

DARKTRACE

Le guide du DSI pour investir dans l'IA

Évaluer avant d'investir :
commencez votre parcours d'adoption de l'IA



Secure
What AI
Can Do

Contenu

02	Introduction
03	Quels éléments prendre en compte avant d'investir dans l'IA ?
03	Ce qu'il faut savoir avant d'évaluer un AISP
04	Comprendre les limites de l'IA en cybersécurité
05	Comment évaluer les fournisseurs d'IA
06	Gouvernance, sécurité et contrôle des données
07	Collecte de données et entraînement
07	Choix du modèle et de la technique
08	Validation des performances et de la précision
08	Interprétabilité, ajustabilité et transparence
09	Darktrace en tant que fournisseur de cybersécurité basée sur l'IA
10	Conclusion

Introduction

L'IA s'est rapidement intégrée aux produits de sécurité et continuera de jouer un rôle majeur dans la manière dont ces produits détectent, analysent et traitent les menaces. En permettant de gagner du temps et d'économiser des ressources, d'extraire des renseignements et des enseignements de données précieuses et d'effectuer des analyses plus approfondies et plus complexes qu'une solution de sécurité traditionnelle, l'IA renforce les capacités et la posture de sécurité d'une organisation.

Face à l'accélération de l'adoption de l'IA, accompagnée de l'essor récent de l'IA agentique et des agents autonomes, les DSI doivent aller au-delà des apparences pour comprendre comment les différents modèles sont conçus, entraînés et validés, ainsi que leur comportement lorsqu'ils sont confrontés à des données nouvelles ou inconnues.

Les fournisseurs utilisent un large éventail d'approches, et chaque choix de conception influe sur la précision, la fiabilité, et la probabilité d'introduire des angles morts. Comprendre clairement le fonctionnement d'une fonctionnalité d'IA, et déterminer si elle est véritablement adaptée à la tâche qu'elle prétend résoudre, permet aux responsables de la sécurité de poser des questions plus précises et de s'assurer que la technologie est appliquée de manière adéquate, sans ajouter de risque inutile.

Il est tout aussi essentiel de déterminer si l'IA est déployée en toute sécurité et fait l'objet d'une gouvernance responsable. Un manque de visibilité, des contrôles insuffisants et une maintenance irrégulière peuvent entraîner une dégradation des performances, une dérive des modèles, des biais et des vulnérabilités exploitables.

Les solutions bien conçues intègrent de solides mesures de protection, des processus de raisonnement transparents et des contrôles clairs concernant le traitement et la protection des données. Les DSI doivent pouvoir interroger les fournisseurs en toute confiance sur leurs cadres de gouvernance, leurs processus de test et de validation, leur résilience face aux entrées malveillantes et les mesures de protection mises en place pour empêcher toute utilisation abusive ou fuite de données. Ce niveau de diligence raisonnable est particulièrement important compte tenu de la récente envolée des investissements, du développement et des déclarations des fournisseurs dans le domaine de l'IA.

Si l'IA peut apporter des améliorations significatives en matière de productivité, de génération d'informations exploitables et d'efficacité de la sécurité, le marché est également marqué par le battage médiatique, les idées reçues et un manque de transparence.

Ce guide est conçu pour aider les RSSI à faire la part des choses, en distinguant les capacités éprouvées des simples effets de mode, afin qu'ils puissent évaluer en toute confiance les outils de sécurité basés sur l'IA et veiller à ce que toute adoption soit fiable, responsable et conforme aux exigences de gouvernance et de gestion des risques de leur organisation.



Quels éléments prendre en compte avant d'investir dans l'IA ?

Un fournisseur de services d'IA (AISP) est un prestataire qui développe et propose de l'IA en tant que service. Il prend en charge l'ensemble du processus, du choix et de la conception à l'entraînement des modèles, en passant par leur perfectionnement, leur validation et leur gouvernance continue. Les solutions d'un AISP peuvent reposer sur un modèle unique ou sur une combinaison de modèles agencés de manière à produire un résultat plus complexe et plus complet.

