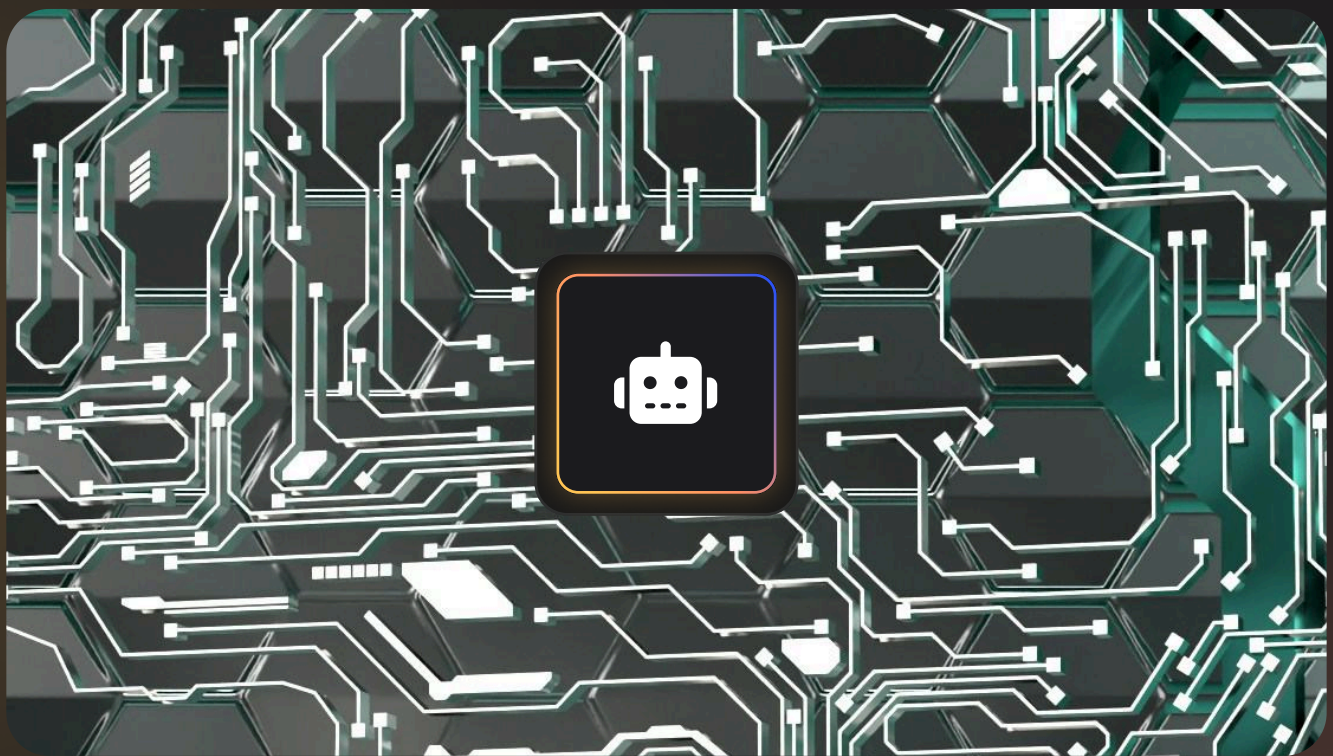
LIVRE BLANC

Cybersécurité à l'ère de l'IA

Entre menaces accrues et défenses augmentées



MAILINBLACK

Contexte

L'IA redessine le paysage des cybermenaces

L'intelligence artificielle (IA) a profondément transformé la sécurité numérique. En quelques mois, des modèles génératifs comme ChatGPT ont fait exploser le volume et la crédibilité des attaques par ingénierie sociale :

41 fois
c'est le nombre par lequel a été multiplié les campagnes de phishing entre fin 2022 et 2024

54%
c'est le taux de clic moyen sur les emails frauduleux générés par IA, contre 12% pour les tentatives manuelles

La quasi-totalité des **organisations** sont désormais concernées :

94%
des organisations ont subi au moins une attaque de phishing en 2024

67,4%
des attaques observées utilisent déjà des outils d'IA

Cette menace ne touche pas uniquement les grandes entreprises. Les **petites et moyennes entreprises (PME)** sont particulièrement exposées :

46%
des brèches de sécurité recensées impliquent des sociétés de moins de 1000 employés

350%
pourcentage d'attaques que subissent les collaborateurs des PME en comparaison avec ceux des grandes organisations

Pourtant plus d'un dirigeant de PME sur deux n'a mis en place aucune mesure de cybersécurité, souvent parce qu'il se croit « trop petit pour être attaqué ».

Dans ce contexte, les nouveaux outils d'IA constituent à la fois des cyberarmes redoutables et des solutions de défense indispensables.

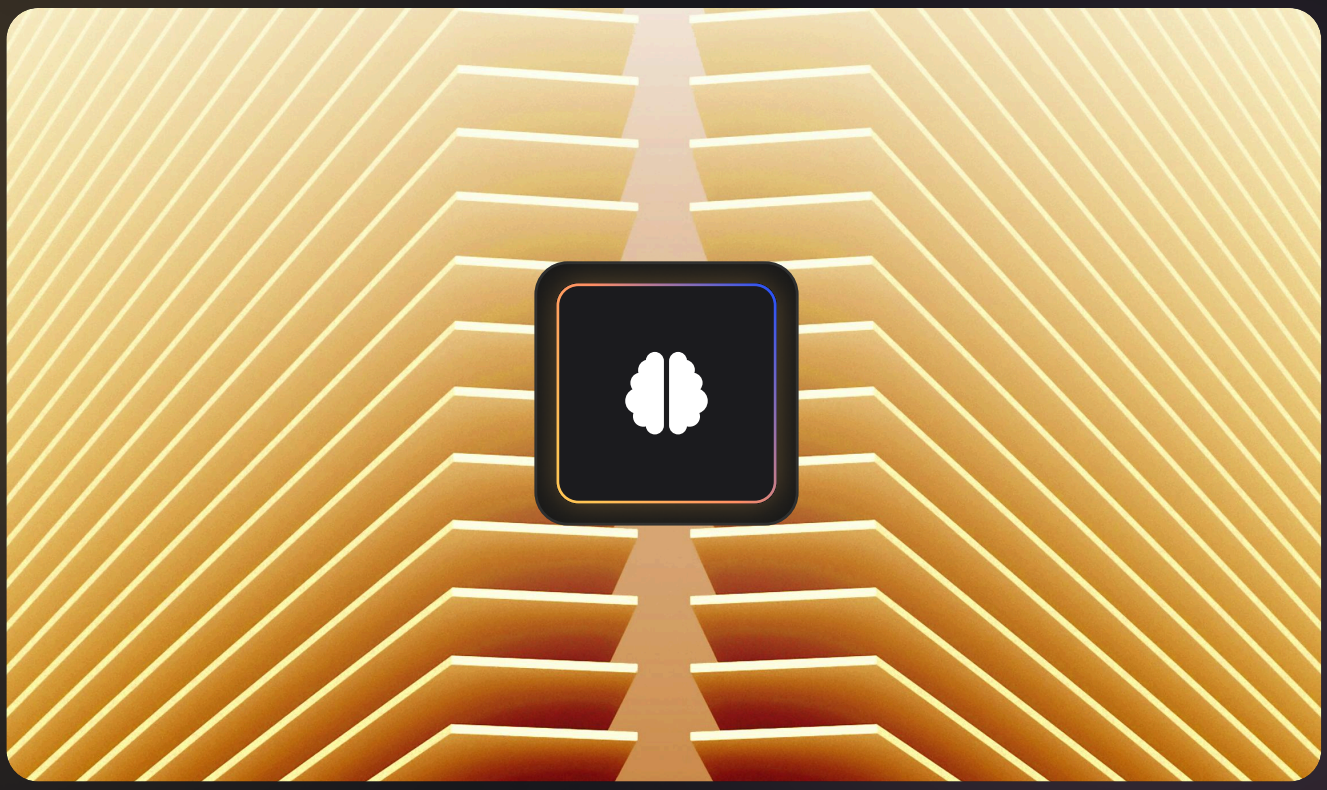
Dans ce livre blanc, nous faisons le point sur **l'usage de l'IA par les attaquants et par les défenseurs** en cybersécurité. Nous illustrons comment des attaques sophistiquées – comme le phishing vocal par deepfake ou le spear-phishing assisté par IA exploitent ces nouvelles technologies, et comment les outils défensifs s'adaptent.

