

IA, Machine Learning & Graphs

Benoit Simard (@logisima)

Bonjour a tous !

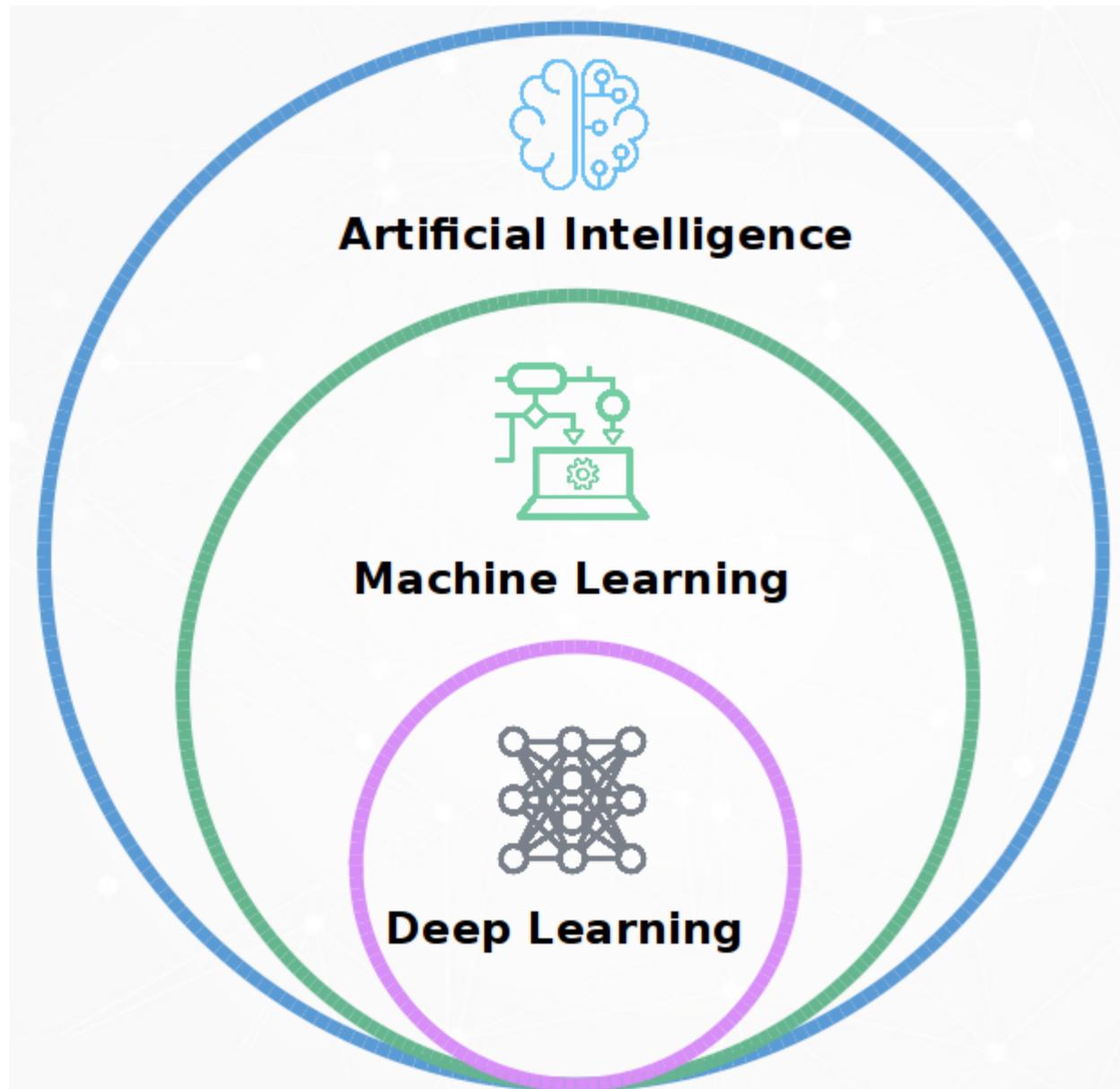
Benoit Simard



- Consultant Neo4j
- Addicte aux graphes
- Libérateur de données
- @logisima
- benoit@neo4j.com