Presque tous les fournisseurs de solutions de sécurité commencent à utiliser l'IA dans le cadre de leurs solutions. Dans le contexte de la cybersécurité, l'IA peut être extrêmement précieuse. Alors que les systèmes de sécurité traditionnels s'appuient sur une détection basée sur les signatures, les systèmes d'IA sont capables d'identifier des schémas et des tendances, et de généraliser afin d'identifier des attaques inédites.

Ce qu'il faut savoir avant d'évaluer un AISP



Les AISP peuvent utiliser plusieurs modèles différents au sein d'une même solution



L'IA englobe un large éventail de technologies (y compris, mais sans s'y limiter, l'IA générative)



L'IA n'est pas toujours exacte et doit être optimisée



Tous les systèmes d'IA introduisent leurs propres risques, nécessitant une gouvernance

L'IA évolue constamment, ce qui rend difficile le maintien d'une définition claire de ce qui constitue « un système d'IA ». L'Organisation de coopération et de développement économiques (OCDE) définit l'IA comme un système fondé sur une machine capable de déduire des résultats à partir d'entrées sans nécessiter de programmation explicite.

Le domaine de l'IA est vaste et englobe de nombreuses techniques, dont le regroupement, la classification, la régression, la vision par ordinateur, les grands modèles de langage (LLM), les modèles de langage spécifiques à un domaine (DSLM) et l'IA agentique.

LLM et DSLM

Les grands modèles de langage (LLM) constituent l'un des domaines de l'IA ayant récemment connu un fort essor. Les LLM sont des modèles généralistes conçus pour traiter des entrées en texte libre et sont particulièrement adaptés aux tâches de traitement du langage naturel.

De même, **les modèles de langage spécifiques à un domaine (DSLM)** sont des LLM conçus pour une tâche spécifique, ou pour opérer dans un domaine précis, tel que le modèle de langage DEMIST-2 de Darktrace, qui a été spécifiquement entraîné pour des tâches liées à la sécurité.

IA générative

L'IA générative désigne un système d'IA capable de produire du contenu original, notamment du texte en langage naturel, des images et du contenu audio. Ces systèmes ont gagné en popularité avec l'essor des chatbots d'IA basés sur le cloud et reposant sur des LLM. L'IA générative ne désigne toutefois pas exclusivement les LLM. Il existe de nombreux autres systèmes d'IA générative, tels que des modèles de génération d'images et de vidéos (comme DALL-E 2 d'OpenAI ou Veo de Google), qui reposent sur les mêmes technologies que les LLM ainsi que sur d'autres techniques statistiques génératives.

Il est également important de noter que les LLM ne doivent pas nécessairement être utilisés de manière générative. Il existe un large éventail d'autres approches basées sur les LLM qui peuvent être mieux adaptées que l'IA générative pour certaines tâches.

Agentic AI

Depuis quelque temps, certains AISP accordent une grande importance au développement de solutions d'IA agentique. Le terme « agentique » est actuellement très utilisé dans le secteur de l'IA. De nombreuses définitions circulent et beaucoup d'AISP présentent leurs systèmes comme étant agentiques.

La définition de l'IA agentique varie considérablement d'un fournisseur à l'autre : elle peut désigner n'importe quel système autonome, tout système d'IA générative capable d'utiliser des outils, voire simplement un terme marketing appliqué à n'importe quel système d'IA.

Compte tenu de ces différences, il est nécessaire de demander à chaque AISP potentiel comment il définit l'IA agentique, comment celle-ci est intégrée à sa solution et quels risques sont susceptibles d'être introduits par cette mise en œuvre.

IA composite

Une autre technique clé que les AISP peuvent utiliser est l'IA composite : une combinaison de plusieurs techniques d'IA différentes au sein d'une solution unique. Cela permet à plusieurs systèmes de se soutenir mutuellement, réduisant les angles morts, et élargissant la portée de la solution.

Comprendre les limites de l'IA en cybersécurité

Bien que l'intégration de systèmes d'IA dans une solution puisse être bénéfique, en apportant des éclairages uniques et des gains d'efficacité grâce à des actions autonomes, il est important de garder à l'esprit que tous les systèmes d'IA comportent leurs propres inconvénients et risques potentiels.

Par exemple, bien que les systèmes d'IA générative soient de plus en plus populaires en raison de leur flexibilité et de leur capacité à traiter et à produire des données non structurées en langage naturel, ils sont plus performants pour certaines tâches que pour d'autres.