Enfin, nous présenterons comment Mailinblack intègre l'IA dans ses solutions afin d'aider les entreprises, notamment les PME, à faire face à ces menaces de nouvelle génération.

Sommaire

1

L’IA au service des attaquants : des cyberarmes « intelligentes » redoutables



L'IA offre aux pirates informatiques un arsenal inédit pour affiner et amplifier leurs attaques. L'attaquant d'aujourd'hui peut utiliser des modèles génératifs pour créer des contenus malveillants extrêmement convaincants, à grande échelle et à moindre coût.

Voici les principaux usages de l'IA côté cybercriminels :

1.1 Phishing nouvelle génération

Les modèles de langage permettent aujourd'hui de rédiger des courriels frauduleux presque parfaits. Leur grammaire est impeccable et le style imite celui de la marque ou du dirigeant ciblé. Cette qualité renforce considérablement l'efficacité des hameçonnages : selon SlashNext, les campagnes de phishing générées par IA présentent un taux de réussite environ quatre fois supérieur à celui des tentatives classiques. L'IA permet en plus d'adapter dynamiquement le contenu en fonction du profil de la cible ou de l'actualité, rendant le leurre encore plus crédible.

En outre, les campagnes automatisées par IA reviennent beaucoup moins cher et sont plus rapides à déployer.

95%

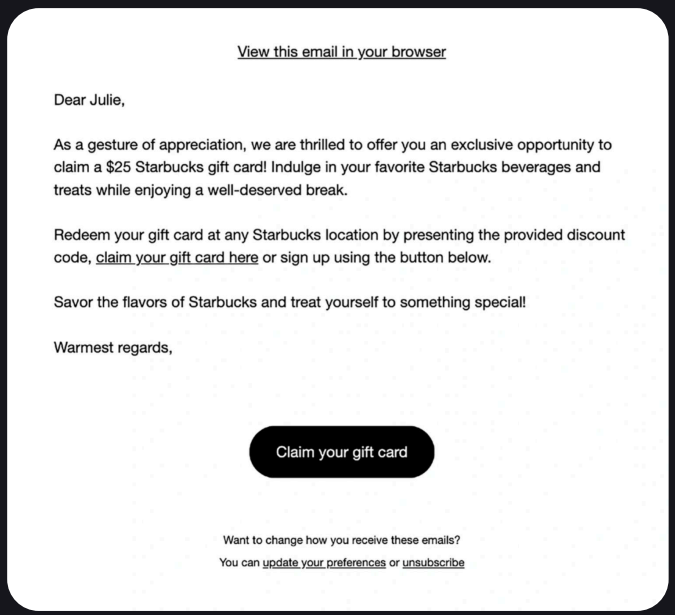
de coût en moins peuvent suffire à atteindre un taux de succès équivalent grâce à l'automatisation IA, selon une recherche de Harvard

Enfin, la génération d'images par IA permet de fabriquer de faux logos ou faux documents parfaitement ressemblants pour renforcer l'arnaque.

EXEMPLE

Voici un email de phishing généré par une IA, proposant une fausse carte-cadeau dans un style professionnel crédible. Ces messages factices dépourvus de fautes de langue sont bien plus difficiles à repérer pour les employés non vigilants (cf. photo à droite)

Source : **Exploding Topics**



View this email in your browser

Dear Julie,

As a gesture of appreciation, we are thrilled to offer you an exclusive opportunity to claim a $25 Starbucks gift card! Indulge in your favorite Starbucks beverages and treats while enjoying a well-deserved break.

Redeem your gift card at any Starbucks location by presenting the provided discount code, claim your gift card here or sign up using the button below.

Savor the flavors of Starbucks and treat yourself to something special!

Warmest regards,

Claim your gift card

Want to change how you receive these emails?
You can update your preferences or unsubscribe

1.2 Spear-phishing automatisé et attaques personnalisées

L'IA aide aussi à cibler individuellement les victimes. En analysant les données publiques (réseaux sociaux, organigrammes, communications antérieures), un attaquant peut personnaliser ses messages avec une précision chirurgicale. Par exemple, l'IA peut imiter la tournure de phrase du PDG dans un email visant la comptabilité.

67,4 %
des attaques observées **utilisent une forme d'IA** pour peaufiner la grammaire et copier les habitudes d'écriture des cibles

Les assaillants peuvent ainsi se faire passer pour un collègue ou un supérieur hiérarchique de façon très crédible, augmentant drastiquement les chances que la victime obéisse aux instructions malveillantes. On parle de *Business Email Compromise* (BEC) assisté par IA, où l'objectif est souvent de convaincre un employé d'effectuer un virement frauduleux ou de divulguer des accès sensibles.

1.3 Attaques polymorphes et malwares adaptatifs

Un autre apport inquiétant de l'IA est la création de malwares polymorphes, c'est-à-dire capables de changer de forme à chaque infection. En s'aidant du machine learning, les pirates génèrent automatiquement des variantes uniques de leurs virus ou ransomwares pour chaque cible, ce qui rend inefficaces les méthodes de détection traditionnelles. À chaque itération, le code malveillant est légèrement modifié par l'IA pour éviter d'être reconnu par les antivirus traditionnels.

Les attaques deviennent mouvantes et furtives, compliquant la tâche des défenseurs.

INFO

Des outils d'IA permettent de trouver plus facilement des vulnérabilités inédites dans les systèmes ou logiciels, ouvrant la voie à des attaques "zero-day" (inconnues à ce jour) plus nombreuses.

En abaissant la barrière de compétence technique, l'IA permet même à des cybercriminels peu expérimentés de lancer des attaques sophistiquées, simplement en décrivant ce qu'ils veulent faire.

1.4 Démocratisation des cyberattaques

L'IA a transformé la cybercriminalité en un véritable service en ligne. Sur le darknet, on voit apparaître des IA hostiles proposées par abonnement, comme *WormGPT* ou *FraudGPT*, des clones de ChatGPT débridés et entraînés pour faciliter les activités criminelles. Par exemple, *FraudGPT* était proposé à partir de 200 $ par mois sur des forums underground.

INFO

Ces plateformes génèrent sur demande du code malveillant, des mails d'hameçonnage ultra-ciblés ou même des voix synthétiques pour des arnaques téléphoniques. Elles rencontrent un succès inquiétant : WormGPT a produit des e-mails qualifiés de « remarquablement persuasifs et stratégiquement astucieux » par des experts en sécurité.

+ 3 000 ventes

que FraudGPT aurait revendiqué auprès de cybercriminels de tous niveaux, en quelques mois

Cela signifie qu'aujourd'hui, n'importe quel individu malintentionné avec une simple carte bancaire peut accéder à des capacités offensives dignes d'un groupe de hackers chevronnés. Le crime numérique s'industrialise via l'IA, mettant des outils redoutables entre toutes les mains.

