SOPHOS

L'intégration de l'IA dans le monde de la cybersécurité

Comment tirer parti de l'IA en toute sécurité pour renforcer les cyberdéfenses de votre entreprise

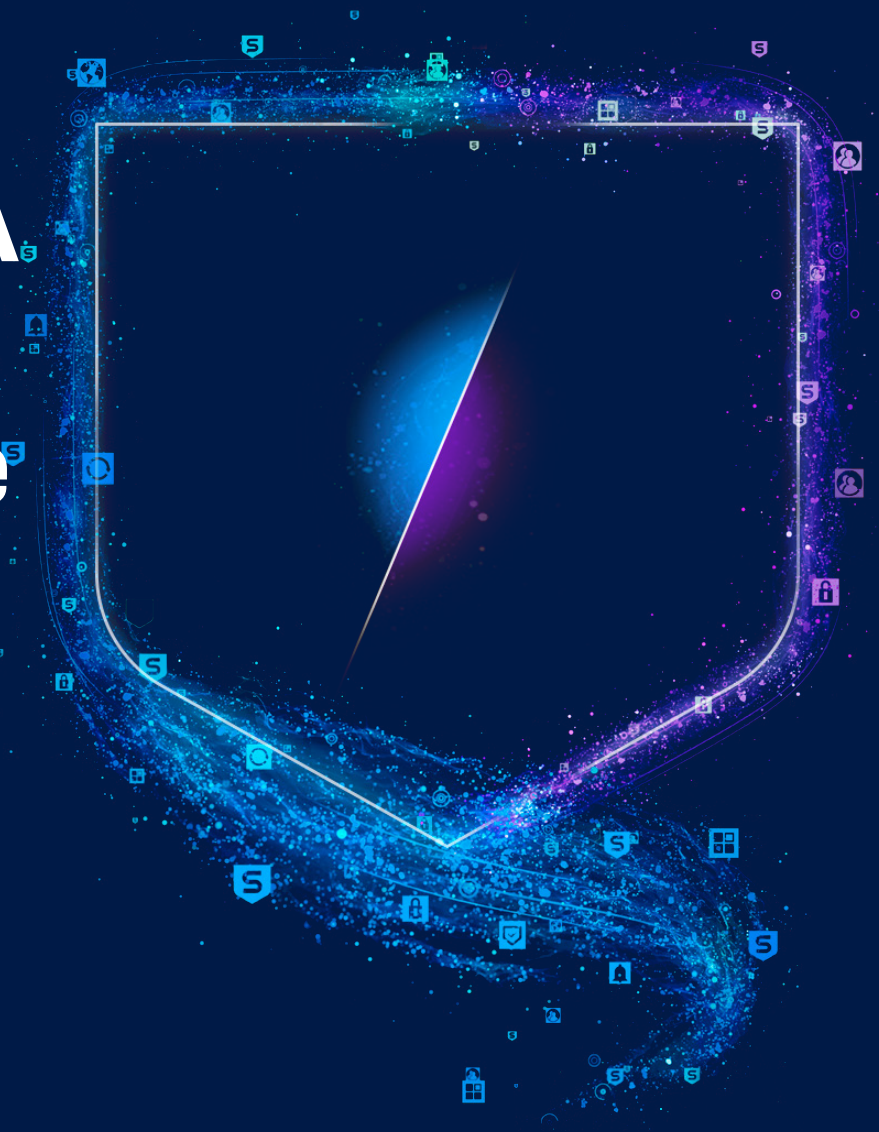

Table des matières

Introduction

La cybersécurité est saturée d'informations autour de l'IA. Les entreprises sont bombardées de promesses alléchantes de transformation des activités de cybersécurité par l'IA — protection accrue, réduction des coûts, diminution des besoins en personnel spécialisé — et de mises en garde alarmistes sur le fait que l'IA va inaugurer une toute nouvelle ère de cyberattaques.

Ce guide est conçu pour aider les entreprises à s'y retrouver dans l'engouement et les idées fausses autour de l'IA dans le domaine de la cybersécurité. Il explique ce que l'IA peut (et ne peut pas) faire pour renforcer les cyberdéfenses des entreprises et explore les risques opérationnels et de cybersécurité que l'IA a fait émerger. Il fournit également des conseils sur la manière d'atténuer ces risques afin d'exploiter l'IA en toute sécurité pour améliorer à la fois la cyberprotection et le retour sur investissement.

