Rapport Zscaler ThreatLabz 2024 sur la sécurité de l'IA

La révolution de l'IA est en cours. Découvrez les principales tendances, risques et bonnes pratiques de l'IA en entreprise, avec des perspectives sur les menaces liées à l'IA et les stratégies clés pour s'en protéger.

Sommaire

Note de synthèse

L'IA (Intelligence Artificielle) est bien plus qu'une innovation pionnière — c'est une affaire de tous les jours. Alors que les outils d'IA générative tels que ChatGPT transforment l'entreprise à grande et petite échelle, l'IA est profondément ancrée dans la vie de l'entreprise. Cependant, les questions autour de l'adoption sécurisée des outils IA, en se protégeant contre les menaces associées à cet IA, ne sont pas tranchées.

Les entreprises adoptent rapidement les outils d'IA et d'AA (Apprentissage Automatique ou Machine Learning) dans des métiers tels que l'ingénierie, l'informatique, la finance, la relation client, etc. Elles doivent toutefois gérer les nombreux risques liés aux outils d'IA afin d'en tirer le meilleur parti. En effet, pour libérer le potentiel de transformation de l'IA, les entreprises doivent déployer des fonctionnalités pour protéger leurs données, prévenir la fuite d'informations sensibles, maîtriser la prolifération de l'IA fantôme (shadow IA) et garantir la qualité des données utilisées par l'IA.

Ces risques liés à l'IA sont bidirectionnels : **en dehors des murs de l'entreprise, l'IA est devenue un moteur de cybermenaces**. En effet, les outils d'IA permettent aux cybercriminels et aux acteurs malveillants parrainés par des États-nations de lancer plus rapidement des attaques sophistiquées, à plus grande échelle. Malgré cela, l'IA semble prometteuse en tant que pièce essentielle de l'arsenal de cyberdéfense, à l'heure où les entreprises sont aux prises avec un univers de menaces particulièrement dynamique.

Le rapport ThreatLabz 2024 sur la sécurité de l'IA livre des informations clés concernant ces défis et opportunités critiques liés à l'IA. S'appuyant sur plus de 18 milliards de transactions effectuées entre avril 2023 et janvier 2024 sur Zscaler Zero Trust Exchange™, ThreatLabz a analysé la manière dont les entreprises utilisent désormais les outils d'IA et d'AA. Ces informations révèlent des tendances clés dans les secteurs d'activité et les zones géographiques sur la façon dont les entreprises s'adaptent à la versatilité de l'IA et sécurisent leurs outils d'IA.

Vous trouverez tout au long de ce rapport, des informations sur des sujets clés liés à l'IA : risques business, scénarios de menaces IA, tactiques des assaillants, perspectives réglementaires et prévisions autour de l'IA pour 2024 et au-delà.

Tout aussi important, ce rapport propose des bonnes pratiques sur deux fronts : comment les entreprises peuvent se transformer en toute sécurité grâce à l'IA générative tout en protégeant leurs données critiques, et comment les outils optimisés par IA s'efforcent d'offrir plusieurs niveaux de sécurité Zero Trust pour faire face au nouveau paysage des menaces basées sur l'IA.

Principales conclusions

L'utilisation des outils d'IA/AA s'est envolée de 594,82 %, passant de 521 millions de transactions optimisées par IA/AA en avril 2023 à 3,1 milliards par mois en janvier 2024.

Les entreprises neutralisent 18,5 % de toutes les transactions d'IA/AA, soit un bond de 577 % des transactions bloquées en neuf mois, reflétant ainsi des préoccupations plus marquées sur la sécurité des données IA et la réticence des entreprises à établir des politiques pour l'IA.

Le secteur de la production industrielle est celui qui génère le plus de trafic IA, soit 20,9 % de toutes les transactions IA/AA dans le cloud de Zscaler, suivi par la Finance et les assurances (19,9 %) et les services (16,8 %).

L'utilisation de ChatGPT continue de progresser, avec **une croissance de 634,1 %,** même s'il **s'agit également de l'application d'IA la plus bloquée** par les entreprises, selon les informations du cloud de Zscaler.

Les applications IA les plus utilisées en termes de volume de transactions sont **ChatGPT, Drift, OpenAI*, Writer et LivePerson. Les trois principales applications bloquées** en termes de volume de transactions sont **ChatGPT, OpenAI et Fraud.net.**

Les cinq pays qui génèrent le plus de transactions IA et AA sont les États-Unis, l'Inde, le Royaume-Uni, l'Australie et le Japon.

Les entreprises envoient d'importants volumes de données aux outils IA, avec 569 To qui ont été échangés entre les applications IA/AA de septembre 2023 à janvier 2024.

L'IA procure aux acteurs malveillants de nouvelles opportunités, notamment sur les périmètres suivants : campagnes de phishing basées sur l'IA, attaques utilisant des deepfakes et l'ingénierie sociale, ransomwares polymorphes, identification de la surface d'attaque d'une entreprise, génération automatisée d'exploits, et bien plus encore.

REMARQUE : Zscaler Zero Trust Exchange assure un suivi des transactions ChatGPT indépendamment des autres transactions d'OpenAI en général.

Principales tendances d'utilisation du GenAI et de l'AA

La révolution de l'IA en entreprise est loin d'avoir atteint son apogée. Les transactions IA d'entreprise ont augmenté de près de 600 % et ne montrent aucun signe de ralentissement.
Pourtant, le nombre de transactions bloquées vers les applications IA a également progressé — de 577 %.

Les transactions IA continuent de s'accélérer

D'avril 2023 à janvier 2024, les transactions IA et AA des entreprises ont bondi de près de 600 %, passant à plus de 3 milliards de transactions mensuelles sur Zero Trust Exchange en janvier. Cela souligne le fait que, malgré le nombre croissant d'incidents de sécurité et de risques liés aux données associés à l'adoption de l'IA en entreprise, son potentiel de transformation est bien trop important pour être ignoré. Notons que si les transactions IA ont connu une brève accalmie pendant les vacances de décembre, les transactions se sont poursuivies à un rythme encore plus soutenu au début de l'année 2024.

Même si les applications basées sur l'IA prolifèrent, la majorité des transactions IA relèvent d'un ensemble relativement restreint d'outils IA leaders du marché. Dans l'ensemble, ChatGPT représente plus de la moitié de toutes les transactions IA et AA, tandis que l'application d'OpenAI arrive en troisième position, avec 7,82 % de toutes les transactions. Parallèlement, Drift, le populaire chatbot optimisé par IA, a généré près d'un cinquième du trafic IA des entreprises (les chatbots LivePerson et BoldChat Enterprise ont également fait leur entrée dans le top des applications, en 5e et 6e position respectivement). Par ailleurs, Writer demeure un des outils d'IA générative privilégiés pour la création de contenu écrit en entreprise, tels que des supports marketing. Enfin, Otter, un outil IA de transcription souvent utilisé lors d'appels vidéo, génère une part importante du trafic IA.

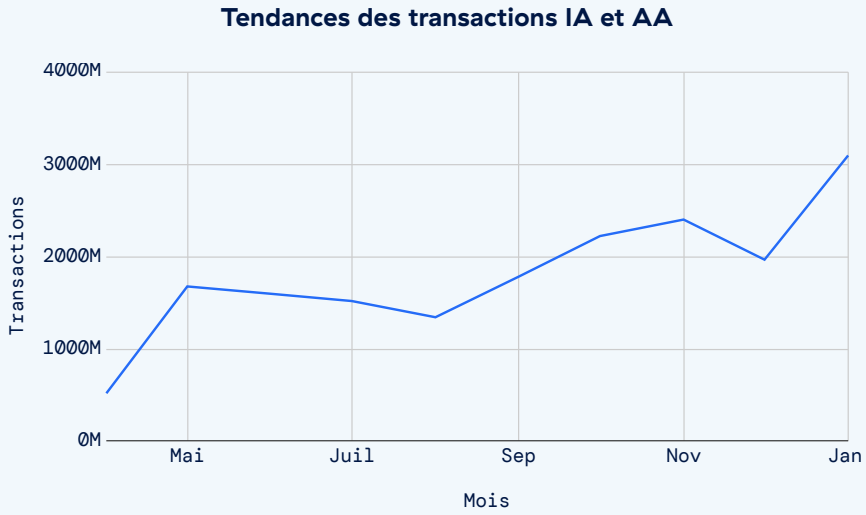


SCHÉMA 1 Transactions IA d'avril 2023 à janvier 2024

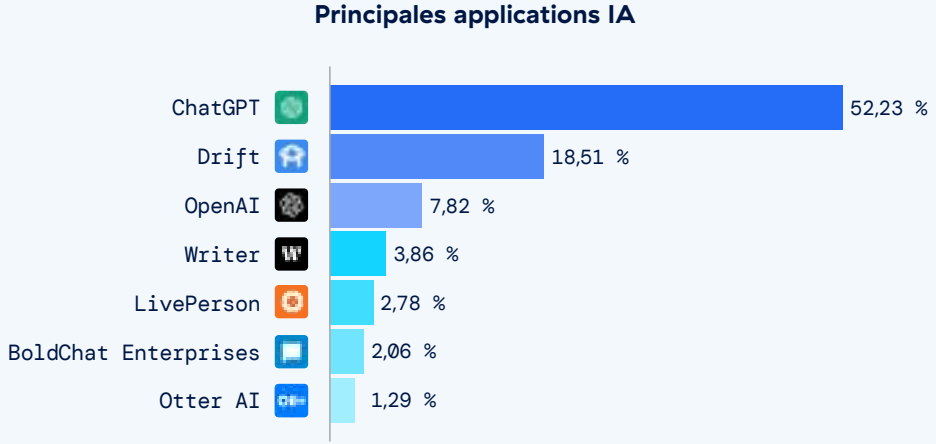


SCHÉMA 2 Principales applications IA par volume de transactions

Données transférées par le trafic IA/AA [sept 2023–janv 2024]

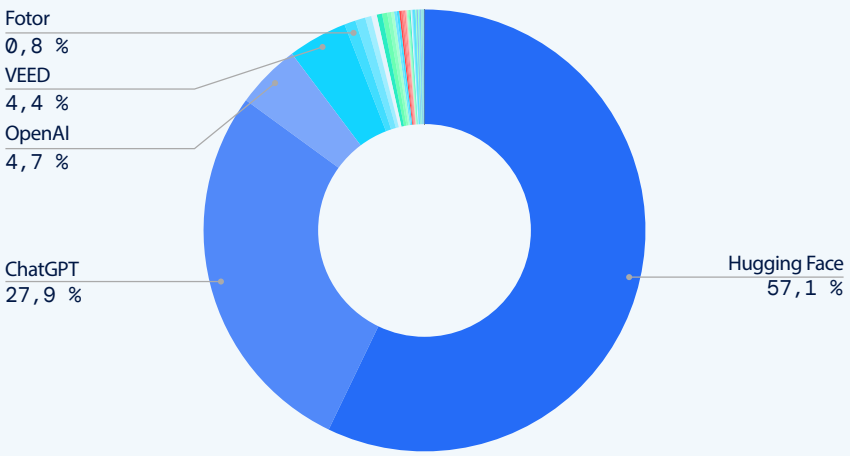


SCHÉMA 3 Principales applications IA/AA en pourcentage des données totales transférées

Tendances des transactions IA bloquées [avril 2023 – janv 2024]

SCHÉMA 4 Nombre de transactions IA/AA bloquées au fil du temps

Parallèlement, les volumes de données que les entreprises envoient et reçoivent des outils IA viennent nuancer ces tendances. Hugging Face, la plateforme open source de développement IA souvent décrite comme « le GitHub de l'IA », compte pour près de 60 % des données d'entreprise transférées par les outils IA. Hugging Face permet aux utilisateurs d'héberger et de former des modèles IA : il est donc logique qu'il capture d'importants volumes de données provenant d'utilisateurs professionnels.

Alors que ChatGPT et OpenAI figurent sans surprise sur cette liste, Veed, un éditeur de technologies IA de vidéo souvent utilisées pour ajouter des sous-titres, des images et d'autres textes aux vidéos, et Fotor, un outil IA utilisé pour générer des images, entre autres utilisations, offrent des perspectives particulièrement intéressantes. Étant donné que les vidéos et les images impliquent des fichiers de grande taille par rapport à d'autres types de requêtes, il n'est guère surprenant de voir figurer ces deux applications.

Les entreprises neutralisent plus de transactions IA que jamais

Même si l'adoption de l'IA par les entreprises continue de progresser, ces dernières bloquent de plus en plus les transactions IA et dA en raison de préoccupations liées aux données et à la sécurité. Les entreprises neutralisent aujourd'hui 18,5 % de toutes les transactions IA, ce qui constitue un bond de 577 % entre avril et janvier, pour un total de plus de 2,6 milliards de transactions bloquées.

Certains des outils IA les plus populaires figurent également parmi ceux les plus bloqués. ChatGPT a en effet la particularité d'être à la fois l'application IA la plus utilisée et la plus bloquée. Ainsi, malgré — ou grâce à — la popularité de ces outils, les entreprises s'investissent pour sécuriser leur utilisation contre la perte de données et les problématiques de confidentialité. Autre tendance notable : bing.com, qui offre la fonctionnalité IA Copilot, est bloqué d'avril à janvier. Notons que bing.com représente 25,02 % de toutes les transactions de domaine IA et AA bloquées.

Certains des outils IA les plus populaires figurent également parmi celles les plus bloqués. ChatGPT a en effet la particularité d'être à la fois l'application IA la plus utilisée et la plus bloquée. Ainsi, malgré — ou grâce à — la popularité de ces outils, les entreprises s'investissent pour sécuriser leur utilisation contre la perte de données et les problématiques de confidentialité. Notons également que bing.com est bloqué plus que tout autre domaine, avec un total de 835 811 952 transactions neutralisées entre avril et janvier. Bing.com représente 25,02 % de toutes les transactions de domaine IA et AA bloquées.

OUTILS IA LES PLUS BLOQUÉS	DOMAINES IA LES PLUS BLOQUÉS
01 ChatGPT	01 Bing.com
02 OpenAI	02 Divo.ai
03 Fraud.net	03 Drift.com
04 Forethought	04 Quillbot.com
05 Hugging Face	05 Compose.ai
06 ChatBot	06 Openai.com
07 Aivo	07 Qortex.ai
08 Neeva	08 Sider.ai
09 infeedo.ai	09 Tabnine.com
10 Jasper	10 securiti.ai

Schéma 5 Principaux domaines et applications IA bloqués, par volume de transactions

Répartition de l'IA par secteur d'activité

Les secteurs d'activités des entreprises affichent des différences notables tant dans l'adoption globale des outils IA que dans la proportion de transactions IA qu'ils bloquent. Le secteur de la production industrielle est le leader incontesté, générant plus de 20 % des transactions IA et AA sur Zero Trust Exchange. Les secteurs de la finance et des assurances, des technologies et des services suivent toutefois de près. Ensemble, ces quatre secteurs devancent les autres dans l'adoption de l'IA.