Intelligence Artificielle

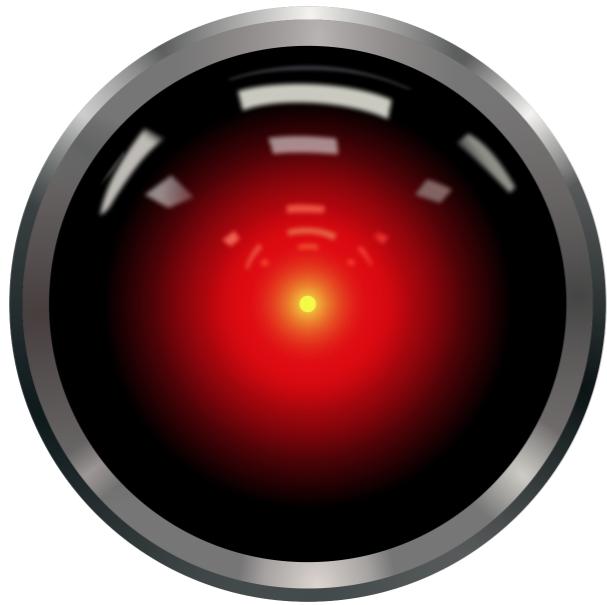
IA, ML, DL



L'IA est l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de **simuler l'intelligence**.

Le machine learning se base sur des approches **statistiques** pour donner aux ordinateurs la capacité d'apprendre à partir de données, c'est-à-dire d'améliorer leurs performances sans être **explicitement programmés**.

Le soulèvement des machines

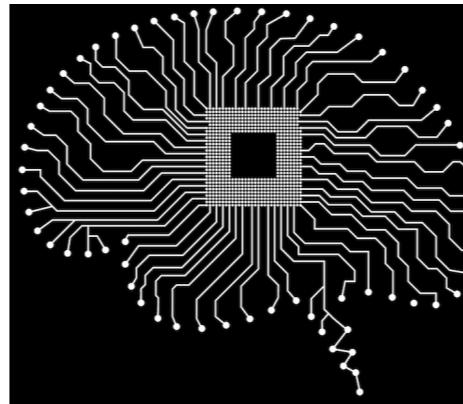


- Capable d'effectuer des millions de calculs
- Se souviennent de tout
- Grâce à ces capacités, elles simulent l'intelligence
- Hyper-spécialisées

C'est pas pour demain !

Pourquoi on parle d'IA aujourd'hui ?

L'intelligence artificielle en tant que domaine de recherche a été créé en **1959**



- **Deep learning** : 2012 - ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
- **Loi de Moore** : Puissance de calcul et la parallélisation
- **Big Data** : base de données volumineuse pour l'apprentissage
- **Eco-système** : Microsoft Cognitive Toolkit (2016, MIT), TensorFlow (2017, Apache), ...

Mise à contribution

Select all images with cars

The interface includes a 'Verify' button and links for 'Report a problem', 'Skip', and 'End'.

