MINISTÈRE
DES ARMÉES

Liberté
Égalité
Fraternité

RESSI 2020

Application de méthodes d'apprentissage automatique à la Cybersécurité

DGA
DIRECTION GÉNÉRALE
DE L'ARMEMENT

MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

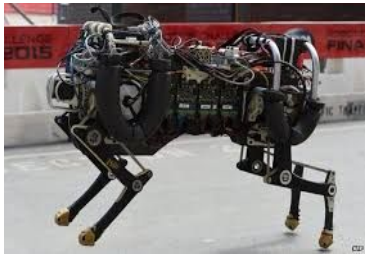
DGA MI - Département IA^{3}D

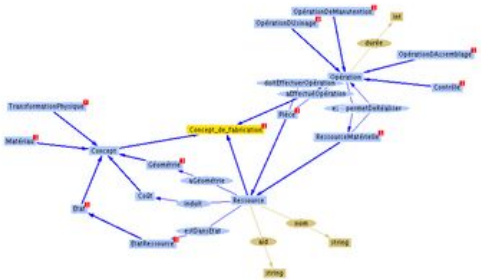
Intelligence Artificielle, Agents Autonomes et Data-sciences

Adeline Bailly
Chargée d'expertise
confirmée en IA

Samuel Hangouët
Expert Technique Référent
en IA

L'Intelligence Artificielle (IA)





Aide à la décision

Traitement du langage naturel

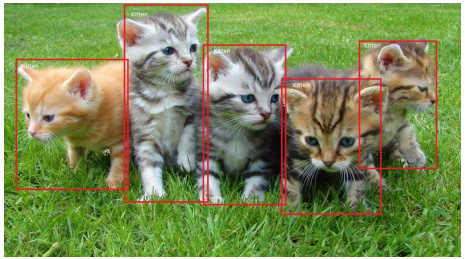
Robotique

Raisonnement et représentation des connaissances

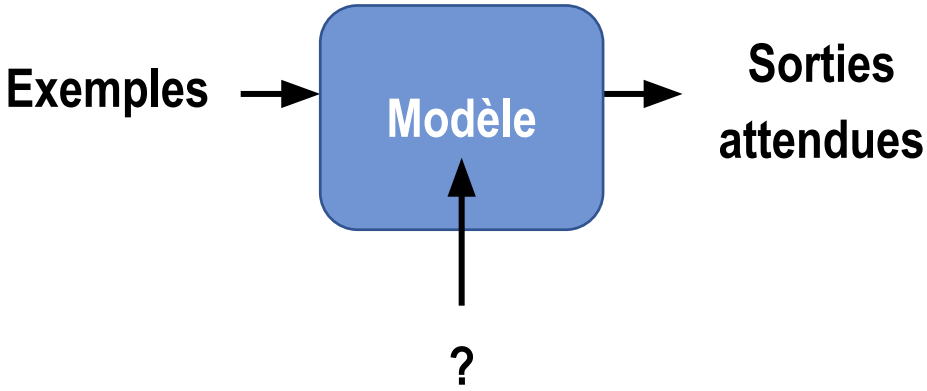
Reconnaissance d'images

Intelligence artificielle

Systèmes multi agents

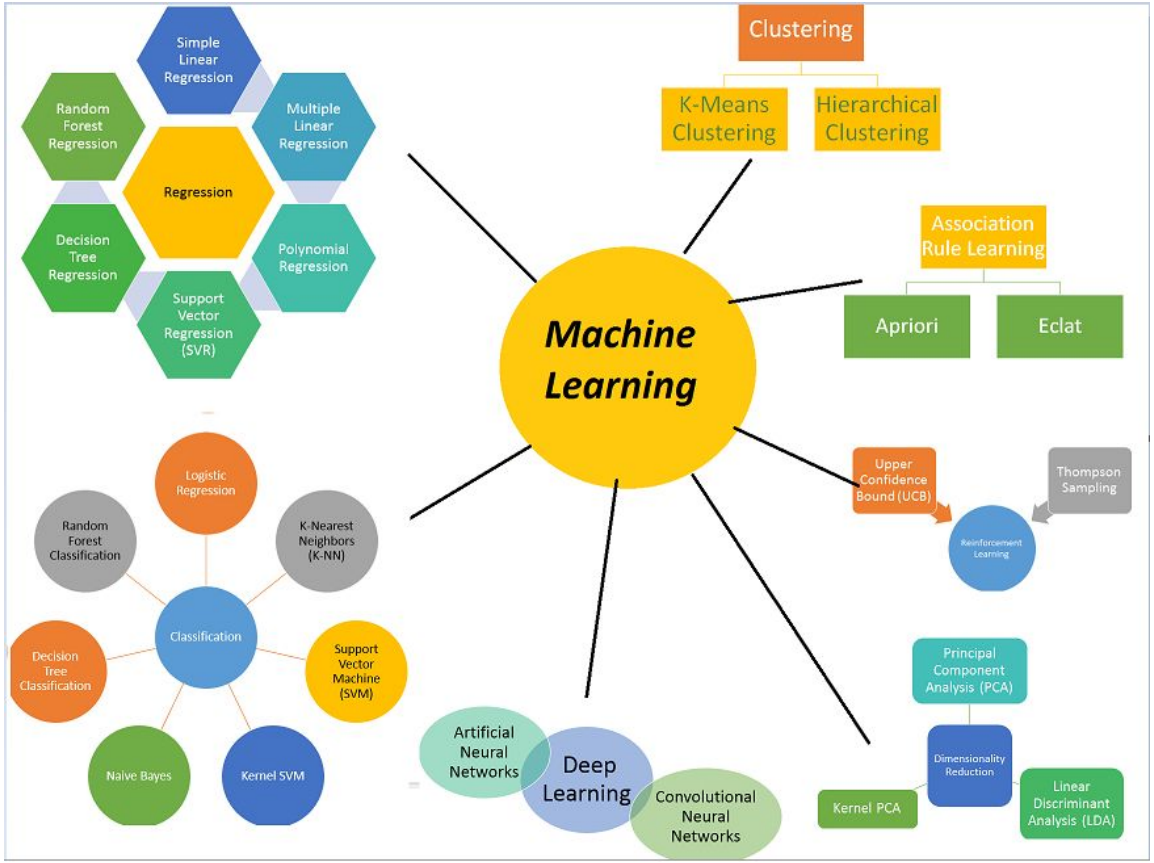


Le ML, c’est quoi ?



MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

Le ML, c'est quoi ?