Lorsqu'ils envisagent d'investir dans une solution d'IA, les acheteurs doivent évaluer les avantages que l'IA offre par rapport à d'autres approches et examiner la diversité des méthodologies d'IA disponibles, ainsi que les avantages uniques que chacune d'elles procure. Dans ce livre blanc, nous aborderons certaines des principales questions que vous pourriez souhaiter poser à votre AISP afin de comprendre son approche en matière de développement, de mise en œuvre et de gouvernance de l'IA.

Enfin, si l'adoption des systèmes d'IA s'accélère rapidement, leur gouvernance et la gestion des risques associés n'en sont encore qu'à leurs débuts. De plus, les exigences en matière de gouvernance et de gestion des risques ne font qu'augmenter à mesure que les systèmes se complexifient. Comprendre les contrôles mis en place par les prestataires à ce niveau peut être déterminant pour distinguer des démonstrations techniques impressionnantes de systèmes capables de répondre aux exigences de fiabilité et de robustesse des entreprises.

Pour l'IA générative en particulier, Gartner suggère qu'elle est :



Très utile pour la génération de contenu, les interfaces utilisateur conversationnelles, la découverte de connaissances



Assez utile pour la segmentation/classification, les systèmes de recommandation, la perception, l'automatisation intelligente, la détection d'anomalies/la surveillance, les systèmes autonomes



Peu utile pour la prédiction/prévision, la planification, l'intelligence décisionnelle

En outre, certaines recherches de Gartner sont parvenues à la conclusion suivante :

Pour obtenir des solutions d'IA plus précises et plus robustes, les responsables de l'IA doivent aller au-delà de l'utilisation d'un seul modèle ou d'une seule technique, adopter des pratiques d'IA composite et adopter une vision holistique des systèmes d'IA.

Comment évaluer les fournisseurs d'IA

Lors d'une première sélection d'AISP, les acheteurs peuvent s'appuyer sur plusieurs éléments pour réduire le nombre de candidats. Cela peut se faire en recherchant des preuves concrètes et des témoignages clients.

Ceux-ci prouveront que l'outil d'IA fonctionne, et qu'il peut également fonctionner pour votre organisation. Recherchez en particulier les retours d'expérience de clients se trouvant dans des situations similaires à celle de votre organisation, qu'il s'agisse du secteur d'activité, de la taille de l'organisation, de la région ou d'autres attributs.

Lors de l'évaluation d'un AISP, il est également utile d'évaluer si le fournisseur dispose d'une culture de bonne gouvernance. Il s'agit de comprendre comment la gouvernance est intégrée aux processus de conception, de test et de déploiement de l'IA au sein de l'organisation.

Les recherches de notre rapport Darktrace : état des lieux de la cybersécurité basée sur l'IA ont révélé que seulement 37 % des organisations disposent d'une politique de gouvernance de l'IA.



Il est important de garder à l'esprit que même lorsqu'une entreprise dispose d'une politique de gouvernance de l'IA ou affirme respecter une norme reconnue dans ce domaine, il faut avant tout déterminer comment elle met concrètement en œuvre ces politiques et processus. Plus la culture et l'adoption de la gouvernance de l'IA sont solides au sein d'une entreprise, plus elles sont susceptibles de se refléter dans ses produits.

Les acheteurs doivent interroger les AISP potentiels sur leurs processus de gouvernance et de sécurité des données, sur les technologies qu'ils utilisent ainsi que sur leurs processus de test, d'évaluation, de validation et de vérification. Ces questions permettent aux acheteurs de recueillir davantage d'informations afin de vérifier qu'une solution répond aux exigences et aux politiques de l'entreprise.

Dans les sections suivantes, nous présentons les questions que les acheteurs peuvent poser autour de 5 catégories clés :

- 01 Gouvernance, sécurité et contrôle des données
- 02 Collecte de données et entraînement
- 03 Choix du modèle et de la technique
- 04 Validation des performances et de la précision
- 05 Interprétabilité, ajustabilité et transparence



Gouvernance, sécurité et contrôle des données

La gouvernance de l'IA est essentielle pour garantir que tous les systèmes et modèles d'IA d'une solution sont sécurisés, précis et correctement entretenus, ainsi que pour vérifier que toutes les données transmises à un système sont utilisées de manière responsable, éthique et légale. Un système de gouvernance solide offre une visibilité suffisante sur l'ensemble des systèmes internes, ce qui permet à une organisation de garantir la sûreté, la sécurité et l'intégrité des données.