Report a problem

Verify

Skip

End

L'apprentissage humain

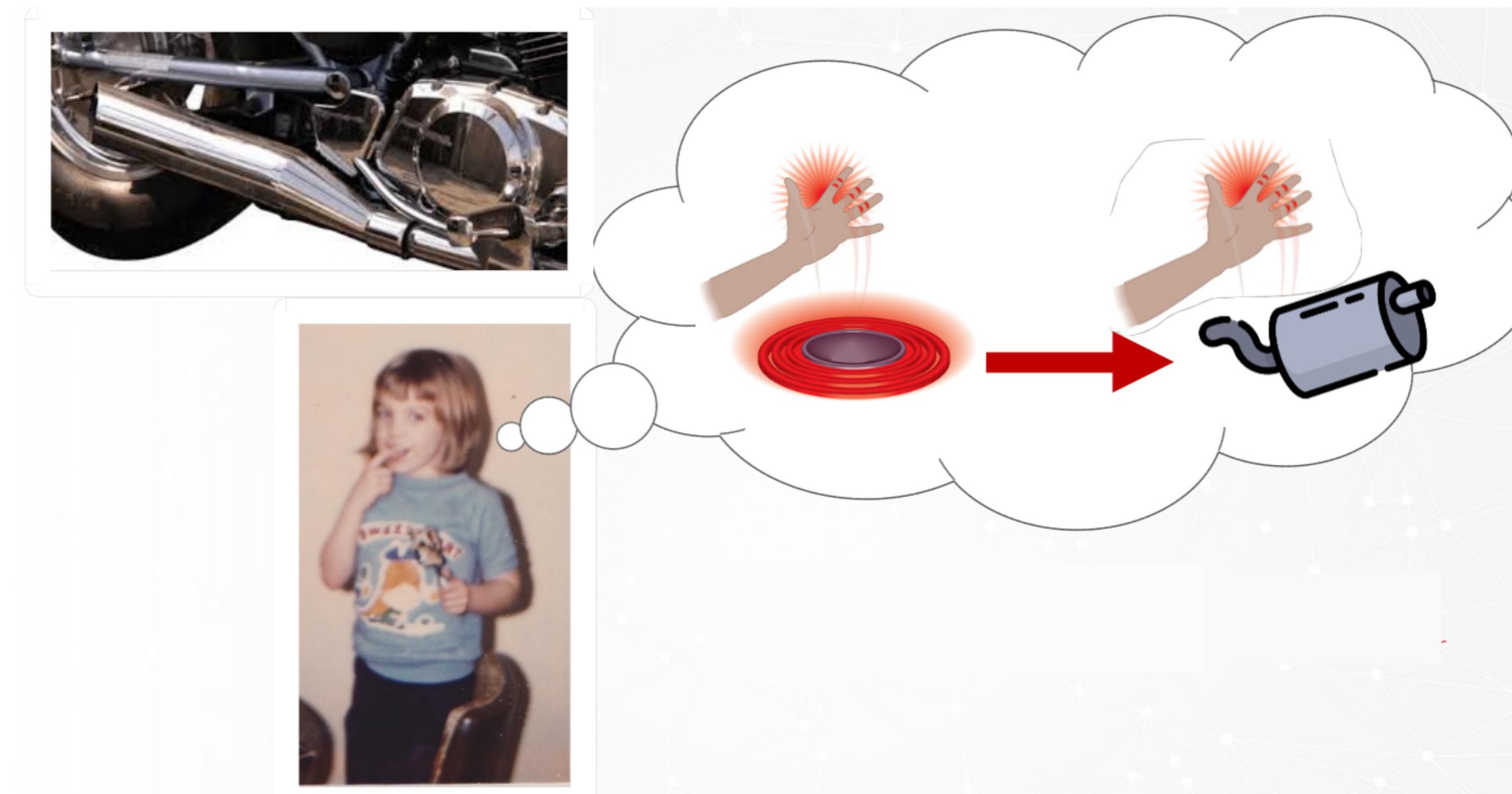


VS



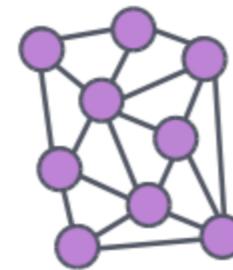
L'apprentissage contextuel

Une notion de "bon sens" ou de "connaissance naturelle"

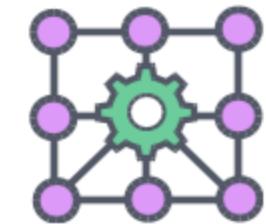


Le graphe & IA

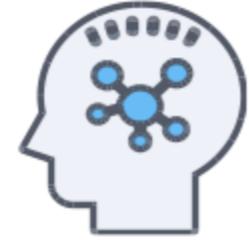
*mais on n'en est pas là ... **encore que ...***