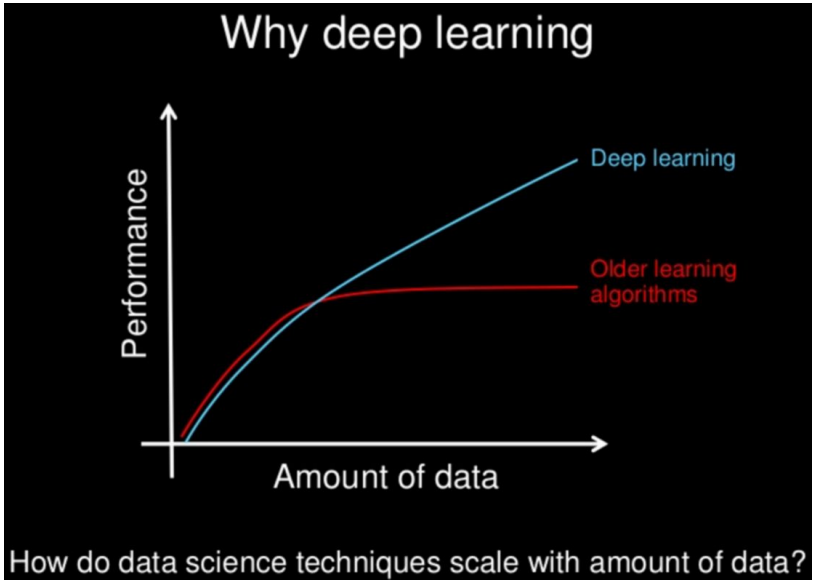


.

Le Deep Learning : une technique de ML

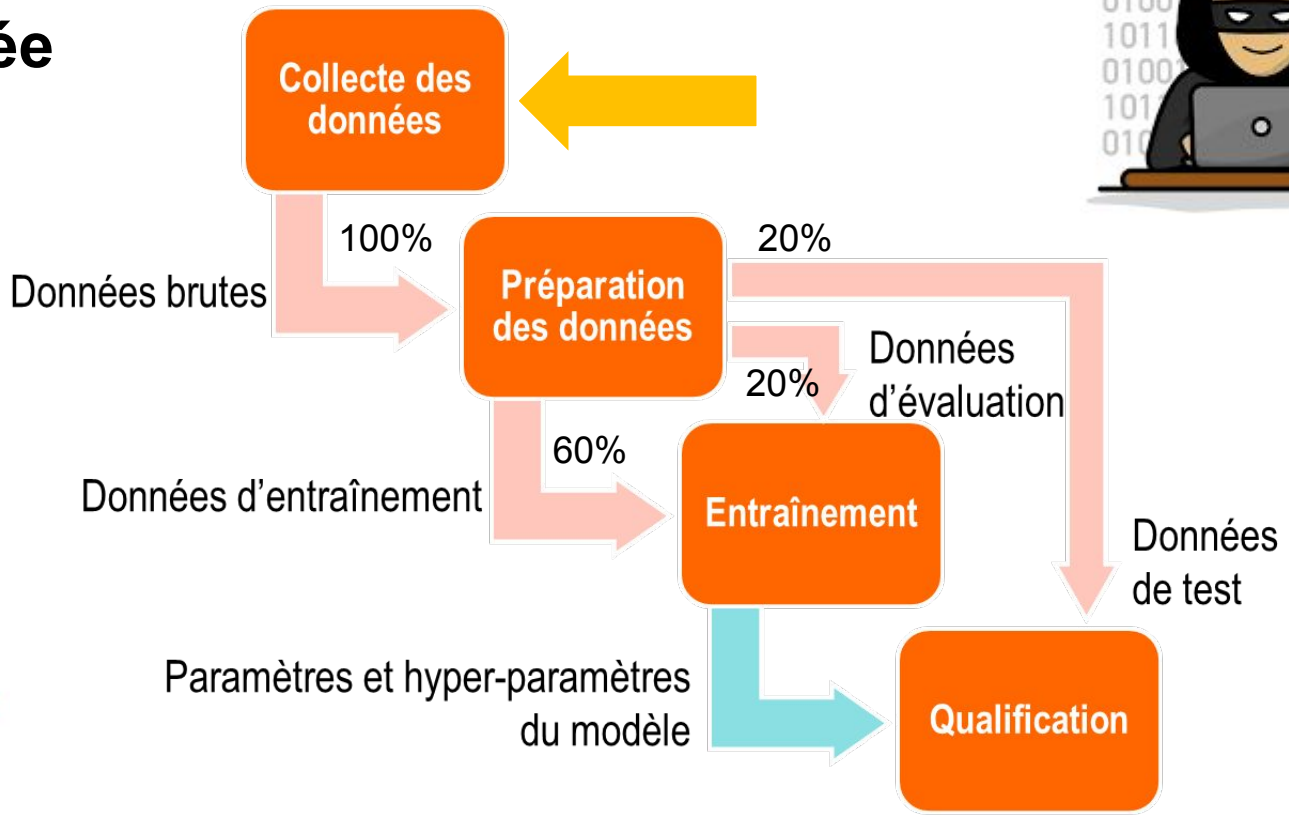
Une sous partie de l'apprentissage automatique (ML)

- DL ⊂ NN ⊂ ML
- Approche de bout en bout
- Volume de données
- Puissance de calcul (GPU)



La donnée


Data Scientist


Collecte des données
Données brutes
100%
Préparation des données
20%
Données d'évaluation
20%
60%
Données d'entraînement
Entraînement
Données de test
Paramètres et hyper-paramètres du modèle
Qualification

MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

Cyber Défense Factory

Appel à projets permanent

Un dispositif qui fournit gratuitement :

