

Foire aux questions sur le Code de conduite volontaire visant les systèmes d'IA générative avancés

De : [Innovation, Sciences et Développement économique Canada](#)

Quelle est la portée de l'application de ce code?

Le Code doit s'appliquer à tous les systèmes d'IA générative qui possèdent des capacités avancées leur permettant de s'adapter à diverses utilisations dans différents contextes, y compris des utilisations pour lesquelles ils n'ont pas été spécialement entraînés. Les systèmes comme ChatGPT, Midjourney, Bard et Llama sont de bons modèles des types de systèmes qui pourraient être assujettis aux mesures de ce code. Toutefois, les systèmes d'IA dans l'ensemble de l'économie intègrent de plus en plus de capacités génératives, et les mises en œuvre ne présentent pas toutes les mêmes risques. Il est important de souligner que les outils génératifs conçus pour accomplir des fonctions précises à faibles risques, comme la correction grammaticale, ne s'inscrivent pas dans la portée, bien que certaines des mesures définies pourraient être utilisées par des sociétés qui cherchent un niveau d'assurance supérieur.

Comment ce code est-il lié à la *Loi sur l'intelligence artificielle et les données* et aux règlements subséquents?

La *Loi sur l'intelligence artificielle et les données* (LIAD) est encore à l'étude au Parlement et ne devrait pas recevoir de sanction royale avant 2024. Le Code vise à servir de pont avec les règlements au titre de la LIAD en fournissant un ensemble clair de lignes directrices que les sociétés peuvent mettre en œuvre immédiatement. Les règlements liés aux systèmes généraux au titre de la LIAD doivent résoudre des problèmes semblables à ceux relevés dans ce code volontaire, et feront l'objet de vastes consultations et d'études publiques avant d'être mis en vigueur.

De quelle façon les développeuses et les développeurs et les gestionnaires des systèmes d'IA générative avancés devraient-ils collaborer pour assurer la sécurité des systèmes?

Loading [Contrib]/a11y/accessibility-menu.js

Le principe « Surveillance humaine » prévoit deux mesures symétriques permettant la gestion efficace d'incidents, la production de rapports et l'atténuation des risques. Les gestionnaires de systèmes ont pour but de surveiller les incidents liés à des utilisations ou à des conséquences nuisibles, par exemple un système d'IA qui fournit des conseils de santé inexacts. Selon la nature de l'occurrence, les gestionnaires de systèmes pourraient être en mesure de modifier leurs mesures d'atténuation en réponse. S'il est possible que l'incident nécessite l'intervention de la développeuse ou du développeur, le gestionnaire du système doit signaler l'incident à la développeuse ou au développeur. La développeuse ou le développeur ferait ensuite le suivi de ces incidents et apportera des mises à jour au système, au besoin pour assurer l'efficacité des mesures d'atténuation des risques. Il est important de noter que les risques à la cybersécurité et les incidents sont traités séparément au titre du Code et qu'ils devraient être réglés conformément à la mesure établie au titre du principe « Validité et fiabilité ».

Qui a été consulté pendant l'élaboration du Code et quels commentaires ont été reçus?

Innovation, Sciences et Développement économique Canada a effectué une consultation publique concernant le Code, du début d'août jusqu'à la mi-septembre 2023. Le gouvernement du Canada a reçu des recommandations et de la rétroaction de la part de 92 participantes et participants, y compris des membres du Conseil consultatif sur l'intelligence artificielle, des entreprises de toute taille, des universitaires, des organisations de la société civile et des membres du public. Les recommandations et la rétroaction étaient fournies au moyen d'observations écrites, de discussions en table ronde des parties prenantes, et de réunions bilatérales avec les personnes représentant le gouvernement organisées à la demande des parties prenantes. Un résumé des points de vue exprimés se trouve dans le rapport du gouvernement « [Ce que nous avons entendu](#) » dans le cadre de cette consultation.