Au fil des pages, ce guide offre des informations sur la réalité des usages, des attentes et des préoccupations en matière d'IA, sur la base d'une enquête indépendante menée fin 2024 auprès de 400 responsables IT/cybersécurité. Ces avis de première ligne offrent un contexte très utile et servent de point de comparaison pour les entreprises qui cherchent à se positionner sur la question de l'IA. Pour les résultats complets, consultez Au-delà de l'engouement : la réalité de l'IA dans le domaine de la cybersécurité

Au bout du compte, avec ou sans IA, l'objectif reste le même : délivrer de manière optimale le niveau de cyber-résilience dont votre entreprise a besoin pour réussir, tout en minimisant vos dépenses. En d'autres termes, utiliser au mieux le budget de cybersécurité (inévitablement limité) pour soutenir l'activité de votre entreprise. Ce guide vous aidera à y parvenir à l'ère de l'IA.

La réalité de l'IA dans le domaine de la cybersécurité

L'IA est un acronyme qui décrit une gamme de capacités susceptibles de soutenir et d'accélérer la cybersécurité de plusieurs manières différentes. La bonne nouvelle est que l'IA apporte plus d'avantages aux défenseurs qu'aux adversaires. Deux approches de l'IA couramment utilisées en cybersécurité sont : les modèles de Deep Learning et d'IA Générative.

Deep learning

Les modèles de Deep Learning (DL) APPLIQUENT des apprentissages pour effectuer des tâches. Ils peuvent accélérer l'application des connaissances bien au-delà des possibilités humaines. Par exemple, des modèles de DL correctement entraînés peuvent identifier si un fichier est malveillant ou bien inoffensif en une fraction de seconde, et ce sans jamais avoir vu ce fichier auparavant.

Le Deep Learning est parfait pour effectuer des tâches répétitives à grande échelle. Il permet de créer un modèle « statistique » qui analyse les nouveaux éléments en fonction de tout ce qu'il a appris à partir de son vaste ensemble de données d'apprentissage. C'est ainsi que les modèles de DL peuvent évaluer des millions d'échantillons de fichiers sans faiblir afin d'identifier la présence de logiciels malveillants. Ces modèles sont donc largement utilisés pour améliorer les capacités de protection des solutions de cybersécurité.

Les modèles de DL permettent aux défenseurs de faire face aux énormes volumes de menaces créées par les adversaires utilisant l'automatisation et la cybercriminalité en tant que service. Les modèles de DL peuvent également être améliorés et adaptés à mesure que les attaques évoluent, ce qui leur permet de rester à jour avec l'environnement par rapport des menaces.

IA générative

Les modèles d'IA Générative (Generative AI/GenAI) assimilent les entrées et les utilisent pour CRÉER du nouveau contenu Parmi les exemples d'applications, citons :

- La création d'un résumé à jour en langage naturel de l'activité des menaces et des étapes recommandées pour l'analyste.
- La mise en évidence du comportement des attaquants par l'analyse des commandes à l'origine des détections.
- La possibilité pour les analystes de faire des recherches en langage naturel plutôt que des requêtes basées sur du code pour investiguer les détections suspectes.
- La possibilité de prioriser l'application de correctifs en fonction de la propension d'une vulnérabilité à être exploitée.

L'IA générative est un outil puissant qui permet d'accélérer les opérations de sécurité. En traitant une grande partie des données les plus lourdes, elle permet aux analystes de prendre rapidement des décisions éclairées et de se consacrer aux tâches qui auront le plus d'impact. De cette manière, l'IA générative permet souvent de réduire la pression exercée sur les analystes, diminuant ainsi le risque d'épuisement professionnel et du départ des employés. Grâce à ces modèles, la barrière technologique des opérations de sécurité peut également être abaissée. Les analystes moins expérimentés peuvent ainsi apporter rapidement une contribution positive et monter plus rapidement en compétences.

L'origine de l'IA générative

La base de l'IA générative moderne est le « transformer », un réseau neuronal d'apprentissage profond qui analyse le contexte et les relations entre les entrées (par exemple, les mots d'une phrase) et utilise ces connaissances pour produire des résultats pertinents. Les ***transformers*** sont souvent utilisés dans des tâches de traitement du langage naturel (NLP), telles que la traduction de textes et la fourniture de réponses à des questions. D'ailleurs, le 'T' de ChatGPT correspond à *Transformer*.

Bien que les ***transformers*** soient largement utilisés dans la GenAI, tous ces réseaux ne sont pas génératifs. Ainsi, le modèle BERT (Bidirectional Encoder Representations from Transformers) est un cadre de Machine Learning à code source ouvert pour le NLP qui est en mesure de lire un texte d'entrée de manière bidirectionnelle, c'est-à-dire à la fois de gauche à droite et de droite à gauche. Cette approche lui permet d'améliorer considérablement la compréhension contextuelle des textes non étiquetés. Chez Sophos, nous utilisons BERT depuis de nombreuses années pour identifier et défendre les entreprises contre les attaques Business Email Compromise.