Lorsqu'il s'agit de systèmes d'IA composites et plus complexes comportant un grand nombre de modèles, la gouvernance est essentielle. Pour les solutions qui intègrent plusieurs systèmes d'IA, il serait difficile de s'assurer que tous les modèles sont tenus à jour, sécurisés et continuellement perfectionnés pour en améliorer la précision sans une solution de gouvernance suffisamment robuste pour suivre et documenter l'ensemble des systèmes.

Une maintenance irrégulière et un manque de visibilité sur les systèmes internes peuvent entraîner l'apparition de vulnérabilités, de faiblesses et de biais susceptibles de nuire aux performances et à la précision de la solution.

Voici quelques exemples de questions qu'un acheteur devrait poser aux fournisseurs de services d'IA potentiels lorsqu'il recueille des informations sur leurs politiques et processus de gouvernance.

QUESTIONS POUR LES FOURNISSEURS

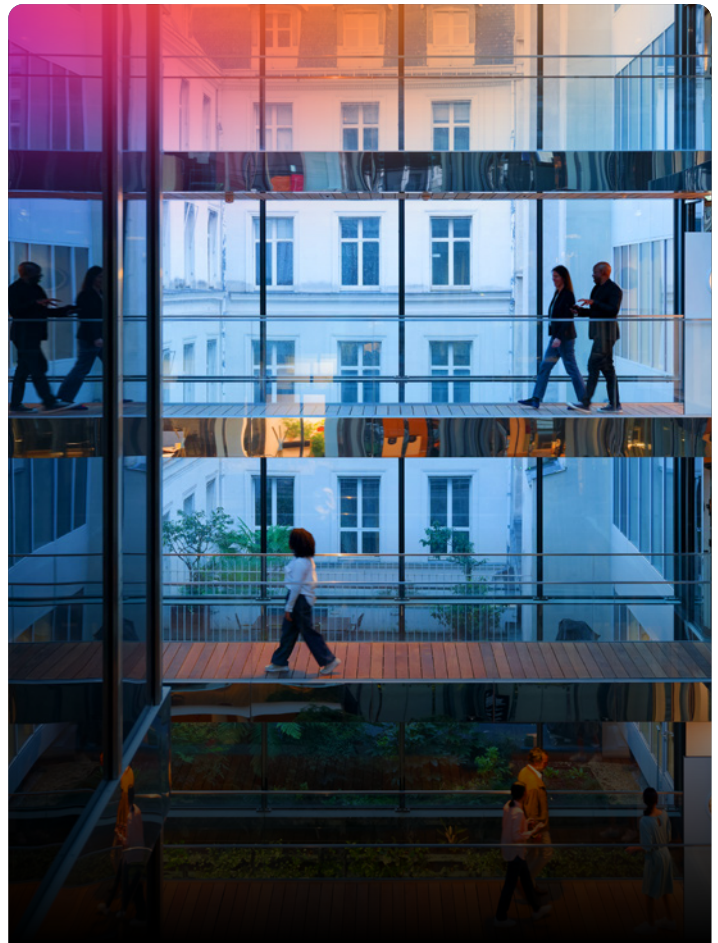
Quelles politiques et quels cadres de gouvernance de l'IA suivez-vous, et/ou quelles certifications détenez-vous actuellement ?

Comment garantissez-vous la sécurité et la confidentialité des données ainsi que le cloisonnement des informations des clients (propriété intellectuelle, données à caractère personnel, etc.) au sein de vos solutions d'IA ?

Comment protégez-vous vos systèmes d'IA contre les entrées malveillantes, l'empoisonnement des données et les autres attaques adversariales ?

Utilisez-vous des modèles ou plateformes tiers hébergés en externe pour traiter ou manipuler les données de vos clients ou de votre entreprise ? Et si oui, comment garantissez-vous la sûreté, la sécurité et la confidentialité de ces systèmes ?