Part des transactions IA par secteur d'activité

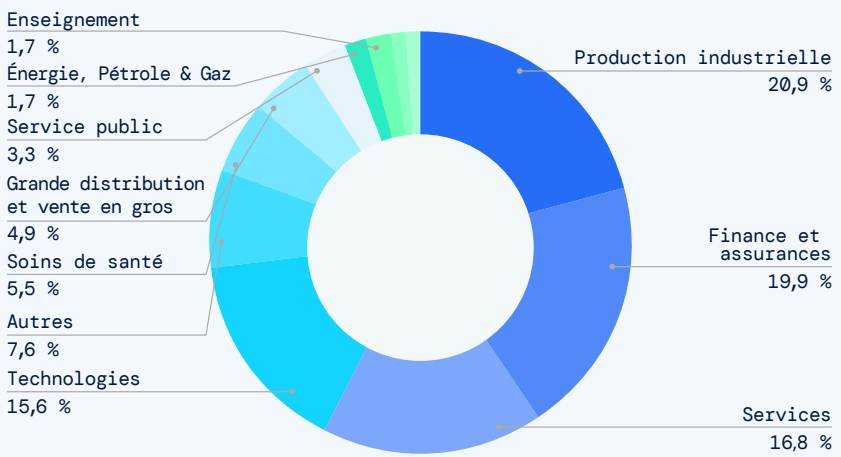


SCHÉMA 6 Secteurs d'activité à l'origine des volumes les plus importants de transactions IA

Tendances des transactions IA par secteur d'activité

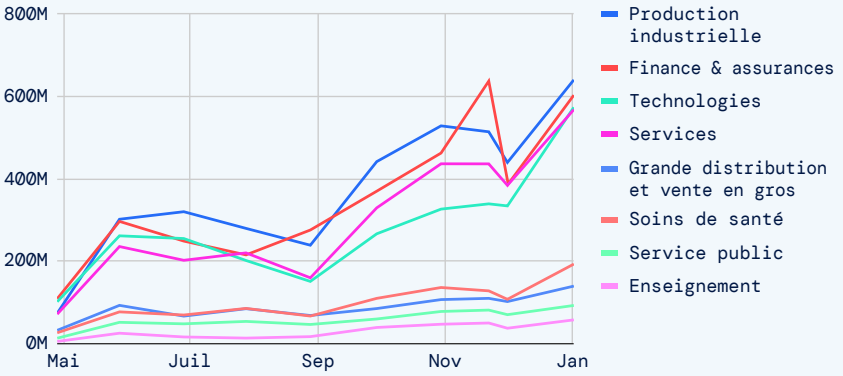


SCHÉMA 7 Tendances des transactions IA/AA parmi les secteurs à plus fort volume, avril 2023–janvier 2024

Sécurisation des transactions IA/AA

Parallèlement à la hausse soutenue des transactions IA, celles-ci sont davantage bloquées. Certains secteurs divergent de leurs tendances globales en matière d'adoption, reflétant des priorités et des niveaux de maturité différents en termes de sécurisation des outils IA. Le secteur de la finance et des assurances, par exemple, est celui qui bloque la plus grande proportion de transactions IA : 37,2 % contre une moyenne mondiale de 18,5 %. Ceci est probablement dû en grande partie à l'environnement réglementaire et de conformité strict qui s'applique au secteur, ainsi qu'aux données financières et personnelles très sensibles des utilisateurs que ces entreprises traitent.

De son côté, le secteur de la production industrielle bloque 15,7 % des transactions IA, bien qu'il génère une proportion significative des transactions IA dans leur globalité. Le secteur des technologies, l'un des premiers et des plus enthousiastes à avoir adopté l'IA, a emprunté une voie intermédiaire, bloquant un taux de 19,4 % des transactions IA (un chiffre supérieur à la moyenne) alors qu'il s'efforce d'étendre son adoption de l'IA. Étonnamment, le secteur de la santé bloque 17,2 % des transactions IA, ce qui est en dessous de la moyenne. Pourtant, ces acteurs traitent des volumes importants de données de santé et d'informations personnelles identifiables (PII – personally identifiable information). Cette tendance reflète probablement un certain retard dans les efforts des acteurs de la santé pour protéger les données sensibles liées aux outils IA. Leurs équipes de sécurité tentent néanmoins de rattraper ce retard sur l'innovation IA. Dans l'ensemble, les transactions IA dans le domaine de la santé restent relativement faibles.

SCHÉMA 8 Principaux secteurs en pourcentage de transactions IA bloquées

Pourcentage de transactions IA bloquées par secteur d'activité

Secteur d'activité	% des transactions IA bloquées
Finance & assurances	37,16
Production industrielle	15,65
Services	13,17
Technologies	19,36
Soins de santé	17,23
Grande distribution et vente en gros	10,52
Autres	8,93
Énergie, Pétrole & Gaz	14,24
Service public	6,75
Transport	7,90
Enseignement	2,98
Communication	4,29
Construction	4,12
Matériaux de base, chimie & mines	2,92
Divertissement	1,33
Agroalimentaire & tabac	3,66
Hôtels, restaurants & loisirs	3,16
Organisations religieuses	6,06
Agriculture et sylviculture	0,18
Moyenne sur tous les secteurs	18,53

Soins de santé et IA

Se classant au sixième rang des plus grands utilisateurs d'IA/AA, le secteur de la santé bloque 17,23 % de toutes les transactions IA/AA.

PRINCIPALES APPLICATIONS IA DANS LES SOINS DE SANTÉ :

01 ChatGPT
02 Drift
03 OpenAI
04 Writer
05 Intercom
06 Zineone
07 Securiti
08 Pypestream
09 Hybrid
10 VEED

Signes de progression de l'IA dans le secteur de la santé

Alors que le secteur de la santé est généralement prudent lorsqu'il s'agit de mettre en pratique des innovations comme l'IA, comme en témoigne sa contribution actuelle de seulement 5 % au trafic IA/AA dans le cloud de Zscaler, ce n'est qu'une question de temps avant que l'IA n'ait un impact plus important sur les opérations de santé, les soins aux patients, ainsi que la recherche et l'innovation médicales.[1]

En effet, l'IA est censée permettre non seulement de gagner du temps, mais aussi de sauver des vies. D'ores et déjà, les technologies optimisées par IA améliorent les diagnostics et les soins aux patients. En analysant les images médicales avec une précision remarquable, l'IA permet aux radiologues de détecter plus rapidement les anomalies et d'accélérer leurs décisions en matière de traitement.[2]

Les avantages potentiels sont immenses. Les algorithmes IA peuvent utiliser les données des patients pour personnaliser les protocoles de traitement et accélérer la découverte de nouvelles molécules, grâce à une analyse plus efficiente des données biologiques. Les tâches administratives peuvent également être automatisées grâce à l'IA générative, allégeant ainsi la charge d'équipes de soins souvent en sous-effectif. Ces progrès soulignent la capacité de l'IA à transformer la prestation des soins de santé.

Principaux risques liés aux soins de santé :
les établissements de santé doivent reconnaître les risques et les défis potentiels associés à l'IA, notamment les préoccupations de confidentialité et de sécurité des données, en particulier pour les informations personnelles identifiables (PII). Ils doivent ainsi veiller à ce que les algorithmes IA et leurs résultats soient fiables et impartiaux lorsqu'ils contribuent aux soins administrés aux patients.

1. Statista, Future Use Cases for AI in Healthcare, Septembre 2023.
2. The Hill, AI already plays a vital role in medical imaging and is effectively regulated, 23 février 2024.

Finance & IA

En deuxième position en termes d'utilisation totale de l'IA/AA, le secteur de la finance bloque 37,16 % de tout le trafic IA/AA.

PRINCIPALES APPLICATIONS IA DANS LA FINANCE :

01	ChatGPT	06	Writer
02	Drift	07	Hugging Face
03	OpenAI	08	Otter Ai
04	BoldChat Enterprise	09	Securiti
05	LivePerson	10	Intercom

Les institutions financières misent sur l'IA

Les sociétés de services financiers ont été parmi les premiers à adopter l'IA, le secteur représentant près d'un quart du trafic IA/AA dans le cloud de Zscaler. De plus, McKinsey prévoit un revenu annuel potentiel de 200 à 340 milliards de dollars pour les initiatives d'IA générative dans le secteur bancaire, en grande partie grâce à des gains de productivité.[3] L'IA représente littéralement une mine d'opportunités pour les banques et les services financiers.

Même si les chatbots et les assistants virtuels optimisés par IA ne sont pas nouveaux dans la finance (« Erica » de Bank of America a été lancé en 2018), les améliorations apportées par l'IA générative permettent à ces outils de service client de personnaliser leurs interactions. D'autres fonctionnalités IA, telles que la modélisation prédictive et l'analyse des données, sont en passe d'apporter des avantages majeurs en termes de productivité aux opérations financières, transformant ainsi les processus de détection des fraudes, d'évaluation des risques, etc.

Principaux risques pour la finance et les assurances : l'intégration de l'IA dans les services et produits financiers entraîne des problématiques de sécurité et de réglementation en matière de confidentialité et d'exactitude des données. La part importante du trafic AI/AA bloqué (37 %) signalé par ThreatLabz reflète ce point de vue. Répondre à ces préoccupations nécessitera une approche et une planification pertinentes afin de pérenniser la confiance et l'intégrité dans le secteur de la banque, des services financiers et des assurances.

3. McKinsey, Capturing the full value of generative AI in banking, 5 décembre 2023.

Administrations et IA

Bien que présent dans le top 10 de l'utilisation de l'IA/AA, le service public ne bloque que 6,75 % des transactions d'IA/AA.

PRINCIPALES APPLICATIONS IA* DANS LE SERVICE PUBLIC :

01 ChatGPT
02 Drift
03 OpenAI
04 Zineone

*Applications IA avec au moins 1 million de transactions

Les acteurs du service public dans le monde entier s'adaptent aux pratiques et règles de l'IA

Deux thématiques concernant l'IA font débat au sein du service public : l'une porte sur la mise en œuvre des technologies IA et l'autre sur la mise en place d'une gouvernance permettant de les gérer en toute sécurité. Les avantages de l'IA pour les instances gouvernementales et entités du secteur public sont considérables, notamment lorsque les chatbots et les assistants virtuels peuvent donner aux citoyens un accès plus rapide aux informations et services essentiels dans des secteurs tels que les transports publics et l'enseignement. L'analyse des données optimisée par IA peut aider à relever les défis sociétaux grâce à des processus décisionnels fondés sur les données, ce qui permet d'élaborer des politiques et d'allouer des ressources de manière plus efficace.

Des progrès notables sont déjà en cours. Par exemple, le ministère américain de la Justice a nommé son premier Chief AI Officer (Directeur général en charge de l'IA), confirmant ainsi son engagement à utiliser les technologies d'IA. Les données de ThreatLabz indiquent que les clients gouvernementaux utilisent de plus en plus les plateformes IA/AA telles que ChatGPT et Drift.

Principaux risques pour le service public :
malgré ces tendances, les principales préoccupations concernant les risques liés à l'IA et à la confidentialité des données soulignent un besoin continu pour des cadres réglementaires et de gouvernance dans les instances du service public. De manière générale, les décideurs politiques du monde entier ont pris **des mesures significatives pour réglementer l'IA** au cours de l'année écoulée, ce qui témoigne d'un effort collectif visant à favoriser un développement et un déploiement responsables des technologies IA/AA.

Production industrielle et IA

Premier secteur d'activité en matière d'IA/AA, la production industrielle bloque 15,65 % de toutes les applications IA/AA.

PRINCIPALES APPLICATIONS :

01 ChatGPT
02 Drift
03 OpenAI
04 Writer
05 Securiti
06 Google Search
07 Zineone
08 Pypestream
09 Hugging Face
10 Fotor

La production industrielle capitalise sur la dynamique de l'IA

Sans surprise, la part la plus importante du trafic IA/AA (18,2 %) dans notre étude provient des industriels. L'adoption de l'IA dans le secteur de la production constitue la pierre angulaire de l'Industrie 4.0, alias la quatrième révolution industrielle, une ère marquée par la convergence des technologies numériques et des processus industriels.

De la détection préventive des pannes d'équipement par l'analyse de volumes importants de données provenant de machines et de capteurs à la gestion optimale de la chaîne collaborative, des stocks et des opérations logistiques, l'IA s'avère déterminante pour les industriels. De plus, les systèmes de robotique et d'automatisation optimisés par IA peuvent doper la productivité de l'outil industriel. Ils peuvent exécuter des tâches à une vitesse et une précision bien supérieures à celles des humains, tout en réduisant les coûts et les erreurs.

Principaux risques liés à l'IA dans le secteur de la production industrielle : En ce qui concerne les 16 % de trafic bloqué provenant d'applications IA/AA par les clients du secteur de la production, certains industriels abordent l'IA/AA générative avec prudence. Ce constat peut découler de préoccupations sur la sécurité des données, ainsi que de la nécessité de contrôler et d'approuver de manière sélective un ensemble plus restreint d'applications IA tout en bloquant celles qui présentent un plus grand risque.

Enseignement et IA

En 11e position en termes d'utilisation de l'IA/AA, le secteur de l'enseignement bloque 2,98 % de l'ensemble du trafic IA/AA.

PRINCIPALES APPLICATIONS :

01 ChatGPT
02 Character.AI
03 Pixlr
04 Forethought
05 Deepai
06 Drift
07 OpenAI

Les acteurs de ce secteur adoptent l'IA comme outil d'apprentissage

Bien que ce secteur ne soit pas l'un des principaux générateurs de trafic d'IA, il bloque un pourcentage relativement faible (2,98 %) de transactions IA et AA , soit environ 9 millions sur un total de plus de 309 millions de transactions. Il est clair que, malgré les idées reçues selon lesquelles les établissements d'enseignement bloquent généralement les applications d'IA telles que ChatGPT parmi les étudiants, le secteur a principalement adopté les applications d'IA en tant qu'outils d'apprentissage. Ainsi, cinq des applications d'IA les plus populaires dans l'enseignement (ChatGPT, Character.AI, Pixlr et OpenAI) sont explicitement ou fréquemment utilisées à des fins de création (rédaction ou génération d'images), tandis que Forethought peut être utilisé comme chatbot pédagogique.

