. Plus l'autonomie est élevée, plus le profil de consommation peut devenir difficile à anticiper.

Le plafond de consommation vise à limiter l'exposition financière du pilote, tout en permettant de suivre les usages réels. Il peut être défini par période, par agent, par cas d'usage, par environnement, par utilisateur ou par exécution, selon le niveau de risque et la capacité de suivi disponible. Des seuils d'alerte devraient permettre d'intervenir avant un dépassement significatif, notamment lorsqu'un agent multiplie les appels, traite un volume inhabituel de données, répète inutilement certaines étapes ou adopte un comportement imprévu.

Avant tout déploiement, l'organisation doit comprendre le modèle de facturation, l'unité de coût applicable, les facteurs de consommation, les seuils d'alerte, le coût maximal acceptable par exécution, le plafond du pilote et les mesures prévues en cas de dépassement. Le passage à l'échelle ne devrait pas être autorisé si les coûts ne peuvent pas être suivis par agent, par cas d'usage, par environnement et, lorsque possible, par type de tâche.

Point de vigilance – Le contrôle des coûts fait partie de la gestion des risques : un agent autonome peut multiplier les appels, les étapes et la consommation.

11. EXPLOITATION, SUIVI ET CYCLE DE VIE TECHNIQUE

La mise en place d'un agent d'IA doit s'accompagner de pratiques d'exploitation adaptées à son niveau d'autonomie, aux données utilisées, aux systèmes touchés et aux effets possibles de ses actions. Ces pratiques, parfois regroupées sous le terme « AgentOps », visent à rendre les changements contrôlables, les comportements observables et les interventions reproductibles tout au long du cycle de vie de l'agent.

Un agent d'IA devrait être traité comme un actif numérique. À ce titre, l'organisation devrait :

- tenir un registre des agents, incluant le propriétaire, la finalité, les systèmes connectés, les données accessibles, le niveau d'autonomie et l'état de déploiement.
- gérer les versions des modèles, instructions, règles, connecteurs, outils et configurations qui influencent le comportement de l'agent.
- tester les changements dans un environnement contrôlé avant toute mise en production.
- surveiller les performances, les erreurs, les coûts, les incidents, les dérives et les comportements inattendus.
- documenter les changements importants et réévaluer les risques lorsqu'un modèle, une source de données, un outil ou un processus d'affaires évolue.
- prévoir les conditions de suspension, de retrait, d'archivage et de déclassement de l'agent.
- documenter l'orchestration, les rôles, les transferts, les validations humaines et la journalisation détaillée lorsqu'un processus mobilise plusieurs agents ou composants automatisés (système multiagent).

Point de vigilance – Un agent doit être géré comme un actif numérique : registre, versions, configuration, surveillance, incidents, changements et retrait doivent être documentés.

12. PASSAGE EN PRODUCTION

Le passage en production doit être graduel, documenté et approuvé par une autorité imputable. Il suppose notamment une validation de sécurité, une validation de protection des renseignements personnels, une analyse juridique lorsque requise, une architecture documentée, une liste des données, des outils, des systèmes et des accès, une journalisation complète, un mécanisme d'arrêt testé, une procédure de retour manuel, du soutien aux utilisateurs, une formation minimale, un suivi des coûts et une décision documentée sur les risques résiduels.

Le maintien en production ne doit pas être présumé. Un agent devrait pouvoir être suspendu ou retiré si les incidents se répètent, si les coûts deviennent imprévisibles, si les journaux sont insuffisants, si la supervision réelle diminue, si l'environnement ne répond plus aux exigences ou si le processus d'affaires évolue au point de rendre l'usage non pertinent ou non sécuritaire.

Une revue post-déploiement devrait être prévue à intervalles définis. Elle permet de confirmer que les usages réels correspondent aux usages autorisés, que les contrôles demeurent efficaces, que les incidents sont traités, que les coûts restent maîtrisés et que les risques résiduels demeurent acceptables. Cette revue devrait aussi vérifier si l'agent continue de répondre à un besoin réel ou si une solution plus simple peut maintenant être privilégiée.

Point de vigilance – Le passage en production n'est pas une fin : l'agent doit être surveillé, révisé périodiquement et retiré si les risques ou les coûts deviennent inacceptables.

