Cyberattaques basées sur l'IA visant les données sensibles : risques et

Découvrez comment l'IA aggrave les cybermenaces et apprenez des stratégies concrètes pour protéger les informations sensibles dans tous les secteurs.

Younoussa SANFO
Expert judiciaire Cybersécurité et investigations numériques

L'IA accélère la fraude : industrialise et banalise

Les données sensibles alimentent des attaques basées sur l'IA de plus en plus sophistiquées

« L'IA n'invente pas la fraude. Elle lui apporte rapidité, l’échelle… et une voix. »

Nos données sensibles sont devenues le carburant des attaques basées sur l'IA

Les violations existaient déjà auparavant : hameçonnage, usurpation d'identité,

L'IA rend les attaques plus réalistes, moins coûteuses et plus rapides

Qu'est-ce qui définit une donnée sensible ?

Découvrez les principales catégories et exemples dans les domaines personnel, professionnel et gouvernemental.

Données personnelles

- Informations **d'identité** telles que les noms et identifiants
- Données **biométriques et vocales** utilisées à des fins d'authentification
- Dossiers **médicaux** et informations médicales
- Données **de géolocalisation** permettant de suivre la position physique

Données d'entreprise

- **Code source** essentiel au développement de logiciels
- **Contrats** régissant les accords commerciaux
- **Feuilles de route** décrivant la planification stratégique
- **Secrets industriels,** y compris les connaissances exclusives

Données gouvernementales

- Informations personnelles et démographiques **des citoyens**
- **Infrastructures critiques** essentielles à la sécurité nationale
- **Doctrine,** politiques officielles et directives stratégiques
- Détails **des marchés publics** relatifs aux contrats gouvernementaux

Point clé

« Une donnée sensible est une bombe à retardement numérique. »

Avant l'IA : techniques d'attaque manuelles

Comprendre les méthodes traditionnelles de cyberattaque et leurs contraintes inhérentes