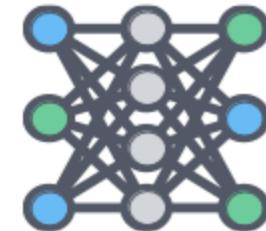
Modélisation et
Algorithmes



Accélère le ML



Graphe de
connaissance



Compréhension
des AI

Jours d'un avenir passé

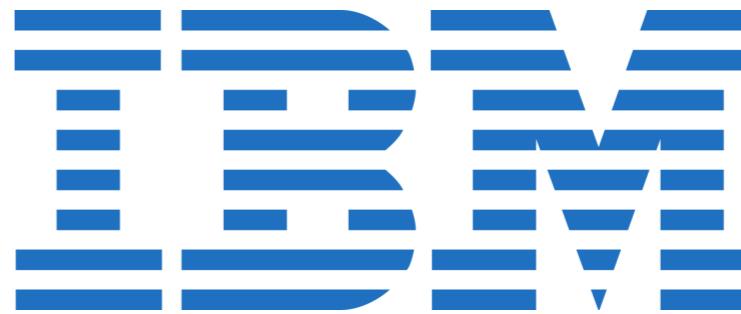


Deep Blue

1997 - Deep Blue vs Kasparov



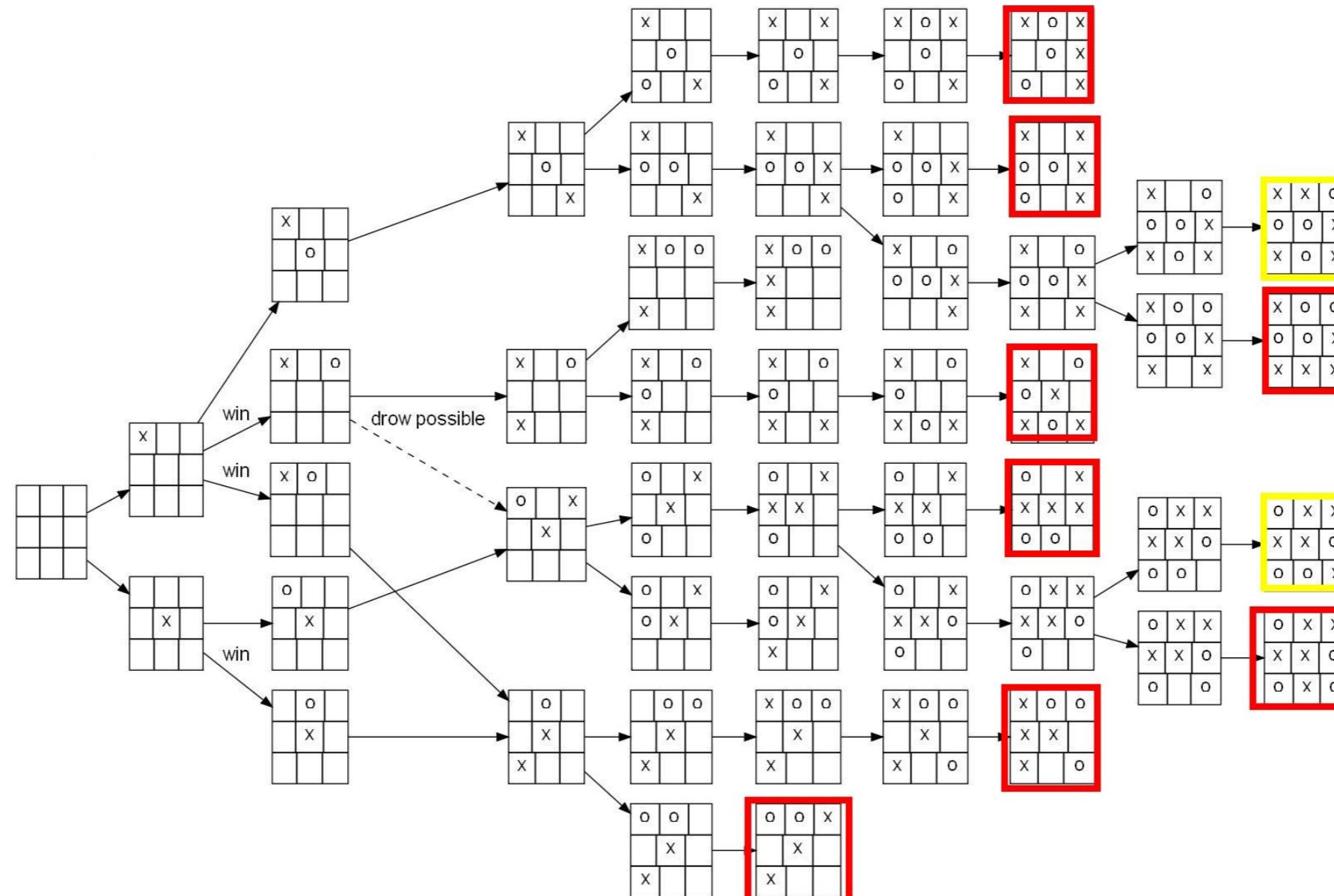
30 cores à 120 MHz, c'est le **259e** ordinateur le plus puissant en **1997**.