13. RÔLES ET RESPONSABILITÉS

La gouvernance d'un agent d'IA doit être multidisciplinaire. Le propriétaire d'affaires porte le besoin, la valeur attendue et l'imputabilité. Le responsable technologique encadre l'architecture, les accès, les environnements et l'exploitation. Les responsables de la sécurité, de la protection des renseignements personnels, des données, du juridique, des finances, de l'approvisionnement, de l'exploitation et de la gestion du changement doivent être impliqués selon la nature du cas d'usage et le niveau de risque.

- **Propriétaire de l'agent** : approuve la finalité, les limites d'autorité, les actions permises, les accès, les indicateurs de succès, les risques résiduels et les décisions de poursuite, suspension ou retrait.

- **Responsable technologique** : encadre l'architecture, les connecteurs, la configuration, les environnements, la journalisation, la surveillance et la gestion des changements.

- **Responsables sécurité, PRP et données** : valident les contrôles de sécurité, la protection des renseignements personnels, la qualité des données, les droits d'accès, la provenance et les exigences de conservation.

- **Supervision opérationnelle** : vérifie les résultats, traite les exceptions, applique les seuils d'escalade, documente les incidents et contribue aux améliorations.

- **Utilisateurs** : appliquent une posture critique, respectent les limites d'usage, évitent les données non autorisées et signalent les résultats inattendus.

La personne responsable de la supervision opérationnelle joue un rôle clé. Elle doit surveiller les interventions de l'agent, valider les actions aux points de contrôle prévus, traiter les exceptions, déclencher l'escalade au besoin et contribuer à l'amélioration continue. Les personnes utilisatrices doivent aussi adopter une posture critique : ne pas présumer que l'agent est fiable, vérifier les résultats avant réutilisation, éviter les données non autorisées et signaler rapidement tout comportement inattendu ou hors périmètre.

14. MESSAGES CLÉS

1. Tous les besoins ne nécessitent pas un agent d'IA. Une solution moins autonome doit être envisagée lorsque celle-ci permet d'atteindre adéquatement l'objectif poursuivi.
2. L'IA agentique peut soutenir l'expérimentation et l'amélioration des services publics, mais son utilisation doit être justifiée par une valeur démontrable, un cas d'usage circonscrit et des conditions de contrôle suffisantes.
3. Plus le niveau d'autonomie augmente, plus les exigences en matière de gouvernance, de sécurité, de supervision humaine, de traçabilité, de protection des renseignements personnels et de contrôle des coûts doivent être renforcées.
4. Les risques propres aux agents doivent être identifiés séparément des risques généraux de l'IA générative, car ils découlent de leur capacité d'action, de leurs accès, de leurs outils et de leurs interactions avec des systèmes.
5. Le parcours d'autorisation devrait confirmer successivement la pertinence du cas d'usage, la préparation du pilote et la préparation à la production.
6. Aucun agent ne devrait être lancé sans propriétaire désigné, limites d'accès, supervision, journalisation, mécanisme d'arrêt et plafond de consommation.
7. Un agent d'IA doit être géré comme un actif numérique : les pratiques d'exploitation doivent permettre de suivre les versions, configurations, incidents, coûts, changements et conditions de retrait.

8. Les premiers pilotes doivent être simples, internes, réversibles et mesurables.
9. Le passage en production doit être approuvé, progressif et révisé périodiquement.
10. L’organisme public demeure responsable des résultats, actions et impacts associés à l’agent.
11. Les rôles d’affaires, technologiques, de sécurité, de protection des renseignements personnels, de gestion des données, de finances, d’exploitation et de supervision doivent être clarifiés avant l’autorisation.

