L'intelligence artificielle, nouvelle arme des cybercriminels

Deepfakes, phishing automatisé et IA malveillantes : comprendre les menaces et s'en protéger

2025 - 2026 • Veille technologique • Cybersécurité

La veille informatique permet de suivre l'évolution des technologies numériques et de rester informé des nouveautés dans un domaine en perpétuelle mutation. En cybersécurité, cette veille est devenue indispensable : les menaces évoluent aussi vite que les technologies censées nous protéger. Ce dossier analyse un phénomène majeur de ces dernières années : le détournement de l'intelligence artificielle par les cybercriminels pour créer des attaques d'un genre nouveau.

PROBLÉMATIQUE

En quoi l'intelligence artificielle représente-t-elle un nouveau danger pour la cybersécurité des entreprises et des particuliers ?

Contexte : l'IA, outil à double tranchant

L'intelligence artificielle s'est imposée comme une technologie incontournable. Dans le domaine de la cybersécurité, elle est utilisée pour détecter les intrusions, analyser les comportements suspects et automatiser la réponse aux incidents. Les solutions de sécurité modernes s'appuient massivement sur le machine learning pour identifier les menaces.

Mais cette même technologie est désormais exploitée par les attaquants. Depuis l'explosion de l'IA générative en 2022-2023 (ChatGPT, Midjourney, ElevenLabs), les cybercriminels disposent d'outils puissants pour automatiser leurs attaques, générer des contenus frauduleux ultra-réalistes et industrialiser des techniques qui nécessitaient auparavant des compétences pointues.

Le résultat : des attaques plus rapides, plus personnalisées, plus difficiles à détecter, et accessibles à un nombre croissant d'acteurs malveillants.

Méthodologie de veille

Cette veille s'appuie sur le croisement de plusieurs types de sources pour garantir la fiabilité des informations :

- Rapports d'éditeurs de sécurité (Kaspersky, ESET, Trend Micro, CrowdStrike)
- Publications institutionnelles (ANSSI, CNIL, Europol, FBI)
- Sites d'actualité spécialisés (BleepingComputer, The Hacker News, Krebs on Security)
- Veille sur les réseaux sociaux professionnels et forums de sécurité
- Analyse des outils et techniques émergents (WormGPT, DeepFaceLive, etc.)

MENACE N°1

Deepfakes : quand l'IA falsifie la réalité

Les **deepfakes** sont des contenus synthétiques (vidéo, audio, image) générés par intelligence artificielle pour imiter une personne réelle. Cette technologie, initialement cantonnée aux laboratoires de recherche, est aujourd'hui accessible à n'importe qui grâce à des outils grand public.

En cybersécurité, les deepfakes posent un problème majeur : ils permettent l'**usurpation d'identité** à un niveau de réalisme jamais atteint. Un attaquant peut désormais cloner la voix d'un dirigeant avec quelques secondes d'enregistrement, ou superposer son visage sur celui d'une autre personne en temps réel lors d'un appel vidéo.

Types de deepfakes utilisés par les cybercriminels

- → **Deepfake vocal** : clonage de voix pour des appels frauduleux (arnaque au président)
- → **Deepfake vidéo** : usurpation d'identité en visioconférence
- → **Deepfake temps réel** : transformation faciale en direct via webcam
- → **Deepfake audio-vidéo combiné** : usurpation complète voix + visage

CAS ARUP - HONG KONG (2024)

Un employé du cabinet d'ingénierie britannique Arup a participé à ce qu'il pensait être une visioconférence avec plusieurs dirigeants de l'entreprise. Tous étaient des deepfakes. Convaincu par le réalisme de la scène, il a validé des virements vers des comptes frauduleux. **Préjudice : 25 millions de dollars.** Ce cas illustre la sophistication atteinte par ces attaques.

ARNAQUES AU PRÉSIDENT MULTIPLIÉES

Depuis 2023, les cas de "CEO fraud" utilisant des deepfakes vocaux se multiplient. Un employé reçoit un appel de son "patron" lui demandant un virement urgent. La voix est parfaitement imitée grâce à des outils comme ElevenLabs. En 2024, **le FBI a alerté sur l'augmentation de 400% de ces fraudes** utilisant l'IA vocale.