- Des locaux
- Un datalake de données cyber
- La proximité des opérationnels

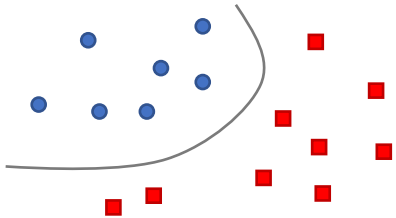
COMCYBER
COMMANDEMENT DE LA CYBERDÉFENSE

AGENCE
INNOVATION
DÉFENSE

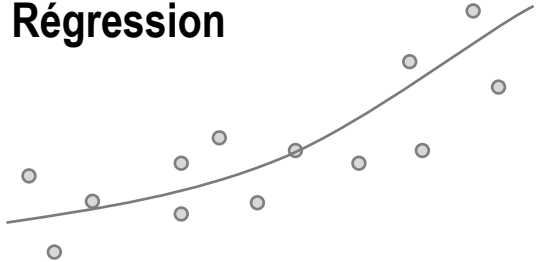
DGA
DIRECTION GÉNÉRALE
DE L'ARMEMENT

Les types de problèmes

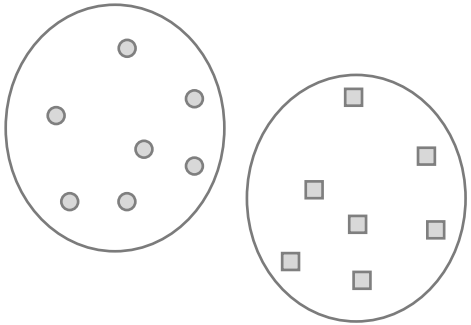
Classification



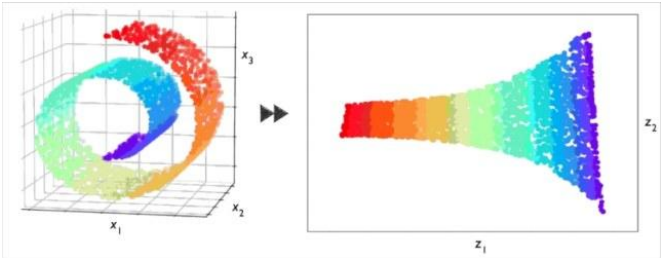
Régression



Partitionnement



Réduction de dimension



Supervisé

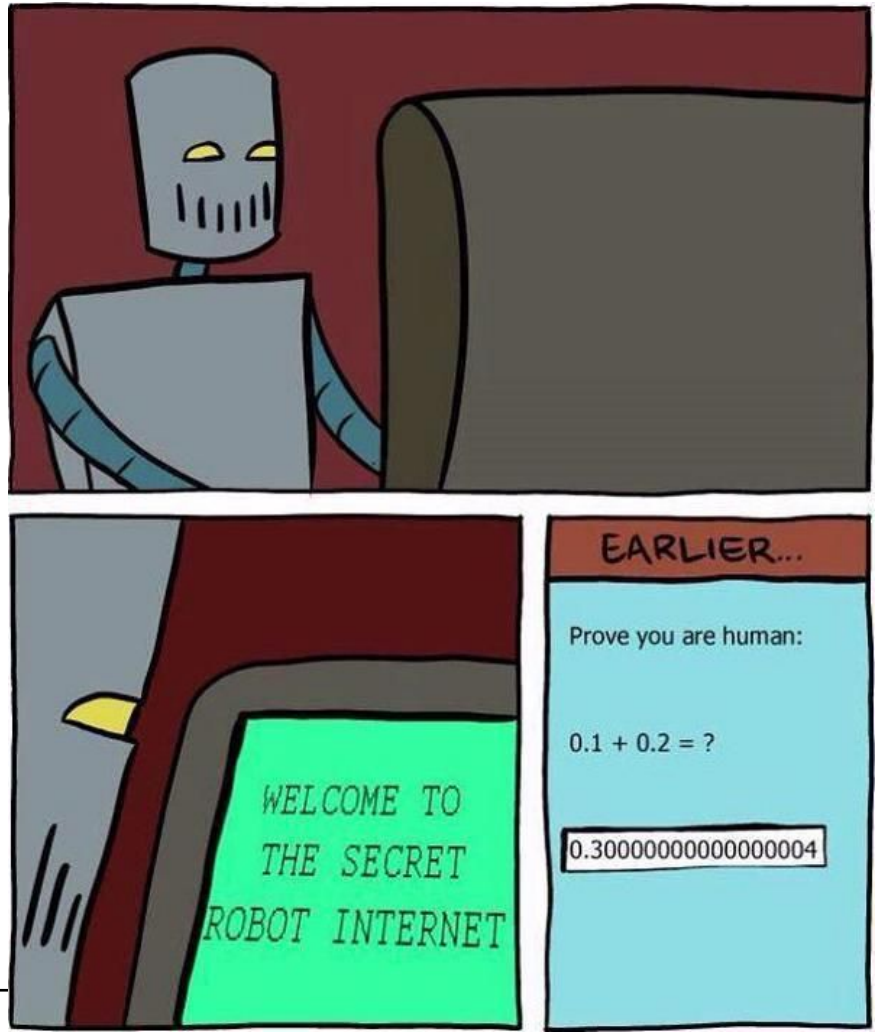
Non-supervisé

MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

Classification

Captchas

- Détection de robots
- Attaque de captcha



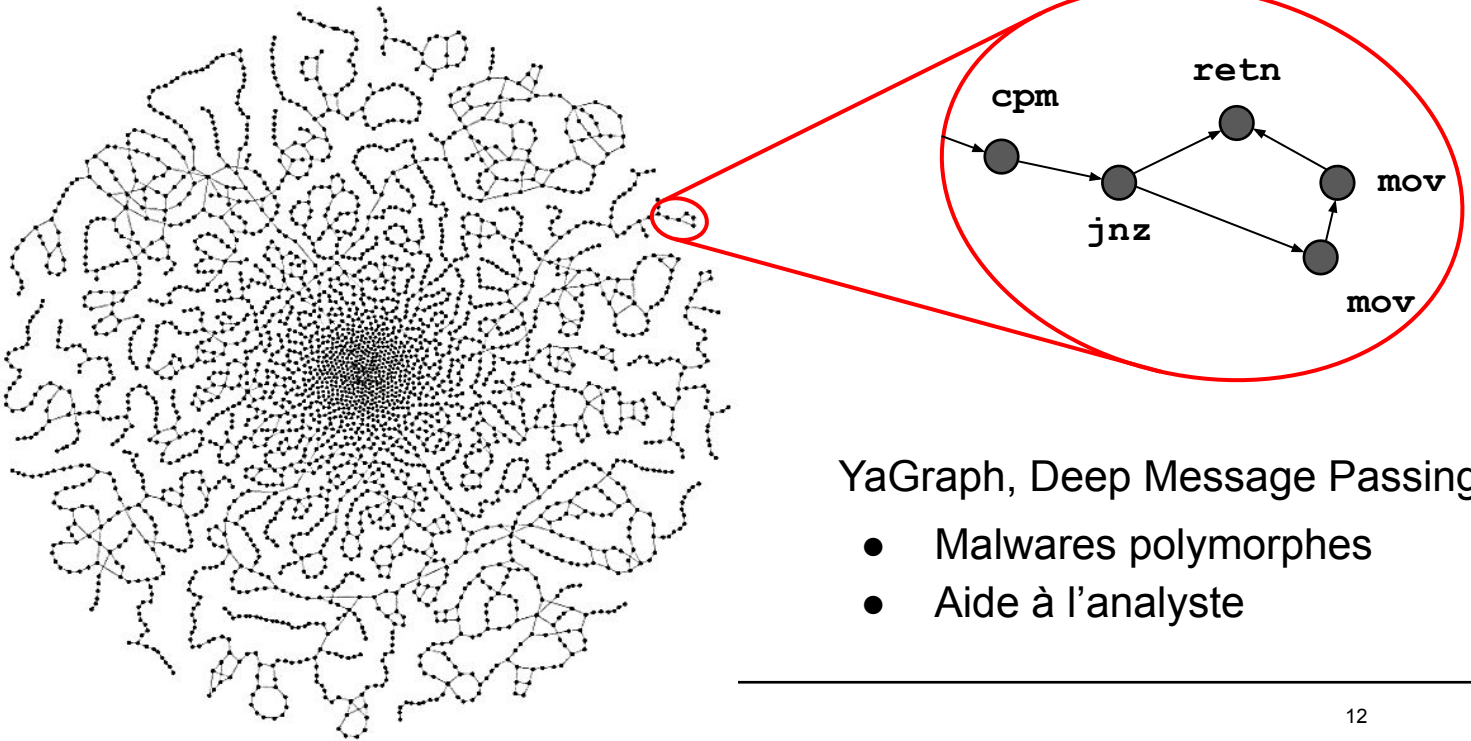
MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

