INTELLIGENCE ARTIFICIELLE

AOÛT 2020

ITSAP.00.040

L'intelligence artificielle (IA) transforme le monde dans lequel nous vivons. Sans elle, il vous serait impossible, par exemple, d'utiliser la reconnaissance faciale pour accéder à votre appareil mobile ou de demander à votre haut-parleur intelligent de vous fournir les dernières prévisions météorologiques. En termes simples, cette technologie en développement fait appel à des programmes informatiques intelligents, comme les algorithmes d'apprentissage, pour trouver les schémas complexes dans les données en vue de résoudre un problème. Les données permettent à l'outil d'apprentissage machine d'adopter un comportement qui pourrait sembler intelligent s'il était observé chez un être humain (p. ex. comprendre une langue ou tirer certaines leçons).

QUE FAIT LE MIEUX L'IA À L'HEURE ACTUELLE?

Les outils d'apprentissage machine peuvent résoudre efficacement les problèmes lorsque la solution se trouve dans les données fournies. Pour obtenir les meilleures solutions, l'entraînement de l'outil doit reposer sur des données pertinentes et précises. Comme les outils d'apprentissage machine continuent d'apprendre à la lumière des données et de la rétroaction que vous leur fournissez, il est inutile de les remplacer.

QUELLES SONT LES LIMITES ACTUELLES DE L'IA?

Les outils d'apprentissage machine ont de la difficulté à résoudre des problèmes pour lesquels il faut faire appel au raisonnement ou sens commun (c.-à-d., obtenir plus de contexte). Souvent, il est trop coûteux et difficile de fournir à l'outil d'apprentissage machine toutes les données nécessaires pour lui permettre d'en venir à une solution basée sur le bon sens. Grâce à leur jugement et leur intuition, les êtres humains arrivent à gérer plus efficacement ces types de situations de décisions.

DE QUELLES FAÇONS LES ORGANISATIONS UTILISENT-ELLES L'IA?

Reconnaissance faciale : L'IA (en particulier la discipline axée sur les réseaux de neurones artificiels) a grandement contribué aux améliorations apportées à la reconnaissance faciale et a permis de rendre le processus plus rapide et précis.

Optimisation des processus : Un outil d'apprentissage machine bien entraîné (au moyen de données précises) peut utiliser les données pour fournir des solutions plus justes et accomplir des tâches banales plus rapidement que pourrait le faire un être humain.

Agent conversationnel (*chatbot*) : Ces outils d'apprentissage machine utilisent un processus de langage naturel dans le but d'automatiser le clavardage pour les services à la clientèle. Un client peut poser une question à l'agent conversationnel et obtenir une réponse en quelques secondes, et ce, 24 heures par jour, sept jours par semaine.

Diagnostic assisté par ordinateur : L'industrie médicale fait appel aux outils d'apprentissage machine pour aider au diagnostic des patients. Les outils tirent des leçons des cas précédents et fournissent une deuxième opinion basée sur ce qu'ils ont appris.

Détection de la fraude : Les outils d'apprentissage machine sophistiqués peuvent détecter les courriels frauduleux plus rapidement qu'un être humain. Ils analysent le contenu de votre boîte de réception et déplacent les pourriels et les courriels d'hameçonnage dans votre dossier Courrier indésirable.

Assistance dans les centres d'appel : Les outils d'apprentissage machine peuvent changer la façon dont les centres d'appel mènent leurs activités. L'IA permet de réduire le temps d'attente et de réaffecter les employés à des postes moins stressants.

© Gouvernement du Canada
Le présent document est la propriété exclusive du gouvernement du Canada. Toute modification, diffusion à un public autre que celui visé, production, reproduction ou publication, en tout ou en partie, est strictement interdite sans l'autorisation expresse du CST.