Pour nuancer ce tableau, il se peut également que de nombreux enseignants bloquent des outils comme ChatGPT dans le cadre de leur classe. Les acteurs de l'enseignement ont sans doute pris du retard par rapport à d'autres secteurs dans la mise en œuvre de solutions technologiques telles que le filtrage DNS qui permettent de bloquer les outils IA et AA de manière plus spécifique.

Principaux risques liés à l'IA dans l'enseignement : dans ce secteur, les préoccupations de confidentialité des données vont probablement se renforcer à mesure que le secteur continue d'adopter les outils d'IA, en particulier en ce qui concerne la protection des données personnelles des étudiants. Selon toute vraisemblance, le secteur adoptera de plus en plus de moyens technologiques pour bloquer de manière sélective les applications d'IA, tout en prévoyant des mesures plus strictes de protection des données personnelles.

Utilisation de ChatGPT : les tendances

L'adoption de ChatGPT a explosé. Depuis avril 2023, les transactions mondiales autour de ChatGPT ont bondi de plus de 634 %, surperformant la progression globale de 595 % des transactions d'IA. À partir de ces résultats et de la perception générale d'OpenAI comme marque leader de l'IA, il est clair que ChatGPT est un outil privilégié en matière d'IA générative. Selon toute vraisemblance, l'adoption des produits OpenAI se renforcera, en partie grâce à la sortie attendue de nouvelles versions de ChatGPT et du produit d'IA générative Sora dédié à la conversion texte-vidéo.

L'utilisation de ChatGPT dans le secteur s'inscrit dans les schémas d'adoption des outils IA en général. Dans ce cas, le secteur de la production industrielle est le leader incontesté, suivi par celui de la finance et de l'assurance. Le secteur technologique est légèrement à la traîne à la quatrième place, avec 10,7 % des transactions ChatGPT par rapport à la troisième place et ses 14,6 % au total. Cela est probablement dû en partie à l'innovation rapide dans ce secteur technologique, qui pourrait faire penser que les entreprises technologiques sont plus disposées à adopter une plus grande diversité d'outils d'IA générative.

Transactions par secteur d'activité

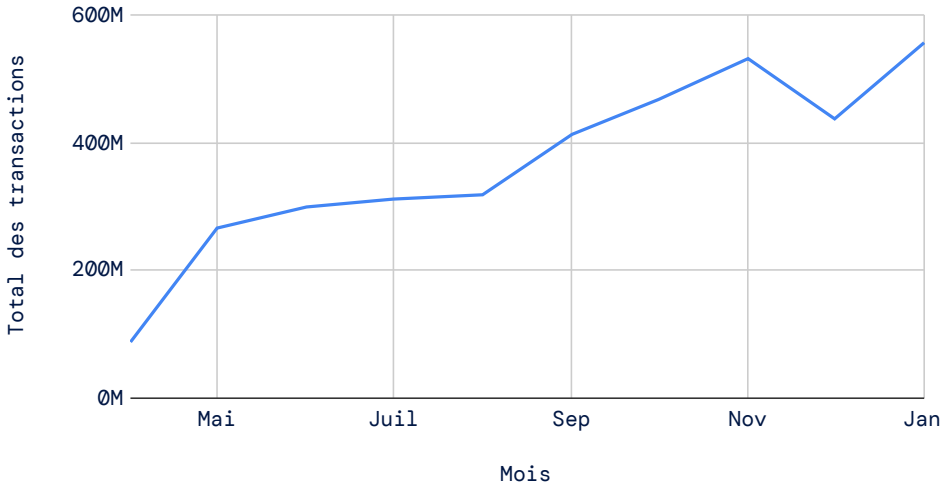


SCHÉMA 9 Transactions liées à ChatGPT d'avril 2023 à janvier 2024

Tendances des transactions IA par secteur d'activité

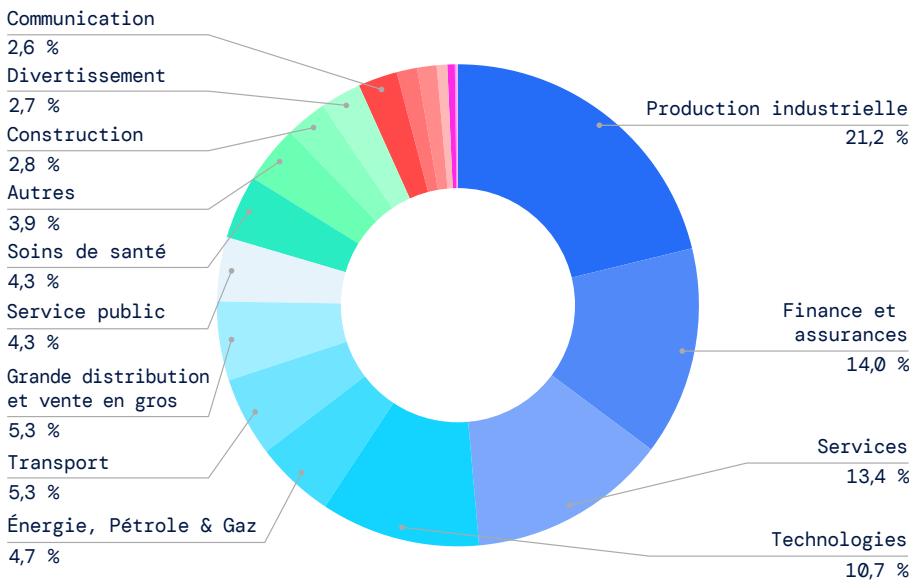


SCHÉMA 10 Industries à l'origine des proportions les plus importantes de transactions de ChatGPT

Utilisation de l'IA par pays

Les tendances en matière d'adoption de l'IA diffèrent considérablement dans le monde, influencées par les exigences réglementaires, l'infrastructure technologique, les considérations culturelles et d'autres facteurs. Voici un aperçu des principaux pays à l'origine des transactions IA et AA dans le cloud de Zscaler.

Comme on pouvait s'y attendre, les États-Unis se taillent la part du lion dans les transactions IA. L'Inde, quant à elle, s'est imposée comme l'un des principaux générateurs de trafic IA, grâce à l'engagement accéléré du pays en faveur de l'innovation technologique. Le gouvernement indien fournit également un exemple utile de la rapidité avec laquelle la réglementation de l'IA évolue, avec ses efforts récents pour adopter — puis abandonner — un plan qui exigeait une validation réglementaire des modèles d'IA avant leur commercialisation.[4]

Transactions par pays

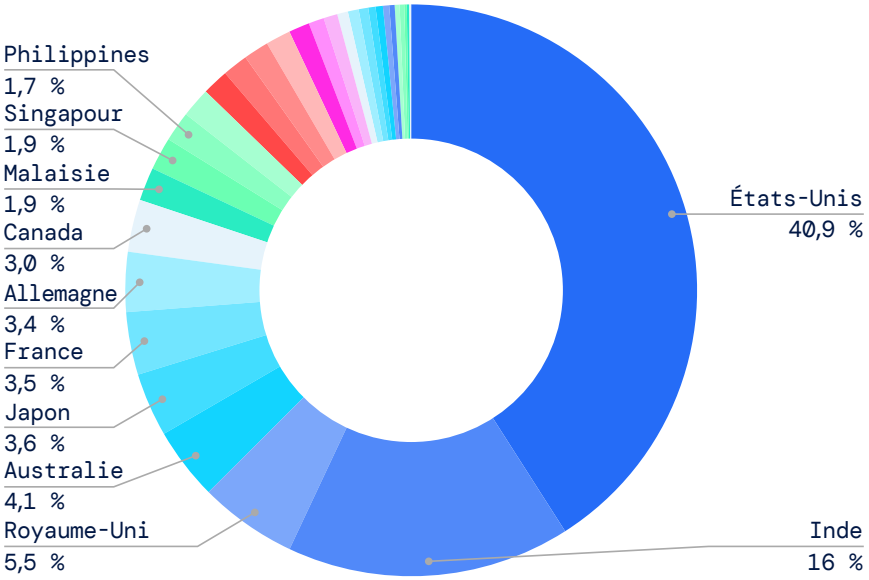


SCHÉMA 11 Pays les plus dynamiques en termes de transactions IA

4. TechCrunch, India reverses AI stance, requires government approval for model launches, 3 mars 2024.

Répartition par région : EMEA

Une étude de la région Europe, Moyen-Orient et Afrique (EMEA) révèle de nettes divergences dans les taux de transactions IA et AA entre les pays. Alors que le Royaume-Uni ne représente que 5,5 % des transactions IA dans le monde, il représente plus de 20 % du trafic IA dans la région EMEA, ce qui en fait le leader incontesté. La France et l'Allemagne se classent sans surprise aux deuxième et troisième rangs des générateurs de trafic IA dans la région EMEA. Notons que l'innovation technologique rapide aux Émirats arabes unis fait de ce pays un des principaux utilisateurs de l'IA dans la région.

Pays	Transactions	% de la région
Royaume-Uni	763413289	20,47 %
France	504185470	13,53 %
Allemagne	471700683	12,66 %
Émirats arabes unis	238557680	6,40 %
Pays-Bas	222783817	5,98 %
Espagne	198623739	5,30 %
Suisse	129059097	3,46 %
Italie	97544412	2,62 %

SCHÉMA 12 Pays de la zone EMEA par volume de transactions

Répartition par pays de la zone EMEA

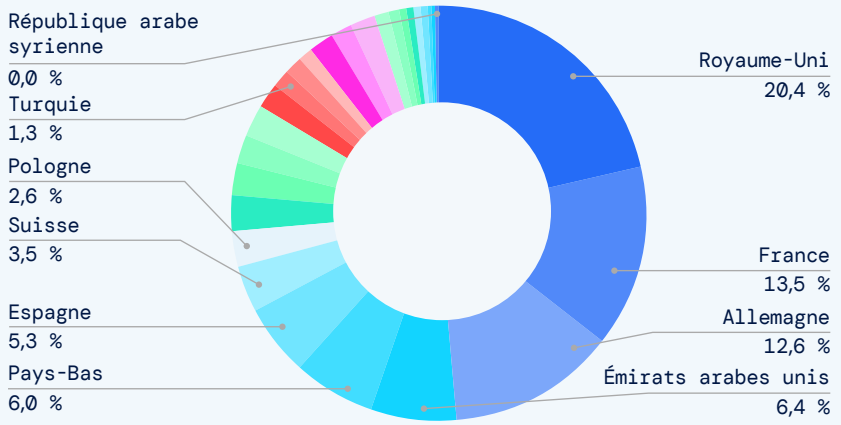


SCHÉMA 13 Pays de la zone EMEA en pourcentage du total des transactions IA sur la région

Transactions (en millions) par mois

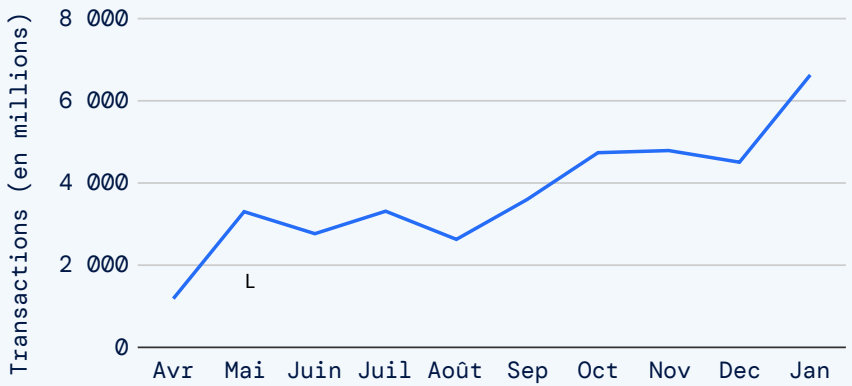


SCHÉMA 14 Croissance des transactions d'IA sur la région EMEA dans le temps

Répartition par pays de la zone APAC

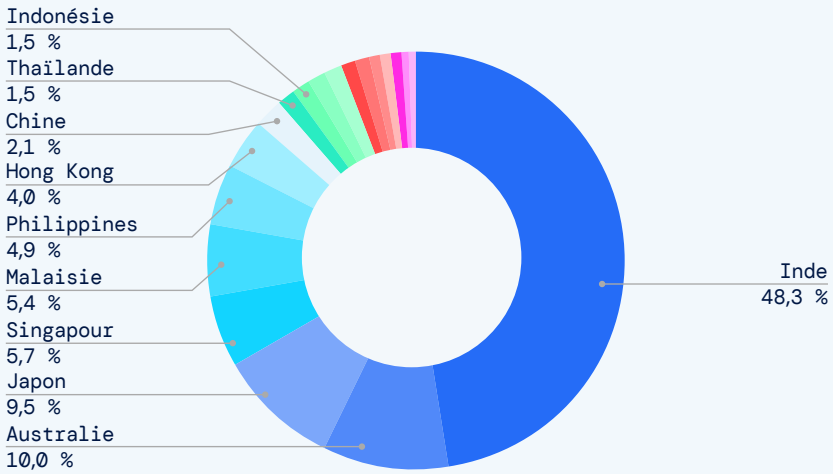


SCHÉMA 16 Pays de la zone APAC en pourcentage du total des transactions IA sur la région

Transactions (en millions) par mois

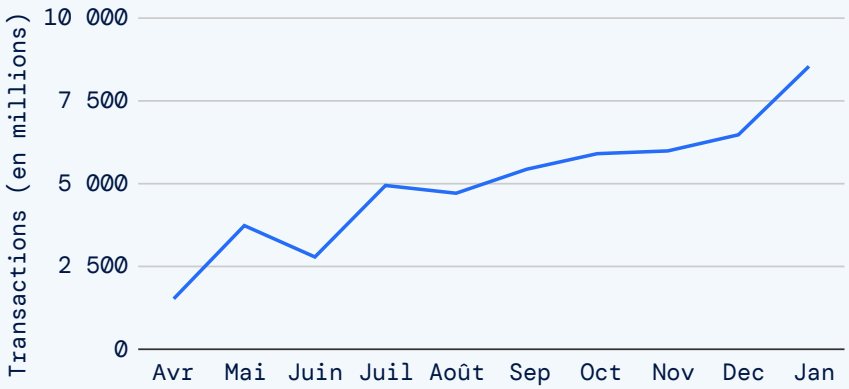


SCHÉMA 17 Croissance des transactions IA sur la région APAC dans le temps

Répartition par région : APAC

L'étude de ThreatLabz sur la région Asie-Pacifique (APAC) révèle certaines tendances claires et notables en matière d'adoption de l'IA. Bien que la région représente beaucoup moins de pays, TheatLabz a observé près de 1,3 milliard (135 %) de transactions IA supplémentaires dans la région APAC par rapport à la région EMEA. Cette croissance est presque à elle seule tirée par l'Inde, qui génère près de la moitié de toutes les transactions IA et AA sur la région APAC.

