Comment l’IA peut nous défendre des cyberattaques?

GenAI Meetup - Morocco

Tristan Bilot - Ph.D. student @ Université Paris-Saclay x Isep x Iriguard

3 mai 2025

université PARIS-SACLAY isep École d'ingénieurs du numérique iriguard

Plan

- Comment détecte-t-on les cyberattaques?
- Pourquoi l’IA peut-elle servir?
- Qui utilise aujourd’hui l’IA pour la cybersécurité?

Comment détecte-t-on les cyberattaques?

Quelques attaques connues

Phishing

Exemple : Faux e-mails de connexion demandant des mots de passe.

Fréquence : 84 % des entreprises touchées en 2024 [1].

Détection:

- Filtres anti-spam (règles prédéfinies, listes noires)
- Analyse des en-têtes d’e-mails.

How a phishing attack works

Akamai

Quelques attaques connues

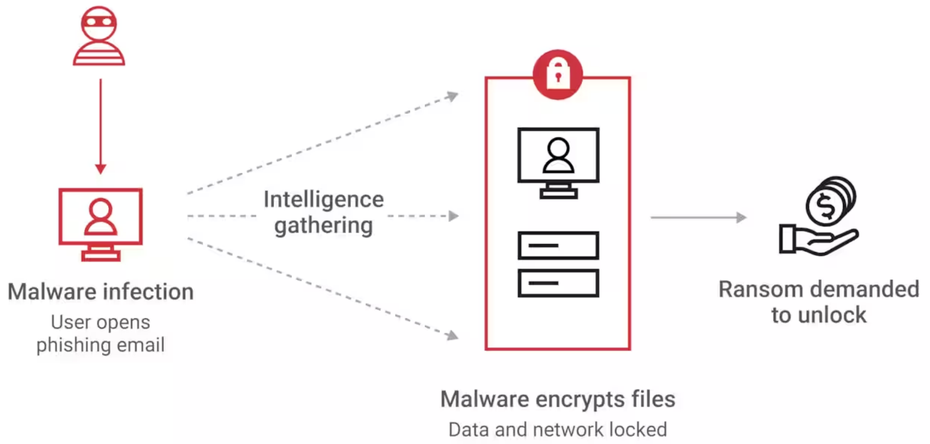
Ransomware

Exemple : Exécution d’un fichier puis chiffrement complet du système.

Fréquence : 72 % des cyberattaques en 2023 [2].

Détection:

- Antivirus (signatures, heuristique)
- Systèmes de détection d’intrusion (IDS)



How ransomware works

Akamai

Quelques attaques connues

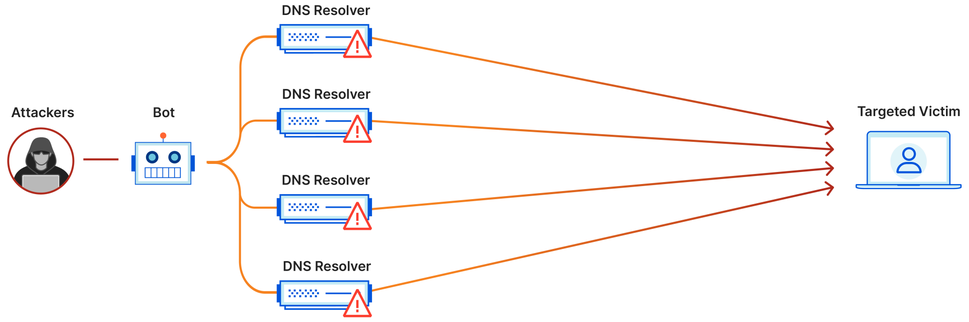
DDoS

Exemple : Inondation de requêtes sur un serveur pour le rendre inaccessible.

Fréquence : En hausse de 53% en 2024 [3].

Détection:

- Pare-feu (nombre de requêtes /s > un seuil)
- Systèmes IDS (analyse des ports, IPs, etc.).



Quelques attaques moins connues

Advanced Persistent Threat (APT)

Exemple : Cyberattaques sophistiquées, ciblées et prolongées, menées par des états ou groupes de hackers

Fréquence : 74% d'augmentation des tentatives APT en 2024 [4].

Quelques chiffres:

- 150 jours en moyenne avant d'être détectées [5].
- 60% sont attribuées à des États (ex: Chine, Russie, Corée du Nord) en 2024 [6].
- 89% sont associées à de l'espionnage [5].