Il n'existe pas de solution unique

Les modèles d'IA varient fortement en termes de taille. **Les modèles massifs**, tels que Microsoft Copilot et Google Gemini, sont de grands modèles de langage (LLM) entraînés sur un ensemble très complet de données qui peuvent effectuer un large éventail de tâches. En revanche, les petits modèles sont généralement conçus et entraînés sur un ensemble de données très spécifiques pour effectuer une seule tâche, comme détecter des URL ou des exécutables malveillants. Bien que leur portée soit plus limitée, les **petits modèles** présentent des avantages en termes de coût, de vitesse et de performance par rapport aux LLM.

Les limites de l'IA

L'IA à elle seule ne peut résoudre tous les problèmes, du moins pas dans un avenir prévisible. L'IA vient compléter l'expertise humaine, mais ne saurait la supplanter complètement. Les menaces sont extrêmement complexes et la réalisation d'opérations de sécurité efficaces requiert à la fois des compétences techniques et la capacité d'appliquer les connaissances acquises en fonction des spécificités de l'entreprise. À elle seule, l'IA ne permet pas aux entreprises de garder une longueur d'avance sur les organisations cybercriminelles bien financées et expérimentées d'aujourd'hui.

TYPE

Deep Learning IA — *Appliquer*

Utilise des réseaux neuronaux artificiels pour reconnaître des modèles et prendre des décisions d'une manière qui imite le cerveau humain. Il APPLIQUE les connaissances acquises pour effectuer des tâches.

Exemple : **Détecter les URL malveillantes**
Un modèle d'IA est entraîné pour identifier les sites web malveillants, permettant ainsi aux solutions de sécurité d'en bloquer l'accès.

IA générative — *Créer*

Utilise la structure et les modèles de données existantes pour CRÉER (générer) des contenus totalement nouveaux.

Exemple : **Résumé des Dossiers menaces**
Le modèle d'IA crée un résumé de l'activité de la menace et recommande aux analystes les étapes à suivre

TAILLE

Modèles d'IA massifs

Outils polyvalents entraînés sur de vastes volumes de donnée publiques afin de faciliter l'accomplissement d'un large éventail de tâches.

Exemple : **Microsoft Copilot, Google Gemini**

Petits modèles d'IA

Ces modèles axés sur les résultats sont conçus, entraînés et construits pour des cas d'usage spécifiques.

Exemple : **Modèle de détection de malwares sous Android**

Les taux d'adoption de l'IA

L'IA est déjà largement intégrée dans l'infrastructure de cybersécurité de la plupart des entreprises :

- 73 % déclarent que leurs solutions de cybersécurité incluent des modèles de Deep Learning
- 65 % déclarent que leurs solutions de cybersécurité exploitent des capacités d'IA générative

Les applications de l'IA en matière de cybersécurité ne se limitent pas aux éditeurs externes puisque 34 % des entreprises utilisent déjà l'IA générative en interne pour améliorer leur cybersécurité, par exemple pour aider à générer des emails de test de phishing.

L'adoption de l'IA est susceptible de devenir quasi universelle assez rapidement, les capacités de l'IA figurant désormais sur la liste des exigences de 99 % (arrondi) des entreprises lors de la sélection d'une plateforme de cybersécurité :

- 57 % déclarent que les capacités d'IA sont essentielles/extrêmement importantes
- 41 % déclarent que les capacités d'IA sont importantes

Avec ce niveau d'adoption et d'utilisation future, comprendre les risques et les mitigations associées à l'IA en matière de cybersécurité est une priorité pour les entreprises de toutes tailles, quels que soient leurs secteurs d'activité.

73 %
Utilisent des outils de cybersécurité basés sur des modèles de Deep Learning

65 %
Utilisent des outils de cybersécurité dotés de capacités de GenAI

99 %
Choisissent des plateformes de cybersécurité dotées de capacités d'IA

L'IA générative : de grandes attentes

L'engouement autour de l'IA générative a suscité de grandes espérances quant aux possibilités offertes par cette technologie pour améliorer les résultats en matière de cybersécurité. Notre enquête a révélé le principal avantage que les entreprises souhaitent voir en matière de capacités d'IA générative intégrées aux outils de cybersécurité, comme indiqué dans le tableau ci-dessous.

Avantage souhaité N°1 de l'IA générative Réponses classées en première position	
1=	Protection renforcée contre les cybermenaces (20 %)
1=	Retour sur investissement accru en matière de cybersécurité (ROI) (20 %)
3	Accroître l'efficacité et l'impact des analystes IT (17 %)
4	L'assurance de suivre l'évolution des innovations en matière de cybersécurité (15 %)
5=	Une plus grande tranquillité d'esprit en sachant que notre entreprise est bien protégée contre les attaques (14 %)
5=	Diminution de l'épuisement des employés (grâce à l'automatisation des tâches qui libère du temps pour les équipes de cybersécurité) (14 %)

Quels avantages, le cas échéant, souhaitez-vous que les capacités d'IA générative intégrées aux outils de cybersécurité vous apportent ? Réponses classées en première position (n=400)

Le large éventail de réponses révèle qu'il n'y a pas d'avantage unique et remarquable attendu concernant l'IA générative en matière de cybersécurité. Parallèlement, les gains souhaités le plus couramment concernent l'amélioration de la cyberprotection ou des performances de l'entreprise (tant financières qu'opérationnelles). Les données suggèrent également que l'intégration des capacités d'IA générative dans les solutions de cybersécurité offre une sérénité et l'assurance qu'une organisation soit informée des dernières capacités en matière de protection.