1.5 Phishing vocal, deefakes audio/vidéo et arnaques multimédias

Les progrès de l'IA ne se limitent pas aux textes. La synthèse vocale et les *deepfakes* (trucages audio/vidéo hyperréalistes) offrent aux attaquants de nouveaux angles d'attaque.

On voit émerger le vishing (phishing vocal), où les escrocs vous appellent en imitant parfaitement la voix d'une personne de confiance (votre PDG, un partenaire, un proche). Cette technique a explosé récemment.

+442%

d'attaques par vishing en seulement six mois (seconde moitié de 2024) stimulées par les outils de clonage vocal IA

3 secondes

d'enregistrement audio d'une personne suffisent pour qu'un logiciel génère une voix synthétique imitant cette personne avec **85%** de similarité

Par exemple, des escrocs ont utilisé un deepfake lors d'une visioconférence pour se faire passer pour le directeur financier d'une entreprise : convaincu d'avoir son patron en face de lui, un employé a validé des transferts de fonds pour un montant de 25 millions de dollars vers les comptes des fraudeurs.

MGM RESORTS INTERNATIONAL™

2023

Une attaque de ce type a piégé un employé de MGM Resorts et conduit à un incident cyber coûtant 100 millions de dollars à l'entreprise.

INFO

Même les gouvernements n'y échappent pas : en 2025 **le FBI a confirmé** que des malfaiteurs avaient utilisé des **messages audio générés par IA** pour se faire passer pour de **hauts responsables officiels américains**.
Ces deepfakes vocaux ou vidéo sont extrêmement difficiles à repérer à l'oreille nue, ce qui leur donne un taux de réussite très élevé.

1.6 Automatisation et vitesse d'exécution

L'IA permet enfin aux attaquants d'opérer à des vitesses et volumes jamais vus. Là où un humain mettrait du temps à concevoir un site de phishing ou à envoyer des messages uniques à des milliers de cibles, l'IA peut tout faire quasi instantanément.

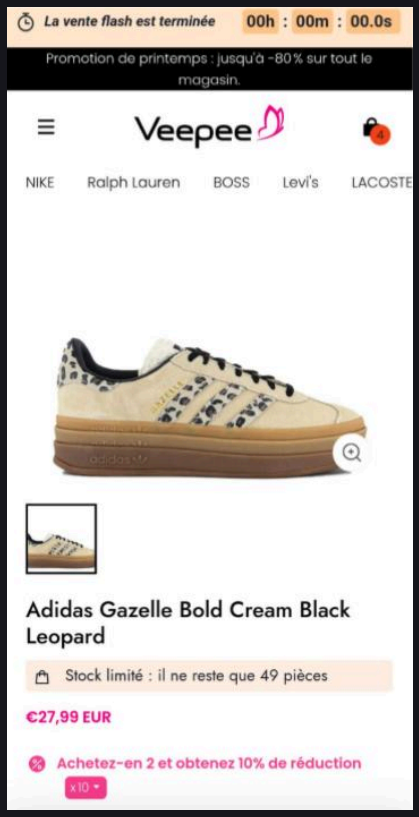
30 secondes
suffisent pour créer un site web de phishing complet à l'aide d'un outil génératif, ont montré des chercheurs

okta

2025

Okta, spécialiste de la gestion des identités, a récemment révélé qu'un outil de développement IA (nommé v0) a été détourné par des hackers pour générer automatiquement de faux sites de connexion, dont une copie conforme de la page de login d'Okta elle-même. Cette automatisation permet d'essaimer des pièges numériques à grande échelle en un temps record.

On note par exemple la fausse campagne de promotion Veepee (cf. photo à droite).



De plus, les modèles d'IA peuvent traduire instantanément des contenus d'hameçonnage dans de multiples langues parfaites, supprimant la barrière linguistique pour les cybercriminels. L'effet est clair :

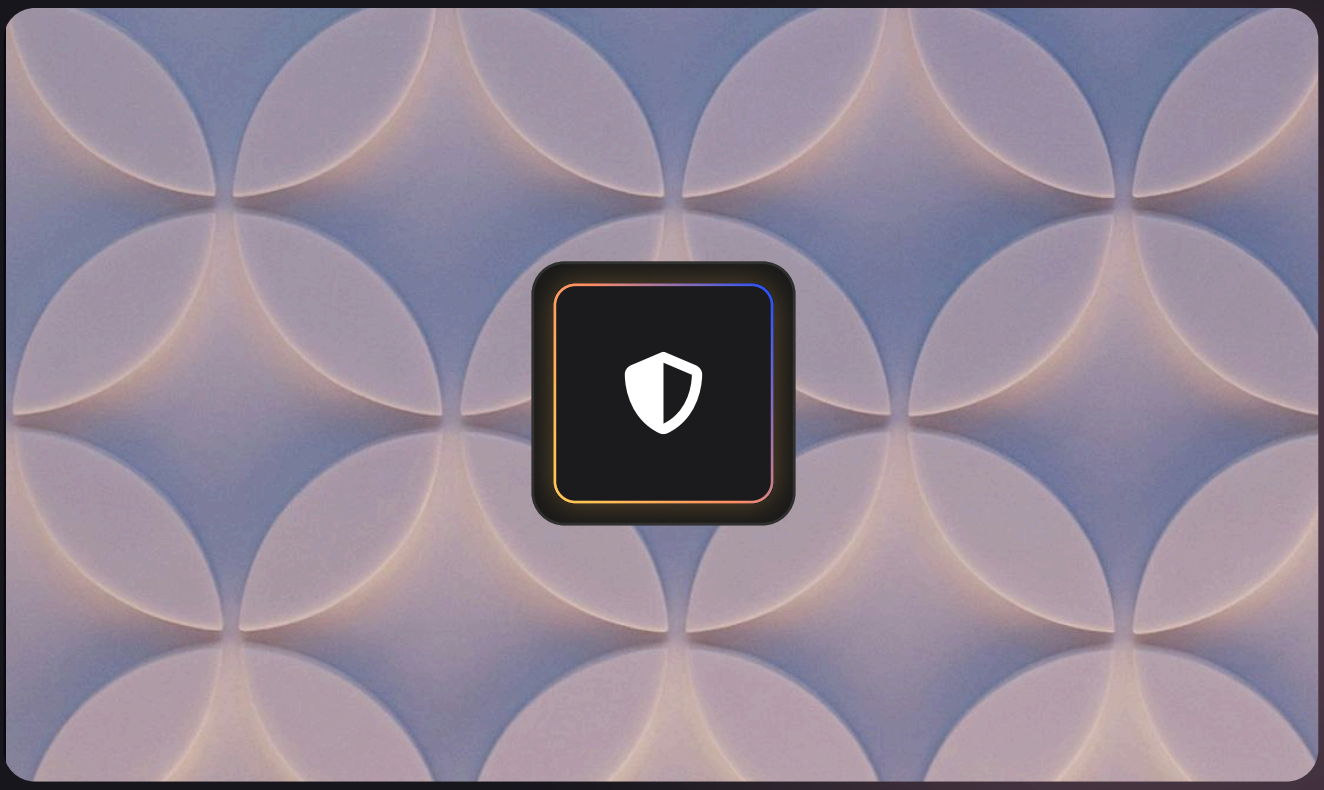
87%
des organisations mondiales ont fait face
à au moins une cyberattaque exploitant l'IA en 2023

Les pirates innovent désormais plus vite que les défenseurs ne peuvent réagir, enchaînant les attaques comme un jeu sans fin de «*whack-a-mole*» (*tape-taupe*) numérique où chaque taupe éliminée est remplacée aussitôt par une autre plus sophistiquée.