Quelques attaques moins connues

Advanced Persistent Threat (APT)

Principaux groupes :

- **APT28** - Fancy Bear (Russie 🇷🇺) : Fuite électorale américaine via WikiLeaks (2016)
- **APT38** - Lazarus (Corée du Nord 🇰🇵) : Ransomware mondial WannaCry (2017)
- **APT41** - Wicked Panda (Chine 🇨🇳) : Espionnage et cybercrime financier (2021)

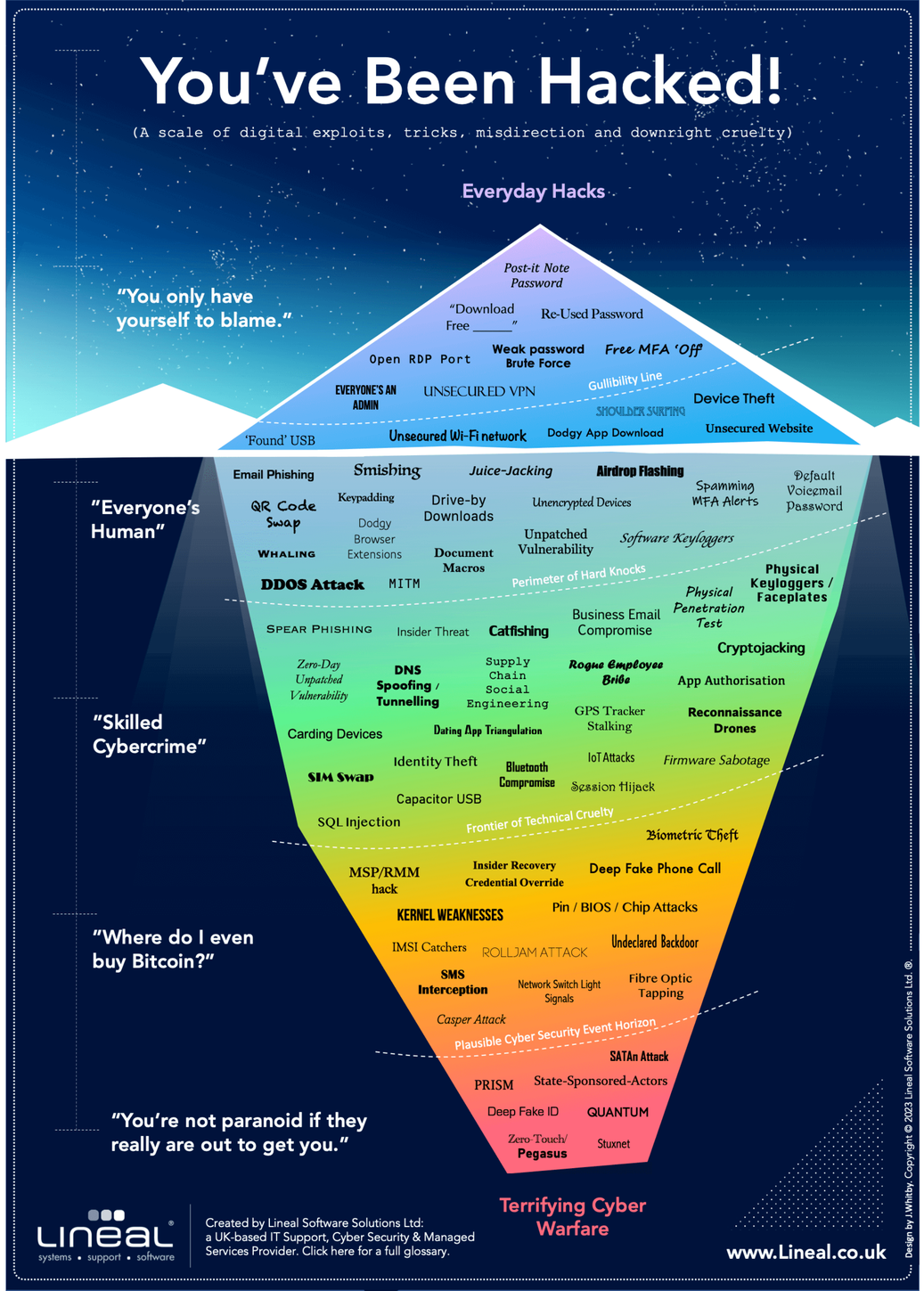
Page affichée par le ransomware Wannacry

Pourquoi l’IA peut-elle servir?

Pourquoi l'IA peut-elle servir?

