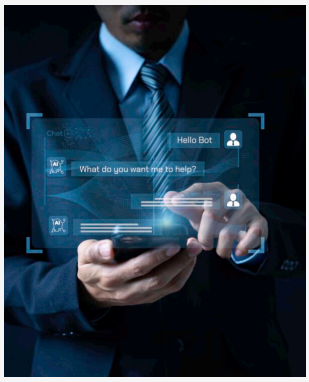
OSLER

Code de conduite volontaire visant un développement et une gestion responsables des systèmes d'IA générative avancés d'ISDE



2 OCTOBRE 2023 8 MIN DE LECTURE

Expertises Connexes

- Gestion de risques et réponse aux crises
- Intelligence artificielle
- Propriété intellectuelle
- Respect de la vie privée et gestion de l'information
- Technologie

Auteurs(trice): Simon Hodgett, Sam Ip, Alannah Safnuk

Le 27 septembre 2023, l'honorable François-Philippe Champagne, ministre de l'Innovation, des Sciences et de l'Industrie, a présenté le Code de conduite volontaire visant un développement et une gestion responsables des systèmes d'IA générative avancés du Canada (le code de conduite).

Faisant suite à la prolifération de systèmes d'IA novateurs capables de générer du contenu – tels que ChatGPT, DALL-E 2 et Midjourney –, le code de conduite est un ensemble d'engagements volontaires visant à inciter les développeurs et les gestionnaires de systèmes génératifs avancés à s'engager à prendre des mesures pour repérer et atténuer les risques connexes. Dans le communiqué annonçant la présentation du code de conduite, Innovation, Sciences et Développement économique Canada (ISDE) a déclaré que le code fournirait un encadrement jusqu'à l'entrée en vigueur du projet de *Loi sur l'intelligence artificielle et les données* (LIAD), qui fait partie du projet de loi C-27, déposé en juin 2022.

Le lancement du code de conduite fait suite à une brève période au cours de laquelle une première ébauche a été publiée. Le 30 août, Osler a notamment tenu un atelier AccessPrivacy destiné à aider les organisations à comprendre la portée, la signification et l'incidence du code de conduite sur cette première ébauche que l'on peut trouver par l'intermédiaire de cette page. Une copie du compte rendu de cet atelier a été soumise au ministre.

Principales caractéristiques

Le code de conduite est propre aux systèmes génératifs avancés, bien que bon nombre des mesures puissent être appliquées en grande partie à divers systèmes, y compris des systèmes d'IA à incidence élevée.

Contrairement à la première ébauche, la version finale du code de conduite établit une distinction entre les mesures applicables à tous les systèmes d'IA générative avancés et celles applicables aux systèmes d'IA générative avancés qui sont accessibles au public, qui peuvent donner lieu à un risque supérieur d'utilisation potentiellement nuisible ou inappropriée. C'est pourquoi le code de conduite suggère que des mesures supplémentaires s'appliquent dans ces cas. En outre, la version finale du code de conduite s'applique à des acteurs précis, appelés développeurs et gestionnaires, reconnaissant que ces participants à l'écosystème de

l'IA ont des responsabilités différentes.

La version finale du code de conduite fournit également plus de détails sur les mesures à prendre pour atteindre les six résultats suivants et sur les personnes qui, au sein d'une entreprise, sont chargées de veiller à ce qu'elles soient prises. Voici, en bref, de quoi il s'agit.

1. Responsabilité

Le code de conduite exige des entreprises qu'elles mettent en œuvre un cadre clair de gestion des risques adapté à l'échelle et à l'incidence de leurs activités. Le code de conduite incite les développeurs et les gestionnaires à mettre en œuvre, entre autres, des cadres complets de gestion des risques, qui comprennent des politiques, des procédures et des formations pour s'assurer que les employés comprennent leurs responsabilités et les pratiques de gestion des risques de l'organisation. Les entreprises s'engagent également à transmettre l'information et les pratiques exemplaires visant la gestion des risques aux entreprises qui jouent des rôles complémentaires dans l'écosystème. En outre, les développeurs des entreprises dont les systèmes d'IA sont mis à la disposition du public s'engagent à utiliser plusieurs lignes de défense, notamment des vérifications par des tiers, pour garantir la sécurité de leurs systèmes d'IA avant leur lancement.

2. Sécurité

Le code de conduite souligne l'importance de l'évaluation et de l'atténuation des risques pour garantir la sécurité des systèmes d'IA avant leur déploiement. Les mesures mentionnées dans le code de conduite pour garantir la sécurité des systèmes d'IA comprennent la réalisation d'une évaluation complète des répercussions négatives potentielles raisonnablement prévisibles, l'engagement des développeurs de systèmes d'IA à mettre en œuvre des mesures adaptées pour atténuer les risques de préjudice et la mise à disposition des développeurs et des gestionnaires en aval des conseils sur l'utilisation appropriée du système.