1

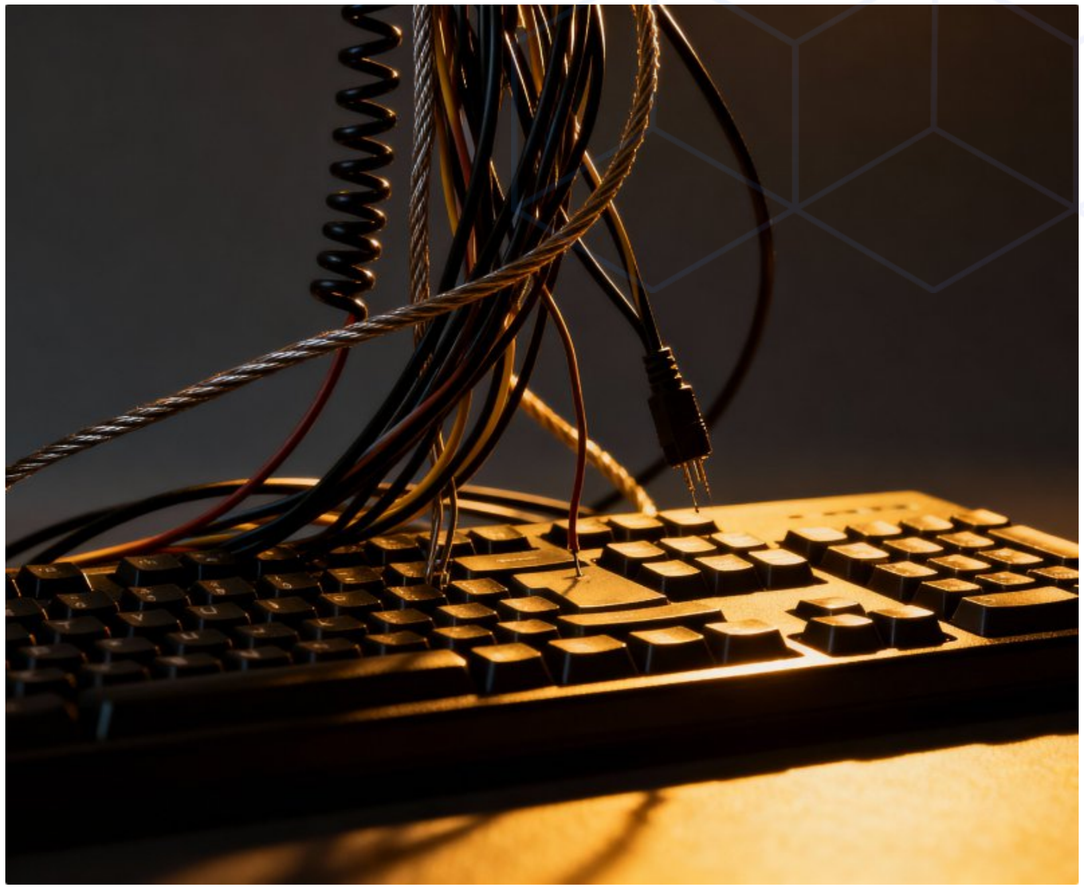
Types d'attaques

- **Hameçonnage basique** avec fautes d'orthographe et anglais approximatif
- **Ingénierie sociale par téléphone** limitée en volume, sans imitation de voix
- **Malware artisanal** nécessitant du temps et de l'expertise
- **Fraudes telles que les faux ordres de virement bancaire** nécessitant une préparation importante

2

Limites

- **Faible réalisme** dans l'exécution des attaques
- **Faible volume** d'attaques possibles
- **Nécessité de faire appel à des experts** pour élaborer les attaques



Après l'IA : l'industrialisation des cyberattaques

Comment l'IA amplifie les capacités d'attaque, accélérant la fraude et la tromperie

Campagnes de phishing améliorées par l'IA

- Générées dans toutes les langues et tous les tons
- Entraînant une augmentation significative des taux de clics

Usurpation d'identité ultra-crédible grâce à la voix et à la vidéo

- Voix clonées et vidéos deepfake
- Cibles : PDG, ministres et proches collaborateurs

Malwares et exploits zero-day alimentés par l'IA

- Assistés par des modèles d'IA non filtrés suggérant du code
- Techniques de dissimulation améliorées pour échapper à la détection

Profilage automatisé des cibles via le scraping et les LLM

- Collecte rapide des profils cibles en quelques minutes
- Combine des outils de scraping avec des modèles linguistiques de

Usurpation d'identité à l'aide de documents synthétiques et

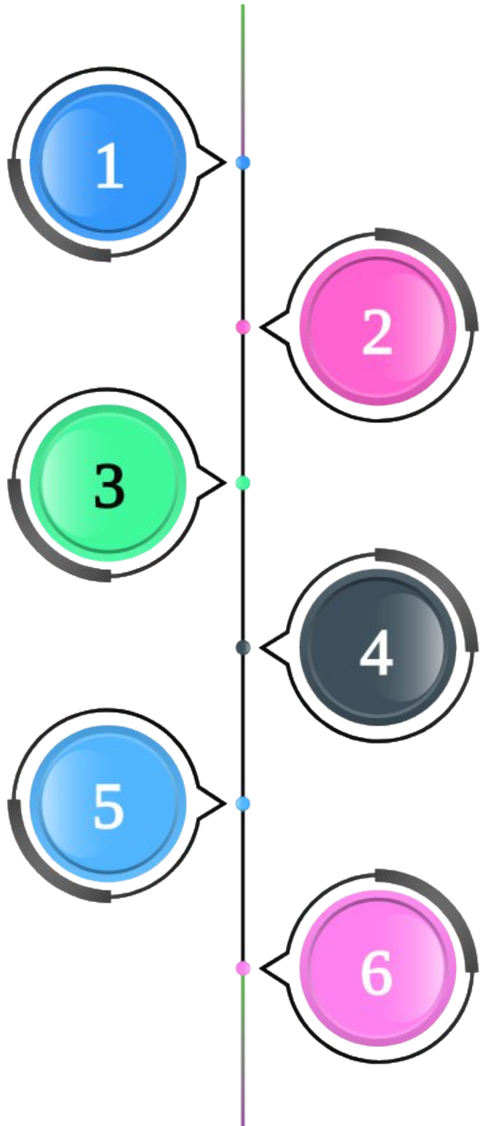
- Utilisation de documents synthétiques et de selfies manipulés
- Contournement des processus KYC (Know Your Customer)

Punchline illustrant l'impact de l'IA sur la cybercriminalité

« Ce que 10 experts faisaient en 10 jours, un acteur moyen le fait désormais en 10 minutes. »

Chaîne d'attaque assistée par l'IA : des données à la monétisation

Décomposition étape par étape du processus automatisé de cyberattaque utilisant des outils d'IA



Collection

LLM et ses outils explorent les réseaux et les archives publiques afin de collecter des données.