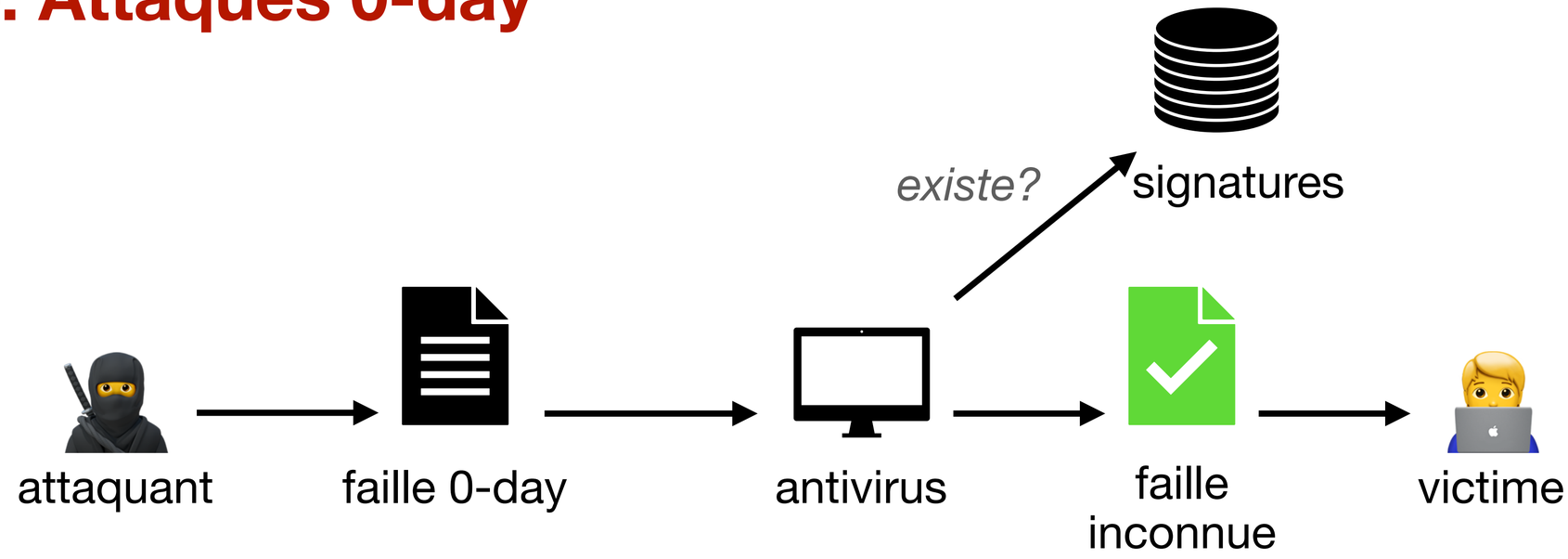
Limitation des techniques traditionnelles

1. Les attaques deviennent de plus en plus **complexes**
2. De simples règles et signatures ne **suffisent plus**
3. L'attaquant a toujours l'avantage



Pourquoi l’IA peut-elle servir?