- **1996** : 2 - 4 (gagne une partie)
- **1997** : 3,5 - 2,5 (gagne 3 parties et le match)

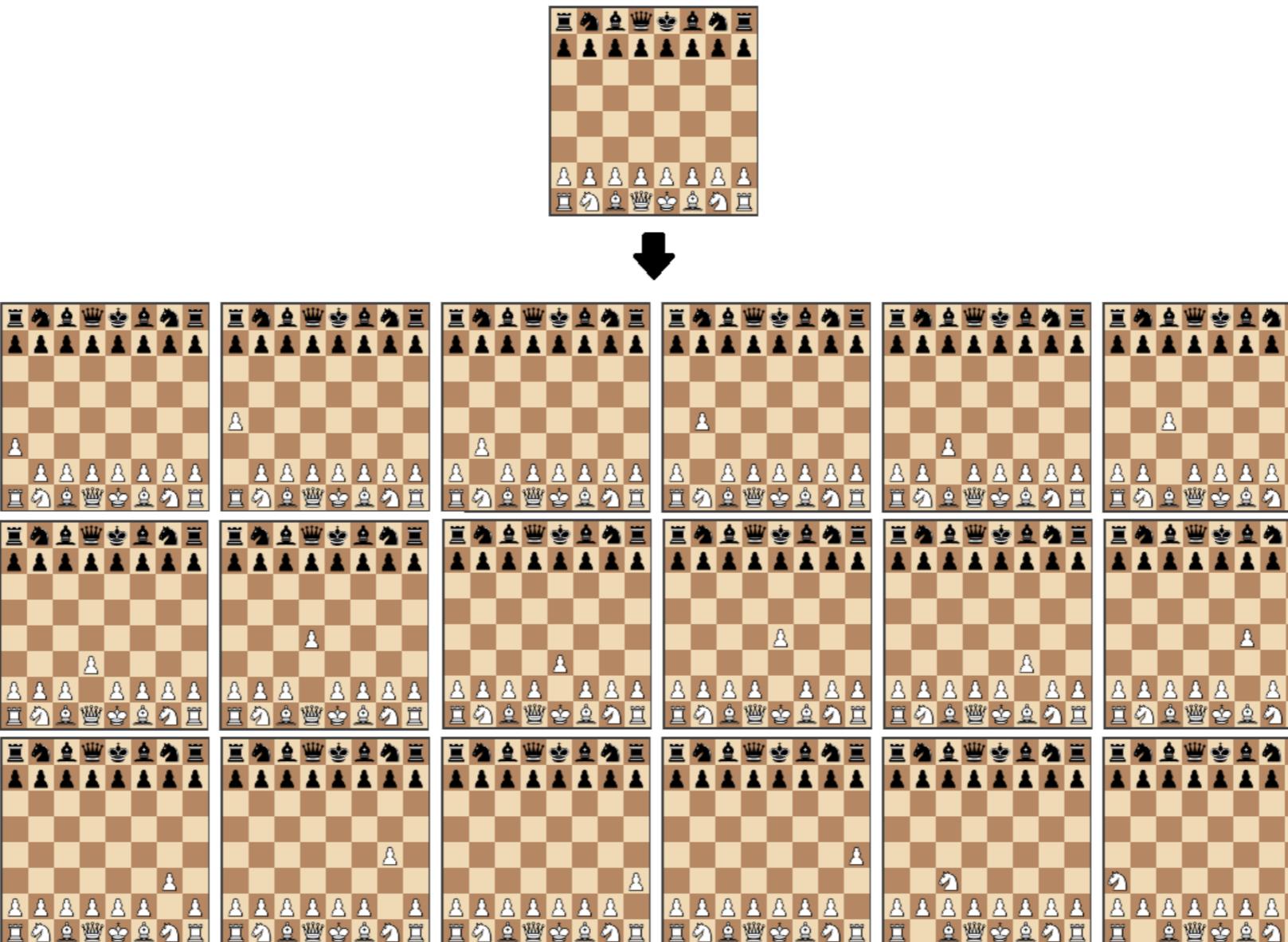
Brute Force - Morpion

Calcul de l'arbre de **tous les jeux** et utilisation de l'algorithme **Min-Max**



Le jeu d'échecs

Il y a environ **10¹²⁰** possible parties (10⁸⁰ est le nombre des particules dans l'univers).



L'algorithme

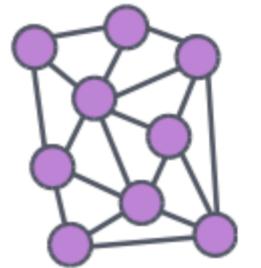


Génération de l'arbre de jeux (en profondeur) sur
les 20 prochains coups.

Mise en place d'une **fonction d'évaluation**
permettant de calculer la probabilité de gagner
selon l'état du plateau.

**SCORE = valeur des pièces blanches - valeur
des pièces noires**

Le graphe



Modélisation et
Algorithmes

- Modélisation de l'arbre
- Traversé du graphe pour les calculs statistiques

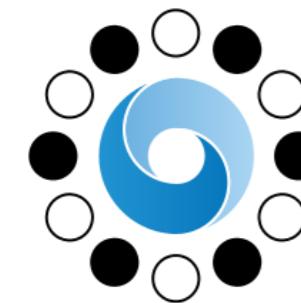
AlphaGO