3. Justice et équité

Le code de conduite reconnaît que les systèmes d'IA peuvent entraîner des répercussions négatives sur la justice et l'équité, notamment en perpétuant des préjugés, et encourage l'évaluation et l'atténuation à différents stades du développement et du déploiement des systèmes. À cette fin, le code de conduite incite les développeurs à évaluer et à organiser les ensembles de données utilisés pour la formation afin de gérer la qualité des données et les biais potentiels. Les développeurs s'engagent également à mettre en œuvre diverses méthodes et mesures d'essai pour évaluer et atténuer le risque d'obtenir des résultats biaisés avant le lancement du système d'IA.

4. Transparence

Le code de conduite reconnaît que les individus ont besoin d'informations suffisantes pour prendre des décisions éclairées et évaluer la manière dont les risques sont traités. Lorsque les systèmes sont mis à la disposition du public, le code de conduite incite les développeurs à publier de l'information sur le système d'IA, notamment : (1) de l'information sur les capacités et les limites du système d'IA, et (2) une description des types de données d'entraînement utilisées pour développer le système d'IA. Les développeurs de systèmes d'IA mis à la disposition du public s'engagent également à fournir une méthode fiable et disponible gratuitement pour détecter le contenu généré par le système. Enfin, les gestionnaires de

systèmes d'IA à usage public et de systèmes d'IA à usage privé s'engagent à veiller à ce que, pour les systèmes d'IA qui pourraient être pris pour des humains, il soit indiqué de façon claire et visible qu'ils sont des systèmes d'IA.

5. Surveillance humaine

Le code de conduite souligne l'importance de la surveillance humaine. Les gestionnaires s'engagent à surveiller le fonctionnement du système d'IA pour s'assurer qu'il n'est pas utilisé à des fins nuisibles ou qu'il n'a pas de répercussions néfastes après qu'on l'ait rendu accessible. Ils s'engagent également à informer le développeur ou à mettre en œuvre des contrôles d'utilisation au besoin pour atténuer les préjudices. Le code de conduite incite les développeurs de systèmes d'IA à conserver une base de données sur les incidents signalés après le déploiement et à fournir des mises à jour au besoin pour veiller à l'efficacité des mesures d'atténuation.

6. Validité et fiabilité

Le code de conduite souligne l'importance que les systèmes fonctionnent efficacement et, comme prévu, soient protégés contre les cyberattaques. Les développeurs de toutes les entreprises doivent veiller à ce que les systèmes d'IA fonctionnent efficacement et soient protégés contre les cyberattaques. Le code de conduite incite les développeurs à utiliser, avant le déploiement, une grande variété de méthodes d'essai dans un ensemble de tâches et de contextes pour mesurer le rendement et garantir la fiabilité. Les développeurs s'engagent également à recourir à des essais axés sur des positions antagonistes (c'est-à-dire la méthode de l'équipe rouge) pour cerner les vulnérabilités, et à effectuer une évaluation des risques liés à la cybersécurité et à mettre en œuvre des mesures adaptées pour atténuer les risques. Afin d'évaluer la validité et la fiabilité des systèmes d'IA, les développeurs s'engagent à effectuer des analyses comparatives pour mesurer le rendement du système d'IA par rapport aux normes reconnues.

Enfin, les signataires s'engagent aussi de manière générale à soutenir le développement continu d'un écosystème d'IA fiable et responsable au Canada, notamment en contribuant à l'élaboration et à l'application de normes, en assurant la transmission d'information et de pratiques exemplaires à d'autres membres de l'écosystème de l'IA, en collaborant avec des chercheurs qui travaillent pour l'avancement de l'IA responsable et en collaborant avec d'autres intervenants, y compris les gouvernements, pour appuyer la sensibilisation et l'éducation du public à l'égard de l'IA.

Prochaines étapes

Alors que les organisations canadiennes envisagent d'adopter le code de conduite, nous nous attendons à ce qu'il y ait d'importantes discussions, notamment sur le champ d'application, la délimitation des acteurs, la signification de certaines définitions (par exemple, les développeurs, les gestionnaires, les systèmes à incidence élevée, etc.), ainsi que des instructions sur la mise en œuvre pratique de certaines des mesures prescrites, et sur l'interaction avec la LIAD à venir. ISDE a indiqué que le code intégrerait les commentaires de divers participants à l'écosystème de l'IA et qu'il publierait un résumé de ces commentaires dans les prochains jours, ce qui pourrait éclairer la logique de certaines des mesures prescrites.

Bien que le code soit volontaire par nature, nous prévoyons que son utilisation se répandra à grande échelle dans de multiples secteurs d'activité au Canada aux fins du développement et de la gestion des systèmes d'IA.