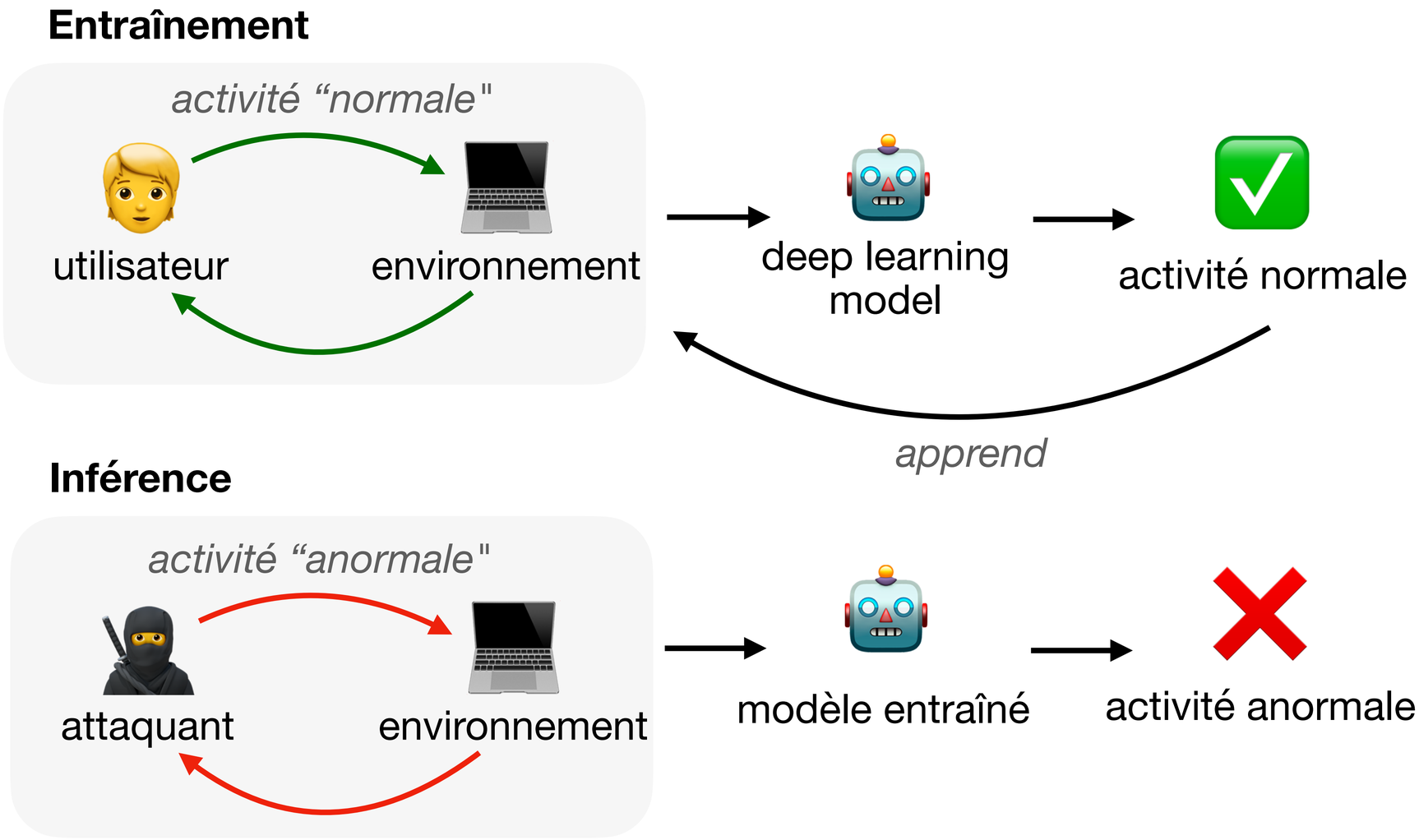
Exemple: Attaques 0-day



- La détection peut être évitée si l’attaque n’est pas déjà connue (signature)
- Ou si l’attaque est suffisamment masquée (obfuscation, modifications, etc.)

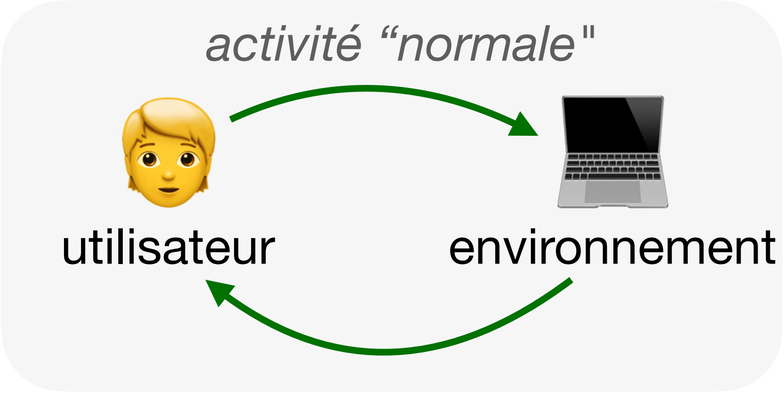
La où l’IA intervient

Un changement de paradigme



La où l’IA intervient

Collecte de données



Différents niveaux de granularité

Réseau

Data: addresses IP, paquets réseaux, …
Attaques: DDoS, mouvement latéral, …

Fichier

Data: code source, code compilé, assembleur, …
Attaques: malware, ransomware, site de phishing, …

Système

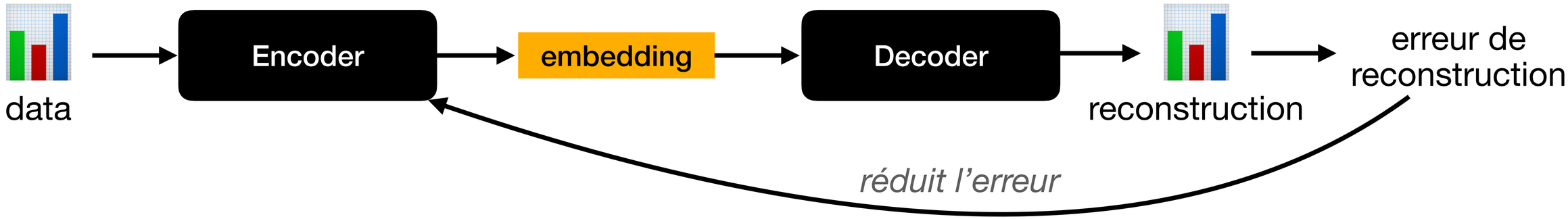
Data: appels systèmes au niveau du kernel
Attaques: intrusion, élévation de privilège, malware, …