Le jeu de GO

Beaucoup plus compliqué que les échecs, le nombre de parties possibles est estimé à **10⁶⁰⁰**



De 2015 à 2017



AlphaGo

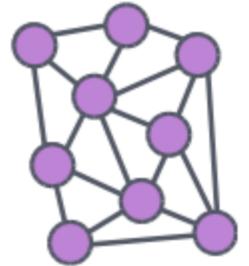
- **2015** : Gagne contre un joueur professionnel
- **2016** : Gagne contre le neuvième joueur mondial
- **2017** : Gagne contre le meilleur joueur mondial

L'algorithme

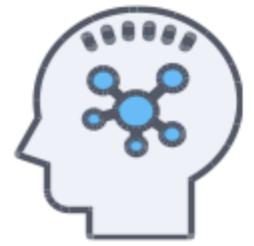


- Génération d'un l'arbre de jeux restreint
 - Sur les X prochains coups
 - Uniquement avec des coups "non-stupide" et de façon aléatoire
- Mise en place d'une fonction d'évaluation avec du **Machine learning**

Le graphe



Modélisation et
Algorithmes



Graphe de
connaissance



Accélère le ML

- Modélisation de l'arbre
- Traversé du graphe pour le calcul statistique
- Base de connaissance pour les coups "non-stupide"
- Stockage graphe des parties passées

