



MÉMORANDUM : Éthique et IA

Rédigé par Benjamin Veyrac, Conseiller en infrastructure,
automatisation et technologie à l'Observatoire global du
Saint-Laurent.

À : L'ensemble des membres de l'équipe travaillant sur "Building Bridges Project"

OBJET : Considérations éthiques relatives à l'intelligence artificielle

Introduction

Dans le cadre du projet "Building Bridges", nous développerons une série d'applications utilisant l'intelligence artificielle (IA) pour analyser et manipuler de vastes quantités de données. L'IA offre des possibilités sans précédent, mais elle soulève également des questions éthiques cruciales que nous devons prendre en compte à chaque étape du développement. Ce mémorandum a pour but de rappeler les principaux points éthiques à considérer afin de garantir que nos solutions respectent les normes éthiques les plus élevées.

1. Transparence et Explicabilité

L'une des préoccupations majeures dans l'utilisation de l'IA est la transparence. Les décisions prises par nos modèles d'IA doivent être compréhensibles par les utilisateurs. Cela signifie que les algorithmes utilisés doivent être documentés de manière claire, et les raisons pour lesquelles certaines décisions sont prises doivent être explicables.

Actions recommandées :

- Documenter les algorithmes et les modèles utilisés de manière claire et accessible.
- Rendre les décisions de l'IA traçables et compréhensibles.
- Fournir des explications claires sur les résultats produits par l'IA.
- Assurer l'accès à des descriptions des modèles pour les utilisateurs finaux.
- Intégrer des mécanismes permettant de comprendre les critères de décision de l'IA.

2. Équité et Non-discrimination

L'intelligence artificielle doit être développée de manière à traiter toutes les données et situations de manière équitable, sans introduire de biais ou de discrimination. Les

décisions et les prédictions issues de nos modèles doivent être impartiales et justes, indépendamment des caractéristiques des données en entrée.

Actions recommandées :

- Identifier les biais potentiels dans les jeux de données.
- Éliminer les biais identifiés dans les modèles d'IA.
- Tester les résultats pour détecter des comportements discriminatoires.
- S'assurer que les modèles d'IA sont représentatifs de la diversité des situations.
- Effectuer des audits réguliers pour vérifier l'équité des décisions prises par l'IA.

3. Protection des Données et Vie Privée

Nous manipulerons une grande quantité de données, dont certaines peuvent être sensibles. Il est crucial de garantir que toutes les données sont protégées conformément aux normes les plus strictes en matière de confidentialité.

Actions recommandées :

- Crypter toutes les données sensibles manipulées par l'IA.
- Anonymiser les données personnelles avant leur utilisation dans les modèles.
- Mettre en place des politiques claires de conservation des données.
- Limiter l'accès aux données aux seules personnes autorisées.
- Établir des procédures de destruction sécurisée des données obsolètes.
- Effectuer des audits réguliers de la gestion des données pour garantir la conformité.

4. Responsabilité et Gouvernance

Chaque membre de l'équipe est responsable de s'assurer que l'IA est utilisée de manière éthique. Cela inclut la conception, le déploiement et l'utilisation des systèmes d'IA.

Actions recommandées :

- Définir clairement les responsabilités éthiques pour chaque membre de l'équipe.
- Établir une structure de gouvernance pour superviser les décisions liées à l'IA.
- Mettre en place des processus pour identifier et corriger les erreurs de l'IA.
- Assurer la formation continue des équipes sur les enjeux éthiques de l'IA.

- Documenter les décisions prises concernant l'IA et leur justification.

5. Autonomie et Contrôle Humain

Il est crucial de déterminer jusqu'à quel point l'IA peut fonctionner de manière autonome et d'assurer que les humains conservent un contrôle suffisant sur les décisions importantes.

Actions recommandées :

- Définir clairement les limites de l'autonomie des systèmes d'IA.
- Assurer que les décisions critiques nécessitent une validation humaine.
- Mettre en place des mécanismes pour permettre une supervision humaine active.
- Fournir une option de désactivation de l'IA pour reprendre un contrôle manuel.
- Établir des procédures pour réviser les décisions de l'IA en cas de besoin.

6. Impact Environnemental

Nos applications IA, bien que virtuelles, ont un impact environnemental, notamment en raison de la consommation d'énergie nécessaire pour les entraîner et les exécuter. Il est important de minimiser cet impact.

Actions recommandées :

- Optimiser les algorithmes pour réduire la consommation d'énergie.
- Évaluer régulièrement l'empreinte carbone des modèles d'IA.
- Adopter des pratiques de développement durable dans la gestion des ressources informatiques.
- Favoriser l'utilisation de matériel et d'infrastructures à faible consommation énergétique.
- Compenser l'empreinte carbone liée au développement et au déploiement des solutions IA.

7. Conformité Légale et Réglementaire

Les applications d'IA doivent respecter les lois et réglementations en vigueur, qui peuvent varier selon les régions et les contextes d'utilisation.

Actions recommandées :

- S'assurer que les solutions IA respectent les lois sur la protection des données.
- Collaborer avec des experts juridiques pour valider la conformité des applications.
- Effectuer une veille régulière des évolutions législatives en matière d'IA.
- Documenter les cadres réglementaires applicables à chaque projet.
- Adapter les systèmes en fonction des exigences légales régionales.

Conclusion

Le respect des considérations éthiques est un élément essentiel du projet "Building Bridges". Les principes énoncés dans ce document sont destinés à guider notre réflexion et nos actions tout au long du développement des solutions d'intelligence artificielle. Ils regroupent les points clés à garder à l'esprit afin de s'assurer que nos pratiques respectent les normes éthiques. Ce document ne doit pas être perçu comme une série d'injonctions, mais plutôt comme un cadre de référence pour orienter nos décisions et renforcer la confiance de nos utilisateurs et partenaires dans l'intégrité de notre travail.