Face à cette avalanche de menaces dopées à l'intelligence artificielle, le risque cyber n'a jamais été aussi élevé. Cependant, l'IA n'est pas que l'alliée des attaquants. Heureusement, elle fournit aussi aux défenseurs de nouvelles armes pour renforcer la sécurité.

2

L’IA au service des défenseurs : vers une cyberdéfense augmentée



Si l'IA donne du fil à retordre aux experts en sécurité, elle peut aussi leur apporter des solutions inédites pour contrer ces menaces.

Du côté de la défense, l'IA est exploitée pour détecter plus finement les attaques, réagir plus vite et soulager les équipes humaines des tâches pénibles.

Voici comment les professionnels de la cybersécurité mettent l'IA à profit :

2.1 Détection prédictive des menaces

Les systèmes de sécurité traditionnels (antivirus, pare-feu) s'appuient souvent sur des bases de signatures ou des règles connues. Cela devient insuffisant face à des attaques inédites ou polymorphes.

L'apprentissage automatique (machine learning) permet d'analyser en temps réel des volumes gigantesques de données et de repérer des anomalies subtiles.

EXEMPLE

Des algorithmes IA scrutent le trafic réseau et les logs d'activité pour détecter des comportements suspects (connexion inhabituelle, transfert de données anormal, pièce jointe malveillante déguisée).

La force de l'IA est de pouvoir corréler des milliers d'indices faibles et d'identifier des attaques naissantes avant même qu'elles ne soient répertoriées.

On parle parfois de détection « prédictive » ou de systèmes UEBA (analyse du comportement des utilisateurs) dopés à l'IA, capables de dire :

> *"Attention, cet employé n'agit pas comme d'habitude, il se pourrait que son compte* soit compromis. "

Cette approche proactive permet de détecter par exemple des tentatives de phishing avant que l'utilisateur ne clique, ou de repérer un malware inconnu parce qu'il se comporte de façon anormale dans le système.

2.2 Réponse automatisée et orchestrée

L'IA est également utilisée pour accélérer la réaction face à une intrusion avérée. Dans les centres opérationnels de sécurité (SOC), on voit émerger des outils d'automation et d'orchestration (SOAR) boostés à l'IA.

Concrètement, dès qu'une alerte critique est déclenchée, une intelligence logicielle peut enclencher automatiquement une série d'actions de réponse :

- isoler un poste infecté du réseau
- bloquer une adresse expéditeur suspecte sur la messagerie
- désactiver temporairement les accès d'un compte compromis etc...

INFO

Ces réponses automatisées se déroulent en quelques secondes ou minutes, là où une intervention manuelle aurait peut-être pris des heures, limitant ainsi drastiquement les dégâts potentiels.

L'IA peut aussi générer des rapports d'incident synthétiques pour aider les analystes humains à comprendre rapidement ce qui s'est passé et quelles mesures correctives prendre. Pour les PME qui n'ont pas d'équipe dédiée 24/7, ces automatisations intelligentes jouent un rôle de « cyber vigie » réactive en permanence.

2.3 Filtres de contenu et analyse avancée

De nombreux éditeurs intègrent désormais l'IA dans les filtres anti-spam, anti-phishing et antivirus.

EXEMPLE

Des modèles de deep learning analysent la sémantique des emails et des liens URL pour déterminer s'ils sont légitimes ou malveillants. Contrairement aux filtres statiques qui se basent sur des listes noires d'adresses ou de mots-clés, une IA entraînée peut comprendre le contexte d'un message et déjouer des variantes inconnues d'attaques.

Les bénéfices sont concrets : les meilleures solutions sur le marché interceptent aujourd'hui la quasi-totalité des pourriels et courriels malveillants avant qu'ils n'atteignent la boîte de réception.

78%
des liens malveillants sont anticipés par l'analyse prédictive des URLs par IA avant même qu'ils ne causent du tort

Couplée à des bases de menace constamment enrichies, l'IA peut vérifier en temps réel la dangerosité d'une pièce jointe ou d'un fichier entrant, même s'il s'agit d'un malware inédit. Tout cela se passe en arrière-plan, sans intervention humaine, afin d'alléger la charge sur les équipes informatiques.

En évitant aux employés d'être exposés à l'énorme volume de spams (qui représente plus de 75 % du trafic email global), ces filtres intelligents augmentent à la fois la sécurité et la productivité (moins de temps perdu à trier les messages indésirables).

2.4 Authentification renforcée et détection de fraude

L'IA aide également à sécuriser l'accès aux systèmes. Par exemple, des algorithmes de reconnaissance de comportement (vitesse de frappe au clavier, façon de déplacer la souris) peuvent servir de biométrie comportementale pour détecter qu'un utilisateur n'est peut-être pas légitime, même s'il a saisi le bon mot de passe.

INFO

En finance, on utilise des IA pour repérer des transactions bancaires frauduleuses en analysant des schémas d'utilisation (lieux, montants, horaires inhabituels).

Ces systèmes apprennent ce qui constitue la "normalité" pour un utilisateur ou une entreprise, et déclenchent des alertes ou des vérifications supplémentaires dès qu'une activité sort de ce cadre.

Dans le contexte des deepfakes, on voit même apparaître des IA détectrices de voix ou de visages synthétiques : certaines entreprises proposent déjà des services capables d'analyser un enregistrement audio ou une vidéo pour estimer s'il s'agit d'une voix humaine naturelle ou d'une voix générée artificiellement. Ces détecteurs, eux-mêmes basés sur du machine learning, comparent des milliers de caractéristiques subtiles, pour tenter de discerner le vrai du faux avec :

- l'intonation
- grain de voix
- pixels d'une image vidéo etc...

Bien que la course soit engagée entre créateurs de deepfakes et détecteurs, ces outils donnent une chance de plus aux défenseurs de démasquer les impostures avant qu'il ne soit trop tard.

2.5 Allègement de la charge pour les équipes de sécurité

Un aspect souvent sous-estimé de l'IA en cybersécurité est son rôle de copilote pour les experts humains. Les centres de sécurité font face à des flots d'alertes quotidiens, dont beaucoup sont de fausses alarmes ou des événements mineurs.

L'IA peut prioriser ces alertes en évaluant leur criticité, en agrégeant celles qui se ressemblent (réduisant le bruit), et même en proposant un diagnostic préliminaire. Cela permet aux analystes de se concentrer sur les incidents réellement importants.

De plus, les chatbots ou assistants virtuels peuvent aider à guider les employés en cas de suspicion

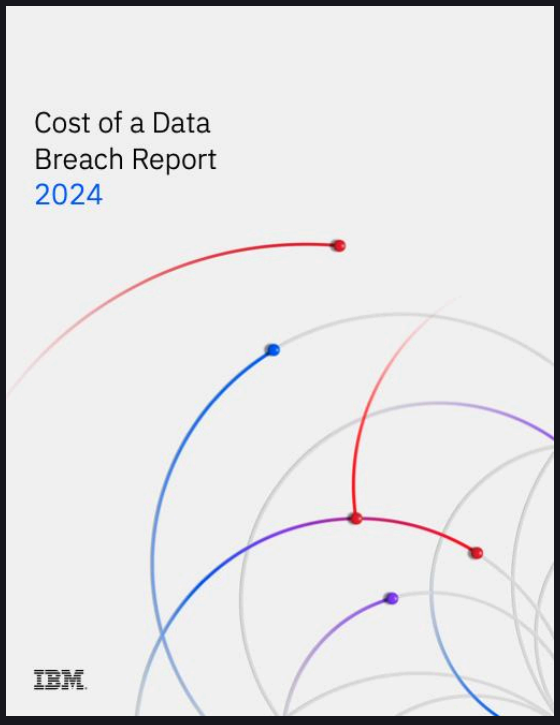
EXEMPLE

Une IA intégrée à la messagerie pourrait avertir un utilisateur en temps réel que le mail qu'il s'apprête à ouvrir est potentiellement frauduleux, ou répondre à ses questions (« Comment vérifier si ce mail est légitime ? »).