AlphaGO Zero

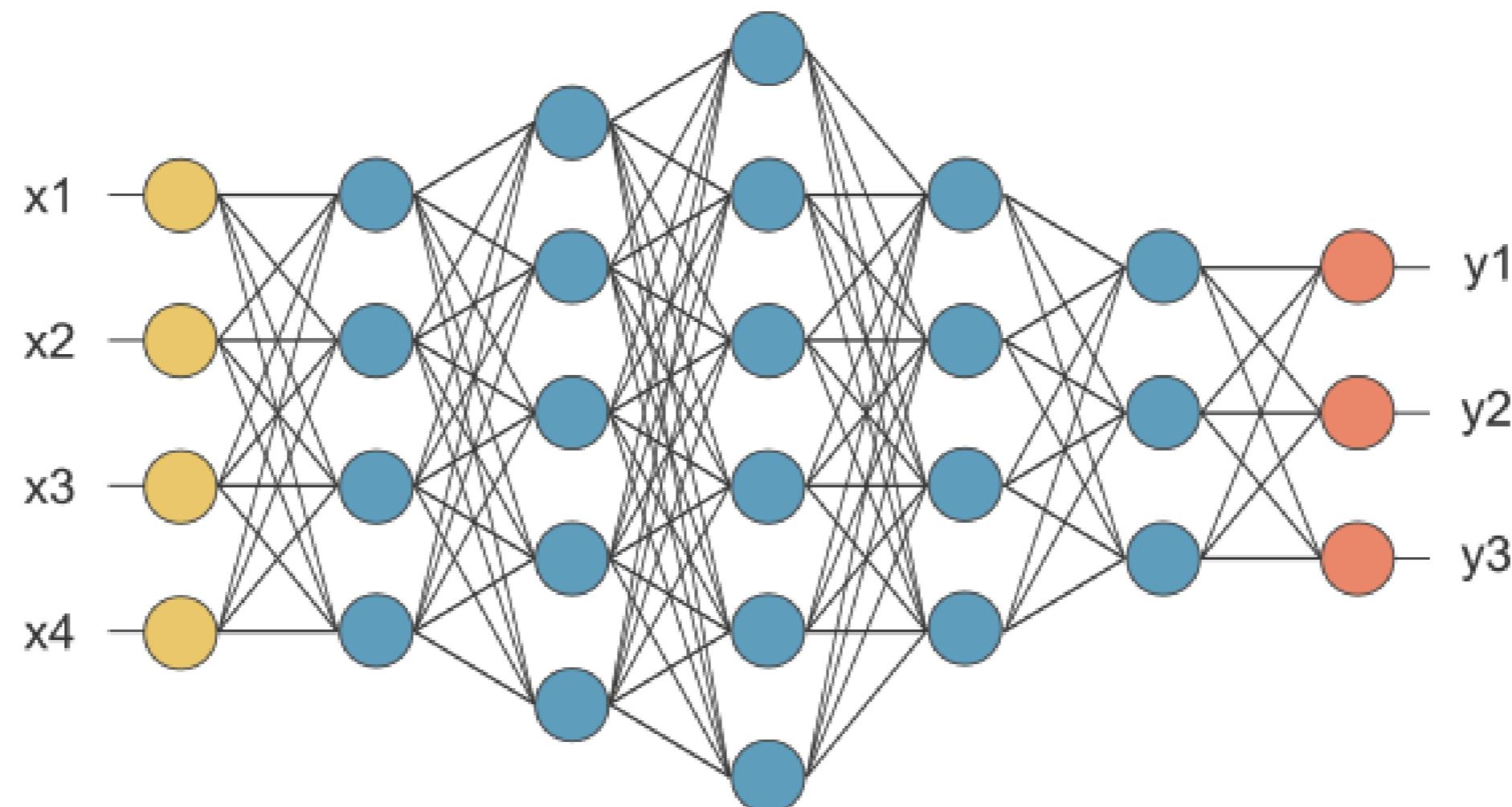
Version développée sans l'aide de données provenant de parties jouées, ni de connaissances autres que les règles du jeu.