Classification

Protection contre le phishing

Classification

Détection de malwares par analyse statique

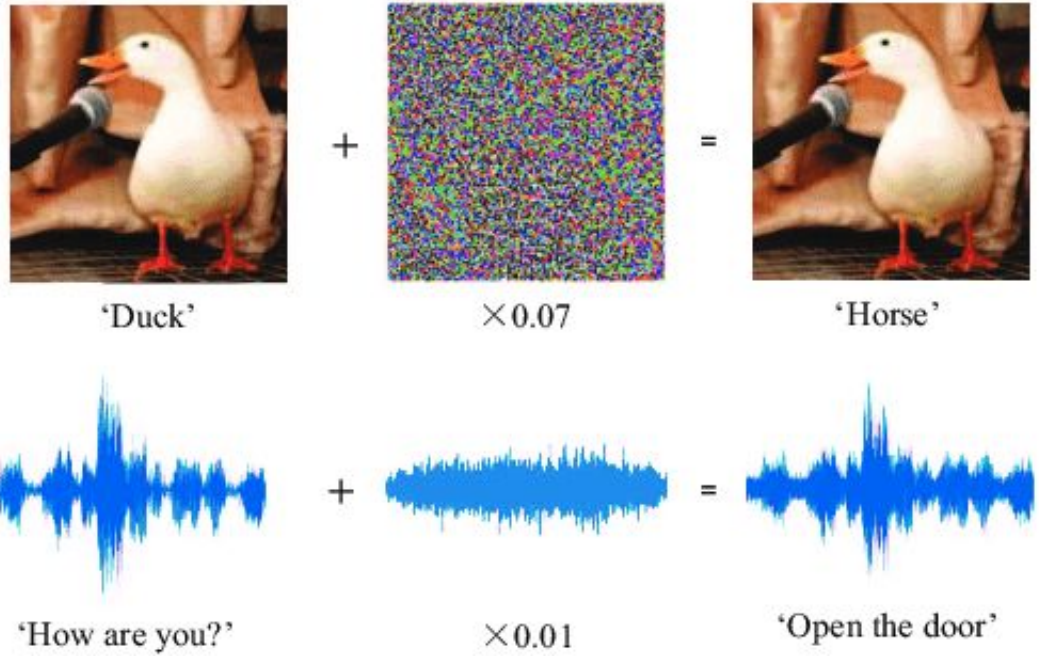


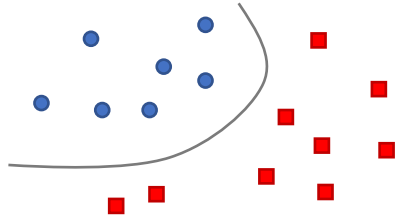
YaGraph, Deep Message Passing :