La réduction des risques de burnout des employés, visible en bas du classement, suggère que les entreprises sont moins conscientes ou moins préoccupées par le potentiel de l'IA générative pour soutenir les utilisateurs. Face à la pénurie de personnel en charge de la cybersécurité, la réduction des risques de départ est un domaine important où l'IA peut aider.

L'amélioration de la **protection** et du **retour sur investissement** sont les principaux avantages que les entreprises attendent de l'IA générative.

Les risques de l'IA pour la cybersécurité

L'utilisation de l'IA dans les activités de cybersécurité est à double tranchant. Si l'IA offre des avantages considérables aux défenseurs dans leur lutte contre les adversaires, elle présente également un certain nombre de risques :

1. **Risque en matière de menaces :** utilisation de l'IA dans les cyberattaques
2. **Risque en matière de défense :** une IA de mauvaise qualité et mal déployée
3. **Risque opérationnel :** dépendance excessive à l'IA
4. **Risque financier :** retour sur investissement faible
5. **Risque de détournement :** compromission de modèles d'IA publics par des adversaires

1. Risque en matière de menaces : utilisation de l'IA dans les cyberattaques

Beaucoup ont affirmé sans retenue que l'IA était en train de créer un paysage de menaces inédit, mais la réalité est bien moins spectaculaire. Les discussions autour l'IA dans les forums sur la cybercriminalité sont peu nombreuses et de nombreux acteurs malveillants restent sceptiques au sujet de cette technologie. Lorsqu'elles sont observées, les tentatives de développement de malwares, d'outils d'attaques ou d'exploits à l'aide de l'IA sont généralement rudimentaires et de faible qualité.

Tout comme les organisations légitimes, les adversaires utilisent principalement l'IA pour améliorer la qualité de leur contenu et l'efficacité de leurs opérations, mais avec des objectifs totalement différents. Pour plus de détails sur les dernières menaces et les attaques basées sur l'IA, consultez notre article de blog Sophos News.

Améliorer la qualité des contenus

L'une des applications les plus rapides, faciles et accessibles de l'IA pour les adversaires consiste à améliorer la qualité et la crédibilité des contenus des emails de phishing et des tentatives d'escroquerie.

Les signes caractéristiques associés aux tentatives de phishing, tels que les fautes de grammaire, d'orthographe et les problèmes de mise en forme, sont facilement corrigés par les outils d'IA. Ainsi, un email bien rédigé utilisé pour des campagnes de phishing peut être créé par des LLM publics en moins d'une minute. De même, il est désormais aisé d'accéder, dans n'importe quelle langue, à des textes et des posts pour les réseaux sociaux convaincants et bien rédigés, qui visent à inciter les destinataires à cliquer sur des liens ou à communiquer des informations personnelles. Les LLM permettent également aux attaquants d'intégrer aisément des informations actuelles à leurs contenus, ce qui augmente d'autant plus la propension des victimes à tomber dans leur piège.

Les outils d'IA générative ont également ouvert la voie à de nouvelles formes d'escroquerie consistant à se faire passer pour un cadre supérieur afin d'inciter les cibles à effectuer des virements bancaires. La technologie de clonage de voix a atteint un tel niveau qu'avec suffisamment d'entraînement, les adversaires peuvent faire croire à quelqu'un qu'il parle à une vraie personne. Ces tentatives de phishing vocal (ou « vishing ») consistent souvent pour l'acteur malveillant à appeler un membre du personnel en se faisant passer pour un cadre supérieur afin de lui « demander » d'effectuer un achat de carte-cadeau, un paiement bancaire ou un transfert de fichier illicite.

Les adversaires utilisent également les technologies de deepfake basées sur l'IA pour usurper l'identité visuelle d'une personne. Des vidéos deepfake ont ainsi été exploitées pour inciter des employés peu méfiants à effectuer des paiements importants ou pour tromper des programmes de reconnaissance faciale dans le cadre de demandes de prêts ou d'ouvertures de comptes bancaires.