2,2 millions $

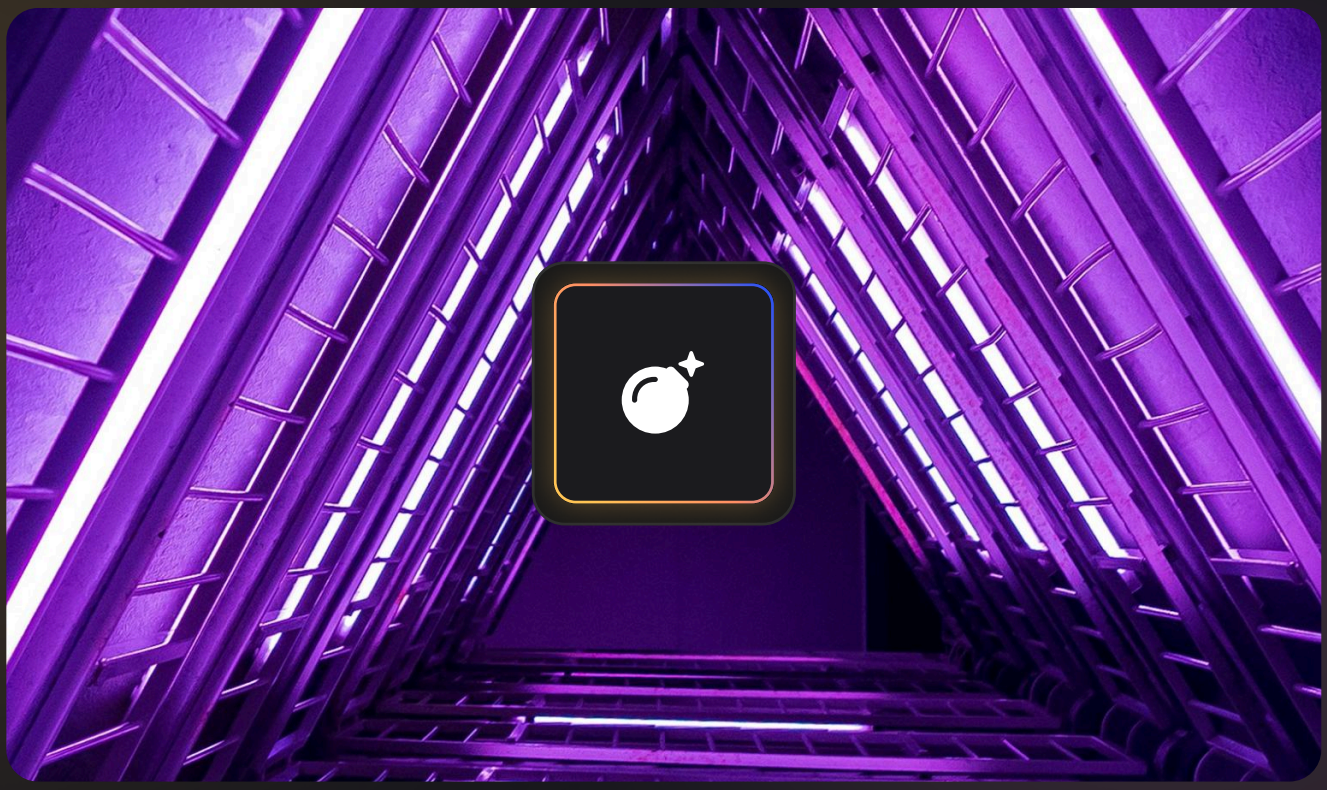
d'économies en moyenne pour les entreprises qui déploient largement l'IA et l'automatisation en sécurité, selon le **IBM Cost of a Data Breach Report 2024**

Dans un contexte de pénurie de talents en cybersécurité, l'IA agit comme un multiplicateur de force pour les équipes réduites, en prenant en charge les tâches fastidieuses 24h/24.



3

Attaques sophistiquées : illustrations de nouveaux vecteurs



Après avoir dressé le panorama général, penchons-nous sur quelques exemples concrets d'attaques sophistiquées rendues possibles (ou amplifiées) par l'intelligence artificielle. Ces cas illustrent à la fois l'ingéniosité des attaquants et les défis auxquels font face les défenseurs.

3.1 Deepfake vocal : l'arnaque téléphonique qui coûte des millions

Imaginez recevoir un appel téléphonique de votre PDG, visiblement paniqué, vous demandant un virement urgent pour régler une affaire confidentielle. La voix est la sienne, indiscernable de l'originale. Il semble pressé, insistant, et crédible. Dans le stress de l'instant, qui penserait à douter ?

C'est exactement ce qui est arrivé dans plusieurs entreprises ces dernières années.

Contexte	En réalité	Conséquences
En 2022, un employé d'une société d'ingénierie a participé à ce qu'il croyait être une réunion virtuelle avec son directeur financier et d'autres collègues.	Il s'agissait d'un deepfake vidéo en direct : chaque participant était un avatar synthétique, imitant à la perfection un véritable cadre de l'entreprise, y compris le DAF.	Convaincu par cette supercherie, l'employé a validé 25 millions de dollars de virements vers des comptes frauduleux à l'étranger.

D'autres attaques s'appuient uniquement sur le voix cloning, désormais accessible via des outils IA génératifs.

220 000€

ont été volés en 2019 à un dirigeant britannique après un appel téléphonique imitant la voix de son PDG

Depuis, cette méthode s'est perfectionnée et largement répandue. En 2023, le piratage du groupe MGM Resorts aurait également été déclenché par un appel vocal frauduleux imitant un technicien du support, ce qui a conduit à un incident estimé à 100 millions de dollars de pertes.

80%

des attaques de type "vishing" (phishing vocal) impliquent aujourd'hui l'usage d'une IA de clonage vocal, d'après plusieurs sources

Exemple européen récent : deepfake vocal en Suisse romande

David Sadigh
Fondateur de Digital Luxury Group
Genève

En février 2025, David Sadigh, a été victime d'une tentative de fraude ciblée. Des cybercriminels ont utilisé un clone vocal IA de sa voix pour envoyer des messages WhatsApp à un de ses collaborateurs.

Conséquences

Le numéro provenait de France, le profil usurpait sa photo LinkedIn, et les messages vocaux imitaient de manière bluffante son timbre et son intonation. Fort heureusement, le collaborateur ciblé a suspecté une anomalie et a pris le soin de vérifier par un autre canal, ce qui a permis de déjouer l'arnaque. (Source : Tribune de Genève)

Ce cas illustre que ces attaques ne touchent pas uniquement les grands groupes ou les entreprises américaines : les PME, les dirigeants exposés sur les réseaux, ou les structures européennes peuvent aussi être ciblés.