- Malwares polymorphes
- Aide à l'analyste

Classification

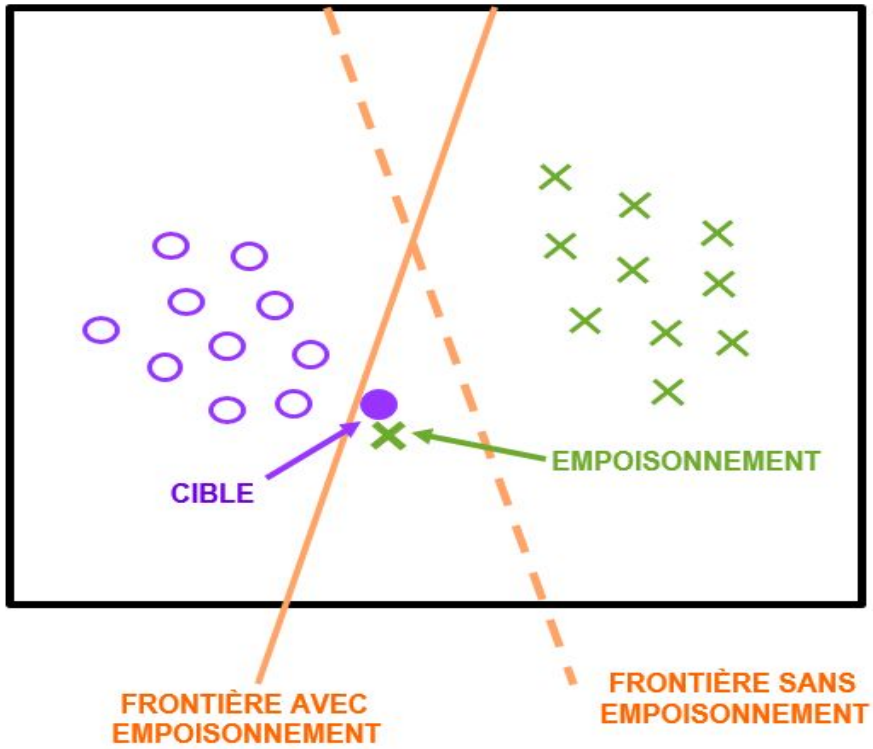
Leurrage



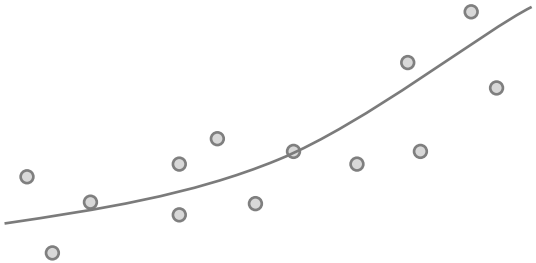


Classification

Empoisonnement
des données



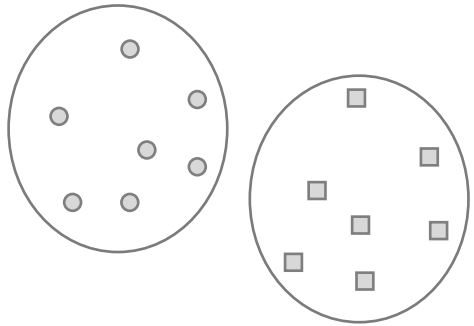
Régression

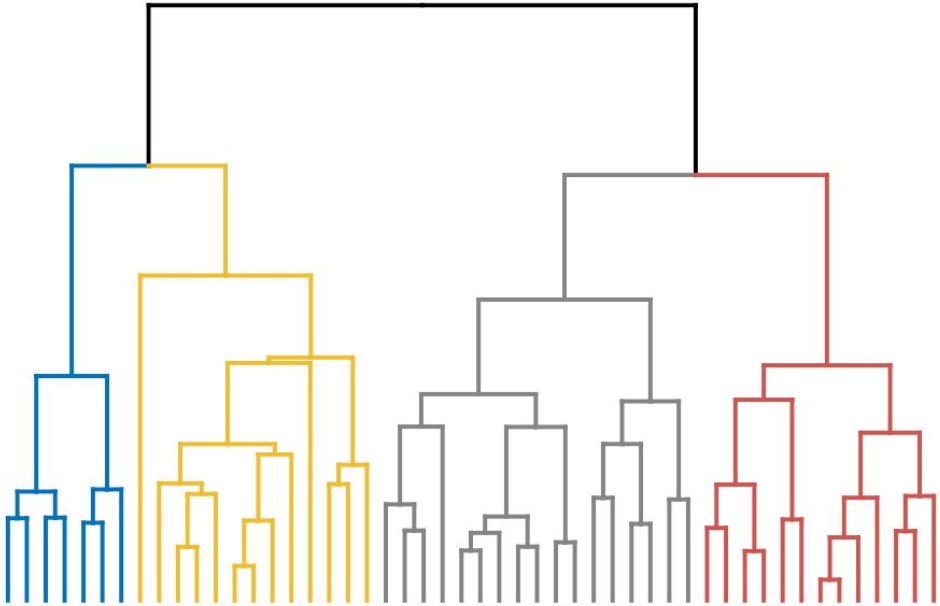


→ Traiter un problème de classification comme de la régression :

Problème	Grandeur prédite
Détecter du SPAM par analyse sémantique du texte	Probabilité d'appartenir à une classe
Trier des alertes dans un SOC	Niveau de risque
Attribuer une attaque à un APT	Degré de similarité
Détecter les dérivés d'un malware	

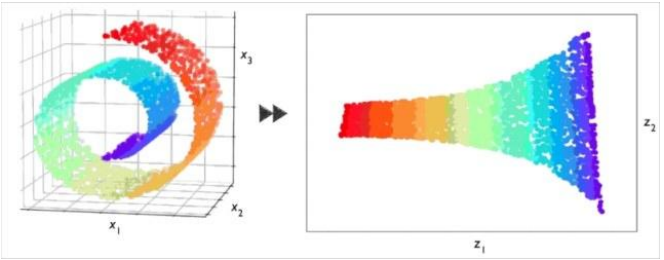
Partionnement



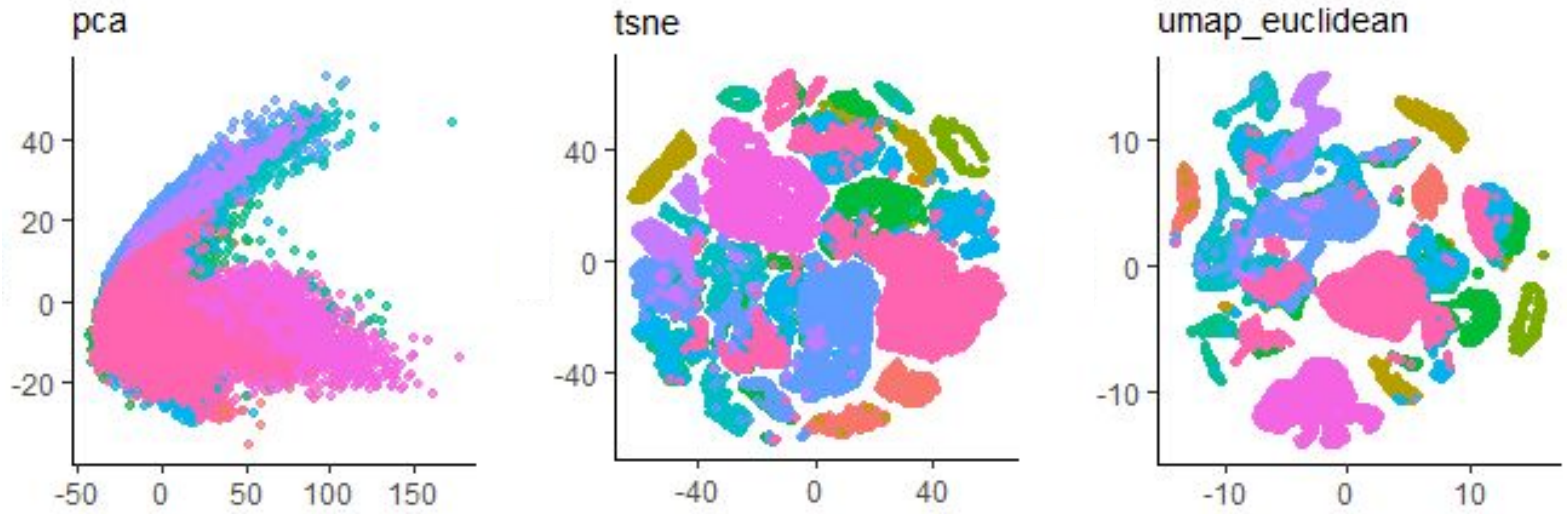


Partionnement

Réduction de dimension



Ex : Visualisation de données

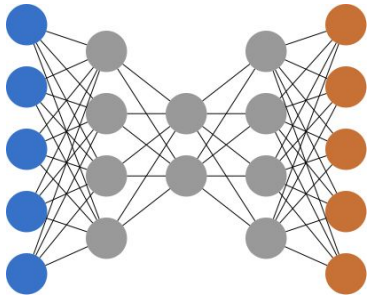


Cas d'usage : Aide au Hunting

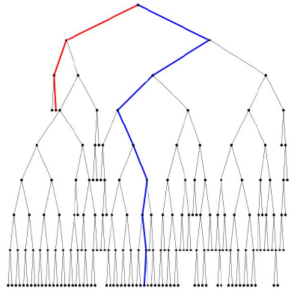
Détection dans les logs – Blue Team

Aider l'analyste à prioriser ses recherches, en identifiant les éléments anormaux :

- Extraction de caractéristiques
- Algorithmes non supervisés (AE, Forêt d'isolation)
- Utilisation de la structure intrinsèque des logs, pour y appliquer des modèles de langage (TAL) ou les transformer en graphe



Auto-encodeur



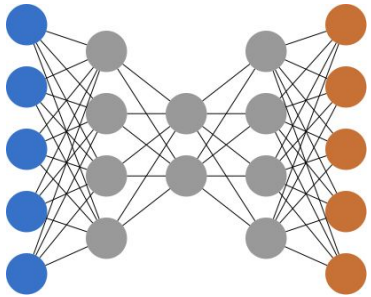
Forêt d'isolation

Cas d’usage : Aide au Hunting

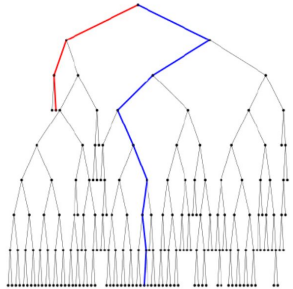
Détection dans les logs – Blue Team

Verrous technologiques :

- Explicabilité
- Réseaux complexes et dynamiques



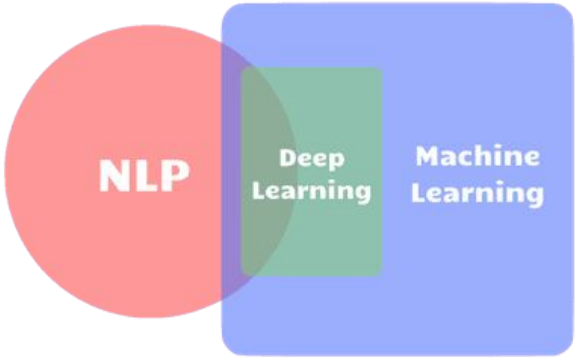
Auto-encodeur



Forêt d’isolation

Et le TAL ?