Lors de l'examen des réponses aux questions sur la gouvernance et la sécurité des données, les acheteurs doivent rechercher des références à des certifications, des flux de travail ou des cadres spécifiques que leur AISP utilise afin de garantir une gouvernance solide.



Darktrace est certifiée conforme à la norme ISO/IEC 42001. La norme ISO/IEC 42001 est la première norme mondiale relative aux systèmes de management de l'IA (AIMS) et fournit des orientations précieuses sur les défis spécifiques posés par l'IA. Pour les organisations, elle établit une approche structurée pour gérer les risques et les opportunités associés à l'IA, favorisant une gouvernance solide et durable de tous les systèmes d'IA.

Il convient de noter que la norme ISO/IEC 42001 n'est pas le seul cadre de gouvernance de l'IA disponible. D'autres cadres, tels que le NIST AI Risk Management Framework, sont également largement utilisés. La norme ISO/IEC 42001 offre toutefois un niveau d'assurance supplémentaire grâce à un audit externe, au lieu de s'appuyer uniquement sur une déclaration de conformité.

Outre les politiques de gouvernance de l'IA, la sécurité et la confidentialité des données sont essentielles pour tous les systèmes, et l'identification de cadres et de certifications de sécurité des données peut également donner aux acheteurs l'assurance qu'un système est suffisamment sécurisé et capable de traiter leurs données en toute sécurité.

Collecte de données et entraînement

L'une des étapes les plus importantes de la création d'un système d'IA consiste à recueillir les données d'entraînement initiales. Si un système d'IA est entraîné à partir de données de mauvaise qualité, incohérentes, inexacts ou biaisées, il ne pourra jamais fonctionner de manière optimale.

La norme ISO/IEC 42001 et d'autres cadres de gouvernance prévoient des contrôles à respecter lors de la collecte de données destinées aux systèmes d'IA. Ils comprennent notamment l'identification des sources et la tenue de journaux d'audit afin de garantir la traçabilité des données d'entraînement et de vérifier qu'elles n'ont pas été altérées, une désidentification ou une abstraction suffisante des données utilisateur collectées en fonction du contexte de risque, ainsi qu'une vérification approfondie des éventuels biais lorsque cela est pertinent.

La liste suivante fournit aux acheteurs un point de départ pour les questions relatives à la collecte de données.

QUESTIONS POUR LES FOURNISSEURS

Quelles mesures prenez-vous pour prévenir les biais dans vos modèles d'IA et vos données d'entraînement ?

Comment réduisez-vous les risques associés aux données d'entraînement (par exemple, la régurgitation et la provenance des données) ?

Lors de l'examen des réponses, les acheteurs doivent évaluer si le fournisseur a mis en place un système de gestion de l'IA qui régit l'approvisionnement et le traitement des données d'entraînement, et si des processus clairs existent pour identifier et atténuer les biais ou les problèmes systémiques.

Lorsque des données à caractère personnel figurent dans les données d'entraînement de l'IA (hors systèmes auto-apprenants propres à une organisation), les AISP doivent pouvoir présenter les contrôles et processus spécifiques utilisés, le cas échéant, pour désidentifier les données en supprimant ces informations, ou décrire tout autre processus d'anonymisation employé.

Choix du modèle et de la technique

Différents modèles et différentes techniques d'IA conviennent à différentes tâches. Sélectionner le modèle adapté à une tâche donnée constitue une étape importante pour garantir la précision et l'efficacité d'une solution.

Les modèles et les techniques nécessitent également une gouvernance solide. Il ne s'agit pas seulement de conformité aux normes et certifications, mais aussi d'adopter les modèles et processus les plus efficaces et efficaces pour atteindre l'objectif visé. Cela implique de tenir des journaux détaillés des modèles et techniques utilisés et de veiller à ce qu'ils fassent continuellement l'objet d'analyses visant à identifier les risques et vulnérabilités, dont les risques liés à la chaîne d'approvisionnement des modèles tiers et internes.

Afin de s'assurer que les fournisseurs de services d'IA suivent des procédures de gouvernance appropriées lorsqu'ils sélectionnent des modèles et des services hébergés en interne et en externe, voici quelques questions que les acheteurs devraient poser :

QUESTIONS POUR LES FOURNISSEURS

Quel(s) type(s) de modèles d'IA utilisez-vous dans votre solution ?

