Intelligence artificielle

Juillet 2023 **ITSAP.00.040**

L'intelligence artificielle (IA) transforme le monde dans lequel nous vivons. 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 faire des prédictions ou à des fins de classification. De nos jours, on emploie l'IA pour effectuer des tâches particulières, comme utiliser la reconnaissance faciale pour accéder à votre appareil mobile ou demander à votre haut-parleur intelligent quelles sont les prévisions météorologiques. Sous-ensemble de l'intelligence artificielle, l'apprentissage machine a recours à des instructions, connues sous le nom d'algorithmes, et à des données pour comprendre les langages afin d'aider le système informatique à apprendre et à s'améliorer en fonction de sa propre expérience. Découlant de l'apprentissage machine, l'apprentissage profond utilise de grandes quantités de données et une structure d'algorithmes en couche (les algorithmes de réseaux de neurones artificiels) pour enseigner à un modèle comment prendre des décisions intelligentes par lui-même.

Que peut faire l'IA?

L'IA joue déjà un rôle considérable dans notre quotidien. Des moteurs de recherche au magasinage en ligne, des assistants vocaux à votre appareil mobile ou haut-parleur intelligent, l'IA formule des recommandations, fournit de l'information, répond aux questions et aide à organiser nos horaires. Les outils d'apprentissage machine tirent avantage des données et des rétroactions créées dans le cadre de ces applications quotidiennes pour apprendre et s'améliorer.

Quelles sont les limitations actuelles de l'IA?

De nos jours, l'IA est encore sujette à certaines limitations fondamentales. Il lui est toujours difficile de raisonner, de faire preuve de bon sens, de s'adapter à différentes situations et de comprendre le lien de cause à effet. Grâce à leur jugement et à leur intuition, les êtres humains arrivent à gérer plus efficacement ces situations exigeant des capacités décisionnelles et des compétences en résolution de problèmes.

De quelles façons les organisations utilisent-elles l'IA?

Reconnaissance faciale: Principale application de l'IA qui consiste à analyser les caractéristiques faciales dans une image ou une photo afin d'identifier ou de vérifier l'identité d'un individu.

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.

Assistants numériques: Les agents conversationnels peuvent améliorer le service à la clientèle et réduire les coûts de soutien. Les clients peuvent obtenir de l'assistance en quelques secondes, et ce, 24 heures par jour, sept jours par semaine. Ces services sont souvent hautement personnalisés et peuvent être basés sur les préférences des utilisateurs et leurs interactions passées avec l'organisation.

Soins de santé: Dans l'industrie des soins de santé, l'IA facilite le diagnostic et le traitement des patients de différentes façons. Des systèmes de diagnostic automatisés peuvent, par exemple, être utilisés pour former un diagnostic. L'apprentissage machine est un autre outil fort utile en médecine personnalisée. On peut l'utiliser pour aider à prédire les traitements les plus susceptibles d'avoir des effets positifs sur un patient.

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.

Analyse de données: Ayant recours aux algorithmes de l'apprentissage machine, l'IA peut analyser de grandes quantités de données et relever les nouveaux schémas. Il est ainsi possible de réduire le temps qu'un analyste des données doit consacrer à leur traitement, misant sur l'automatisation, et d'améliorer le rendement opérationnel.

Cybersécurité: L'IA fait appel à l'automatisation pour détecter les nouvelles menaces qui pèsent sur les organisations. Grâce à des algorithmes sophistiqués, l'IA peut automatiser la détection des menaces, comme les maliciels, procéder à la reconnaissance des schémas pour déterminer les liens entre les différents vecteurs d'attaque et fournir des renseignements prédicatifs supérieurs.

SÉRIE SENSIBILISATION

No de cat. D97-1/00-040-2022F-PDF
ISBN 978-0-660-44249-5

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 ou automatisés.

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 révélera 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.

L'intelligence artificielle generative (IA)

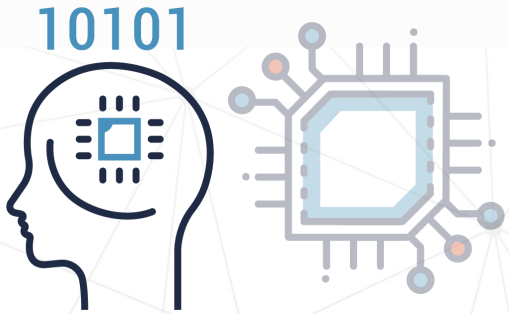
Ceci est un type d'intelligence artificielle qui génère du nouveau contenu en modélisant les caractéristiques des données tirées des grands jeux de données qui alimentent le modèle. Alors que les systèmes d'IA traditionnels peuvent reconnaître les modèles ou classifier le contenu existant, l'IA générative peut créer du nouveau contenu sous plusieurs formes, comme du texte, une image, un fichier audio ou du code logiciel.

Les modèles de langage de grande taille (LLM pour Large Language Model) sont une catégorie de l'IA générative qui s'est grandement améliorée au cours des dernières années. Depuis le début de 2022, plusieurs LLM (ChatGPT d'OpenAI et LaMDA de Google) et services ayant recours à des LLM (Bard de Google et Bing de Microsoft) ont retenu l'attention du public à travers le monde.

Prière de consulter la publication L'intelligence artificielle générative (ITSAP.00.041)

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 fausser l'exactitude de ceux qui sont découverts
    - 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



Vous avez des questions ou vous avez besoin d'aide? Vous voulez en savoir plus sur les questions de cybersécurité? Consultez le site Web du Centre canadien pour la cybersécurité (Centre pour la cybersécurité) à **cyber.gc.ca**.