Traitement Automatique des Langues



Applications cyber :

- Traduction automatique
- Génération de texte (ex, phishing : SNAP_R)
- Détection de sentiments
- Reconnaissance de langues
- Génération de mots de passe (PassGAN)
- Identification de Domain Generation Algorithm
- Threat Intelligence (dark net mining…)

Et le TAL ?

GPT3

→ Avancées :

- Création de textes longs
- Génération de code

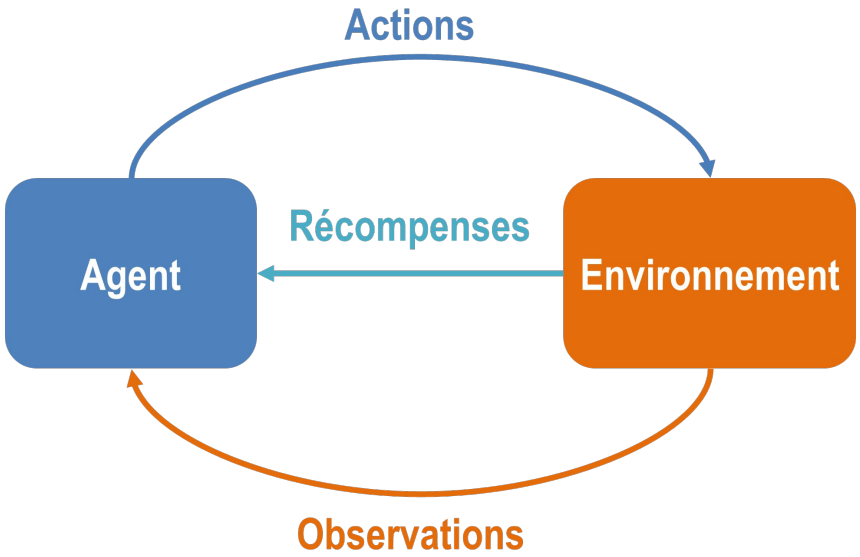
→ Verrous restants :

- Taille du modèle : 350 Go
- Propos racistes, haineux et sexistes
- Incohérences

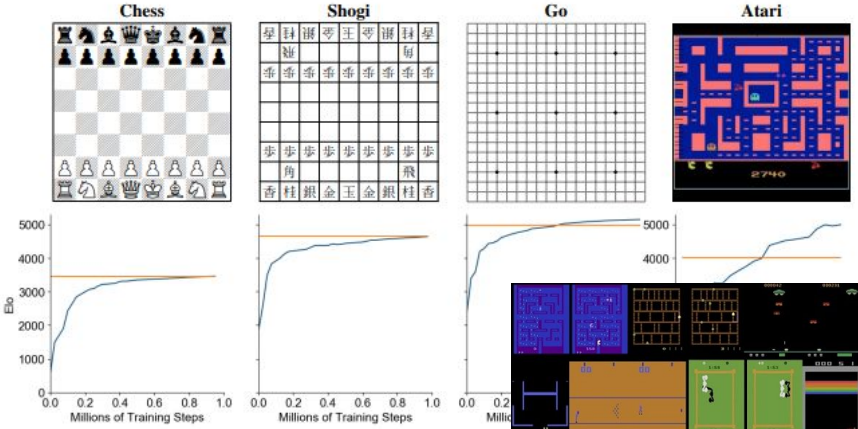
```
Q: How many eyes does a giraffe have?
A: A giraffe has two eyes.

Q: How many eyes does my foot have?
A: Your foot has two eyes.

Q: How many eyes does a spider have?
A: A spider has eight eyes.

Q: How many eyes does the sun have?
A: The sun has one eye.

Q: How many eyes does a blade of grass have?
A: A blade of grass has one eye.
```

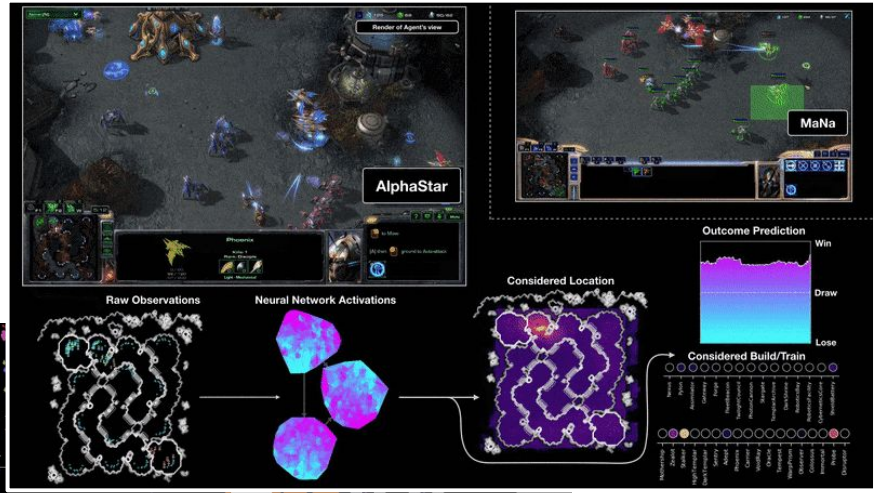
Et le RL ? (Reinforcement Learning)

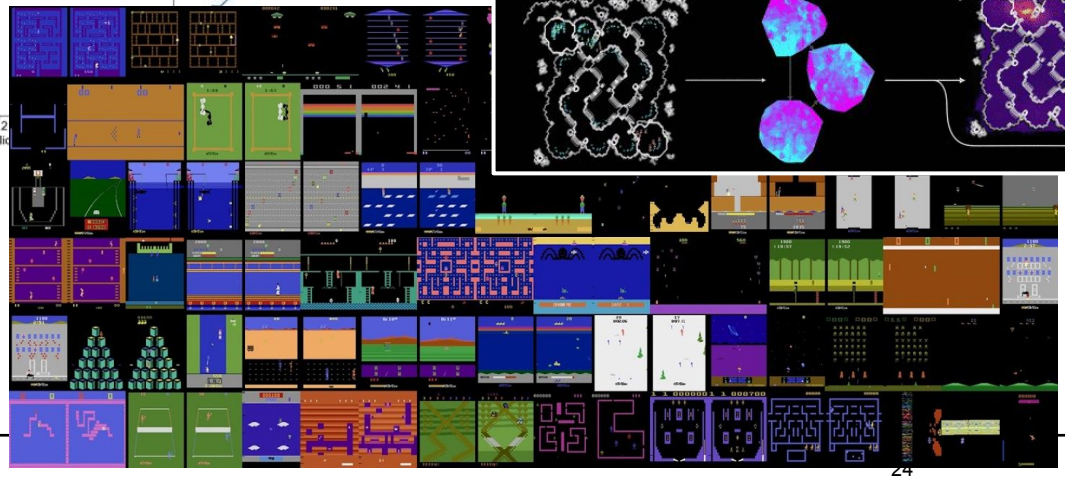
- Optimiser une politique d'action
- Apprendre en générant de la donnée
- Récompenses différées
- Capable de généraliser