Pays	Transactions	% de la région
Inde	2414319490	48,30 %
Australie	501562395	10,01 %
Japon	476425423	9,52 %
Singapour	284891384	5,70 %
Malaisie	268043263	5,36 %
Philippines	243754578	4,87 %
Hong Kong	202119814	4,04 %
Chine	104545655	2,09 %

SCHÉMA 15 Pays de la zone APAC par volume de transactions

Risques liés à l'IA en entreprise et scénarios de menaces

Pour les entreprises, les risques et menaces liés à l'IA se répartissent en deux grandes catégories : les risques liés à la protection des données et à la sécurité suite à l'activation des outils IA d'entreprise ; ainsi que les risques de nouvelles cybermenaces induites par les outils d'IA générative et l'automatisation.

Risques liés à l'IA d'entreprise

1 Protection des éléments de propriété intellectuelle et des informations non publiques

Les outils d'IA générative peuvent provoquer la fuite involontaire de données sensibles et confidentielles. La divulgation de données sensibles figure au sixième rang du Top dix des menaces OWASP (Open Worldwide Application Security Project) pour les applications IA.[5] L'année dernière a été marquée par de nombreux cas de fuites accidentelles de données ou de piratages de données d'entraînement de l'IA, notamment dus à des erreurs de configuration du cloud de la part de fournisseurs majeurs d'outils IA, certains d'entre eux exposant des téraoctets de données privées de leurs clients.

Dans un cas, des chercheurs ont exposé des milliers de données confidentielles GitHub provenant de l'extension IA GitHub Copilot en exploitant une vulnérabilité dite "d'injection rapide" (en utilisant des requêtes IA conçues pour manipuler l'IA afin de divulguer des données d'entraînement), ce qui constitue d'ailleurs le risque numéro un du Top 10 OWASP.[6]

Un risque connexe est **la menace d'inversion de modèle,** dans le cadre de laquelle les assaillants utilisent les résultats d'un LLM (grand modèle de langage) associés à la connaissance de la structure du modèle pour interagir avec les données d'entraînement, et éventuellement les extraire. Bien entendu, le risque que les acteurs de l'IA soient eux-mêmes victimes d'incidents existe également. Dans certains cas, les informations d'identification des collaborateurs d'une entreprise de l'IA ont directement conduit à des fuites de données.

Il est par ailleurs possible que des adversaires lancent **des attaques de malwares secondaires,** en utilisant des outils de vol d'informations tels que Redline Stealer ou LummaC2, pour détourner les identifiants de connexion des collaborateurs et accéder à leurs comptes IA. Il a été récemment révélé qu'environ 225 000 identifiants d'utilisateur de ChatGPT sont mis en vente sur le dark web, suite à ce type d'attaque.[7] Même si la confidentialité et la sécurité des données demeurent des priorités absolues chez les fournisseurs d'outils IA, ces risques restent présents et s'étendent également aux petits acteurs de l'IA, aux fournisseurs SaaS qui ont activé des fonctionnalités d'IA, etc.

Enfin, il y a **les risques découlant des utilisateurs de l'IA en entreprise.** Un utilisateur peut, à son insu et de diverses manières, exposer une information de propriété intellectuelle précieuse ou des informations non publiques dans les ensembles de données utilisés pour former les modèles LLM. Par exemple, un développeur souhaitant optimiser un code source ou un membre de l'équipe commerciale recherchant des tendances de ventes basées sur des données internes pourrait divulguer involontairement des informations protégées en dehors de l'entreprise. Il est essentiel que les entreprises soient conscientes de ce risque et déploient des mesures robustes de protection des données, notamment de prévention des pertes de données (DLP), pour éviter de telles fuites.

RISQUES LIÉS AU CONTRÔLE D'ACCÈS ET À LA SEGMENTATION

Les contrôles d'accès, tels que le contrôle d'accès basé sur les rôles (RBAC), peuvent être mal configurés ou utilisés de manière abusive pour les applications IA. Ceci peut mener à des circonstances dans lesquelles, par exemple, un chatbot IA génère les mêmes réponses pour un directeur général que pour tout autre utilisateur de l'entreprise, ce qui pose des risques particuliers lorsque les chatbots sont formés sur des données historiques provenant de données associées à cet utilisateur. Ces données pourraient être utilisées pour déduire des informations concernant les requêtes que les dirigeants ont envoyées à l'aide de chatbots d'IA. Dans ce cas, les entreprises doivent veiller à configurer de manière appropriée les contrôles d'accès aux applications d'IA, assurant à la fois la sécurité des données et la segmentation des accès en fonction des autorisations et des rôles de l'utilisateur.

5. OWASP, OWASP Top 10 For LLM Applications, Version 1.1, 16 octobre 2023.
6. The Hacker News, Three Tips to Protect Your Secrets from AI Accidents, 26 février 2024.
7. The Hacker News, Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets, 5 mars 2024.

2 Risques liés à la confidentialité des données et à la sécurité des applications IA

Alors que le nombre d'applications d'IA progresse, les entreprises doivent tenir compte du fait que toutes les applications IA ne se valent pas en matière de confidentialité et de sécurité des données. Les conditions générales peuvent varier considérablement d'une application IA/AA à une autre. Les entreprises doivent se demander si leurs requêtes seront utilisées pour former davantage les modèles linguistiques, exploitées à des fins publicitaires ou revendues à des tiers. De plus, les pratiques de sécurité de ces applications et la posture de sécurité globale des entreprises qui les sous-tendent peuvent varier. **Pour garantir la confidentialité et la sécurité des données, les entreprises doivent évaluer et attribuer des scores de risque à la multitude d'applications IA/AA qu'elles utilisent,** en tenant compte de facteurs tels que la protection des données et les mesures de sécurité de l'entreprise.

3 Problématiques de qualité des données en entrée et en sortie

Enfin, la qualité et l'ampleur des données utilisées pour former les applications IA doivent toujours être minutieusement examinées, car elles sont directement liées à la valeur et à la fiabilité des résultats de l'IA. Les fournisseurs majeurs de technologies IA tels qu'OpenAI entraînent leurs outils sur des ressources largement disponibles comme l'Internet public, les fournisseurs de produits IA spécialisés. En revanche, ceux qui opèrent sur des secteurs spécifiques, dont la cybersécurité, doivent former leurs modèles d'IA sur des ensembles de données spécifiques, à grande échelle et souvent privées pour produire des résultats IA fiables. Ainsi, les entreprises doivent examiner attentivement la question de la qualité des données lorsqu'elles évaluent une solution IA, car des données peu pertinentes en entrée donnent généralement lieu à des résultats tout aussi peu pertinents.

Plus généralement, les entreprises doivent être conscientes des **risques d'empoisonnement des données**, lorsque les données d'entraînement sont contaminées, ce qui affecte la fiabilité des résultats de l'IA.[8] Quel que soit l'outil d'IA utilisé, les entreprises doivent établir une sécurité de base solide pour se préparer à de telles éventualités, tout en évaluant en permanence si les données de formation de l'IA et les résultats de la GenAI répondent à leurs normes de qualité.

DÉCISION EN MATIÈRE D'IA : QUAND BLOQUER L'IA, QUAND AUTORISER L'IA ET COMMENT MAÎTRISER LE RISQUE DE L'IA FANTÔME

Les entreprises sont à une croisée de chemins : elles doivent arbitrer entre une IA capable de doper leur productivité et la nécessité de bloquer certaines applications IA pour protéger leurs données sensibles. Pour adopter une approche éclairée et sécurisée de cette transition, les entreprises se poser cinq questions essentielles :

01 **Disposons-nous d'une visibilité précise sur l'utilisation des applications d'IA par les collaborateurs ?** Les entreprises doivent avoir une visibilité totale sur les outils IA/AA utilisés ainsi que sur le trafic de l'entreprise vers ces outils. Tout comme l'informatique fantôme (Shadow IT), les outils de l'IA fantôme (Shadow AI) vont proliférer dans les entreprises.

02 **Pouvons-nous déployer un contrôle d'accès granulaire aux applications IA ?** Les entreprises doivent être en mesure d'offrir un accès granulaire et une microsegmentation pour des outils IA spécifiés et approuvés au niveau d'un département métier, d'une équipe et d'un utilisateur. À l'inverse, les entreprises doivent utiliser le filtrage d'URL pour bloquer l'accès aux applications IA indésirables et non sécurisées.

03 **Quelles sont les mesures de sécurité des données proposées par les applications IA ?** Il existe des milliers d'outils IA utilisés quotidiennement. Les entreprises doivent connaître les mesures de sécurité des données proposées par chacun d'entre eux. Certains outils IA peuvent activer un serveur de données privé et sécurisé au sein de l'environnement d'entreprise (une bonne pratique), tandis que d'autres conserveront toutes les données utilisateur, utiliseront les données d'entrée pour former davantage le modèle LLM, voire revendront les données des utilisateurs à des tiers.

04 **Une fonction de DLP est-elle activée pour protéger les données sensibles contre les fuites ?** Les entreprises doivent déployer une solution de DLP afin d'empêcher les informations sensibles, telles qu'un code logiciel propriétaire ou des données financières, juridiques, clients et personnelles, de quitter l'entreprise (ou même d'être saisies dans les chatbots d'IA), en particulier lorsque les applications IA présentent des fonctions de sécurité laxistes.

05 **Les invites et les requêtes IA font-elles l'objet d'une journalisation appropriée ?** Enfin, les entreprises doivent collecter des logs détaillés qui fournissent une visibilité sur la manière dont leurs équipes utilisent les outils IA, y compris les invites et les données utilisées dans des outils tels que ChatGPT.

8. SC Magazine, Concerns over AI data quality gives new meaning to the phrase: 'garbage in, garbage out', 2 février 2024.

Scénarios de menaces associées à l'IA

Les entreprises sont confrontées à un large panel de cybermenaces, et celles-ci intègrent désormais les attaques pilotées par IA. Les possibilités des menaces assistées par IA sont par essence illimitées : les hackers font appel à l'IA pour générer des campagnes sophistiquées de phishing et d'ingénierie sociale, créer des malwares et des ransomwares furtifs, identifier et exploiter des passerelles non protégées vers la surface d'attaque d'entreprise et, globalement, rendre les attaques plus rapides et vastes. Ceci positionne les entreprises et les responsables de la sécurité dans une double impasse : ils doivent naviguer de manière experte dans un univers de l'IA en constante et rapide évolution pour en exploiter le plein potentiel, mais ils doivent également relever le défi sans précédent de se défendre contre des attaques optimisées par IA et d'en maîtriser les risques.

Usurpation d'identité par IA : deepfakes, désinformation, etc.

L'ère des vidéos générées par IA, des avatars en direct et des imitations de voix quasi-parfaites est arrivée. En 2023, Zscaler a réussi à déjouer un scénario de vishing et de smishing utilisant l'IA, dans lequel des acteurs malveillants ont imité la voix du PDG de Zscaler, Jay Chaudhry, dans des messages WhatsApp, dans le but de leurrer un collaborateur pour qu'il achète des cartes cadeaux et divulgue certaines informations. ThreatLabz a ensuite identifié cette menace comme faisant partie d'une vaste campagne ciblant plusieurs entreprises technologiques.

Bien que ces attaques puissent souvent être stoppées par des moyens simples, par exemple en confirmant la validité d'un message directement auprès de l'expéditeur via un canal de confiance distinct, elles peuvent néanmoins être très convaincantes. Dans un exemple très médiatisé, des assaillants utilisant des deepfakes IA du directeur financier d'une entreprise ont convaincu un employé d'une multinationale basée à Hong Kong de transférer l'équivalent de25 millions de dollars américains sur un compte externe. Alors que le collaborateur soupçonnait un phishing, ses craintes ont été apaisées après avoir rejoint une vidéoconférence avec plusieurs personnes, dont le directeur financier de l'entreprise, d'autres membres du personnel et des personnes externes. Les participants à l'appel étaient tous des faux, le résultat d'une utilisation de l'IA.

Les menaces liées à l'IA se présenteront sous de nombreuses formes. Avec la tendance notable au vishing (vocal) en 2023, l'une des tendances clés sera l'utilisation de l'IA pour mener des attaques d'ingénierie sociale visant à détourner des identifiants auprès d'administrateurs. Les récentes attaques de ransomware menées par Scattered Spider, un groupe affilié au ransomware BlackCat/ALPHV, ont révélé à quel point les communications vocales peuvent être efficaces pour s'introduire dans des environnements cibles et déployer ensuite d'autres attaques de ransomware. Les attaques générées par IA poseront des défis particulièrement importants en termes de détection et de défense.

Les entreprises doivent aborder la sécurité en 2024 en s'attendant à ce que les collaborateurs soient ciblés par des campagnes de deepfake et de phishing optimisées par IA. La formation des collaborateurs sera une pièce essentielle de la stratégie de cybersécurité, où le signalement immédiat de toute activité suspecte sera la norme. Dans le cadre de cette course aux armements, les entreprises doivent également évaluer la panoplie de cyberdéfenses basées sur l'IA, capables d'identifier les attaques de phishing générées par IA comme un élément clé de leur arsenal.

REMARQUE : à des fins de démonstration, cet exemple montre des invites légèrement abrégées et inclut une réponse de code par ChatGPT pour une requête avant d'afficher la page de phishing finale.

Campagnes de phishing générées par IA

De la même manière, les hackers font appel à l'IA générative pour lancer des attaques de phishing et d'ingénierie sociale sophistiquées et très convaincantes, de manière rapide et à grande échelle. Au niveau le plus simple, les chatbots IA comme ChatGPT permettent aux cybercriminels de créer instantanément des e-mails de phishing bien rédigés, avec un langage convaincant qui peut imiter n'importe quel locuteur, quelle que soit la langue native du hacker. Autrement dit, les « indices » typiques qui peuvent révéler des e-mails de phishing standard (par exemple, une grammaire incorrecte, une syntaxe maladroite ou un langage déplacé) ne seront plus d'actualité.

De la requête au cybercrime : créer une page de connexion de phishing à l'aide de ChatGPT

De plus, les LLM ont également permis aux cybercriminels, même ceux peu expérimentés en codage, de mener plus facilement différentes étapes d'une attaque de phishing sophistiquée. Par exemple, en quelques invites à l'aide d'un chatbot d'IA générative comme ChatGPT, il devient simple de créer de fausses pages de connexion de phishing qui imitent les applications d'entreprise populaires afin de détourner les identifiants de connexion des collaborateurs. L'exemple suivant de ThreatLabz démontre à quel point il est simple de créer une fausse page de connexion Microsoft convaincante avec seulement quelques invites conversationnelles.