Ces attaques vocales visent souvent des transferts d'argent (arnaque au président), mais pas uniquement. D'autres escrocs se font passer pour des :

- banques (pour confirmer de faux virements)
- techniciens informatiques (pour obtenir des identifiants)
- recruteurs (pour soutirer des données personnelles sensibles)

Ce qui rend ces attaques particulièrement redoutables, c'est leur efficacité psychologique. La plupart des gens ne s'attendent pas à devoir douter de la voix de leurs collègues.

25%
des employés reconnaissent avoir du mal à distinguer une voix réelle d'une voix générée par IA, selon une étude menée en 2024.

6,5%
des personnes testées finissent par divulguer des informations sensibles à un faux interlocuteur vocal.

Une réponse nécessaire : procédures et vigilance humaine

Les agences comme le FBI, l'ANSSI, ou l'ENISA alertent aujourd'hui sur la montée en puissance de cette menace. Face à ces attaques, les outils de détection vocale IA commencent à émerger, mais ils restent loin d'être infaillibles, surtout en temps réel.

La meilleure protection aujourd'hui reste procédurale et comportementale :

1. **Vérification systématique hors bande**
   Rappeler la personne via un numéro déjà connu, ou demander une double confirmation écrite.
2. **Sensibilisation des équipes**
   Aux nouvelles formes de fraudes (exercices, mises en situation, tests réguliers).
3. **Réduction de la surface d'exposition publique**
   Limiter les publications vidéo/audio publiques, en particulier des dirigeants.
4. **Utilisation de solutions de filtrage et d'analyse comportementale en amont**
   Pour réduire les risques d'entrée initiale (emails, SMS, contacts suspects).

3.2 Phishing « augmenté » et sites frauduleux automatisés

Le courrier électronique reste un vecteur d'attaque majeur, et l'IA y a introduit un niveau de sophistication inédit. Les campagnes de phishing assistées par IA combinent volume, personnalisation et réalisme visuel pour piéger leurs cibles.

De même, en ce qui concerne le contenu des e-mails, les attaquants utilisent l'IA pour générer des messages non seulement bien écrits, mais aussi hautement personnalisés. On parle de spear-phishing 2.0.

Qu'est-ce que c'est ?

Le **spear-phishing 2.0** est une IA qui peut rédiger un mail se faisant passer pour un fournisseur avec qui vous traitez habituellement, en mentionnant une facture en attente, avec un ton et un style correspondant à vos échanges passés. Les liens et documents joints renvoient évidemment vers des pièges (site clone, malware, etc.).

Un autre phénomène récent est l'usage d'IA pour automatiser l'infrastructure d'attaque. Au lieu de rédiger manuellement chaque mail ou de configurer chaque site factice, tout peut être orchestré par des modèles génératifs.

EXEMPLE

Okta et Vercel, mentionné plus haut, en est l'illustration : un outil d'IA conçu à la base pour aider les développeurs web a été détourné pour bâtir des pages de connexion frauduleuses en série. Durant l'enquête, les experts ont vu les attaquants générer de nouveaux sites de phishing ciblant des utilisateurs de services de cryptomonnaie ou de Microsoft 365, à la chaîne et en très peu de temps.

Cette industrialisation des attaques rend caduques de nombreuses parades traditionnelles.

EXEMPLE

Par exemple, vérifier l'URL ou la mise en page d'un site ne suffit plus, car les contrefaçons sont quasiment parfaites. De même, filtrer les e-mails par mots-clés ou par langue n'est plus efficace, car l'IA produit des messages sur-mesure sans erreurs flagrantes.

Face à cette menace, les entreprises renforcent leur vigilance et investissent dans des solutions de sécurité plus pointues (on l'a vu, l'IA fait partie de la réponse). Il devient indispensable de combiner technologie et sensibilisation humaine : même l'email le plus convaincant peut être mis en échec si l'utilisateur a un doute et suit les procédures.

EXEMPLE

Appeler le prétendu expéditeur pour vérification, ne jamais payer la facture sur la base d'un simple email inattendu etc...

Hélas, l'ingénierie sociale couplée à l'IA exploite aussi les failles psychologiques :

- en jouant sur l'urgence
- la peur de déplaire
- la récompense (cadeau, promotion)

21 secondes

en moyenne suffisent pour que les campagnes générées par IA incitent leurs cibles à réagir après l'ouverture du mail, rendant la fenêtre de réaction extrêmement courte pour détecter l'arnaque.

3.3 Menace grandissante, défense proactive

Ces exemples (ci-dessous) illustrent bien comment l'IA a fait basculer le rapport de force en faveur des assaillants imprudents :

- deepfakes vocaux
- spear-phishing automatisé
- faux sites instantanés, etc...

Le point commun est la capacité de l'IA à produire du faux plus vrai que nature, en grande quantité, et à exploiter la confiance ou la distraction des cibles.
La menace est protéiforme :

- elle peut viser la trésorerie (fraudes aux faux ordres)
- le vol de données sensibles (vol d'identifiants via sites clones)
- la désinformation (faux contenus générés pour nuire à la réputation d'une personne ou d'une entreprise).

Cependant, mieux connaître ces nouveaux modes opératoires permet de mieux s'en prémunir. La sensibilisation des utilisateurs est plus que jamais cruciale :

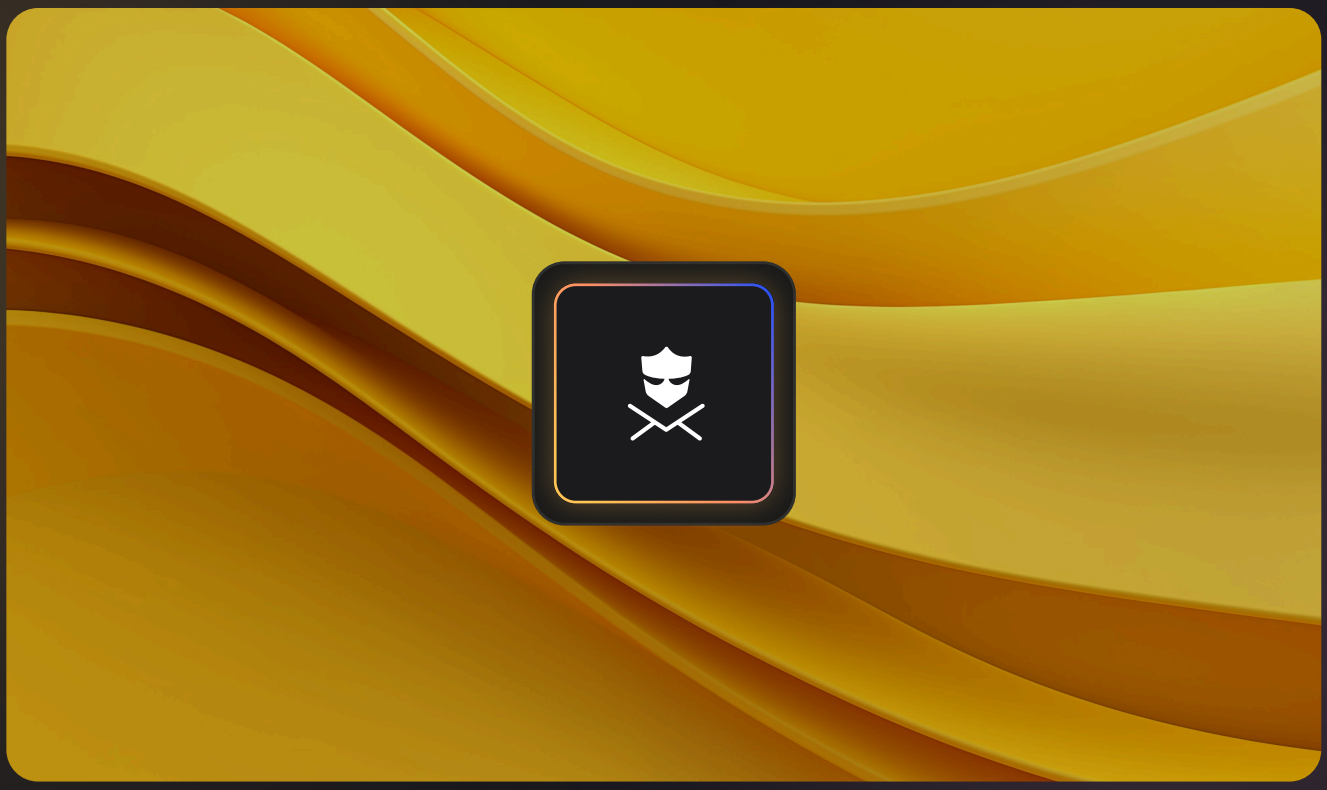
1. **Entraîner les employés à repérer les signaux faibles d'une attaque**
2. **Les encourager à la prudence**
   Par exemple, se méfier des messages trop beaux ou trop alarmants pour être vrais
3. **Instaurer une culture où le doute raisonnable est permis sans reproche**