ANNEXE A – DÉFINITIONS DE TRAVAIL

- **IA générative** : système d'IA qui produit du contenu comme du texte, des images, du son, de la vidéo ou des réponses structurées.
- **Assistant d'IA** : outil qui aide une personne à rechercher, organiser, résumer, analyser ou produire de l'information, sans agir seul dans un processus.
- **Agent d'IA** : système capable de poursuivre un objectif, planifier des étapes, utiliser des outils, consulter des sources et proposer ou exécuter certaines actions dans des limites définies.
- **Autonomie encadrée** : capacité d'un système à agir sans intervention humaine immédiate, mais dans un périmètre préalablement fixé.
- **Supervision humaine** : intervention permettant d'autoriser, valider, corriger, suspendre ou refuser une action proposée ou exécutée par l'agent.
- **Journalisation** : enregistrement structuré des demandes, sources, outils, actions, validations, erreurs, incidents et coûts.
- **Garde-fou** : mesure organisationnelle, fonctionnelle ou technique qui limite les comportements non souhaités d'un agent.
- **Système multiagent** : ensemble d'agents spécialisés qui interagissent ou se coordonnent dans un même processus.

ANNEXE B – CAS D'USAGE DE RÉFÉRENCE

Cas d'usage	Valeur attendue	Garde-fous minimaux
Préparation de dossiers internes	Réduire le temps de compilation et améliorer la cohérence des synthèses.	Sources approuvées, validation humaine, aucun envoi externe automatique.
Triage de demandes internes	Soutenir l'aiguillage initial et réduire les erreurs de routage.	Aucune décision finale automatisée, escalade des cas ambigus.
Contrôle qualité documentaire	Repérer les incohérences, les doublons, les champs manquants et les écarts de format.	Mode suggestion privilégié, corrections limitées, revue humaine.
Centre de services interne	Accélérer l'accès à l'information validée et réduire la charge de soutien.	Base de connaissances approuvée, journalisation, transfert vers un humain au besoin.

ANNEXE C – VÉRIFICATIONS MINIMALES AVANT AUTORISATION

- Le besoin, la valeur attendue et les limites d’usage sont définis.
- Le niveau d’autonomie est documenté et proportionné au risque.
- Les données, sources, classifications et droits d’usage sont connus.
- Les renseignements personnels ont été évalués et l’EFVP est réalisée lorsque requise.
- Les accès, outils, connecteurs et permissions sont limités au strict nécessaire.
- Les risques d’abus, d’exfiltration, d’injection d’instructions et de dérive fonctionnelle ont été examinés.
- La journalisation est suffisante, protégée et exploitable.
- Le mécanisme d’arrêt, la reprise manuelle et la procédure d’incident sont testés.
- Les coûts, seuils d’alerte et plafonds de consommation sont définis.
- Les rôles de supervision, d’escalade, d’approbation et de soutien sont attribués.

ANNEXE D – SOURCES CONSULTÉES

- GOUVERNEMENT DU QUÉBEC. *Utilisation responsable de l'intelligence artificielle dans l'administration publique*, [En ligne], [https://www.quebec.ca/gouvernement/numerique/intelligence-artificielle-administration-publique/utilisation-responsable-intelligence-artificielle].
- GOUVERNEMENT DU QUÉBEC. *Stratégie d'intégration de l'intelligence artificielle dans l'administration publique 2021-2026,* [En ligne], [https://www.quebec.ca/gouvernement/numerique/intelligence-artificielle-administration-publique].
- GOUVERNEMENT DU QUÉBEC. *Guide d'application des mesures applicables lors de l'utilisation de l'intelligence artificielle générative*, [En ligne], [https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/cybersecurite_numerique/Publications/Strategie_IA/GU_lignes_dir_IAG_2025.pdf].
- GOUVERNEMENT DU QUÉBEC. *Mesures applicables lors de l'utilisation de l'intelligence artificielle générative,* [En ligne], [https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/cybersecurite_numerique/Publications/Strategie_IA/Indication_app_IAG.pdf].
- GOUVERNEMENT DU CANADA, *Guide sur l'utilisation de l'intelligence artificielle agentive,* [En ligne], [https://www.canada.ca/fr/gouvernement/systeme/gouvernement-numerique/innovations-gouvernementales-numeriques/utilisation-responsable-ai/guide-utilisation-intelligence-artificielle-agentive.html].
- DIGITAL NSW, *Guide to using AI Agents in NSW Government,* [En ligne], [https://www.digital.nsw.gov.au/policy/artificial-intelligence/guide-to-using-ai-agents-nsw-government].
- Digital NSW, *AI agent usage and deployment guidance,* [En ligne], https://www.digital.nsw.gov.au/sites/default/files/2025-10/ai-agent-usage-and-deployment-guidance.pdf.

Québec