Utilisez-vous une IA générative, (hébergée en interne ou en externe), et si oui, quelles mesures de protection et quels garde-fous avez-vous mis en place ?

Comment vous assurez-vous que votre choix de modèles est optimal pour chaque tâche requise ?

Utilisez-vous l'IA agentique ? Si oui, comment et dans quels domaines est-elle utilisée ?

Utilisez-vous des modèles de fondation ou généralistes ? Si oui, comment sont-ils hébergés ?

Même si les informations détaillées sur les systèmes personnalisés utilisés par les AISP sont susceptibles d'être propriétaires, les acheteurs doivent pouvoir attendre des prestataires une présentation générale des principales techniques utilisées. La transparence concernant l'utilisation de modèles de fondation tiers ou hébergés dans le cloud, les API associées et la localisation des données reste un critère important pour les acheteurs qui souhaitent obtenir des garanties concernant le traitement et l'intégrité des données.

Les AISP doivent pouvoir démontrer qu'ils ont procédé à une diligence raisonnable ainsi qu'à une étude et une analyse approfondies des tiers avant d'intégrer leurs systèmes externes afin d'en garantir la sûreté et la sécurité.

Validation des performances et de la précision

Les systèmes d'IA doivent faire l'objet de contrôles approfondis de test, d'évaluation, de vérification et de validation (TEVV) avant leur utilisation générale. L'évaluation d'un système, suivie d'optimisations, d'ajustements et d'améliorations, est essentielle pour garantir que les systèmes produits sont calibrés de manière appropriée et sécurisée afin de fournir des informations pertinentes pour la tâche prévue.

La TEVV ne se limite pas à valider la précision. De nombreuses techniques doivent être appliquées afin de vérifier qu'un système est exempt de biais, qu'il ne présente ni surapprentissage ni sous-apprentissage des données et qu'il fonctionne de manière efficace autant que précise. Une évaluation continue après la mise en production est également essentielle pour éviter toute dérive indésirable des données et des modèles. Les tests doivent également porter sur les performances physiques du modèle afin de vérifier qu'il peut produire des résultats dans un délai approprié et fonctionner avec une consommation raisonnable de ressources.

La norme ISO/IEC 42001 exige la création d'un flux de travail permettant d'évaluer la pertinence et la précision d'un système. Celui-ci doit être appliqué à chaque système d'IA avant sa mise en production.

Vous trouverez ci-dessous une liste de questions que les acheteurs pourraient poser à leur fournisseur de services d'IA pour s'assurer qu'il mène les activités de TEVV appropriées et que la gouvernance est suffisante en matière de validation des performances et de la précision.

QUESTIONS POUR LES FOURNISSEURS

Comment auditez-vous, testez-vous, évaluez-vous, vérifiez-vous et validez-vous les résultats de vos modèles d'IA ?

Comment vos modèles s'adaptent-ils à un paysage des menaces externes en constante évolution, et comment réduisez-vous le risque de dérive des données ?

Votre flux de travail TEVV comprend-il des tests de performances physiques telles que la latence, la consommation de mémoire, et les exigences en matière d'énergie et de ressources ?

Les réponses à ces questions doivent comporter des références directes à un flux de travail TEVV, qu'il fasse partie d'un AIMS existant (ou d'un autre système de gouvernance) ou d'un cadre distinct.

Interprétabilité, ajustabilité et transparence

Un aspect important de tout système d'IA consiste à garantir que les résultats produits sont interprétables, que les étapes suivies pour parvenir à ces résultats sont compréhensibles et que les utilisateurs finaux peuvent être sûrs que les résultats générés par l'IA sont corrects. Bien qu'un système d'IA soit souvent une boîte noire, un utilisateur devrait tout de même pouvoir comprendre pourquoi un système a pu parvenir à une conclusion donnée, même si les étapes exactes sont masquées.

