

Le National Institute of Standards and Technology publie une ébauche de son cadre de gestion des risques liés à l'IA, pour commentaires

24 NOVEMBRE 2022 4 MIN DE LECTURE

Expertises Connexes

- [Intelligence artificielle](#)
- [Technologie](#)

Auteurs(trice): [Sam Ip](#), [Simon Hodgett](#), Sam Dobbin

Le 29 septembre 2022, le National Institute of Standards and Technology (NIST) des États-Unis a publié la [deuxième ébauche de son cadre de gestion des risques liés à l'IA](#) [PDF; en anglais seulement] (CGRIA), ainsi que la version préliminaire du [manuel d'accompagnement](#), pour commentaires. Par la suite, les 18 et 19 octobre 2022, le NIST a tenu en ligne un atelier à l'intention d'experts en IA de premier plan afin de discuter des commentaires de la communauté sur la deuxième ébauche du CGRIA et des prochaines étapes.

Cadre de gestion des risques liés à l'IA

Le CGRIA a pour but d'améliorer la gestion des risques pour les particuliers, les organisations et la société touchés par l'IA. Il est volontaire et vise à améliorer l'intégration de la fiabilité dans la conception, le développement, l'utilisation et l'évaluation des produits, systèmes et services d'IA.

Tout logiciel ou système informatique comporte des risques communs, tels que la cybersécurité, la protection des renseignements personnels, etc., mais il existe aussi des risques propres à l'IA qui sont difficiles à prévoir et à gérer. C'est l'objectif du CGRIA : prendre en compte et gérer les risques complexes liés à l'IA tels que l'amplification des résultats inéquitables et les conséquences involontaires pour les particuliers et les collectivités. Le CGRIA n'est pas une liste de contrôle destinée à être utilisée dans des situations isolées; il doit plutôt être largement intégré dans le processus décisionnel et être utilisé parallèlement aux lois et aux règlements existants, et non comme un outil destiné à les remplacer.

Le CGRIA organise la gestion des risques liés à l'IA en fonctions qui gouvernent, cartographient, mesurent et gèrent les risques liés à l'IA :

- **Gouverner** : Cette fonction vise à cultiver la culture de gestion des risques liés à l'IA au sein des organisations qui développent et déploient des systèmes d'IA. Cette fonction garantit que les risques et les répercussions potentiels sont décelés, mesurés et gérés de manière efficace et cohérente.
- **Cartographier** : Cette fonction établit le contexte lié à un système d'IA, et les risques liés au contexte sont décelés. À l'issue de la fonction Cartographier, les utilisateurs devraient disposer de connaissances contextuelles suffisantes pour prendre une décision d'acceptation ou de refus concernant la conception, le développement ou le déploiement

d'un système d'IA.

- **Mesurer :** Cette fonction permet de mesurer, d'analyser, d'évaluer et de suivre les risques liés à l'IA et les répercussions connexes en utilisant des techniques et des outils quantitatifs, qualitatifs ou mixtes.
- **Gérer :** Cette fonction affecte des ressources de gestion des risques aux risques cartographiés et mesurés. Ces risques sont ensuite classés par ordre de priorité et traités en fonction de leur incidence sur le projet.

Le CGRIA introduit également le concept de profils, qui sont utilisés pour illustrer et analyser les processus de gestion des risques liés à l'IA dans le cadre de situations particulières. Deux profils dignes d'intérêt sont mentionnés :

- **les profils de cas d'utilisation**, qui peuvent être utilisés pour offrir un aperçu de la façon de gérer le risque à différentes étapes du cycle de vie de l'IA dans un secteur précis, comme l'embauche ou le logement équitable.
- **les profils temporels**, qui peuvent être utilisés pour décrire l'état actuel ou souhaité d'activités particulières de gestion des risques liés à l'IA dans un secteur ou une industrie, ce qui peut révéler les écarts existants entre les processus actuels de gestion des risques et ceux qui sont visés.

Prochaines étapes

Les commentaires sur la deuxième ébauche du CGRIA, ainsi que les commentaires sollicités lors de l'atelier, seront examinés et intégrés. Le NIST prévoit de publier la version finale du CGRIA en janvier 2023. Bien que le CGRIA soit volontaire, une fois publié, nous nous attendons à ce qu'il soit largement mentionné au sein de diverses industries au cours de la conception, du développement et du déploiement des systèmes d'IA.