Et le RL ? (Reinforcement Learning)







Et le RL ? (Reinforcement Learning)

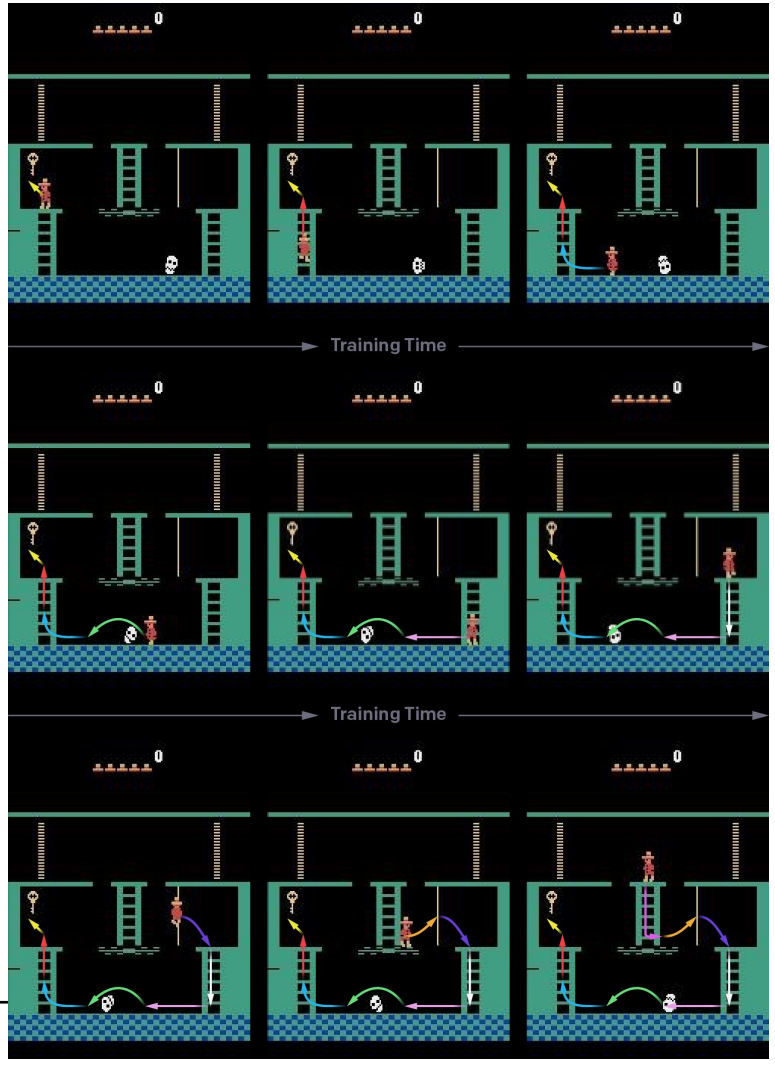
Nombreux cas d'usages cyber :

- Pentest automatisé
- Réaction automatique
- Configuration de réseaux
- Fuzzing / recherche de vulnérabilité
- Génération d'exploits
- Command & Control furtif

