Google

Cadre d'IA sécurisé de Google

L'IA évolue rapidement. Il est donc important que les stratégies efficaces de gestion des risques suivent cette évolution. C'est dans cette optique que nous avons élaboré le cadre d'IA sécurisé, un cadre conceptuel assurant des systèmes d'IA sécurisés. Il s'appuie sur six axes principaux :

1. Le déploiement de bases de sécurité solides dans l'écosystème de l'IA

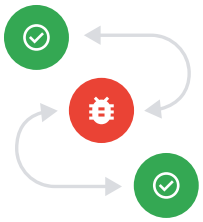
Tirez profit des protections d'infrastructure sécurisées par défaut et de l'expertise acquise ces 20 dernières années pour protéger les systèmes d'IA, les applications et les utilisateurs. En parallèle, développez une expertise organisationnelle pour suivre la cadence des avancées en matière d'IA. Commencez à déployer les protections d'infrastructure et adaptez-les aux enjeux spécifiques de l'IA et aux modèles de menace en constante évolution. Par exemple, les techniques d'injection telles que l'injection SQL existent depuis un certain temps et les organisations peuvent adopter des mesures d'atténuation, comme le nettoyage et la limitation des intrants pour se protéger contre les attaques soudaines par injection.

2. L'élargissement de la portée de détection et de réponse pour inclure l'IA lors de l'évaluation des menaces pesant sur les organisations

Détectez les cyberincidents liés à l'IA et répondez-y à temps en développant une veille sur les menaces et d'autres initiatives. Pour les organisations, il s'agit notamment de surveiller les intrants et les extrants des systèmes d'IA générative afin de détecter des anomalies et d'utiliser les informations recueillies sur les menaces pour anticiper les attaques. Pour ce faire, il est généralement nécessaire de collaborer avec les équipes responsables de la confiance et de la sécurité, des renseignements sur les menaces et de la lutte contre les abus.

3. L'automatisation des systèmes de défense pour suivre l'évolution des menaces

Tirez profit des dernières innovations en matière d'IA pour accroître et accélérer les efforts de réponse aux incidents de sécurité. Les ennemis utiliseront probablement l'IA pour maximiser leur impact; il est donc important d'exploiter cette même technologie et toutes les nouvelles fonctionnalités pour se protéger de façon agile et économique.



4. L'harmonisation des contrôles sur les plateformes pour offrir un niveau de sécurité homogène à l'échelle de l'organisation

Alignez les systèmes de contrôle pour soutenir l'atténuation des risques liés à l'IA et développez des protections sur l'ensemble des plateformes et outils pour assurer une sécurité optimale, évolutive, économique et efficace de toutes les applications d'IA. Chez Google, cela implique d'étendre les protections sécurisées par défaut à des plateformes d'IA comme Vertex AI et Security AI Workbench, et d'intégrer des points de contrôle et des protections dans le cycle de développement des logiciels. Les outils qui répondent à des cas d'utilisation généraux, comme l'API Perspective, permettent à l'ensemble de l'organisation de bénéficier de protections de pointe.

5. L'adaptation des contrôles pour ajuster les mesures d'atténuation et générer des boucles de rétroaction plus rapides en vue de déployer l'IA

Testez systématiquement les mises en oeuvre grâce à l'apprentissage continu et perfectionnez les mesures de détection et de protection pour faire face à un environnement de menaces en constante évolution. Cela inclut des techniques comme l'apprentissage par renforcement basé sur les incidents et les commentaires des utilisateurs. Cette phase implique également plusieurs étapes, comme la mise à jour des ensembles de données d'entraînement, l'affinage des modèles pour répondre aux attaques de façon stratégique, et l'élaboration de modèles par le logiciel utilisé pour ajouter une couche de sécurité supplémentaire selon le contexte (par exemple, pour la détection de comportements anormaux). Les organisations peuvent aussi organiser des exercices d'équipe rouge afin d'améliorer la sécurité de leurs fonctionnalités et produits optimisés par l'IA.

6. La mise en contexte des risques pesant sur les systèmes d'IA dans les processus opérationnels connexes

Évaluez de bout en bout les risques liés au déploiement de l'IA par les organisations. Il s'agit notamment d'évaluer le risque global pour l'entreprise, notamment en faisant un suivi des données sur certains types d'applications ou en validant et surveillant leur fonctionnement. Par ailleurs, les organisations devraient mettre en place des contrôles automatisés pour valider les performances de l'IA.