VEILLE IA & CYBERCRIMINALITÉ 07

MENACES ÉMERGENTES ET PRINCIPALES UTILISATIONS PAR LES CYBERCRIMINELS

© COMCYBER-MI

Ingénierie sociale et manipulation cognitive

Les campagnes de phishing évoluent vers une industrialisation qui repose largement sur l'IA. Microsoft a récemment documenté des attaques de type « device code phishing » où les messages sont générés et adaptés automatiquement, augmentant fortement leur crédibilité. Cette évolution touche directement les citoyens : multiplication des sollicitations frauduleuses, messages personnalisés et exploitation de la confiance dans des services légitimes. Le risque est une banalisation de la fraude numérique dans la vie quotidienne.

Malware news, 6 avril 2026
Microsoft; 6 avril 2026 - Inside an AI-enabled device code phishing campaign | Microsoft Security Blog

Malwares et exploits générés ou pilotés par IA

L'usage de l'IA ne se limite pas à l'ingénierie sociale mais s'étend à l'aspect offensif en complexifiant les scénarios d'attaque. Un rapport d'Expel montre comment le groupe Lazarus industrialise ses attaques en s'appuyant sur l'IA pour cibler des développeurs et introduire des composants compromis dans des environnements de développement. Cette évolution illustre un déplacement des attaques vers la supply chain logicielle, avec des impacts potentiels à grande échelle sur les utilisateurs finaux.

En parallèle, l'évaluation des LLM par le UK AI Safety Institute met en évidence les capacités croissantes de certains modèles à assister des activités liées à la cybersécurité offensive. Sans être conçus pour un usage malveillant, ces systèmes peuvent expliquer des vulnérabilités ou structurer des scénarios d'attaque.

Expel, avril 2026
AISI, avril 2026

Exfiltration et exploitation de données

Les agents IA connectés à des systèmes d'information deviennent de nouveaux points de fragilité. Les attaques par prompt injection permettent de manipuler ces systèmes pour accéder à des données sensibles ou déclencher des actions à l'insu de l'utilisateur. Cela se traduit par un risque accru de fuite de données personnelles, parfois via des outils du quotidien.

TrueFoundry, avril 2026 ; CERT-FR, 13 avril 2026

Attaques sur les modèles et infrastructures IA

Les modèles d'IA présentent des vulnérabilités persistantes face à des instructions malveillantes. Un article récent montre qu'une simple ligne de code peut permettre de contourner les mécanismes de sécurité de plusieurs grands modèles (dont ChatGPT, Claude ou Gemini). Cette technique repose sur des formulations spécifiques capables de désactiver ou de détourner les garde-fous intégrés.

Ce type de contournement est particulièrement préoccupant car il est simple à reproduire, sans expertise technique avancée, facilement diffusable, notamment via des forums ou communautés en ligne et transposable à grande échelle, sur différents modèles.

En pratique, ces vulnérabilités peuvent être exploitées pour générer des contenus frauduleux, automatiser des scénarios d'escroquerie ou produire des instructions malveillantes.

Cybersecurity News, avril 2026

Menaces hybrides et géopolitiques

Le mois d'avril 2026 confirme l'intégration croissante de l'intelligence artificielle dans les rapports de puissance entre États. Plusieurs signaux récents montrent une montée en tension autour de ces technologies : renforcement des contrôles à l'export sur les composants stratégiques (notamment les puces IA), structuration de blocs technologiques concurrents et multiplication des alertes sur l'usage de l'IA dans des opérations cyber étatiques.

Les autorités britanniques ont notamment mis en garde contre une augmentation des cyberattaques soutenues par des états, avec un recours accru à l'IA pour identifier des vulnérabilités et accélérer les opérations. Par ailleurs, des échanges récents entre entreprises majeures du secteur et autorités américaines soulignent les inquiétudes croissantes quant aux capacités cyber des modèles avancés, notamment vis-à-vis des infrastructures critiques.

Ces évolutions traduisent un basculement : l'IA devient un levier stratégique dans la compétition internationale, tout en augmentant les risques indirects pour les citoyens, exposés aux conséquences de cyberopérations étatiques (perturbations de services, exploitation de données, campagnes d'influence).

Reuters, 22 avril 2026
Axios, 28 avril 2026
ETC Journal, 19 avril 2026

Menaces émergentes et tendances à suivre

Les progrès des deepfakes permettent désormais de contourner certains dispositifs de vérification d'identité en temps réel. Cette évolution représente une menace directe pour les citoyens : usurpation d'identité, fraude bancaire, ouverture de comptes frauduleux. Elle fragilise les mécanismes de confiance numérique qui structurent de nombreux services en ligne.

FrenchBreaches, 9 avril 2026

OPPORTUNITÉS ET OUTILS POTENTIELS POUR LES FSI

© COMCYBER-MI

Une lecture stratégique des menaces

Le rapport "AI-Powered Red Team and Adversarial Testing Platform Market" met en évidence l'essor du marché des plateformes permettant de simuler automatiquement des attaques contre des systèmes d'IA.

Ces outils peuvent être utiles aux enquêteurs pour comprendre concrètement comment des systèmes sont testés et contournés, et ainsi mieux reconstituer les modes opératoires utilisés par des acteurs malveillants dans des environnements réels.

Market intelo , avril 2026

Des outils pour détecter le vrai du faux

Les analyses publiées par Mandiant en avril 2026 mettent en évidence une évolution des pratiques dans les forums cybercriminels, avec l'intégration croissante de l'IA dans les phases de préparation des attaques.

Plusieurs signaux faibles concrets peuvent être observés par les FSI : sur les espaces numériques (forums, messageries, dark web), dans les contenus frauduleux analysés (emails, messages, faux sites), dans les traces techniques et environnements compromis.

Mandiant - M-Trends 2026 Report | Google Cloud

Un cadre structurant national et européen

L'actualisation du cadre d'évaluation des capacités cyber publiée par l'European Union Agency for Cybersecurity en avril 2026 propose une grille structurante pour analyser la maturité des organisations face aux menaces numériques.

Elle peut être utile aux FSI pour replacer un incident impliquant de l'IA dans un contexte plus large (niveau de préparation, exposition, gouvernance), et ainsi affiner la qualification des faits au-delà du seul aspect technique.

ENISA, avril 2026

Commandement du ministère de l'Intérieur dans le cyberespace
#VeilleIA n° 07 mai 2026

Directeur de la publication : GDI P. TOUAK
Responsable éditorial : LCL G. DECKMYN
Rédacteur : M. J. NEVICATO
Conception graphique et maquettage : Mme M. BARREAU

En attendant le prochain numéro de la veille IA & cybercriminalité, suivez notre actualité :

sur LinkedIn et sur X (ex Twitter) : ***@ComCyberMI***