Améliorer l'efficacité opérationnelle

Tout comme de nombreuses entreprises, les attaquants utilisent des agents conversationnels (chatbots) basés sur l'IA pour améliorer leur expérience utilisateur. Certains acteurs malveillants ont recours aux LLM pour améliorer les forums qu'ils fréquentent en créant des chatbots et des réponses automatiques. Dans un exemple relayé par Sophos X-Ops, le forum XSS a créé un chatbot pour répondre aux questions des utilisateurs. L'administrateur annonçait (traduit du russe) :

« Dans cette section, vous pouvez discuter avec une IA (Intelligence Artificielle). Posez votre question, notre bot IA vous répondra.... Cette section et le bot IA sont conçus pour résoudre des problèmes techniques simples, pour offrir un divertissement technique à nos utilisateurs [et] pour donner aux utilisateurs une idée des possibilités de l'IA ».

Construire et entraîner des modèles personnalisés nécessite une expertise étendue en matière d'IA, laquelle est une ressource coûteuse et rare. Si certains groupes de cybercriminels ont acquis une expertise interne en matière d'IA, les acteurs malveillants s'appuient généralement sur les LLM existants pour mener leurs attaques, plutôt que de créer leurs propres LLM.

Le contexte est important

Il est important de replacer dans son contexte l'utilisation de l'IA à des fins malveillantes. L'IA n'est qu'un des nombreux éléments de la boîte à outils des attaquants. Depuis plusieurs années, les acteurs malveillants recourent à l'automatisation et aux modèles de cybercriminalité en tant que service en vue d'accroître l'ampleur et la fréquence de leurs attaques. Pour de nombreuses entreprises, ces capacités vont avoir un impact plus important sur l'exposition aux risques que l'IA.

2. Risque en matière de défense : une IA de mauvaise qualité et mal déployée

Comme nous l'avons vu, les modèles d'IA sont déjà largement intégrés dans les cyberdéfenses des entreprises. Bien que cela parte toujours d'une bonne intention, les modèles d'IA de mauvaise qualité et mal implémenté peuvent engendrer des risques considérables en matière de cybersécurité. La propension des modèles d'IA à engendrer ces risques dépend de plusieurs facteurs :

- **Qualité des données sur lesquelles les modèles sont entraînés**. L'expression anglaise « garbage in, garbage out », que l'on pourrait traduire par « à entrées erronées, sorties erronées », s'applique tout particulièrement à l'IA. De fait, l'utilisation de données de mauvaise qualité dans la phase d'apprentissage des modèles risque d'introduire des erreurs, tandis que l'utilisation de jeux de données déséquilibrés peut fausser les résultats en raison d'une sur-représentation ou d'une sous-représentation de certaines variables.

  Plus vaste est le volume de données d'apprentissage de haute qualité, meilleur est le résultat obtenu.

- **Expertise des équipes qui créent les modèles**. L'élaboration de modèles d'IA efficaces destinés à la cybersécurité nécessite une compréhension approfondie de deux aspects distincts, mais complémentaires :

  - **Les menaces :** Pour déterminer ce que le modèle d'IA doit accomplir pour vous, il faut d'abord comprendre comment les malwares et les adversaires opèrent.
  - **L'IA :** Une fois que vous avez déterminé ce que l'IA doit faire, il vous reste à identifier et à construire le modèle adéquat pour atteindre cet objectif.

  Pour construire des modèles d'IA efficaces ayant un effet concret en matière de cybersécurité, il est essentiel que ces deux ensembles de compétences fonctionnent de concert, en tirant parti de leur expertise réciproque.

- **Qualité du processus de développement et de déploiement des produits**. Au milieu de l'année 2024, le déploiement d'une mise à jour défectueuse du contenu d'une solution de cybersécurité a entraîné des perturbations immédiates pour les entreprises du monde entier. Des outils d'IA mal testés, mal évalués et mal déployés peuvent s'avérer encore plus destructeurs, sans compter le risque que le problème ne soit pas facilement identifié ou corrigé.

Un faux sentiment de (cyber) sécurité

Les entreprises sont largement conscientes du risque d'une IA mal développée et déployée dans les solutions de cybersécurité. La grande majorité (89 %) des professionnels IT/cybersécurité interrogés se disent préoccupés par de potentielles failles dans les capacités d'IA générative au sein des outils de cybersécurité qui pourraient nuire à leur entreprise, 43 % se disant extrêmement inquiets et 46 % plutôt inquiets.

Dès lors, il n'est pas surprenant que 99 % (chiffre arrondi) des répondants déclarent que pour évaluer les capacités de l'IA générative dans les solutions de cybersécurité, les entreprises évaluent la qualité des processus et des contrôles de cybersécurité utilisés dans le développement des modèles :

- 73 % déclarent évaluer pleinement la qualité des processus et des contrôles de cybersécurité
- 27 % déclarent évaluer partiellement la qualité des processus et des contrôles de cybersécurité

Même si le pourcentage élevé de personnes déclarant mener une évaluation complète peut paraître encourageant à première vue, il suggère en réalité que de nombreuses entreprises sont exposées à un angle mort majeur dans ce domaine.