Ciblage

Évaluation automatique des victimes en fonction de leur rôle, de leur budget et de leurs habitudes.

Leurre

E-mails, SMS, messages WhatsApp parfaits et sites web clones générés.

Usurpation

Voix clonée du PDG ou d'un proche utilisée pour créer un sentiment d'urgence dans le paiement.

Implantation

Malware polymorphe généré avec des techniques suggérées pour contourner les systèmes EDR.

Monétisation

Utilisation de mélangeurs de cryptomonnaies, de comptes mules et d'extorsion via le doxing.

Les attaques IA dans le monde réel perturbent la

Des cas concrets révèlent les risques liés au clonage vocal, aux deepfakes et à l'ingénierie sociale

La voix clonée d'un PDG utilisée dans le cadre d'un transfert frauduleux d'environ 200 000 € dans une affaire européenne largement médiatisée

Des deepfakes politiques de dirigeants internationaux ont été utilisés pour semer la confusion pendant les crises

Attaque « sociale » par ransomware contre des casinos américains : un simple appel téléphonique et une manipulation psychologique ont permis d'obtenir l'accès à Okta/IT, suivi d'un cryptage et d'une destruction

Escroqueries avec prise d'otages familiaux aux États-Unis et au Canada à l'aide de voix clonées à partir de quelques secondes d'audio Instagram ou TikTok

Leçon clé : la voix et l'image ne peuvent plus être considérées comme des preuves fiables

Cybermenaces dans les différentes régions

Des cas réels révèlent diverses méthodes d'attaque et vulnérabilités

Afrique du Sud : **attaques par ransomware** visant les infrastructures critiques et le secteur judiciaire

Afrique du Sud : **violations** massives **des données clients** dans le secteur privé

Kenya, Ghana, Nigeria : **fraude à l'argent mobile** intensifiée par les bots et les deepfakes, notamment le contournement des procédures KYC et les échanges de cartes SIM orchestrés

Afrique de l'Ouest : **campagnes de phishing** multilingues ciblant les banques, les opérateurs et les gouvernements avec des tons culturellement crédibles

Leçon : **l'écosystème mobile** et les secteurs informels créent de nombreuses surfaces d'attaque

Menaces cybernétiques émergentes au Burkina Faso : tactiques locales et

Typologies détaillées des attaques et évolution du rôle de l'IA dans les stratagèmes de fraude

Typologies d'attaques locales observées

- **Escroqueries par WhatsApp/SMS :** faux colis, faux paiements, usurpation d'identité d'autorités (administration/ONG/banque)
- **Fraude liée à l'argent mobile :** assistance technique fictive, codes QR et OTP volés
- **Construction/marchés publics :** fausses coordonnées bancaires (RIB), faux bons de commande, appels d'offres frauduleux
- **Secteurs de l'éducation/de la santé :** hameçonnage des comptes de messagerie institutionnels

Comment l'IA transforme la fraude

- Messages parfaitement rédigés en français et dans les langues locales
- Voix clonées de supérieurs hiérarchiques ou de proches
- Faux documents sans aucune erreur
- Sites web miroirs hautement crédibles

Leçon clé

Bien qu'aucun incident national n'ait été largement rapporté, le risque de cyberfraude reste présent et augmente

Concepts de démonstration en direct

Présentation de techniques avancées d'hameçonnage, de clonage et de réponse rapide basées sur l'IA, tout en