Parallèlement, les solutions technologiques doivent être continuellement améliorées pour filtrer, analyser et bloquer ces menaces avant qu'elles n'atteignent l'utilisateur final.

C'est ce que nous allons voir dans la section suivante, en prenant l'exemple des solutions Mailinblack qui combinent IA et approche centrée humain pour contrer ces attaques de nouvelle génération.

4

Mailinblack : l’IA au service d’une cybersécurité accessible aux PME



Depuis plus de 15 ans, Mailinblack s'est donné pour mission de protéger les organisations contre les cyberattaques par email, en particulier les PME qui n'ont pas toujours les moyens de se doter d'équipes et d'outils sophistiqués. À travers ses solutions, Mailinblack intègre les dernières avancées en intelligence artificielle pour offrir une protection de niveau entreprise tout en restant simple d'utilisation et abordable.

Voici comment nos produits phares exploitent l'IA pour sécuriser vos communications et former vos collaborateurs :

4.1 Filtrage intelligent des emails

Protect

Il s'agit de notre solution de protection de la messagerie qui utilise l'IA pour bloquer les menaces avant qu'elles n'atteignent votre boîte de réception.

Concrètement, Protect analyse chaque courriel entrant grâce à des algorithmes de machine learning entraînés sur des millions de messages. Les spams et emails malveillants (phishing, ransomware, virus…) sont détectés et placés en quarantaine automatiquement, sans polluer la boîte principale.

Notre IA propriétaire effectue un classement intelligent :

1. **Les messages légitimes arrivent normalement**
2. **Les newsletters et envois automatiques sont triés à part**
3. **Tout ce qui est suspect ou connu comme nuisible est isolé**

40 minutes par jour

de productivité récupérée en moyenne grâce à la réduction du spam, ce qui permet à vos employés de ne plus perdre de temps à trier des courriers indésirables. Surtout, cela les protège des pièges : un email piégé ne peut causer de dégâts s'il n'est jamais lu

Protect utilise par exemple le deep learning pour reconnaître les tentatives de phishing même lorsqu'elles contournent les filtres traditionnels. Il en résulte un environnement de messagerie assaini et sécurisé par l'IA, où vos collaborateurs peuvent travailler sereinement.

En outre, notre solution s'insère en complément de vos protections existantes (antivirus, Microsoft 365, etc.) sans les perturber, en ajoutant une couche intelligente supplémentaire.

4.2 Analyse avancée des liens URL par IA

+80%

des cyberattaques commencent par un e-mail, dont une grande partie via des liens frauduleux, rendant crucial le déploiement d'une protection efficace dès le moment du clic

Pour cela, l'intelligence artificielle joue un rôle central.

Nous avons développé une technologie qui examine les liens contenus dans les emails en temps réel, dès qu'un utilisateur s'apprête à cliquer. Cette vérification repose sur trois niveaux d'analyse complémentaires :

1. **Analyse morphologique de l'URL**
   Pour détecter les noms de domaine suspects ou les caractères trompeurs (ex. typo-squatting)
2. **Analyse sémantique**
   Qui identifie les mots ou combinaisons de termes fréquemment utilisés dans les campagnes de phishing
3. **Analyse du contenu de la page cible**
   Pour anticiper son degré de dangerosité même si elle n'est pas encore référencée dans une base connue

Avant l'ouverture réelle de la page, le lien est redirigé de manière transparente vers nos serveurs, qui effectuent cette triple évaluation instantanée.

RÉSULTAT

La solution est capable d'**anticiper 78 % des liens malveillants** avant même qu'ils n'aient le temps de nuire.

Lien sain
L'utilisateur est redirigé vers la page prévue, sans interruption

Lien malveillant
L'accès est bloqué et une alerte pédagogique s'affiche, expliquant clairement la menace

Ce système agit comme un véritable filet de sécurité, essentiel dans un contexte où 90 % des incidents liés au phishing impliquent une erreur humaine. Une seconde d'inattention peut suffire pour compromettre un poste, un réseau ou divulguer des données sensibles.

En renforçant la protection au moment précis du clic, l'IA pallie cette fragilité humaine, sans entraver la productivité. Les collaborateurs peuvent ainsi naviguer sereinement, sachant qu'un dispositif intelligent veille discrètement à leur sécurité.

4.3 Formation et simulations d'attaques par IA

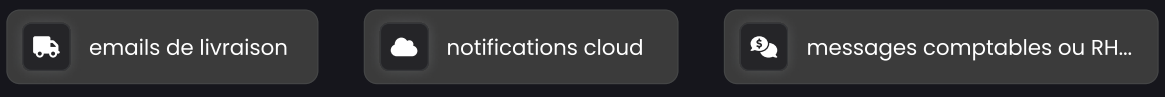
Cyber Coach

La technologie seule ne suffit pas. En cybersécurité, l'humain reste la première ligne de défense, et souvent aussi le maillon le plus vulnérable. C'est pourquoi la formation continue des collaborateurs est devenue un pilier incontournable d'une stratégie de protection efficace.

INFO

Nous avons développé un **programme de sensibilisation automatisée,** enrichi par l'intelligence artificielle, qui permet de former les équipes en conditions réelles, sans perturber leur quotidien.

Ce dispositif repose sur l'envoi régulier de fausses campagnes de phishing inoffensives, conçues pour tester la vigilance des collaborateurs tout en les formant par l'expérience. Ces simulations sont scénarisées, personnalisées et générées par IA, de façon à reproduire fidèlement les méthodes des cybercriminels, tout est pensé pour refléter les attaques actuelles :



Lorsqu'un collaborateur clique sur un lien piégé ou saisit ses identifiants, il est immédiatement redirigé vers une interface pédagogique, qui lui explique l'erreur, les signaux d'alerte qu'il aurait pu repérer, et lui propose des conseils concrets.

Ce retour à chaud est essentiel pour ancrer les bons réflexes.

Ce programme de sensibilisation est complété par des modules e-learning interactifs, eux aussi adaptés par IA selon le niveau de chaque utilisateur.

RÉSULTAT

Un véritable coaching de cybersécurité individualisé, à l'échelle de l'entreprise, sans surcharge administrative.

70%

de réduction des erreurs humaines face aux attaques de phishing en moyenne pour les entreprises ayant déployé cette approche

Un collaborateur formé devient plus alerte, plus confiant, et agit comme un rempart actif plutôt qu'un risque latent.