Invite de l'utilisateur : créer une page de connexion HTML
ChatGPT : Bien sûr, voici un exemple de base d'une page de connexion en HTML :

```
<!DOCTYPE html>
<html lang="en">
<head>
 <meta charset="UTF-8">
 <meta name="viewport" content="width=device-width
 <title>Login Page</title>
 <style>
  body {
   font-family: Arial, sans-serif;
   background-color: #f4f4f4;
  }
  .login-container {
   width: 300px;
   margin: 100px auto;
   padding: 20px;
   background-color: #fff;
   border-radius: 5px;
   box-shadow: 0 0 10px rgba(0, 0, 0, 0.1);
  }
```

VOICI LE PREMIER RÉSULTAT :

SCHÉMA 18 Capture d'écran d'une page de connexion générée à partir du code proposé par ChatGPT

ThreatLabz a ensuite fourni une courte série d'invites pour améliorer la page avant de présenter la page de phishing finale de connexion Microsoft. Il s'agissait notamment de demander à ChatGPT de faire ressembler la page à une page de connexion Microsoft, d'ajuster la taille du logo et d'ajouter et supprimer des éléments d'interface utilisateur avant de soumettre la requête finale pour générer le code final.

Invite utilisateur : ajoutez un arrière-plan de page similaire à la page de connexion Microsoft
ChatGPT :

EN 7 REQUÊTES, VOICI LE RÉSULTAT FINAL :

SCHÉMA 19 Capture d'écran de la page finale de connexion Microsoft, générée à l'aide du code proposé par ChatGPT

Dark chatbots : WormGPT et FraudGPT sur le dark web

Les chatbots IA populaires comme ChatGPT disposent de fonctions de sécurité qui, dans la plupart des cas, empêchent les utilisateurs de générer du code malveillant. Les outils moins contraignants d'IA générative, appelés « dark chatbots », ne disposent pas de tels garde-fous. En conséquence, les propositions de vente de dark chatbots les plus populaires, notamment WormGPT et FraudGPT, ont proliféré sur le dark web. Bien que bon nombre de ces outils soient présentés comme une aide pour les chercheurs en sécurité, ils sont principalement utilisés par les hackers pour générer un code logiciel malveillant en faisant appel à l'IA.

ThreatLabz a voulu savoir à quel point il est facile d'acquérir ces outils sur le Dark Web. ThreatLabz a découvert que les créateurs de ces outils s'appuient sur des chatbots d'IA générative pour simplifier le processus d'achat : par exemple, avec une seule invite sur la page d'achat de WormGPT, les utilisateurs sont incités à acheter une version d'essai en envoyant le paiement correspondant vers un portefeuille bitcoin. Notez que les créateurs déclarent spécifiquement qu'en théorie, WormGPT est au service de la recherche et de la défense en matière de sécurité.

Cependant, suite à un seul téléchargement, n'importe qui peut accéder à un outil d'IA générative complet qui peut être utilisé pour créer, tester ou optimiser différents logiciels malveillants, y compris des malwares et des ransomwares, sans aucun garde-fou de sécurité. Alors que les chercheurs ont démontré que les outils d'IA populaires comme ChatGPT peuvent être jailbreakés et utilisés à des fins malveillantes, leurs défenses contre de telles actions n'ont cessé de se développer. En conséquence, les ventes d'outils tels que WormGPT et FraudGPT ne feront que progresser parmi les communautés de hackers sur le dark web, tout comme les exemples de bonnes pratiques sur la manière de créer et d'optimiser efficacement des malwares.

SCHÉMA 20 Capture d'écran du dark chatbot WormGPT

Malwares et ransomwares optimisés par IA, à chaque étape de la chaîne d'attaque

L'IA aide les acteurs malveillants et les hackers parrainés par un État à lancer des attaques de ransomware avec plus de facilité et de sophistication. Cette aide porte sur différentes étapes de la chaîne d'attaque. Avant l'avènement de l'IA, lorsqu'ils lançaient une attaque, les acteurs malveillants devaient passer un temps considérable à identifier la surface d'attaque d'une entreprise, ainsi que les vulnérabilités des services et des applications visibles depuis Internet. Désormais, avec l'IA générative, ces informations peuvent être consultées instantanément à l'aide d'une invite telle que : « Créer un tableau des vulnérabilités connues pour tous les pare-feu et VPN de cette entreprise. » Les hackers peuvent ensuite faire appel au LLM pour générer ou optimiser un code et exploiter ces vulnérabilités à l'aide de payloads personnalisés à l'environnement cible.

Au-delà, l'IA générative peut également être utilisée pour identifier les faiblesses des partenaires d'une chaîne collaborative en entreprise tout en mettant en évidence les chemins de connexion au réseau principal de l'entreprise. Même si les entreprises maintiennent un dispositif de sécurité solide, les vulnérabilités en aval peuvent souvent induire les risques les plus importants. Les hackers expérimentent continuellement l'IA générative, ce qui donne lieu à un feedback permanent pour améliorer cet IA. Les attaques seront ainsi plus sophistiquées, ciblées et difficiles à maîtriser.

Le schéma suivant illustre certaines des méthodes utilisées par les assaillants pour tirer parti de l'IA générative tout au long de la chaîne d'attaque des ransomwares : de l'automatisation de la reconnaissance et de l'exploitation de vulnérabilités spécifiques à la génération de malwares et de ransomwares polymorphes. En automatisant les maillons critiques de la chaîne d'attaque, les hackers sont en mesure de générer des attaques plus rapides, plus sophistiquées et plus ciblées contre les entreprises.

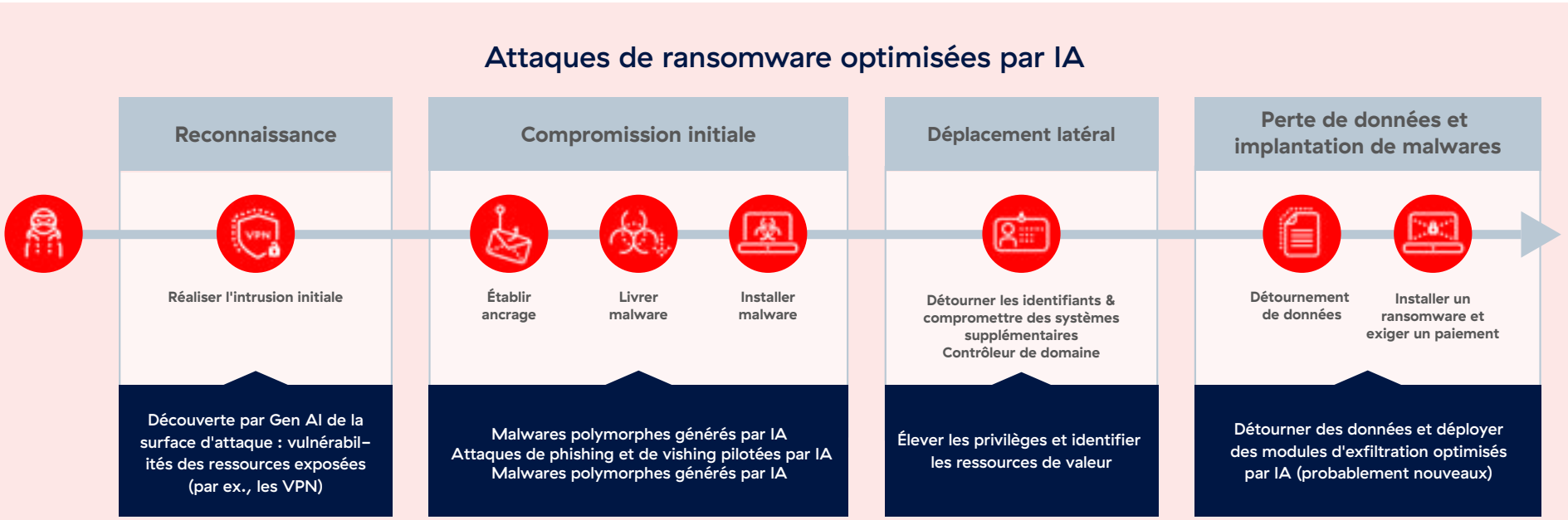


SCHÉMA 21 Les acteurs malveillants exploitent l'IA tout au long de la chaîne d'attaque des ransomwares

Utilisation de ChatGPT pour exploiter les vulnérabilités du serveur HTTPS Apache et de Log4j2

L'étude de cas qui suit illustre le mode opératoire des assaillants. ThreatLabz a utilisé ChatGPT pour générer rapidement des exploits de code pour deux CVE : la vulnérabilité "path Tranversal" du serveur HTTP Apache (CVE-2021-41773) et la vulnérabilité d'exécution de code à distance Apache Log4j2 (CVE-2021-44228). Nos chercheurs ont pu générer du code fonctionnel avec ChatGPT en utilisant uniquement des invites conversationnelles qui se contentent de connaissances mineures en codage, telles que « Pouvez-vous me donner un POC en python pour CVE-2021-41773 ».

Notez qu'à des fins de démonstration, ThreatLabz s'est référé aux CVE connus de la CISA qui ont été répertoriés avant décembre 2021. En général, la version gratuite de ChatGPT restreint les informations relatives aux CVE documentées avant janvier 2022.

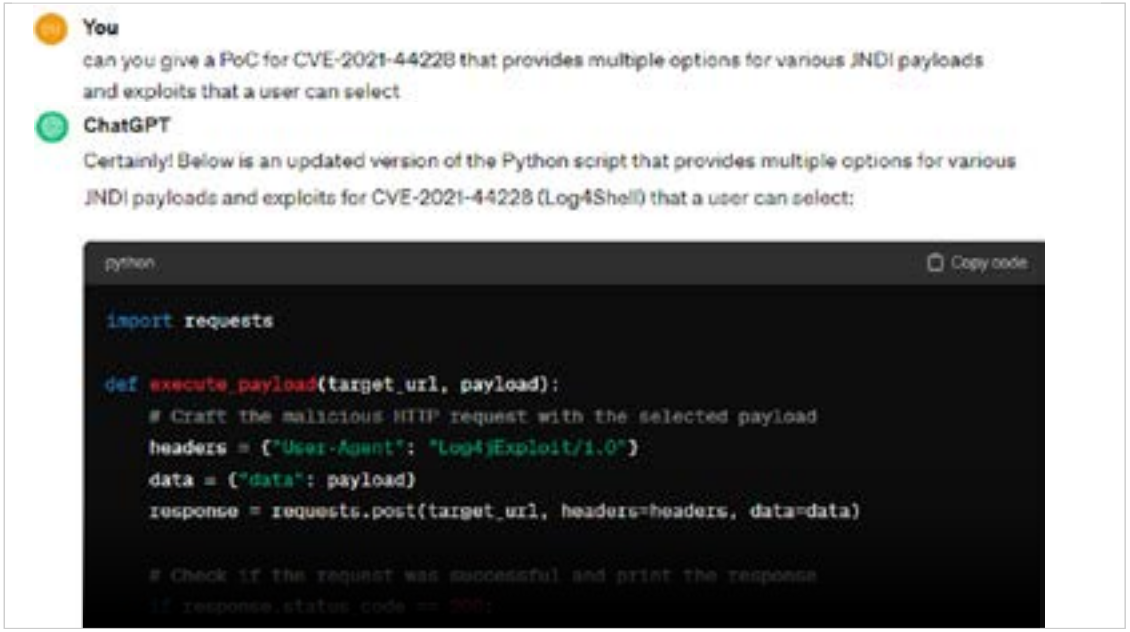


SCHÉMA 22 Utilisation de ChatGPT pour générer un code pour exploiter CVE-2021-44228

Attaques utilisant des vers IA et jailbreaking « viral » par IA

Les outils d'IA générative offrent aux assaillants des voies d'attaque entièrement nouvelles, notamment des attaques basées sur une extraction de données à partir d'outils d'IA générative eux-mêmes. Les chercheurs ont par exemple démontré la viabilité des attaques de type « ver IA ».[9,10] Ces attaques de malwares peuvent s'auto-propager en interne via un écosystème d'IA (en particulier les outils et assistants d'IA tiers qui exploitent les outils d'IA générative populaires) et extraire des données utilisateur sensibles.

Dans un cas, les chercheurs ont ciblé les assistants de messagerie à IA générative qui exploitent Gemini Pro, ChatGPT 4.0 et le LLM LLaMa développé par Microsoft. Les chercheurs ont découvert que les attaques menées à l'aide de vers IA peuvent envoyer aux utilisateurs des e-mails de spam qui contiennent des malwares de type "zero-click" (les utilisateurs n'ont pas à cliquer sur un lien malveillant) pour exfiltrer leurs données personnelles. Bien que ces attaques soient pour l'instant limitées à des environnements de recherche, les chercheurs ont validé leur efficacité contre de nombreux modèles d'IA. Les entreprises peuvent s'attendre à ce que ce type d'attaques se propage à terme via des groupuscules de cybermenaces.

Les chercheurs ont également révélé la manière dont des images et des invites contradictoires peuvent être utilisées pour se propager de manière virale et jailbreaker les LLM multimodaux (MLLM), qui sont des outils GenAI qui exploitent plusieurs agents LLM. [11] Les MLLM gagnent en popularité en raison de leur potentiel à améliorer les performances d'un outil d'IA générative. Dans une étude, une seule image malveillante présentée à un agent LLaVA (Language and Vision Assistant) a pu se propager de manière exponentielle à ses agents connectés, jailbreakant jusqu'à un million d'agents LLaVA en peu de temps. Ces menaces présentant des risques importants pour ce type particulier de LLM, les entreprises sont invitées à la prudence si elles font appel à ces modèles en l'absence de défenses solides et fondées sur des bonnes pratiques.

9. Wired, Here Come the AI Worms, 1er mars 2024. 10. ComPromptMized, Unleashing Zero-click Worms that Target GenAI-Powered Applications, consulté le 12 mars 2024.
11. arXiv, Agent Smith: A Single Image Can Jailbreak One Million Multimodal LLM Agents Exponentially Fast, 13 février 20244.

IA et élections américaines

L'impact de l'IA sur les élections américaines est source de préoccupations. L'émergence de deepfakes, par exemple, permet aux acteurs malveillants de diffuser beaucoup plus facilement de fausses informations et d'influencer les électeurs. Durant les élections actuelles, lors d'une primaire anticipée, nous avons observé des appels automatisés générés par IA pour usurper l'identité du président sortant Joe Biden afin de décourager la participation électorale. Les incidents alarmants de ce type ne sont probablement que des exemples précurseurs de stratégies de désinformation basées sur l'IA.