1. Test A/B de phishing : comparez un e-mail mal rédigé à un e-mail généré par IA avec la même marque et le même ton.
2. Démonstration de voix clonée : extrait de 15 secondes d'un discours public converti en un message vocal urgent et crédible
3. Création de fausses pages web : capturez un site public en 60 secondes et générez une page d'accueil trompeuse.
4. Démonstration de rapidité : de la demande à la génération d'un modèle de logiciel malveillant sans fournir de code, démontrant la capacité d'assistance de l'IA
5. Rappel éthique : toutes les démonstrations utilisent des données fictives afin de garantir des démonstrations responsables et sûres

Identification des vulnérabilités techniques et humaines

Principaux points faibles exploités par les attaquants et les menaces basées sur l'IA

Vulnérabilités humaines

Décisions prises dans l'urgence, réactions émotionnelles et biais d'autorité exploités par l'IA

Faiblesses des processus

- Absence de contre-vérifications ou d'étapes de vérification
- Absence du principe du double contrôle pour les paiements

Failles techniques

- Mauvaise mise en œuvre de l'authentification multifactorielle (MFA)
- Protocoles d'authentification des e-mails manquants : DMARC, DKIM, SPF
- Détection et réponse aux incidents sur les terminaux (EDR) non gérées
- Absence de prévention des pertes de données (DLP)

Risques liés aux tiers

- Sous-traitants non audités
- Comptes partagés avec des contrôles insuffisants

Conclusion

« La meilleure IA offensive cible d'abord... notre cerveau. »

Protection individuelle : renforcez votre sécurité numérique

Habitudes essentielles en matière de cybersécurité personnelle et techniques de vigilance

Hygiène : maintenez vos logiciels à jour, utilisez des mots de passe uniques, employez un gestionnaire de mots de passe, activez l'authentification multifactorielle (MFA) et les clés d'accès (Passkeys).

Scepticisme : appliquez la règle du double appel ; n'autorisez jamais de virements par simple messagerie

Empreinte numérique : limitez l'exposition de votre voix et de votre visage grâce aux paramètres de confidentialité ; désactivez la géolocalisation par défaut

Signaux IA : soyez attentif aux voix trop parfaites, aux incohérences dans les micro-expressions, aux demandes urgentes et aux canaux inhabituels.

N'oubliez pas : « La confiance se vérifie, elle ne se présume pas. »

Stratégies de protection des données d'entreprise

Contrôles de sécurité complets en matière de gouvernance, d'accès, de détection et de sensibilisation continue

Données relatives à la gouvernance

- **Cartographie** des actifs de données
- **Classification** des données : publiques, internes, sensibles, secrètes
- Prévention des pertes de données (**DLP**)

Sécurité des e-mails et du Web

- Protocoles d'authentification des e-mails : **DMARC/DKIM/SPF**
- **Sandboxing** des pièces jointes aux e-mails
- **Isolation du navigateur** pour prévenir les menaces Web

Contrôles d'accès

- **Authentification multifactorielle (MFA)** partout
- Mise en œuvre d'un modèle de sécurité **Zero Trust**
- **Renforcement** des configurations **VPN/SSO**
- Gestion des accès privilégiés (**PAM**) pour les comptes privilégiés

Capacités de détection

- Détection des terminaux et détection
- Journalisation centralisée avec **SIEM**
- Activités **de recherche de menaces**

Processus sécurisés

- Rappel obligatoire pour **les paiements et vérification de l'IBAN**
- **Principe du double contrôle** pour l'approbation

Sensibilisation et formation continues

- **Simulations de phishing** de nouvelle génération avec des textes de type IA
- **Ateliers sur les deepfakes** pour sensibiliser les participants

Utilisation responsable de l'IA

- Instances locales/sur site pour **les invites sensibles**
- Politiques : **ne pas coller d'informations confidentielles dans** les outils **d'IA**

Sécurisation des actifs numériques du secteur public et de l'État

Stratégies de protection complètes pour la souveraineté des données et la cyber-résilience

1

Cadre juridique

- Renforcer les rôles en matière de protection des données (CIL)
- S'aligner sur la Convention de Malabo et les réglementations similaires au RGPD

CERT nationaux et sectoriels

- Partager les indicateurs de
  Surveiller les cybermenaces liées à l'IA
- Organiser des exercices cyber interministériels