Outils de deepfake accessibles au public

- → **ElevenLabs** : clonage vocal en quelques minutes, très réaliste
- → **DeepFaceLive** : transformation faciale en temps réel (open source)
- → **Avatarify** : animation de portraits fixes
- → **FaceSwap** : échange de visages dans les vidéos
- → **HeyGen / Synthesia** : création de vidéos avec avatars IA

MENACE N°2

Phishing assisté par IA : la fin des fautes d'orthographe

Le phishing (hameçonnage) reste la première cause de compromission des systèmes d'information. Traditionnellement, ces attaques étaient reconnaissables à leurs maladresses : fautes d'orthographe, formulations étranges, mises en page approximatives. L'IA a changé la donne.

Grâce aux modèles de langage (LLM), les cybercriminels génèrent désormais des emails parfaitement rédigés, personnalisés selon le profil de la victime. L'IA peut analyser les publications LinkedIn d'une cible pour créer un message sur mesure, mentionnant des projets réels, des collègues existants, des

événements récents.

+135%	91%	60€/mois
Augmentation des attaques phishing en 2024	Des cyberattaques commencent par un email	Prix d'accès à WormGPT sur le dark web

Ce que l'IA apporte au phishing

- → Rédaction parfaite dans toutes les langues, sans fautes
- → Personnalisation automatique basée sur les réseaux sociaux de la cible
- → Génération de faux sites web réalistes en quelques clics
- → Analyse automatisée des données volées pour cibler les victimes à fort potentiel
- → Création de scénarios d'urgence crédibles et contextualisés
- → Traduction instantanée pour des campagnes internationales

WORMGPT ET FRAUDGPT : LES IA DU DARK WEB

Apparus en 2023, ces modèles de langage sont spécialement conçus pour les activités malveillantes. Contrairement à ChatGPT qui refuse de générer du contenu dangereux, **WormGPT n'a aucune restriction éthique**. Il génère des emails de phishing, des scripts d'attaque, du code malveillant. Ces outils sont vendus par abonnement sur les forums cybercriminels et démocratisent l'accès à des attaques sophistiquées.

Le spear phishing (phishing ciblé) atteint désormais un niveau de personnalisation qui le rend quasi indétectable. Un email peut mentionner une réunion réelle à laquelle la victime a participé, un projet en cours dans son entreprise, ou un contact commun. Cette contextualisation augmente drastiquement le taux de réussite des attaques.

MENACE N°3

Autres usages malveillants de l'IA

Au-delà des deepfakes et du phishing, l'IA est exploitée par les cybercriminels dans de nombreux autres domaines, démultipliant leurs capacités d'attaque.

Recherche automatisée de vulnérabilités

- → L'IA analyse le code source pour identifier des failles de sécurité
- → Scan automatisé de milliers de sites web pour trouver des cibles vulnérables
- → Génération automatique d'exploits à partir de CVE publiées
- → Fuzzing intelligent pour découvrir des bugs exploitables

Malwares polymorphes

- → L'IA génère des variantes de malwares pour échapper aux antivirus
- → Code malveillant qui se réécrit automatiquement à chaque exécution
- → Obfuscation intelligente rendant l'analyse difficile
- → Adaptation en temps réel aux défenses détectées

Ingénierie sociale augmentée

- → Chatbots malveillants capables de maintenir des conversations réalistes

- → Analyse comportementale des victimes pour adapter l'approche
- → Création de faux profils sociaux ultra-crédibles
- → Manipulation psychologique assistée par IA

CAMPAGNES DE DÉSINFORMATION AUTOMATISÉES

L'IA permet de générer massivement du contenu faux : articles, commentaires, images, vidéos. Ces contenus sont ensuite diffusés par des réseaux de bots pour manipuler l'opinion publique, déstabiliser des entreprises ou influencer des marchés financiers. **En 2024, plusieurs opérations d'influence utilisant des deepfakes de personnalités politiques ont été détectées.**

SE PROTÉGER

Comment se défendre face à ces nouvelles menaces

Face à l'utilisation croissante de l'IA par les attaquants, les entreprises et les particuliers doivent adapter leurs pratiques de sécurité. La bonne nouvelle : l'IA est également utilisée du côté défensif pour détecter ces nouvelles menaces.