L'évaluation des processus et des contrôles utilisés pour développer les capacités d'IA générative nécessite de la transparence de la part du fournisseur et un degré raisonnable de connaissances en matière d'IA de la part de l'évaluateur. Malheureusement, les deux sont rares. Les fournisseurs de solutions rendent rarement facilement accessible l'intégralité de leurs processus de déploiement et de développement de l'IA générative, et les équipes IT ont souvent une connaissance limitée des bonnes pratiques en matière de développement de l'IA. Pour de nombreuses entreprises, ce résultat suggère qu'elles « ne savent pas ce qu'elles ne savent pas ».

3. Risque opérationnel : dépendance excessive à l'IA

L'IA intervient dans presque tous les aspects de notre vie quotidienne, qu'il s'agisse de calculer le meilleur itinéraire pour se rendre au supermarché ou de recommander des séries télévisées. Sa nature omniprésente fait qu'il est facile de se reposer trop facilement sur celle-ci et de tenir pour acquis que l'IA peut accomplir certaines tâches mieux que les opérateurs humains. Heureusement, la plupart des entreprises sont conscientes et préoccupées par les conséquences sur la cybersécurité d'une dépendance excessive à l'IA :

- 84 % s'inquiètent des pressions exercées pour réduire les effectifs des professionnels de la cybersécurité
- 87 % sont préoccupés par le défaut de responsabilité en matière de cybersécurité qui découle de telles pratiques

La première chose à faire pour atténuer ces risques est d'être attentif à leur existence. Il ne faut jamais perdre de vue que l'IA n'est qu'un outil parmi d'autres dans les cyberdéfenses d'une entreprise ; bien qu'elle soit un élément précieux de la pile de sécurité, elle ne représente pas nécessairement la bonne approche et suffit rarement à elle-seule. Chaque entreprise est différente et l'utilisation de l'IA doit être adaptée à sa structure et ses besoins.

4. Risque financier : retour sur investissement faible

Les capacités d'IA générative de haute qualité présentes dans les solutions de cybersécurité sont coûteuses à développer et à maintenir. Les responsables IT/cybersécurité sont attentifs aux conséquences de telles dépenses, 80 % d'entre eux estimant que l'IA générative augmentera considérablement le coût de leurs produits de cybersécurité.

Malgré ces risques d'augmentation des prix, la plupart des entreprises considèrent l'IA générative comme un moyen de réduire leurs dépenses globales en matière de cybersécurité, avec 87 % des personnes interrogées se disant convaincues que les coûts de l'IA générative dans les outils de cybersécurité seront entièrement compensés par les économies qu'elle génèrera.

En parallèle, les entreprises reconnaissent que la quantification de ces coûts constitue un défi. Les dépenses liées à l'IA générative sont généralement intégrées au prix global des produits et services de cybersécurité, rendant ainsi difficile l'identification du montant que les entreprises dépensent avec l'utilisation de l'IA générative pour la cybersécurité. Reflétant ce manque de visibilité, 75 % conviennent que ces coûts sont difficiles à mesurer (39 % tout à fait d'accord, 36 % plutôt d'accord).

Sans reporting efficace, les entreprises risquent de ne pas obtenir le retour sur investissement souhaité en matière d'IA pour la cybersécurité ou, pire encore, d'orienter vers l'IA des investissements qui auraient pu être utilisés plus efficacement ailleurs.

5. Risque de détournement : compromission des grands modèles de langage (LLM)

Les risques pour la cybersécurité relatifs à l'IA vont au-delà des outils et des applications de cybersécurité. Le développement fulgurant du recours aux LLM publics dans le monde entier a amené des acteurs sophistiqués et bien financés à compromettre les modèles eux-mêmes afin de les aider à atteindre leurs objectifs. Ce problème risque de se manifester de plusieurs manières, notamment :

- **Attaques par empoisonnement (Data Poisoning)**. Dans leur article de 2023 intitulé Poisoning Web-Scale Training Datasets is Practical, Carlini ***et al.*** ont démontré que l'empoisonnement des données (c'est-à-dire l'introduction de données corrompues dans la phase d'entraînement d'un modèle afin d'influencer ses résultats) constitue un risque de menace réel.
- **Portes dérobées au profit d'acteurs étatiques**. De nombreux États-nations sont dotés des ressources nécessaires pour créer de puissants LLM. En intégrant des portes dérobées secrètes puis en mettant les modèles à la disposition du public, les acteurs étatiques peuvent manipuler les LLM à leur avantage si nécessaire.
- **Utilisation illicite de LLM (LLM spoofing)**. Des acteurs malveillants peuvent

compromettre des LLM légitimes (par exemple, en ajoutant des portes dérobées) pour ensuite annoncer les changements comme des « améliorations ». Dans l'optique d'inciter les cibles à utiliser leur outil compromis, ces derniers usurpent le nom d'un fournisseur réputé, par exemple en omettant une lettre ou en remplaçant la lettre O par le chiffre 0.