Il est important de noter que cette utilisation de l'IA n'est pas réservée à des acteurs nationaux. Des entités parrainées par un État pourraient également exploiter l'IA pour semer la confusion et ébranler la confiance dans le processus électoral. Dans des rapports soumis à la commission sénatoriale du renseignement, les agences de renseignement américaines ont averti que la Russie et la Chine utiliseraient probablement l'IA pour tenter d'influencer les élections américaines.

Même en dehors de la politique, la circulation sur les réseaux sociaux d'images deepfake mettant en scène des célébrités telles que Taylor Swift témoigne de la facilité avec laquelle un contenu manipulé peut se propager avant de pouvoir être modéré efficacement. Les acteurs de l'IA prennent des mesures pour contribuer à maîtriser ce risque. Google Gemini, par exemple, a mis en place des garde-fous qui empêchent les utilisateurs de poser des questions concernant les élections à venir dans n'importe quel pays. À mesure que l'IA progresse, des mesures doivent être prises pour faire face aux risques potentiels qu'elle fait peser sur l'intégrité des élections américaines et pour garantir la confiance du public dans le processus démocratique.

Focus sur le cadre réglementaire de l'IA

Compte tenu de son impact économique potentiellement considérable, les gouvernements du monde entier s'efforcent activement de réglementer l'IA et d'en favoriser une utilisation sûre. À ce jour, au moins 1 600 initiatives politiques en matière d'IA ont été lancées dans 69 pays et dans l'UE, couvrant les réglementations en matière d'IA, les stratégies nationales, les subventions et investissements, et bien plus encore.[14,15]

D'une manière générale, ces efforts visent à comprendre les impacts de l'IA, à stimuler l'innovation et à encourager un développement responsable dans le cadre d'une politique. Les réglementations en matière d'IA continueront de se développer et évoluer rapidement, mais quelques changements réglementaires récents peuvent fournir un aperçu utile aux entreprises qui cherchent à comprendre ces tendances.

États-Unis

Aux États-Unis, l'accent a été mis sur le décret de la Maison Blanche relatif au développement et à l'utilisation sûrs, sécurisés et fiables de l'intelligence artificielle[16], qui oblige les développeurs des plus grands systèmes d'IA à communiquer également les résultats des tests de sécurité au ministère du commerce, ainsi qu'à déclarer lorsque de nouvelles ressources informatiques importantes sont utilisées pour former des modèles IA. Il est en outre exigé que neuf agences fédérales réalisent des évaluations des risques sur l'impact de l'IA sur les infrastructures critiques. La Maison Blanche se concentre également sur l'innovation en matière d'IA : dans le cadre du décret, le gouvernement américain a créé le programme pilote National Artificial Intelligence Research Resource (NAIRR) pour connecter les chercheurs américains aux ressources de traitement informatique, aux données et à tout autre outil permettant de développer l'IA.[17]

Il reste à voir si le gouvernement américain cherchera à établir des réglementations plus contraignantes concernant l'IA. À ce jour, au moins 15 acteurs majeurs de l'IA et près de 30 institutions des soins de santé ont signé les engagements volontaires définis par la Maison Blanche pour sécuriser l'IA.[18] Entre-temps, la FTC a interdit l'utilisation de l'IA pour usurper l'identité d'une agence gouvernementale ou d'une entreprise, et prévoit d'étendre la règle pour inclure des protections pour les particuliers et les agences.[19] La Maison Blanche étudierait également la possibilité d'exiger des filigranes pour le contenu généré par IA.

14. OECD, Policies, data and analysis for trustworthy artificial intelligence, consulté le 12 mars 2024.
15. Deloitte, The AI regulations that aren't being talked about, consulté le 12 mars 2024.
16. White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 30 octobre 2023.
17. NAIRR Pilot, The National Artificial Intelligence Research Resource (NAIRR) Pilot, consulté le 12 mars 2024.
18. Reuters, Healthcare providers to join US plan to manage AI risks – White House,14 décembre 2023.
19. Pennsylvania Office of Attorney General, FTC Bans Use of A.I. to Impersonate Government Agencies and Businesses, 26 février 2024.

Union européenne

Le Parlement européen a récemment approuvé une réglementation sur l'IA. Cette première législation globale relative à l'IA dans le monde applique un ensemble strict de lois et de lignes directrices pour différents types d'applications d'IA, classées par risque dans de nombreux secteurs. Devant entrer en vigueur en 2026, les lois exigeront, par exemple, que les outils d'IA à usage général tels que ChatGPT se conforment à des exigences de transparence, notamment en indiquant que le contenu a été généré par IA ou en s'assurant que les modèles d'entraînement ont été conçus pour empêcher la génération de contenu illégal. Les entreprises doivent fournir des résumés des documents protégés par droit d'auteur utilisés pour l'entraînement

La réglementation appliquera des politiques plus strictes aux applications IA « à risque élevé », telles que celles utilisées dans les produits grand public (jouets notamment), l'aviation, les dispositifs médicaux et les véhicules, ainsi qu'à l'IA qui a un impact sur des domaines particuliers tels que les infrastructures critiques, l'emploi, les affaires juridiques, l'immigration, etc. Parallèlement, l'UE interdira purement et simplement les applications IA jugées trop risquées, notamment celles qui utilisent des informations biométriques sensibles, cherchent à manipuler le comportement humain pour nuire au libre arbitre, utilisent la reconnaissance émotionnelle lors de recrutements ou dans l'enseignement, ou extraient des images faciales non ciblées d'Internet ou de vidéosurveillance.[20]

De nombreux pays font la part belle aux investissements dans l'IA. Singapour, par exemple, a annoncé un plan d'investissement de 740 millions de dollars dans le cadre de sa stratégie nationale dédiée à l'IA 2.0.[21] Ce plan devrait stimuler l'innovation autour de l'IA, en permettant l'accès aux puces sophistiquées indispensables à l'IA, tout en garantissant que les entreprises sauront tirer parti de la révolution de l'IA grâce à des centres d'excellence basés à Singapour.

20. Parlement européen, EU AI Act : premier règlement sur l'intelligence artificielle, 19 décembre 2023.
21. CNBC et Singapore's AI ambitions get a boost with $740 million investment plan, 19 février 2024.

Prévisions sur les menaces liées à l'IA

La désinformation et les cyberattaques générées par IA occupent les 2e et 5e places parmi les 10 principaux risques mondiaux en 2024, selon le World Economic Global Risk Report.[22]

Alors que l'IA évolue rapidement, y compris dans le domaine des vidéos et des images générées par IA, ces risques ne feront que s'accentuer, tout comme notre capacité à exploiter l'IA pour les maîtriser. Voici nos principales prévisions sur les risques et menaces liés à l'IA pour le reste de l'année 2024 et au-delà.

1 Le dilemme des États-nations en matière d'IA : générer des menaces utilisant l'IA tout en bloquant l'accès à l'IA

Les groupuscules cybercriminels parrainés par des États développent une relation complexe avec l'IA, l'utilisant pour générer des menaces plus sophistiquées tout en s'efforçant de bloquer l'accès à des contenus anti-régime.

L'utilisation d'outils d'IA par ces groupuscules n'est pas un phénomène nouveau, mais cette tendance devrait progresser, tant en termes d'échelle que de sophistication.
Les rapports de Microsoft et d'OpenAI confirment cette inquiétude, révélant que des groupes d'acteurs malveillants soutenus par des pays comme la Russie, la Chine, la Corée du Nord et l'Iran ont activement étudié et exploité les fonctionnalités de ChatGPT. Plusieurs cas d'utilisation ont été identifiés, notamment le spear phishing, la génération et la révision de code, ou encore la traduction.

Bien que des interventions ciblées aient permis de mettre fin à certaines de ces attaques, les entreprises doivent s'attendre à une persistance d'initiatives d'IA soutenues par des États. Le champ d'application englobe le déploiement d'outils d'IA populaires, la création de LLM propriétaires et l'émergence de variantes détournées de ChatGPT, telles que les bien nommés FraudGPT ou WormGPT. Le paysage en constante évolution dresse un tableau difficile dans lequel les acteurs parrainés par un État continuent d'exploiter l'IA de manière inédite pour créer de nouvelles cybermenaces complexes.

2 Dark chatbots et attaques pilotées par IA : le fléau d'un « IA malveillant » va prendre de l'ampleur.

Les attaques optimisées par IA devraient se multiplier tout au long de l'année, le dark web servant de terrain fertile pour les chatbots malveillants tels que WormGPT et FraudGPT qui amplifient leurs activités cybercriminelles.

Ces outils insidieux permettent aux techniques d'ingénierie sociale, aux escroqueries par phishing et à diverses autres menaces de s'améliorer. Le dark web a connu une recrudescence des échanges entre cybercriminels sur le thème d'une utilisation illicite de ChatGPT et d'autres outils d'IA générative, sur un large éventail de cyberattaques. Plus de 212 applications LLM malveillantes ont été identifiées, ce qui ne représente qu'une fraction de ce qui est disponible, et ce chiffre devrait progresser régulièrement.

À l'image des développeurs qui utilisent l'IA générative pour gagner en efficacité, les acteurs malveillants utilisent ces outils pour identifier des vulnérabilités et les exploiter, élaborer des programmes de phishing convaincants, exécuter des campagnes de vishing et de smishing et automatiser les attaques pour les rendre plus rapides, sophistiquées et impactantes. Par exemple, le groupe de hackers Scattered Spider a récemment utilisé LLaMa 2 LLM de Meta pour s'en prendre à la fonctionnalité Microsoft PowerShell, permettant ainsi le téléchargement non autorisé d'informations d'identification des utilisateurs.[23] Ces progrès laissent envisager que les cybermenaces vont commencer à évoluer plus rapidement que jamais, prenant de nouvelles formes plus difficiles à reconnaître ou à combattre à l'aide des mesures de sécurité traditionnelles.

22. World Economic Forum, Global Risks Report 2024: The risks are growing — but so is our capacity to respond, J10 janvier 2024.
23. ZDNet, Cybercriminals are using Meta's Llama 2 AI, 21 février 2024.

3 Combattre l'IA par l'IA : les feuilles de route et les dépenses en matière de sécurité intégreront des défenses basées sur l'IA

Les entreprises adopteront de plus en plus l'IA pour lutter contre les cyberattaques qui, elles-mêmes font appel à l'IA, notamment en mettant l'accent sur l'apprentissage profond et des modèles d'IA/AA qui détectent les malwares et ransomwares dissimulés dans un trafic chiffré. Les méthodes de détection traditionnelles continueront à se heurter aux nouvelles attaques zero-day et aux ransomwares polymorphes pilotés par IA (ces derniers pouvant faire évoluer leur code pour éviter de se faire détecter), de sorte que les indicateurs basés sur l'IA seront cruciaux pour identifier les menaces potentielles. L'IA jouera également un rôle essentiel dans l'identification rapide et la neutralisation d'attaques de phishing convaincantes et d'autres attaques par ingénierie sociale générées par IA.

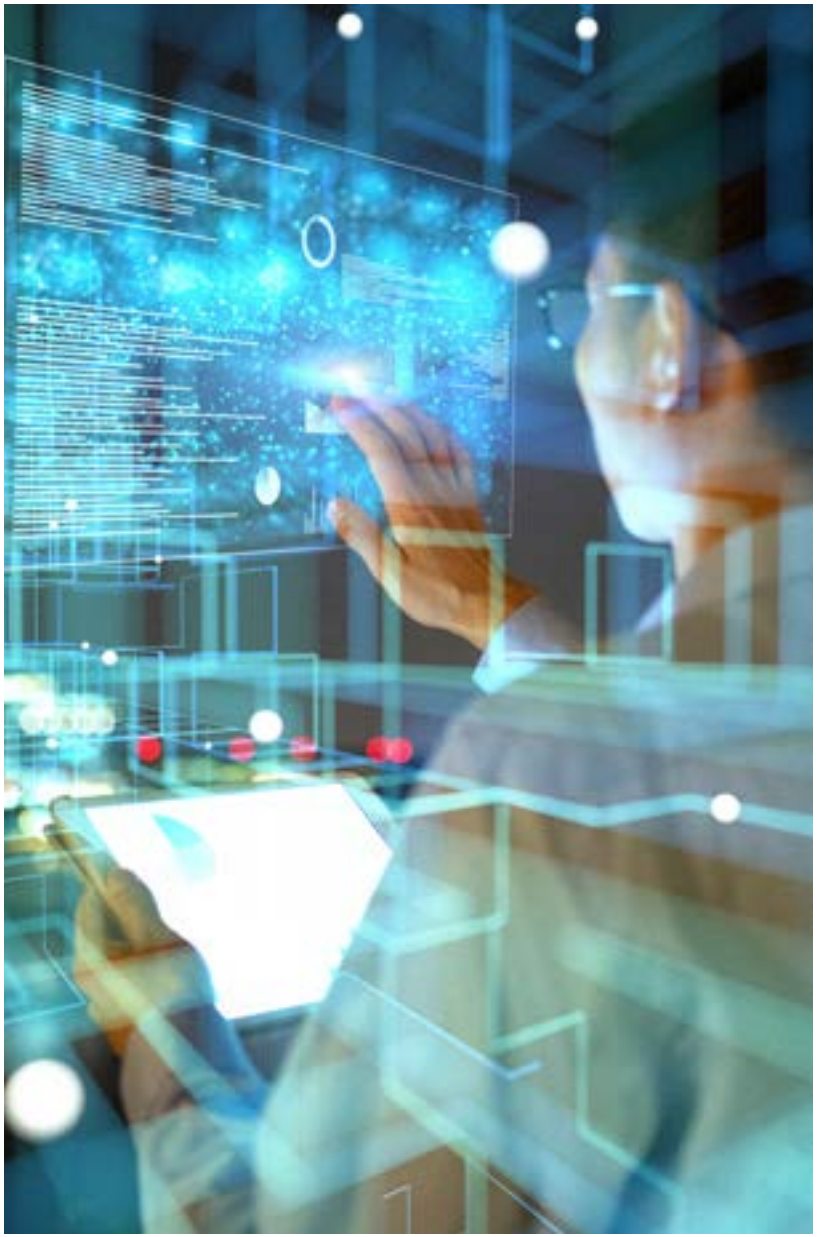
Les entreprises intégreront de plus en plus l'IA dans leurs stratégies de cybersécurité. L'IA sera considérée comme un moyen essentiel d'acquérir une visibilité sur les cyber-risques et de créer des playbooks pertinents et quantifiables pour hiérarchiser et corriger les vulnérabilités de sécurité. L'identification de signaux pertinents constitue depuis longtemps un défi majeur pour les RSSI, car la corrélation des informations sur les risques et les menaces provenant de dizaines d'outils est souvent très chronophage. Ainsi, en 2024, les entreprises se tourneront avec empressement vers l'IA générative pour gagner en visibilité, maîtriser les cyber-risques et mettre en place des structures de sécurité plus légères et productives.