Dans le domaine de la cybersécurité, la confiance accordée à un système d'IA est essentielle. Ces systèmes doivent pouvoir agir en temps réel et de manière autonome pour bloquer des connexions, rejeter des messages et restreindre l'accès au réseau. Si les utilisateurs finaux ne peuvent pas compter sur le système pour exécuter ces actions de manière fiable et précise, ils risquent de désactiver certaines fonctionnalités ou d'être tentés d'annuler ses actions, ce qui réduirait considérablement ses capacités de défense. Les exigences élevées en matière de confiance propres à la cybersécurité rendent de nombreux cas d'usage peu adaptés à l'automatisation par l'IA générative. La possibilité d'ajuster les systèmes d'IA est une fonctionnalité importante.

Chaque organisation possède un ensemble distinct de comportements, de structures et de cultures. La capacité d'affiner les modèles d'IA aidera à atteindre les objectifs globaux et permettra à une solution de s'adapter à un plus large éventail de cas d'usage.

Les questions suivantes constituent un point de départ pour les questions que les acheteurs pourraient poser à leur AISP afin de confirmer que leur système sera suffisamment interprétable, ajustable, et qu'il favorisera une relation de confiance solide entre les analystes humains et les outils d'IA.

QUESTIONS POUR LES FOURNISSEURS

Comment vous assurez-vous que vos systèmes d'IA sont interprétables ?

Comment favorisez-vous une relation de confiance entre les analystes humains et les résultats de l'IA ?

Comment une équipe SOC peut-elle affiner vos modèles afin de les adapter à ses politiques et exigences métier uniques ?

Un AISP doit être en mesure de décrire comment sa solution peut être adaptée à la structure, aux politiques et aux besoins uniques d'une organisation, que ce soit par l'auto-apprentissage et l'adaptation autonome ou par des ajustements manuels mis à la disposition des utilisateurs finaux.

Il est également essentiel de s'assurer que la transparence et l'interprétabilité sont prises en compte. Un AISP doit fournir des informations sur les politiques, les flux de travail et les méthodes qu'il utilise pour assurer une transparence complète et renforcer la confiance dans sa solution.

Darktrace en tant que fournisseur de cybersécurité basée sur l'IA

Darktrace développe et applique l'IA à la cybersécurité depuis plus de dix ans, faisant évoluer ses capacités face à un paysage des menaces toujours plus complexe et en rapide évolution.

Cette expérience a permis de développer une approche mature et multicouche de l'IA, dans laquelle la Self-Learning AI de Darktrace apprend en continu les schémas normaux de chaque organisation afin de comprendre les comportements, d'interpréter le contexte et d'identifier les écarts significatifs, sans s'appuyer sur des règles prédéfinies ni sur des signatures d'attaque connues. Au fil du temps, cela a permis d'acquérir une compréhension comportementale éprouvée qui aide à détecter les signaux subtils de risque qui pourraient autrement passer inaperçus.

Cette approche a été développée et validée dans des environnements dynamiques couvrant plusieurs domaines, où l'activité s'étend aux réseaux, aux identités, au cloud, aux e-mails et aux applications. En comprenant les comportements à mesure qu'ils évoluent entre les différents systèmes, Darktrace offre une visibilité continue sur l'activité à l'échelle de l'organisation, permettant une détection, une investigation et une réponse qui s'adaptent à l'évolution des environnements et des menaces. Cette approche témoigne d'un engagement de longue date en faveur de l'innovation, fondé sur des applications concrètes et des améliorations continues, ainsi que d'une attention particulière portée au développement responsable et transparent de l'IA.

En tant que pionnier dans ce domaine, Darktrace a également placé la responsabilité au cœur de ses processus. Cet engagement se reflète dans sa certification ISO/IEC 42001, qui démontre la capacité de Darktrace à fournir une Self-Learning AI de pointe au service de la résilience de la cybersécurité.

Nos parties prenantes, clients et partenaires peuvent avoir l'assurance que Darktrace développe ses systèmes d'IA de manière responsable, éthique et sûre, et gère l'utilisation de l'IA dans nos opérations quotidiennes de manière conforme, sécurisée et éthique.