Pour une analyse approfondie des compromissions relatives aux LLM, consultez les dernières recherches de l'équipe Sophos AI.

Mesures concrètes pour exploiter au mieux l'IA

Même si l'IA comporte des risques, avec une approche réfléchie, les entreprises peuvent les gérer et tirer parti de l'IA en toute sécurité pour améliorer leurs cyberdéfenses. Bon nombre de ces recommandations peuvent également être utilisées pour faciliter l'implémentation de l'IA dans d'autres domaines.

Risque en matière de menaces : renforcez les cyberdéfenses à l'ère de l'IA

Il convient de chercher avant tout à améliorer la résilience face aux menaces alimentées par l'IA. Étant donné que les adversaires utilisent principalement l'IA pour améliorer la qualité et la crédibilité des emails de phishing et des tentatives d'escroquerie, il est logique de se concentrer sur ces domaines. Voici quelques suggestions :

- **Renforcez la protection de votre messagerie**. Recherchez des solutions capables de détecter les emails de phishing et les tentatives d'escroquerie générés par l'IA, afin d'éviter qu'ils ne se retrouvent dans les boîtes de réception des utilisateurs.
- **Déployez une protection contre les attaques Business Email Compromise/ protection VIP**. Choisissez des solutions de sécurité des messageries qui analysent le ton et le style du contenu pour détecter les escroqueries.
- **Méfiez-vous particulièrement des messages sur les réseaux sociaux** : les utilisateurs ont tendance à baisser leur garde lorsqu'ils naviguent sur les réseaux sociaux, et sont donc plus susceptibles de se faire prendre au piège.
- **Mettez en place des procédures pour limiter les risques posés par le clonage vocal**, tels que des procédures à suivre en cas de réception d'un paiement inattendu ou d'une demande de partage de données. Quelques exemples d'actions :
  - Rappeler l'interlocuteur pour vérifier la demande.
  - Mettre en place des codes ou des phrases/mots de passe.

Risque en matière de défense : évaluez la qualité de l'IA utilisée dans les produits de cybersécurité

Soyez attentifs aux risques et à l'impact d'une IA de mauvaise qualité vis-à-vis de vos investissements de sécurité.

Demandez aux éditeurs des précisions :

- **Données d'entraînement**. Quelle est la qualité, la quantité et la source des données sur lesquelles les modèles sont entraînés ? De meilleures données d'entrée conduisent à de meilleurs résultats.
- **Équipes de développement**. Découvrez les personnes agissant derrière les modèles. Quel niveau d'expertise en IA ont-elles ? Dans quelle mesure connaissent-elles les menaces, les comportements des adversaires et les opérations de sécurité ?
- **Ingénierie et processus de déploiement des produits**. Quelles étapes le fournisseur suit-il lors du développement et du déploiement des fonctionnalités d'IA dans ses solutions ? Quels contrôles et vérifications sont en place ?

Posez-vous la question : Jusqu'à quel point ai-je la certitude que cette organisation maîtrise l'IA et qu'elle met en place les contrôles de qualité et de déploiement qui s'imposent ?

Risque opérationnel : appréhendez l'IA avec une vision axée sur l'humain

Contrairement aux membres de vos équipes, l'IA restera indifférente à une éventuelle violation de vos données. Et si le pire venait à se produire et que vous êtes victime d'une attaque, vous aurez besoin d'une équipe expérimentée, capable de comprendre la situation et d'y remédier en tenant compte des particularités de votre entreprise.

- **Conservez une vue globale**. L'IA n'est qu'un élément de la boîte à outils du défenseur. Servez-vous-en, mais ayez conscience que la responsabilité en matière de cybersécurité incombe en fin de compte à des humains.

- **Ne remplacez pas, mais accélérez plutôt**. La pénurie mondiale de professionnels qualifiés dans le domaine de la cybersécurité est un fait bien connu. Les problèmes majeurs d'épuisement professionnel ne font qu'exacerber la situation. Plutôt que de considérer l'IA comme un moyen de réduire les effectifs, concentrez-vous d'abord sur la façon dont cette technologie peut faciliter le travail de votre personnel. En prenant en charge de nombreuses tâches d'opérations de sécurité répétitives de bas niveau et en fournissant des informations guidées, l'IA peut :
  - Libérer du temps en faveur de tâches plus utiles et ayant un impact sur l'activité de l'entreprise.
  - Réduire la surcharge d'alertes, et donc la fatigue.
  - Accélérer le développement professionnel des analystes qualifiés.
  - Permettre à des analystes moins expérimentés d'effectuer des opérations de sécurité, et créer ainsi un vivier de ressources.