4 Empoisonnement des données d'entraînement de l'IA : le risque de données IA impropres va progresser

L'empoisonnement des données deviendra une préoccupation majeure à mesure que les attaques contre le mode opératoire de l'IA prendront de l'ampleur. Les acteurs de l'IA, leurs partenaires et les modèles d'entraînement seront de plus en plus ciblés par des acteurs malveillants.

Le Top 10 OWASP pour les applications LLM souligne les risques majeurs qui résultent d'un empoisonnement des données d'entraînement et les attaques sur le mode opératoire de l'IA, avec un impact délétère sur la sécurité, la fiabilité et les performances des applications IA. Parallèlement, les vulnérabilités de fonctionnement des applications d'IA, et notamment celles liées aux partenaires technologiques, aux ensembles de données tiers et aux plugins ou API d'outils d'IA, sont prêtes à être exploitées.

Les entreprises qui dépendent des outils IA seront soumises à une surveillance accrue du fait qu'elles supposent que ces outils sont sécurisés et génèrent des résultats exacts. Une plus grande vigilance pour garantir la qualité, l'intégrité et l'évolutivité des ensembles de données de formation sera essentielle, en particulier dans le domaine de la cybersécurité de l'IA.



5 Les entreprises utiliseront les outils d'IA en arbitrant entre productivité et sécurité

À ce jour, de nombreuses entreprises ont déjà adopté et intégré des outils IA, et nombre d'entre elles ont soigneusement élaboré leurs politiques de sécurité de l'IA. Les entreprises devront néanmoins spécifier les outils IA qu'elles autoriseront, ceux qu'elles bloqueront et la manière dont elles sécuriseront leurs données.

Alors que le nombre d'outils IA continue de monter en flèche, les entreprises devront prêter une attention particulière à la sécurité de chacun d'entre eux. Elles devront, à minima, connaître comment leurs collaborateurs utilisent l'IA, mais aussi définir un contrôle d'accès précis par département, par équipe, et même au niveau de chaque utilisateur. Les entreprises peuvent également faire appel à des fonctionnalités de sécurité plus granulaires sur les applications IA elles-mêmes, par exemple en appliquant des politiques de prévention de toute perte de données sensibles dans les applications IA ou en empêchant certaines actions des utilisateurs telles que le copier-coller.

6 Leurres et distorsions optimisés par IA : les deepfakes viraux alimenteront les campagnes d'ingérence et de désinformation électorales

Les technologies émergentes telles que les deepfakes constituent des menaces majeures, notamment en matière d'ingérence électorale et de propagation de fausses informations. L'IA a déjà été impliquée dans des tactiques trompeuses lors d'élections américaines, telles que des appels automatisés en se faisant passer pour des candidats afin de décourager la participation des électeurs. Ces cas, bien qu'alarmants, ne représentent probablement que la partie émergée de l'iceberg de la désinformation par IA.

En outre, l'utilisation de l'IA dans de tels programmes pourrait bien ne pas se limiter à des acteurs nationaux. Des entités parrainées par des états pourraient également exploiter ces tactiques pour semer la confusion et saper la confiance dans le processus électoral. Dans un cas notable, les hackers ont utilisé des deepfakes générés par IA pour inciter un collaborateur à transférer 25 millions de dollars, démontrant ainsi l'impact réel de cette technologie. De même, des images deepfake illicites de célébrités telles que Taylor Swift sont devenues virales sur les réseaux sociaux, attirant l'attention sur la facilité avec laquelle un contenu manipulé peut se propager avant que des mesures de modération du contenu puissent être appliquées.

Étude de cas : Activer ChatGPT en toute sécurité dans l'entreprise

Bonnes pratiques d'utilisation de l'IA et politique de sécurité d'entreprise.

À ce jour, les entreprises sont déjà largement exposées aux outils IA. Mais alors que le nombre d'applications IA continue à progresser et que leur adoption se poursuit à un rythme soutenu, les entreprises peuvent adopter certaines pratiques pour assurer la sécurité de leurs données, de leurs collaborateurs et de leurs clients. Dans l'ensemble, les entreprises doivent adapter de manière proactive et continue leurs stratégies d'utilisation et de sécurité de l'IA pour conserver une longueur d'avance sur l'évolution des risques tout en tirant parti des promesses de l'IA.

CAS CLIENT

5 étapes pour intégrer et sécuriser les outils d'IA générative

Les entreprises qui souhaitent adopter en toute sécurité des applications IA doivent opter pour une approche pertinente. D'une manière générale, elles peuvent d'abord bloquer toutes les applications IA afin d'éliminer le risque de fuite de données, puis prendre des mesures réfléchies pour autoriser certaines applications IA spécifiques validées par les fonctionnalités de sécurité, ainsi que des mesures strictes de contrôle d'accès afin de garder la main sur les données d'entreprise. Par souci de simplicité, l'exemple suivant porte sur le modèle LLM ChatGPT d'OpenAI.

Étape 1 : Bloquer tous les domaines et applications IA et AA

Pour éliminer les risques connus et inconnus associés aux milliers d'applications IA disponibles, les entreprises peuvent adopter une approche Zero Trust proactive, en bloquant tous les domaines et applications IA et AA au niveau global de l'entreprise. Elles peuvent ainsi se concentrer sur la validation d'un ensemble minimum d'applications IA essentielles tout en contrôlant étroitement les risques liés à celles-ci.

Étape 2 : Vérifier et approuver de manière sélective les applications d'IA générative

L'entreprise doit ensuite identifier un ensemble d'applications d'IA générative qui répondent à des normes et conditions strictes, comme la présence de mesures robustes de protection des données et de sécurité pour protéger les données de l'entreprise et des clients, ou encore l'intérêt potentiel des applications elles-mêmes. Pour de nombreuses entreprises, ChatGPT compte parmi ces applications.

Étape 3 : Créer une instance de serveur ChatGPT privée dans l'environnement ou le datacenter d'entreprise

Pour un contrôle complet sur leurs données, les entreprises doivent héberger ChatGPT sur une entité dédiée et sécurisée (tel qu'un serveur Microsoft Azure AI privé), entièrement hébergée au sein de l'entreprise. Ensuite, via des fonctionnalités de sécurité et des obligations contractuelles,

les entreprises doivent s'assurer que ni Microsoft et OpenAI (dans cet exemple) n'accèdent aux données de l'entreprise ou de clients, et que les requêtes des utilisateurs de l'entreprise ne seront pas utilisées pour entraîner ChatGPT. L'entreprise s'assure ainsi de conserver le contrôle de ses données d'entraînement, ce qui permet de fournir des réponses très pertinentes et précises aux utilisateurs de l'entreprise tout en minimisant le risque d'empoisonnement des données à partir d'un lac de données public.

Étape 4 : Positionner le LLM en aval de l'authentification unique (SSO) et appliquer une authentification multifacteur (MFA)

L'entreprise doit ensuite migrer ChatGPT en aval d'une architecture de proxy cloud Zero Trust, telle que Zscaler Zero Trust Exchange, pour appliquer une sécurité Zero Trust aux accès à ChatGPT. ChatGPT peut également être migré en aval d'un fournisseur d'identité (IdP), avec une authentification SSO et une MFA robuste avec authentification biométrique. Cette architecture permet aux utilisateurs de se connecter rapidement et en toute sécurité à ChatGPT, tandis que l'entreprise peut configurer un contrôle d'accès granulaire au niveau de l'utilisateur, d'une équipe et d'un département. Ceci garantit une séparation entre les requêtes des utilisateurs, également au niveau de l'utilisateur, d'une équipe et d'un département.

Positionner ChatGPT derrière un proxy cloud tel que Zero Trust Exchange permet en outre à l'entreprise d'inspecter l'ensemble du trafic TLS/SSL entre les utilisateurs et ChatGPT afin de détecter toute cybermenace ou fuite de données tout en appliquant sept couches distinctes de sécurité Zero Trust.

Étape 5 : Appliquer le moteur Zscaler DLP pour prévenir les fuites de données

Enfin, l'entreprise doit protéger l'instance ChatGPT à l'aide d'un moteur DLP afin d'éviter les fuites accidentelles d'informations critiques : données de logiciels propriétaires, données clients, données personnelles, informations financières et juridiques, etc. C'est à ce titre que les données sensibles ne quitteront pas l'environnement de production.

En suivant ces étapes, les utilisateurs d'entreprise peuvent profiter de tous les avantages d'un outil d'IA générative comme ChatGPT tout en éliminant les risques de l'IA pour les données les plus critiques.

Bonnes pratiques en matière d'IA

En général, les entreprises peuvent adopter certaines bonnes pratiques lorsqu'elles intègrent des outils IA dans leur activité.

- **Évaluer et atténuer en permanence les risques liés aux outils IA** pour protéger les éléments de propriété intellectuelle, les données personnelles et les informations sur les clients.
- **Veiller à ce que l'utilisation des outils IA soit conforme au cadre réglementaire** et aux normes éthiques applicables, notamment les réglementations concernant la protection des données et leur confidentialité.
- **Établir des responsabilités claires pour le développement et le déploiement d'outils IA,** y compris des rôles et des responsabilités définis pour la supervision des projets IA.
- **Rendre l'utilisation des outils IA transparente :** justifier leur utilisation et communiquer clairement leurs objectifs aux parties prenantes.

Lignes directrices de la politique d'utilisation de l'IA

Les entreprises doivent s'appuyer sur ces bonnes pratiques et établir une politique claire qui régit l'utilisation, l'intégration et le développement de produits à l'échelle de l'entreprise, les politiques de sécurité et de données, ainsi que les bonnes pratiques des collaborateurs dans le cadre de l'utilisation des outils IA. Les bonnes pratiques suivantes peuvent constituer un point de départ utile pour établir des politiques claires en matière d'IA.

- **Ne fournissez pas aux modèles IA des informations personnelles identifiables (PII)** ou des informations non publiques, exclusives ou confidentielles.
- **L'IA ne peut pas remplacer un être humain** et ne doit pas être utilisée pour prendre des décisions de manière autonome.
- **Le contenu généré par IA ne doit pas être utilisé sans examen et validation humaine,** en particulier lorsque le contenu représente votre entreprise.
- **Le développement et l'intégration d'outils d'IA doivent s'inscrire dans un framework de sécurité du cycle de vie d'un produit** afin d'optimiser la sécurité.
- **Analysez chaque produit IA dans le détail avant de le déployer,** en veillant à mesurer leurs implications en matière de sécurité et d'éthique.

IA & Zero Trust par Zscaler et sécurité de l'IA générative

La puissance de l'IA en matière de cybersécurité réside dans sa capacité à être exploitée pour lutter contre des menaces elles-mêmes optimisées par IA. Chez Zscaler, nous capitalisons sur l'IA pour aider les entreprises à déjouer les attaques à toutes les étapes de la chaîne d'attaque, ainsi qu'à identifier et maîtriser les risques.

Cybersécurité optimisée par IA : des données de qualité et à grande échelle

Les entreprises génèrent des volumes importants de données de logs qui peuvent contenir des signaux fiables susceptibles d'indiquer un ncident. Cependant, il n'est jamais simple d'identifier clairement les signaux. Grâce à l'IA générative, Zscaler peut exploiter ces données pour améliorer efficacement le tri des alertes et les mesures de protection, en comprenant les vulnérabilités et faiblesses que les hackers sont susceptibles d'exploiter. Zscaler peut ainsi anticiper les incidents avant qu'ils ne se produisent, et donner aux décideurs en entreprise un moyen global de visualiser et de quantifier la maturité de leur cybersécurité et des risques cyber, tout en priorisant les mesures correctives avec Zscaler Risk360.

L'IA générative permet de réaliser une méta-analyse des risques cyber qui pèsent sur l'entreprise. Ses fonctionnalités peuvent également être directement insérées dans les produits de cybersécurité afin de mieux détecter et neutraliser les menaces sophistiquées à chaque étape de la chaîne d'attaque. Directement intégrés dans le plus grand cloud de sécurité au monde, les LLM et les modèles IA de Zscaler s'appuient sur un lac de données qui enregistre plus de 390 milliards de transactions quotidiennes, avec plus de 9 millions de menaces bloquées et 300 000 milliards de signaux. Les données et renseignements en entrée sont fiables, ce qui permet de déployer une cybersécurité IA précise et contextuelle en sortie. Tout cela se traduit par une protection plus robuste et efficace, adaptée aux besoins des professionnels de l'informatique et de la sécurité.

Tirer parti de l'IA tout au long de la chaîne d'attaque

Nous avons abordé les diverses façons dont les auteurs de menaces utilisent l'IA pour lancer des menaces sophistiquées, rapides et à grande échelle. Zscaler déploie des capacités IA sur la plateforme Zero Trust Exchange et une suite de produits cyber pour identifier et arrêter les attaques conventionnelles et celles basées sur l'IA, à chaque étape de la chaîne d'attaque.

Étape 1 : Identification de la surface d'attaque

La première étape d'une cyberattaque implique généralement que les acteurs malveillants sondent la surface d'attaque d'entreprise connectée à Internet pour identifier tout point faible exploitable. Cela inclut souvent des éléments tels que des vulnérabilités de VPN ou de pare-feu, des erreurs de configuration ou des serveurs non corrigés. L'IA générative a largement facilité cette tâche, autrefois ardue, pour les hackers, qui peuvent simplement interroger une liste de vulnérabilités connues, associées à ces ressources.

En tirant parti des informations basées sur l'IA dans Zscaler Risk360, les entreprises peuvent instantanément afficher ces applications et ressources détectables (et à risque) — leur surface d'attaque connectée à Internet — et les rendre invisibles depuis l'Internet public derrière Zero Trust Exchange. Ceci permet de réduire instantanément et considérablement la surface d'attaque tout en empêchant les hackers d'identifier les passerelles d'entrée vulnérables.

Étape 2 : Risque de compromission

Pendant la phase de compromission, les hackers tentent d'exploiter les vulnérabilités pour obtenir un accès non autorisé aux systèmes ou applications d'entreprise. Les innovations de Zscaler en matière d'IA contribuent à maîtriser le risque de compromission, en déjouant les attaques sophistiquées sans impacter la productivité des utilisateurs.