Solutions technologiques

- → **Détection de deepfakes par IA** : outils analysant les vidéos pour repérer les artefacts (Microsoft Video Authenticator, Intel FakeCatcher, Sensity)
- → **Filtres anti-phishing avancés** : solutions utilisant le machine learning pour analyser le contexte des emails, pas seulement les mots-clés
- → **Authentification multi-facteurs (MFA)** : ne jamais se fier à un seul canal de vérification
- → **Analyse comportementale** : détection des anomalies dans les actions utilisateurs
- → **Watermarking des contenus** : marquage invisible des vidéos et audios officiels pour prouver leur authenticité

Bonnes pratiques organisationnelles

- → **Procédures de double validation** : toute demande sensible (virement, accès, modification) doit être confirmée par un autre canal
- → **Mots de passe de vérification** : code convenu à l'avance entre collègues pour les demandes urgentes
- → **Formation continue** : sensibiliser régulièrement les employés aux nouvelles techniques d'attaque
- → **Simulations d'attaques** : tester la vigilance des équipes avec de faux phishing
- → **Politique de méfiance constructive** : encourager les employés à vérifier plutôt qu'à faire confiance aveuglément

Réflexes individuels

- → Ne jamais valider une demande urgente sans vérification par un autre moyen
- → Se méfier des appels ou vidéos inattendus, même de contacts connus
- → Limiter les informations personnelles partagées publiquement (photos, vidéos, voix)
- → Vérifier l'URL des sites avant de saisir des identifiants
- → En cas de doute, raccrocher et rappeler via un numéro officiel

La lutte contre ces menaces est une course permanente : chaque nouvelle protection est étudiée et potentiellement contournée par de nouvelles techniques. La veille technologique et la formation continue

sont donc essentielles pour maintenir un niveau de protection adapté.

Conclusion

L'intelligence artificielle a profondément modifié le paysage de la cybersécurité. Les mêmes technologies qui permettent de détecter les menaces et de protéger les systèmes sont désormais exploitées par les attaquants pour créer des fraudes d'un réalisme sans précédent.

Les deepfakes vocaux et vidéo remettent en question notre capacité à faire confiance à ce que nous voyons et entendons. Le phishing assisté par IA atteint un niveau de personnalisation qui rend les messages frauduleux quasi indiscernables des communications légitimes. Les outils malveillants comme WormGPT démocratisent l'accès à des techniques d'attaque autrefois réservées aux experts.

Face à ces menaces, la réponse doit être globale : technologies de détection basées sur l'IA, procédures de vérification renforcées, et surtout sensibilisation continue des utilisateurs. La confiance aveugle dans les communications numériques n'est plus possible. Chaque demande inhabituelle, chaque urgence suspecte doit déclencher un réflexe de vérification.

Pour les professionnels de l'informatique, maintenir une veille active sur ces sujets n'est plus une option mais une nécessité. L'IA continuera d'évoluer, et avec elle les menaces qu'elle permet de créer. Comprendre ces évolutions est la première étape pour s'en protéger.

Sources

Europol - Rapport "ChatGPT: The impact of Large Language Models on Law Enforcement" (2023)

FBI - Public Service Announcement sur les deepfakes (IC3, 2024)

Kaspersky - Rapport annuel sur les menaces 2024

The Hacker News - Articles sur WormGPT et FraudGPT

CNN Business - "Finance worker pays out $25 million after video call with deepfake CFO"

ANSSI - Recommandations de sécurité (cyber.gouv.fr)

BleepingComputer - Actualités cybersécurité

Krebs on Security - Analyses d'attaques

Microsoft - Documentation Video Authenticator

ESET - WeLiveSecurity blog

CrowdStrike - Global Threat Report 2024

Efflam Connan — BTS SIO — 2025-2026