Risque financier : des décisions en matière d'investissement IA à prendre de manière rigoureuse

C'est l'un des domaines où il est le plus facile pour les entreprises d'atténuer les effets, de nombreux facteurs étant entièrement sous leur contrôle.

- **Fixez-vous des objectifs**. Soyez clair, spécifique et précis quant aux résultats que vous attendez de l'IA.
  - Identifiez vos besoins. Quelles sont vos lacunes ? En quoi l'IA peut-elle vous être utile ?
  - Prenez en compte les gains financiers, les gains de temps et les gains de protection.
- **Quantifiez les avantages**. Comprenez à quel point les investissements en IA feront une différence.
  - Si l'objectif est de réduire le coût total de la cybersécurité, calculez les économies qui en résulteront.
  - Si vous souhaitez que l'IA réduise le taux de départ des équipes IT/cybersécurité, expliquez clairement quel sera l'impact de l'outil d'IA sur les membres du personnel. Quelles tâches seront retirées de leurs listes ? Combien d'heures seront ainsi libérées ?
- **Priorisez les investissements**. L'IA peut être utile de bien des façons ; certaines auront un impact plus important que d'autres. Identifiez les indicateurs importants pour votre entreprise : économies financières, impact sur la rétention du personnel, réduction de l'exposition, etc., et comparez les différentes options.
- **Mesurez l'impact**. Les décisions d'investissement sont généralement fondées sur de bonnes intentions. Assurez-vous que les performances réelles sont conformes aux attentes initiales. Avez-vous obtenu les avantages que vous recherchiez ? Y a-t-il des gains imprévus ? Y a-t-il des domaines dans lesquels vous n'obtenez pas les résultats escomptés ? Utilisez ces informations pour procéder aux ajustements nécessaires.

Demandez-vous si l'IA est le meilleur moyen d'atteindre votre objectif ou si une autre technologie ou approche aurait un impact plus important.

Risque de détournement : restez attentif au danger

Ce risque est le plus difficile à atténuer pour les entreprises. Toutefois, le simple fait d'être attentif à ce risque permet d'en réduire l'impact. Cela dit, lorsque vous décidez d'opter pour un LLM public, vérifiez les points suivants :

- **Modèles provenant de fournisseurs réputés**. Bien qu'ils ne soient pas à l'abri des attaques par empoisonnement des données, ces fournisseurs seront plus susceptibles de rendre publics les éventuels problèmes liés aux données de sortie.
- **Nom du fournisseur**. Attention, les attaquants usurpent le nom de fournisseurs réputés pour faire croire que leurs modèles compromis sont légitimes.

Les spécialistes de l'IA appliquée à la cybersécurité travaillent activement à l'élaboration d'approches visant à neutraliser ce risque.

Conclusion

L'IA présente des avantages considérables pour la cybersécurité. En triant le vrai du faux autour de l'IA et en adoptant une stratégie réfléchie axée sur les résultats, les entreprises peuvent tirer parti de cette technologie pour renforcer leurs cyberdéfenses et accroître les capacités de leurs précieux experts en informatique et en cybersécurité.

À propos de l'enquête

Source : Au-delà de l'engouement : la réalité de l'IA dans le domaine de la cybersécurité

Sophos a chargé le cabinet de recherche indépendant Vanson Bourne de mener une enquête auprès de 400 responsables IT/cybersécurité dans des entreprises comptant entre 50 et 3 000 employés. L'enquête a été conduite en novembre 2024 et les personnes interrogées appartenaient à 13 secteurs d'activité. Afin d'assurer une large représentation du secteur, l'enquête a été réalisée indépendamment des éditeurs, les entreprises interrogées utilisant des solutions de sécurité endpoint émanant de 19 éditeurs distincts.

À propos de Sophos

Sophos est un leader mondial de la cybersécurité offrant une gamme complète de produits et de services de cybersécurité maintes fois primés, dont des pare-feux, des solutions Endpoint, des outils EDR/XDR, ainsi que des services MDR (Managed Detection and Response) et de réponse aux incidents (IR).

Sophos renforce la cybersécurité grâce à l'IA depuis 2017, en alliant IA et expertise humaine pour bloquer le plus grand nombre de menaces possibles, où qu'elles se trouvent. Les capacités de Deep Learning et d'IA générative qui permettent de résoudre les problèmes les plus critiques pour les clients sont intégrées dans nos produits et services et sont fournies depuis la plus grande plateforme de sécurité AI-native de l'industrie. Entraînée à partir de données issues d'attaques perpétrées dans plus de 600 000 environnements clients différents, notre plateforme adaptative AI-native offre une protection inégalée contre les menaces avancées et renforce la capacité des défenseurs.

Pour en savoir plus et découvrir les solutions Sophos, consultez www.sophos.fr

2025-01-15 (WP-MP)

SOPHOS