MINISTÈRE
DES ARMÉES
Liberté
Égalité
Fraternité

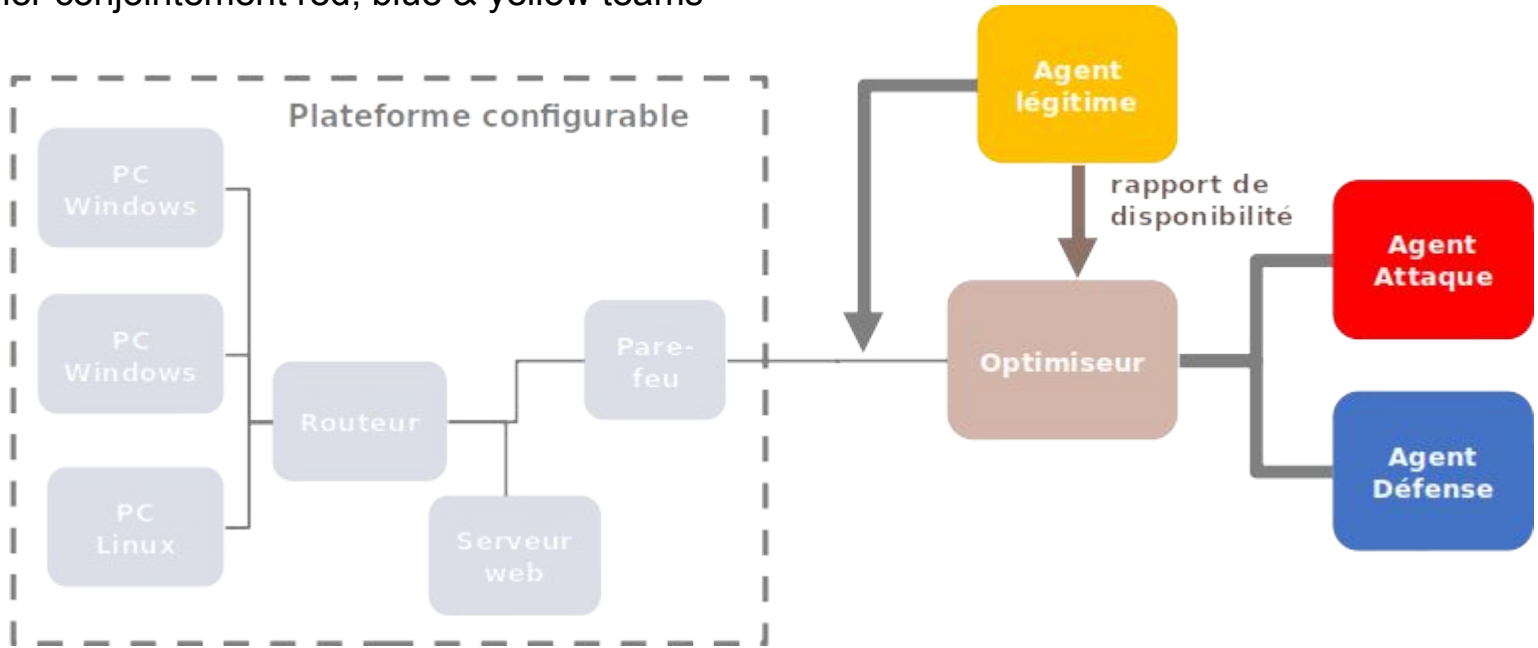
RL & Imitation Learning

- Apprentissage rapide
- Capitalisation
- Performance surhumaines



La perspective antagoniste

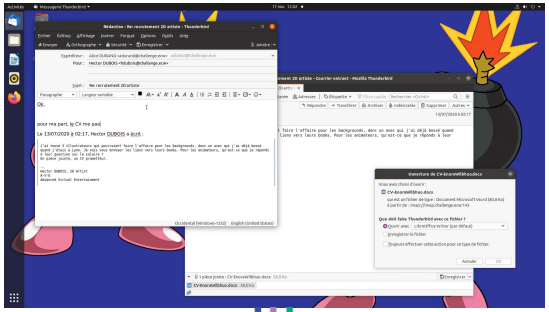
- Attaque et défense indissociables
- Simuler conjointement red, blue & yellow teams

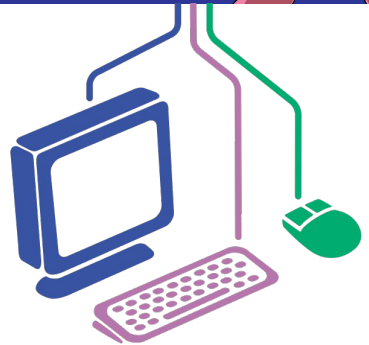


Challenge IA & Cyber

- Edition 2019 : Pentest automatisé
- **Edition 2020 : Simulation de vie**

  → 14 janvier 2021 : proclamation des résultats
- Edition 2021 : **?**





Grand Défi Cyber-sécurité

CONSEIL DE L'INNOVATION

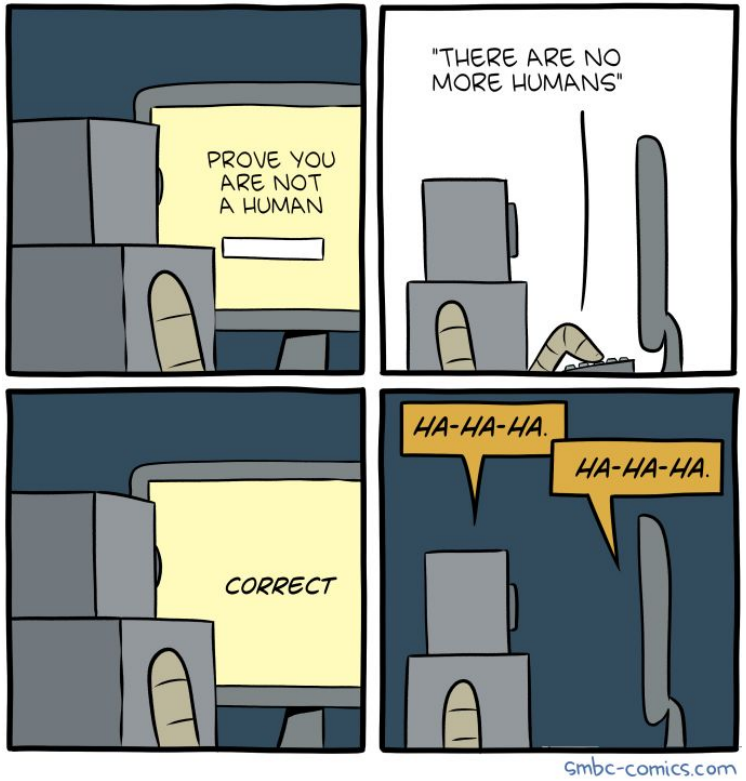
But : rendre nos systèmes durablement résilients aux cyber-attaques

Les axes du Grand Défi Cyber :

- Réseaux dynamiques
- Objets connectés
- Protection des PME contre la cybercriminalité
- Amorçage en cyber-sécurité
- Problématique des données cyber

Conclusion

- ML devenu incontournable
- Besoin de pluridisciplinarité
- Non-(semi-)supervisé
- Explicabilité & collaboration humain/machine essentiels



Ressources

- Appel à projet Cyber Défense Factory : https://www.defense.gouv.fr/aid/appels-a-projets/appel-a-projets-pour-la-cyber-defense-factory
- I am robot:(deep) learning to break semantic image captchas, Sivakorn, 2016
- No Bot Expects the DeepCAPTCHA! Introducing Immutable Adversarial Examples, Osadchy, 2017
- Machine learning based phishing detection from URLs, Sahingoz, 2019
- YaDiff : de l'intelligence artificielle pour comparer les codes binaires, Renault, MISC HS N°18, 2018
- Neural Message Passing for Quantum Chemistry, Gilmer, 2017
- Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples, Athalye, ICML 2018
- Poison Frogs! Targeted Clean-Label Poisoning Attacks on Neural Networks, Shafahi, NeurIPS 2018
- Learning to Deceive with Attention-Based Explanations, Pruthi, 2019
- Algorithms for hierarchical clustering: an overview, Murtagh, 2012
- Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter, Seymour, 2016
- Umap: Uniform manifold approximation and projection for dimension reduction, McInnes, 2018
- L'IA pour la détection, Technoférence #34 : Embarquement pour l'IA, Pôle Images et Réseaux
- Expérimentation et évaluation d'algorithmes de détection d'anomalies appliqués à des logs de proxy pour l'aide au hunting, Godefroy, C&ESAR 2018

Ressources

- Predicting domain generation algorithms with long short-term memory networks, Woodbridge, 2016
- A LSTM based framework for handling multiclass imbalance in DGA botnet detection, Tran, 2018
- Optimizing semantic LSTM for spam detection, Jain, 2019
- Transformers Are Better Than Humans at Identifying Generated Text, Maronikolakis, 2020
- PassGAN: A Deep Learning Approach for Password Guessing, NeurIPS 2018
- Mastering atari, go, chess and shogi by planning with a learned model, Schrittwieser, 2019
- Monte-Carlo tree search as regularized policy optimization, Grill, 2020
- Agent57: Outperforming the atari human benchmark, Badia, 2020
- Grandmaster level in StarCraft II using multi-agent reinforcement learning, Vinyals, Nature 2019
- Deep reinforcement learning for cyber security, Nguyen, 2019
- FuzzerGym: A Competitive Framework for Fuzzing and Learning, Drozd, 2018
- Online cyber-attack detection in smart grid: A reinforcement learning approach, Kurt, 2019
- Learning Montezuma's Revenge from a Single Demonstration, Salimans, 2018
- Artificially intelligent cyberattacks, Uppsala University, Zouave, 2020
- Challenge IA & Cyber 2020, European Cyber Week, https://www.challenge-ia-ecw.eu
- Grand Défi Cyber, https://www.gouvernement.fr