En combinant cette montée en compétence des équipes avec des outils technologiques intelligents de filtrage et d'analyse, les organisations se dotent d'une approche globale, alliant intelligence artificielle et intelligence humaine. Et c'est cette synergie qui fait la différence face à des menaces toujours plus sophistiquées.

Conclusion : se préparer aux cybermenaces de demain

L'essor fulgurant de l'intelligence artificielle a marqué un tournant dans la cyberguerre entre attaquants et défenseurs.

D'un côté

Les criminels disposent d'outils capables de générer des attaques hyperréalistes, massives et rapides comme jamais auparavant.

De l'autre

Les professionnels de la sécurité ont commencé à intégrer l'IA pour reprendre l'avantage, en détectant proactivement les signaux faibles et en automatisant la réponse aux incidents.

Cette course technologique rend le paysage mouvant : il est probable que les attaques de demain, alimentées par l'IA, prendront des formes encore inédites : peut-être des deepfakes en réalité augmentée, des malwares auto-apprenants plus furtifs, ou des campagnes de désinformation ciblée pilotées par des IA toujours plus convaincantes.

Face à cela, les organisations, petites ou grandes, doivent adopter une posture de vigilance permanente et d'amélioration continue. Cela passe par :

- veille sur les nouvelles menaces
- mise à jour régulière des outils de protection
- l'entraînement fréquent des équipes et le partage d'informations au sein de la communauté (CERT,etc...)

L'IA, finalement, n'est ni bonne ni mauvaise en soi : tout dépend de l'usage qu'on en fait. Il appartient aux défenseurs de l'utiliser de manière éthique et responsable pour contrer ceux qui en font un usage malveillant. En combinant innovation technologique et sensibilisation humaine, il est possible de significativement réduire les risques et d'aborder sereinement cette nouvelle ère.

Chez Mailinblack, nous restons engagés à développer des solutions toujours plus performantes et à vous accompagner dans cette démarche. La sécurité informatique est un voyage plutôt qu'une destination : à l'ère de l'IA, ce voyage s'accélère, mais avec les bons partenaires et les bonnes pratiques, vous pouvez garder une longueur d'avance. Restez prudents, outillez-vous intelligemment, et vous ferez de l'IA votre alliée plutôt que votre ennemi.

Sources

- **Egress**
  Statistiques de phishing incontournables pour 2025 - lien

- **DASHLANE**
  Un nouveau chapitre de la cybercriminalité : comment l'IA alimente la sophistication du phishing - lien

- **SOCRadar**
  Phishing en 2024 - lien

- **CybelAngel**
  Le phishing alimenté par l'IA est en hausse - lien

- **Infosecurity Magazine**
  Les marchés du dark web offrent un nouvel outil FraudGPT AI - lien

- **MOUSER Electronics**
  Comprendre les dangers posés par le phishing génératif par l'IA - lien

- **EXPLODING TOPICS**
  7 tendances en matière de cybersécurité de l'IA pour le paysage de la cybercriminalité de 2025 - lien

- **Jericho Security**
  Phishing Deepfake : la menace d'ingénierie sociale alimentée par l'IA mettant les RSSI en alerte en 2025 - lien

- **Jumpcloud**
  Le phising vocal augmente. Voici comment le combattre - lien

- **CFO DIVE**
  Les escrocs siphent 25 millions de dollars de la société d'ingénierie Arup - lien

- **AXIOS**
  Les pirates informatiques abusent de l'outil d'IA génératif pour créer des sites de phishing en 30 secondes - lien

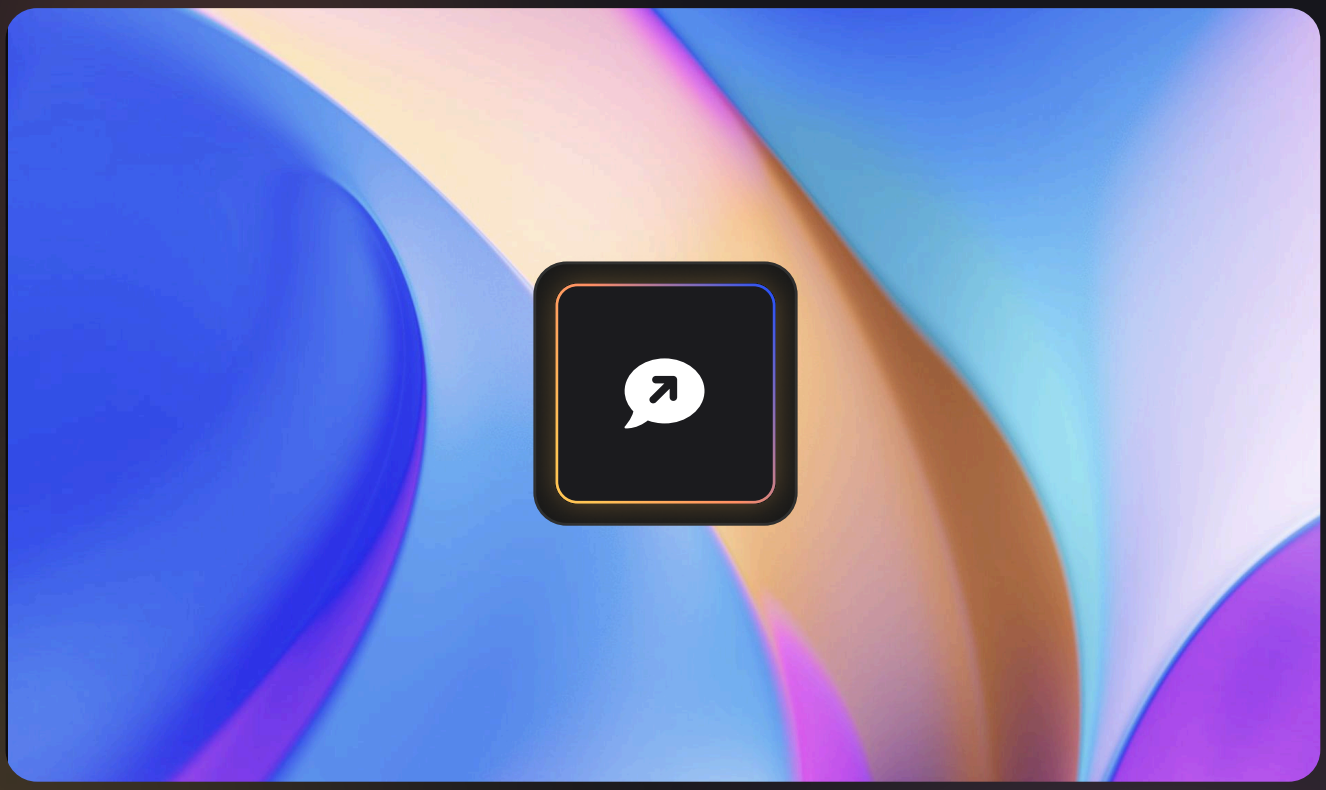
- **BUSINESS INSIDER**
  Les entreprises se démènent pour se protéger contre les menaces croissantes liées à l'IA - lien

- **ActuIA**
  Mailinblack intègre l'intelligence artificielle à Secure Link pour une navigation plus sereine - lien

- **TdG**
  Des escrocs ont cloné la voix d'un entrepreneur grâce à l'IA - lien

- **IBM**
  Cost of a Data Breach Report PDF - lien

Contactez-nous



contact@mailinblack.com
+33 (0)4 88 60 07 80
www.mailinblack.com
4 place Sadi Carnot, 13002 Marseille

MAILINBLACK