Comment Darktrace sécurise les systèmes d'IA

Darktrace offre désormais ces capacités pour surveiller les risques générés par les systèmes d'IA au sein des organisations et y répondre avec Darktrace / SECURE AI. Cette solution analyse la manière dont les prompts, les agents et les systèmes sont utilisés dans le contexte propre à chaque organisation, en regroupant toutes les interactions de l'IA dans une vue unique. Cette approche unique aide les équipes à comprendre les intentions, à évaluer les risques, à protéger les données sensibles et à appliquer les politiques à l'ensemble des activités humaines et de celles des agents IA. Elle offre une visibilité et un contrôle en temps réel sur l'IA générative, les agents IA, les environnements de développement et l'IA fantôme, permettant aux équipes d'identifier les usages abusifs, les mauvaises configurations et les dérives que les approches fondées sur des règles ne détectent souvent pas.

En comprenant le contexte et les intentions des utilisateurs et des systèmes, Darktrace aide les organisations à conserver une visibilité et un contrôle suffisants à mesure que l'adoption de l'IA progresse dans toute l'entreprise.

Inscrivez-vous au programme Secure AI Readiness ici : vous bénéficierez d'un accès exclusif aux dernières actualités sur les menaces liées à l'IA, aux évolutions des approches émergentes qui façonnent la sécurité de l'IA et aux dernières innovations, dont les travaux actuellement menés par Darktrace dans ce domaine.

Vous souhaitez échanger avec un expert Darktrace sur la sécurisation de l'IA ? Inscrivez-vous ici pour obtenir des conseils pratiques sur les risques liés à l'IA qui comptent le plus pour votre entreprise et déterminer clairement les priorités en matière de gouvernance, de visibilité, de réduction des risques et de préparation à long terme.



L'approche de Darktrace en matière d'IA responsable dans la cybersécurité

Découvrez les principes qui sous-tendent l'approche de Darktrace en matière d'IA responsable, élaborée grâce à une collaboration avec des experts internationaux issus du monde universitaire et des pouvoirs publics. Découvrez comment la responsabilité, l'explicabilité et la validation continue sont intégrées à ses technologies de cybersécurité.

Lire la suite [ici](#)

Arsenal de l'IA

Approfondissez vos connaissances sur l'application de l'apprentissage automatique supervisé et non supervisé et des LLM à la cybersécurité, et découvrez comment Darktrace associe différentes techniques au sein d'une architecture multicouche afin de renforcer les équipes de sécurité humaines dans le livre blanc AI Arsenal

Lire la suite [ici](#)



Conclusion

À mesure que l'IA s'intègre davantage à la cybersécurité, le facteur le plus important pour les responsables de la sécurité n'est pas de savoir si un produit utilise l'IA, mais comment cette IA est conçue, gouvernée, validée et contrôlée. Les acheteurs doivent pouvoir obtenir des réponses claires concernant l'origine et la protection des données, le choix des modèles, les pratiques de test et de validation, la résilience face aux entrées malveillantes et la transparence des décisions prises par l'IA.

Les solutions d'IA robustes sont conçues à partir d'une combinaison de techniques soigneusement choisies, d'une évaluation continue et de cadres de gouvernance qui réduisent les risques au fil du temps, garantissant que les systèmes restent précis, adaptables et alignés sur les politiques organisationnelles à mesure que les menaces évoluent.

Lorsqu'elle est mise en œuvre de manière responsable, l'IA offre des avantages significatifs : amélioration de la détection des nouvelles menaces, investigations plus rapides et plus précises, réduction de la charge opérationnelle des équipes de sécurité et confiance accrue dans les réponses autonomes. L'approche multicouche et auto-apprenante de Darktrace montre comment différentes techniques d'IA, une gouvernance solide et des résultats explicables peuvent fonctionner de concert pour fournir une cybersécurité fiable à grande échelle.

Contactez Darktrace pour une démonstration personnalisée

[en savoir plus ↗](#)

■ **À propos de Darktrace**

Darktrace est un leader mondial de la cybersécurité basée sur l'IA, qui aide chaque jour les organisations à garder une longueur d'avance face à l'évolution des menaces. Fondée en 2013 à Cambridge, au Royaume-Uni, Darktrace fournit une plateforme de cybersécurité essentielle qui protège les organisations contre les menaces inconnues grâce à une IA qui apprend en temps réel à partir de l'environnement propre à chaque entreprise. La plateforme et les services Darktrace s'appuient sur plus de 2 700 employés qui protègent près de 10 000 clients dans le monde. Pour en savoir plus, rendez-vous sur www.darktrace.com.