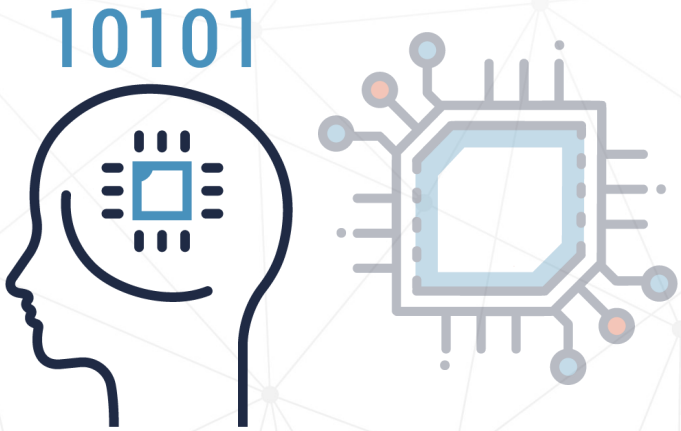
QUELLES MENACES PÈSENT SUR LES OUTILS D'IA?

L'efficacité des outils d'IA dépend en grande partie du modèle de données sur lequel ils reposent. La compromission des données constitue la principale menace qui pèse sur l'IA. Parmi les méthodes de compromission les plus courantes, on retrouve les suivantes :

Attaque par empoisonnement de données : Ce type d'attaque survient au cours de la phase d'entraînement de l'outil d'apprentissage machine. Les outils d'IA dépendent fortement de l'exactitude des données utilisées lors du processus d'entraînement. Si des données empoisonnées (erronées) sont injectées dans le jeu de données, l'empoisonnement pourrait pousser le système d'apprentissage à faire des erreurs.

Exemple nuisible : Ce type d'attaque survient une fois l'entraînement de l'outil d'apprentissage machine terminé. Son but est d'arriver à « duper » l'outil afin qui classe les données saisies incorrectement. Dans un scénario impliquant un véhicule autonome, un exemple nuisible pourrait consister, entre autres, en une modification mineure des panneaux de signalisation dans le monde réel (une légère décoloration ou des étiquettes sur un panneau d'arrêt), faisant en sorte que le système d'AI confonde le panneau d'arrêt avec un panneau de vitesse. Une telle situation risquerait d'avoir de réelles répercussions sur le fonctionnement sécuritaire des véhicules autonomes.

Attaques par inversion de modèle et inférence d'appartenance : Ces deux scénarios se produisent lorsqu'un auteur de menace interroge le modèle de données de votre organisation. Une attaque par inversion de modèle révèlera le jeu de données sous-jacent, permettant ainsi à l'auteur de menace de reproduire les données d'entraînement. Une attaque par inférence d'appartenance confirme qu'un fichier de données en particulier fait partie des données d'entraînement. Ces deux types d'attaques pourraient compromettre la confidentialité de vos données d'entraînement et exposer de l'information sensible.



QUE DEVRIEZ-VOUS SAVOIR D'AUTRE AU SUJET DE L'IA?

- Les outils d'apprentissage machine peuvent détecter les schémas de données.
- Les outils d'apprentissage machine doivent analyser assez de données pour arriver à en extraire les schémas à une fréquence suffisamment élevée.
- Les données utilisées lors du processus d'entraînement devraient être complètes, variées et exactes.
  - La présence de vides dans les données pourrait empêcher la découverte de schémas et ceux qui sont découverts pourraient ne pas être exacts.
  - Un manque de variété dans les données limitera la portée de l'outil.
  - La fiabilité des résultats de l'outil dépendra de l'exactitude des données utilisées lors du processus d'entraînement.
- Les données qui sont enregistrées et recueillies à des fins de « contrôle de la qualité » peuvent contenir tant de l'information sensible que des renseignements personnels.
- Plusieurs organisations adoptent maintenant des politiques en matière d'IA fiables pour veiller à ce que leur utilisation des outils d'IA minimise les biais potentiels et les conséquences imprévues, en particulier pour les outils utilisés pour le traitement de personnes. Les politiques peuvent également faciliter le développement des protocoles appropriés au traitement de l'information sensible et des renseignements personnels. La *Directive sur la prise de décision automatisée* est un exemple de politique en matière d'IA récemment adoptée par le gouvernement du Canada.
- Si votre organisation a l'intention de faire appel à l'IA, elle devrait envisager de demander des conseils juridiques afin de gérer les nombreux éléments à prendre en compte sur le plan juridique, de l'éthique, du respect de la vie privée et des politiques.

Avez-vous besoin d'aide ou des questions? Vous voulez tout savoir sur la cybersécurité?
Visitez le site Web du Centre canadien pour la cybersécurité à **cyber.gc.ca**.