La où l’IA intervient

Comment le modèle “apprend”?

La où l’IA intervient

Architecture Autoencoder



La où l’IA intervient

Use cases avec différents encoders

Détection réseau

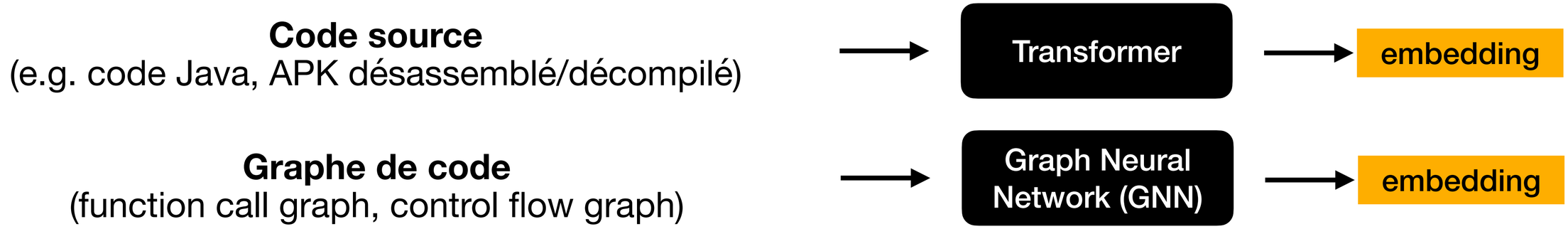
Logs réseau
(192.168.1.8 :22 => TCP => 192.168.1.9 :22)

Graphes réseau
(les nodes sont des IPs, les edges sont des communications)

La où l'IA intervient

Use cases avec différents encoders

Détection fichier



La où l'IA intervient

Use cases avec différents encoders

Détection système

Logs systèmes
(syslog, historique des évènements kernel, etc.)

Graphe de provenance
(les nodes sont des processus/fichiers/sockets,
les edges sont des appels systèmes)

Qui utilise aujourd’hui l’IA pour la cybersécurité?

Qui utilise d’IA pour la cybersécurité?

Google - Gmail

RETVec (Resilient and Efficient Text Vectoriser)

- Gmail filtre les spams en utilisant RETVec
- Convertit le text contenu dans chaque e-mail en un embedding
- Puis détection binaire: spam/non spam

Qui utilise d’IA pour la cybersécurité?

Microsoft - Defender for Endpoint

Détection d’anomalie à large échelle

Microsoft Defender
for Endpoint

- Utilise différents modèles de type autoencoder
- Détecte des anomalies dans les endpoints et le cloud

Qui utilise d’IA pour la cybersécurité?

Vectra AI / DarkTrace

Systèmes de détection d’attaque basés sur l’IA

- Proposent plusieurs modèles basés sur des autoencoders
- Pré-entraîné avec beaucoup de données puis déployé chez des clients

Conclusion

- L’IA offre de nouvelles possibilités pour la détection d’attaques
- Elle possède des capacités d’adaptation, permettant de détecter de nouvelles attaques ou des variantes
- La grande majorité de la littérature scientifique autour de la détection d’attaques est basée sur du deep learning
- De plus en plus de systèmes de détection intègrent des modèles d’IA

The End

Merci!

Contact: Tristan Bilot

E-mail: tristan.bilot@universite-paris-saclay.fr
Website: tristanbilot.github.io

References

- [1] https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2024/cyber-security-breaches-survey-2024
- [2] https://www.sentinelone.com/cybersecurity-101/cybersecurity/cyber-attacks-in-the-united-states/
- [3] https://blog.cloudflare.com/ddos-threat-report-for-2024-q4/
- [4] https://securitybrief.asia/story/advanced-persistent-threats-rise-by-74-in-2024-report
- [5] https://fr.vectra.ai/topics/advanced-persistent-threat
- [6] https://go.crowdstrike.com/2025-global-threat-report.html