culture
Saguenay–Lac-Saint-Jean

PRÉCAUTIONS LORS DE L'UTILISATION D'OUTILS IA DANS LES TÂCHES PROFESSIONNELLES

Gestion sécurisée et responsable de l'Intelligence Artificielle en milieu professionnel

L'évolution de l'Intelligence Artificielle

Type	Description
IA Traditionnelle	Systèmes fonctionnant sur des règles programmées détaillées
Machine Learning	Algorithmes capables d'apprendre à partir de données et d'améliorer leur performance
Deep Learning	Réseaux de neurones multicouches permettant de résoudre des problèmes complexes
IA Générative	Systèmes capables de produire du contenu original (texte, image, code) en s'inspirant des données d'apprentissage

Pourquoi encadrer l'usage ?

Confidentialité

Les services publics d'IA peuvent stocker et traiter les requêtes ainsi que les réponses, exposant ainsi des informations sensibles.

Hallucinations

Les modèles d'IA peuvent générer des réponses plausibles mais fausses, telles que des chiffres inventés ou des sources inexistantes.

Biais

Un système d'IA peut prendre des décisions injustes ou discriminatoires, résultant de biais présents dans ses données d'entraînement.

Conformité

Un risque lié aux droits d'auteur, aux licences des modèles et aux obligations de protection des données personnelles.

culture
Saguenay–Lac-Saint-Jean

Risque 1 : Confidentialité

Exemples

- Copier-coller un contrat ou un numéro d'assurance sociale dans un prompt
- Coller une liste de clients avec numéros de téléphone/adresses courriel

Mesures d'atténuation

- Ne **jamais entrer d'informations personnelles** identifiables (PII) ou **confidentielles** dans un outil public
- Utiliser la **pseudonymisation** ou **l'anonymisation** avant toute saisie
- Favoriser des environnements d'IA internes ou des offres « **Entreprise** » avec des **clauses** de **non-rétention**
- Ajouter une clause « **Do not share** » si l'outil le permet (paramètre ou plan entreprise)

Risque 2 : Exactitude et hallucinations

Exemples

- L'IA fournit une statistique ou une citation qui n'existe pas.
- L'IA génère un résumé inexact d'un rapport technique.

Bonnes pratiques

- Toujours demander les sources/preuves et vérifier les informations en dehors de l'outil IA.
- Prévoir une étape de contrôle qualité (fact-checking) avant toute publication.
- Ne pas utiliser la réponse de l'IA comme unique base de décision.
- Exiger une validation humaine.

Risque 3 : Biais

Définition

Un système d'IA peut prendre des décisions ou générer des résultats qui sont systématiquement injustes ou discriminatoires. Ces biais ne sont pas intentionnels, mais résultent des biais humains présents dans les données utilisées pour entraîner l'IA ou dans sa conception.

Impacts potentiels

Discrimination systématique contre certains groupes

Renforcement de stéréotypes et de préjugés

Décisions inéquitables et injustes

Mesures de mitigation

Formuler des prompts inclusifs

Exemple : « Montrer la diversité »

Cette approche encourage l'IA à générer des résultats équilibrés et inclusifs.

Risque 4 : Conformité

Exemples

Générer une image semblable à une œuvre protégée et l'utiliser dans une campagne commerciale sans obtenir les autorisations nécessaires.

Mesures de mitigation

- Vérifier la licence et les droits d'utilisation auprès de l'éditeur de l'outil IA.
- Consulter le service juridique avant d'utiliser des sorties IA à des fins commerciales.
- Prendre en compte les règles locales (ex. obligations de notification en cas de fuite de données).

culture
Saguenay–Lac-Saint-Jean

Classification des données

Une classification claire des données est essentielle pour déterminer ce qui peut être partagé avec les outils d'IA.

PUBLIC

Aucune contrainte

Exemples : brochures publiques, contenus web, informations générales

INTERNE

Usage interne

Exemples : plannings internes, communications entre employés

CONFIDENTIEL

Données clients, contrats, données RH

Exemples : fichiers clients, contrats commerciaux, données personnelles

SECRET

Secrets commerciaux, codes sources critiques

Exemples: algorithmes propriétaires, stratégies commerciales sensibles

Règle d'utilisation avec l'IA

Seuls les contenus PUBLIC ou pseudonymisés peuvent être saisis dans les IA publiques.

Procédure d'anonymisation

Avant de soumettre des données à des outils d'IA publics, il est crucial de les anonymiser pour protéger l'information sensible. Seuls les contenus PUBLIC ou pseudonymisés peuvent être saisis dans les IA publiques.

1. Identifier

Identifiez les informations sensibles (PII) dans votre document que vous souhaitez anonymiser.

2. Anonymiser

Effectuez un remplacement automatique des informations sensibles (par exemple, via un outil de traitement de texte comme Word). ou via des solutions spécialisées

3. Utiliser

Soumettez le contenu obtenu après anonymisation à l'outil d'IA public.

4. Réintégrer

Adaptez les résultats (sorties IA) avec les données réelles après validation.

Exemples d'outils d'anonymisation

ARX Data Anonymization Tool

Bonnes pratiques et validation

Contrôle qualité

Prévoir une étape de contrôle qualité (fact-checking) avant toute publication.

Demandez les sources

Toujours demander les sources/preuves et vérifier les informations en dehors de l'outil IA.

Validation humaine

- Ne pas utiliser la réponse de l'IA comme unique base de décision
- Exiger une validation humaine.

Importance de la validation humaine

La validation des sorties de l'IA est une étape cruciale pour garantir l'exactitude et la pertinence des informations générées. Cela implique une vérification humaine systématique et, si nécessaire, l'utilisation d'outils de vérification.

- **Vérification des faits**
- **Validation technique**
- **Validation technique**