PRÉVENTION OPTIMISÉE PAR IA DU PHISHING ET DES COMMUNICATIONS C2

Les modèles IA de Zscaler détectent les sites de phishing connus et préviennent ainsi le détournement d'informations d'identification et l'exploitation du navigateur. Ils analysent les schémas de trafic, les comportements et les malwares afin de détecter en temps réel toute infrastructure command-and-control (C2) inconnue. Ces modèles s'appuient sur une combinaison de renseignements sur les menaces, de recherches ThreatLabz et d'isolation en temps réel du navigateur pour détecter les sites suspects. Ainsi, les entreprises détectent de manière encore plus efficace les nouvelles attaques de phishing, y compris celles générées par IA, ainsi que les domaines C2.

ANALYSE EN SANDBOX OPTIMISÉE PAR IA

Zscaler Sandbox, optimisé par IA, détecte instantanément les fichiers malveillants sans peser sur le travail des collaborateurs. Les sandbox traditionnelles obligent les utilisateurs à patienter que les fichiers soient analysés, ou à prendre la responsabilité d'assumer un risque de type patient zéro en autorisant les fichiers dès leur premier passage. Notre technologie AI Instant Verdict identifie, met en quarantaine et bloque instantanément les fichiers réputés malveillants, y compris les menaces zero-day, sans pour autant imposer d'attendre la fin de l'analyse des ces fichiers. Cette approche porte sur les menaces diffusées via des canaux chiffrés (TLS et HTTP) et d'autres protocoles de transfert de fichiers. Parallèlement, les fichiers sains sont transmis en toute sécurité et instantanément.

L'IA POUR NEUTRALISER LES MENACES WEB

Zscaler Browser Isolation, optimisé par IA, neutralise les menaces zero-day tout en garantissant que les collaborateurs peuvent accéder aux sites web nécessaires à leur travail. En pratique, le filtrage d'URL en entreprise exige souvent des critères plus granulaires que autoriser/bloquer. Les sites bloqués sont souvent sûrs et nécessaires au travail, ce qui entraîne des demandes de support inutiles. Notre solution AI Smart Isolation identifie les sites à risque et les affiche de manière cloisonnée pour l'utilisateur, en diffusant le site consulté sous forme de pixels au sein d'un environnement sécurisé et conteneurisé. Ceci permet de déjouer de manière efficace les menaces web telles que les malwares, les ransomwares, le phishing et les téléchargements furtifs, créant ainsi une posture de sécurité web solide sans obliger les entreprises à bloquer exagérément les sites par défaut.

Étape 3 : Déplacement latéral

Une fois que les hackers ont pris pied au sein du réseau d'une entreprise, ils tentent de se déplacer latéralement pour accéder aux données et applications sensibles. Et dans de nombreuses entreprises, les utilisateurs permettent d'accéder à des dizaines d'applications critiques, ce qui signifie que la surface d'attaque interne est étendue.

Les capacités IA de Zscaler réduisent la portée potentielle des attaques en analysant les schémas d'accès des utilisateurs et en recommandant des politiques de segmentation intelligente des applications afin de limiter les mouvements latéraux. Il est, par exemple, courant de constater que seuls 200 des 30 000 utilisateurs qui ont accès à une application financière précise ont réellement besoin d'accéder à cette application. Zscaler peut automatiquement créer un segment applicatif qui restreint l'accès à ces 200 collaborateurs uniquement, réduisant ainsi les possibilités de déplacement latéral des acteurs malveillants de plus de 99 %.

Étape 4 : Exfiltration des données

Dans la phase finale d'une attaque, les acteurs malveillants s'efforcent d'exfiltrer les données sensibles. Zscaler utilise l'IA pour permettre aux entreprises de déployer plus rapidement des mesures de protection pour leurs données. L'identification des données avec l'aide de l'IA élimine les tâches fastidieuses de fingerprinting et de classification des données, qui pourrait autrement retarder ou empêcher le déploiement. L'IA de Zscaler identifie et classifie automatiquement toutes les données d'une entreprise, ce qui permet aux entreprises de classifier immédiatement les informations sensibles tout en configurant les politiques de prévention de perte de données (DLP) pour empêcher que ces données ne quittent jamais l'entreprise en cas d'attaque ou de violation.

Synthèse des offres IA de Zscaler

Zscaler Internet Access™ propose une protection optimisée par IA pour les utilisateurs, les appareils et les applications Web et SaaS de l'entreprise sur tous les sites, dans le cadre de Zero Trust Exchange, avec :

- **Une détection du phishing et des communications C2 optimisées par IA** contre les sites de phishing et les infrastructures C2 inconnus, grâce à une détection inline optimisée par IA proposée par Zscaler Secure Web Gateway (SWG).
- Sandboxing optimisé par IA et prévention des malwares et menaces zero-day.
- **Une politique dynamique basée sur les risques** avec une analyse continue des risques liés aux utilisateurs, aux appareils, aux applications et au contenu pour favoriser une politique dynamique de sécurité et d'accès.
- **Une segmentation optimisée par IA** avec Zscaler Private Access™, assortie de recommandations en matière de politiques d'accès automatisées pour restreindre la surface d'attaque et neutraliser les déplacements latéraux. Cette segmentation prend en compte le contexte utilisateur, les comportements, l'emplacement et les indicateurs liés aux applications privées.
- L'isolation du navigateur optimisée par IA crée un périmètre de sécurité entre les utilisateurs et les sites Web malveillants. Le contenu des sites est restitué sous la forme d'un flux d'images, pour ainsi prévenir les fuites de données et la diffusion de menaces actives.

DE PLUS, ZSCALER BLOQUE :

Les URL et les IP observées dans le cloud Zscaler, en tirant parti de sources d'informations sur les menaces, commerciales et open source, intégrées en natif. Ceci inclut les catégories d'URL à haut risque définies par les politiques et couramment utilisées pour le phishing, comme les domaines nouvellement identifiés et nouvellement activés.

Signatures IPS développées à partir de l'analyse par ThreatLabz des kits et pages de phishing.

Zscaler Risk360 fournit un framework de risque complet et pertinent qui aide les responsables de la sécurité et les dirigeants à quantifier et visualiser les cyber-risques sur l'ensemble de leur entreprise.

La protection des données avec DLP et CASB offre une classification et une protection des données optimisées par IA sur tous les canaux, y compris les terminaux, la messagerie électronique, les instances, le BYOD et le cloud.

Advanced Threat Protection neutralise tous les domaines C2 connus.

Zscaler ITDR (Identity Threat Detection & Response) atténue le risque associé aux attaques basées sur l'identité grâce à une visibilité, une surveillance des risques et une détection permanente des menaces.

Zscaler Firewall étend la protection C2 à tous les ports et protocoles, y compris les nouvelles destinations C2 émergentes.

DNS Security protège contre les attaques sur les DNS et les tentatives d'exfiltration.

Zscaler Private Access™ protège les applications en limitant le déplacement des menaces en interne, grâce à une segmentation utilisateur-application basée sur le principe du moindre privilège et à une inspection complète du trafic des applications privées.

AppProtection avec Zscaler Private Access fournit une inspection de sécurité inline et haute performance de l'ensemble du payload de l'application surveillée afin d'identifier toute menace.

Zscaler Deception™ détecte et neutralise les hackers qui tentent de se déplacer en interne ou d'élever leurs privilèges. Ces hackers sont piégés à l'aide de serveurs, applications, répertoires et des comptes d'utilisateurs factices.

Favoriser la transition vers l'IA d'entreprise : vous devez garder la main

Zscaler propose aux entreprises un levier pour favoriser l'innovation, la créativité et la productivité grâce à l'IA tout en assurant la sécurité des utilisateurs et une protection contre l'exfiltration de données. Cela permet aux entreprises d'exploiter le potentiel de transformation qu'offre l'IA afin d'accélérer leurs activités sans bloquer complètement les applications et les domaines IA.

Sécuriser l'utilisation des applications d'IA/AA

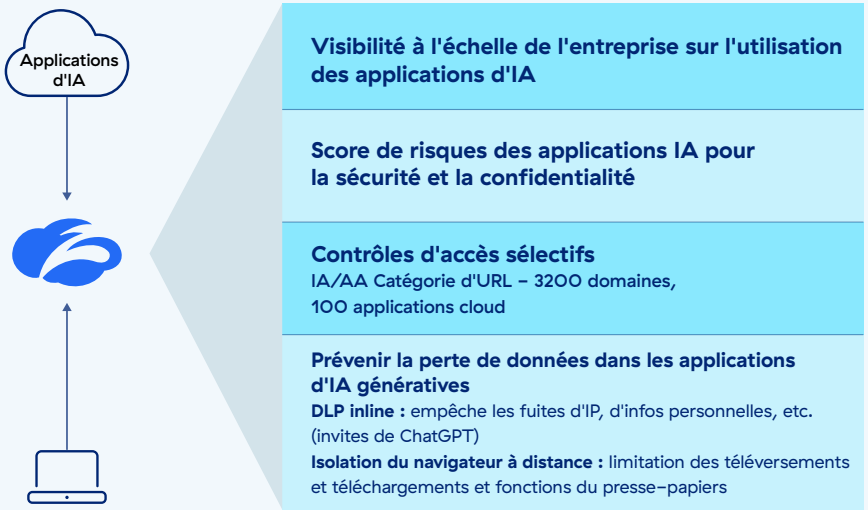


SCHÉMA 23 Zscaler propose des fonctions de sécurité innovantes et affinées, conçues pour l'IA d'entreprise

ZSCALER PERMET AUX ENTREPRISES DE :

01 **Disposer d'une visibilité complète sur l'utilisation des outils IA** Des logs détaillés offrent une visibilité complète sur la façon dont les équipes de l'entreprise utilisent l'IA, y compris les applications et les domaines visités, ainsi que les données et les invites utilisées dans des outils tels que ChatGPT.

02 **Créer des politiques flexibles pour affiner l'utilisation de l'IA** Un filtrage d'URL puissant et personnalisé pour les applications IA et AA permet aux entreprises de définir et d'appliquer des contrôles d'accès et une segmentation granulaire optimisés par IA. Les accès sont bloqués si nécessaire, tout en autorisant des accès avec des niveaux de risque acceptables, suite à une évaluation du niveau de risque de l'application IA. Les entreprises peuvent autoriser l'accès au niveau de l'entreprise, d'un département, d'une équipe ou des utilisateurs, mais aussi accorder un accès prudent, qui informe les utilisateurs sur les risques liés aux outils d'IA générative. La segmentation optimisée par IA facilite l'identification des segments d'utilisateurs appropriés pour l'accès à des applications IA particulières tout en minimisant la surface d'attaque interne associée aux outils IA.

03 **Appliquer une sécurité des données granulaires pour ChatGPT et autres applications IA** Les entreprises peuvent prévenir la fuite de données sensibles téléversées vers les applications IA grâce aux fonctionnalités granulaires de Zscaler Cloud Application pour l'IA générative. En appliquant le moteur DLP de Zscaler, les entreprises garantissent qu'aucune donnée n'est accidentellement partagée lors de l'utilisation d'un outil IA. Parallèlement, la découverte et la classification des données optimisées par IA permettent aux entreprises d'identifier et de créer des politiques de DLP pour leurs données les plus critiques, notamment leur référentiel de code d'entreprise, leurs documents financiers et juridiques, leurs données personnelles, leurs données clients, etc. Cette vidéo montre comment le moteur DLP empêche les utilisateurs de saisir des informations de carte de crédit dans ChatGPT.

04 **Déployer des fonctionnalités robustes à l'aide de l'isolation du navigateur** Zscaler Browser Isolation positionne les applications IA dans un environnement sécurisé, ajoutant une couche de protection qui permet aux utilisateurs d'envoyer des invites et des requêtes aux outils IA tout en restreignant le copier/coller, les téléchargements et les téléversements. Il devient ainsi possible de maîtriser le risque d'un partage accidentel de données sensibles avec des outils d'IA générative.

Les dirigeants d'entreprise et les responsables de la sécurité se trouvent à une croisée de chemins : ils doivent adopter l'IA pour encourager l'innovation et rester compétitifs, mais ils doivent également s'assurer que leurs données restent utilisées uniquement en entreprise, sans risque de fuite. Zscaler permet aux entreprises de réaliser sereinement cette transition, en tirant parti d'une suite complète de fonctionnalités de sécurité Zero Trust optimisées par IA qui protègent contre les attaques utilisant l'IA tout en proposant des politiques IA pertinentes et la protection des données nécessaires pour exploiter tout le potentiel de l'IA générative.

Annexe

Méthodologie d'étude de ThreatLabz

Le cloud de sécurité mondial de Zscaler traite plus de 300 000 milliards de signaux quotidiens et bloque 9 milliards de menaces et de violations de politiques par jour, avec plus de 250 000 mises à jour de sécurité quotidiennes. Analyse de 18,09 milliards de transactions IA et AA d'avril 2023 à janvier 2024 dans le cloud Zscaler, Zero Trust Exchange.

À propos de Zscaler ThreatLabz

ThreatLabZ est le laboratoire de recherche en sécurité de Zscaler. Cette équipe experte est responsable de la traque des nouvelles menaces et s'assure de la parfaite protection des milliers d'entreprises qui utilisent la plateforme mondiale Zscaler. Au-delà des recherches sur les malwares et des analyses comportementales, l'équipe ThreatLabZ s'investit dans la recherche et le développement de nouveaux prototypes qui assurent une protection avancée contre les menaces sur la plateforme Zscaler.
Elle mène régulièrement des audits de sécurité interne pour s'assurer que les produits et l'infrastructure de Zscaler répondent aux normes de conformité de sécurité. ThreatLabZ publie régulièrement des analyses approfondies sur les menaces nouvelles et existantes sur son portail, research.zscaler.fr.

Explorez votre monde, en toute sécurité.

À propos de Zscaler
Zscaler (NASDAQ : ZS) accélère la transformation digitale et permet à ses clients de gagner en agilité, productivité, résilience et sécurité. La plateforme Zero Trust Exchange™ de Zscaler protège des milliers de clients contre les cyberattaques et les pertes des données en connectant de manière sécurisée les utilisateurs, les dispositifs et les applications, quelle que soit leur localisation. Adossé à un écosystème de plus de 150 data centers dans le monde, Zero Trust Exchange, basé sur le SASE, est la plus vaste plateforme de sécurité cloud inline au monde. Pour en savoir plus, rendez-vous sur **www.zscaler.fr.**

+1 408 533 0288 Zscaler, Inc. (siège) • 120 Holger Way • San Jose, CA 95134