Protection des infrastructures critiques

- Segmenter les réseaux OT et IT
- Mettre en œuvre des plans de reprise après sinistre (PRA) et de continuité des activités (PCA)
- Effectuer régulièrement des tests de

Souveraineté numérique

Héberger les données critiques au niveau régional

Adopter des solutions cloud fiables

5

Éducation et sensibilisation

- Intégrer des modules d'IA et de cybersécurité dans les écoles et les administrations
- Mettre en place des programmes

6

Sécurité des marchés publics

- Inclure des clauses de sécurité dans la
- Appliquer les exigences DMARC, MFA et de journalisation

7

Engagement en faveur de la

« La souveraineté numérique n'est pas un slogan : c'est un budget, une loi, une pratique. »

Accélérez la sécurité de votre entreprise en 90 jours

Une feuille de route en quatre phases pour renforcer les défenses et la résilience de votre organisation

1 2 3 4

Semaines 1-2 : Audit de sécurité rapide et cartographie des données

• Réalisation d'un audit rapide : exposition MX/DMARC, statut MFA, inventaire des comptes • Cartographie des emplacements des données

Semaines 3 à 6 : renforcement de l'accès et de la détection des menaces

• Activer l'authentification multifactorielle (MFA)• Configurer DMARC avec une politique de quarantaine/rejet• Déployer un projet pilote de détection et de réponse aux incidents au niveau des terminaux (EDR)/détection et réponse étendues (XDR)•

Semaines 7 à 10 : appliquer la protection des données et les contrôles d'accès

• Établir une politique de prévention des pertes de données (DLP) • Déployer une architecture d'authentification unique (SSO) et de confiance zéro • Améliorer les sauvegardes immuables • Effectuer des tests de restauration

Semaines 11 à 13 : Formation avancée et préparation aux incidents

• Dispenser une formation anti-hameçonnage de type IA • Élaborer un guide pratique pour les procédures relatives aux deepfakes et au clonage

Liste de contrôle pour la réponse aux incidents liés aux

Étapes clés pour détecter, contenir et communiquer efficacement

Mettez fin à l'urgence et imposez un contre-canal via un appel connu

Vérifiez les métadonnées, les horodatages et les informations contextuelles

Collectez les traces : journaux de messagerie/VoIP et informations whois des sites miroirs

Si un paiement est impliqué : appliquez un gel temporaire et informez la banque, le SOC et le CERT

Communiquer en interne par messages courts ; éviter les reproches et encourager le partage d'informations

Principales conclusions sur l'impact et la responsabilité de l'IA

Trois idées essentielles pour guider une utilisation éthique et efficace de l'IA

- **L'IA transforme l'échelle :** la vitesse, le volume et la plausibilité augmentent considérablement
- **Les données sensibles sont essentielles :** il est crucial de comprendre quelles informations sont exposées
- **La solution est holistique :** elle combine le jugement humain, les processus, la technologie et les cadres juridiques

Démonstrations pratiques de sécurité IA et meilleures pratiques

Exemples concrets et conseils d'experts pour détecter et prévenir les menaces liées à l'IA

Démonstrations

E-mail IA vs e-mail humain : même marque, ton local → comparez les taux de détection en direct

Voix clonée sur un échantillon public générique de personnalité → lire un message urgent puis expliquer les contre-mesures

Maquette de page web statique clonée → démontrer l'impossibilité de distinguer le vrai du faux

Arbre de décision pour les transferts urgents : 3 vérifications obligatoires (rappel, IBAN validé, principe du double contrôle)

Conseils de sécurité individuels

Gestionnaire de mots de passe + MFA/clés d'accès

Rappel pour toute demande d'argent ou d'IBAN

Paramètres de confidentialité pour la voix, le visage et la géolocalisation

Méfiez-vous des tactiques d'urgence et de culpabilisation

Conseils au niveau organisationnel et

Entreprises : DMARC/DKIM/SPF, MFA partout, EDR/XDR, DLP

Processus à quatre yeux + comptes sur liste blanche

Formation trimestrielle de sensibilisation au phishing par IA

Inventaire et classification des données sensibles

États : CERT national, exercices sectoriels