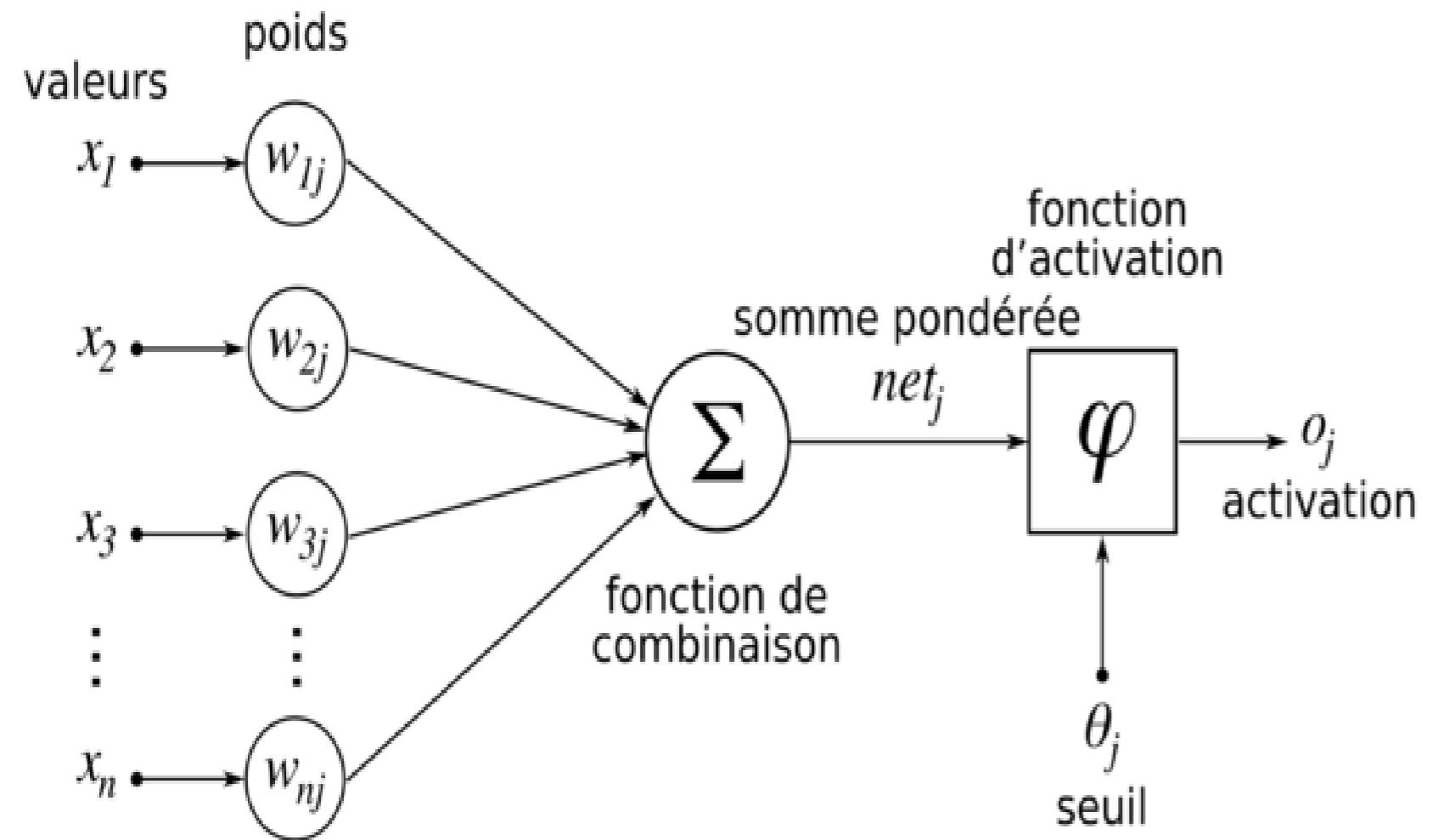
- Réseaux de neurones
- Apprentissage par renforcement (apprendre par l'expérience) en jouant contre lui-même

Les réseaux de neurones

Un réseau de neurones est composé d'une succession de couches dont chacune prend ses entrées sur les sorties de la précédente.



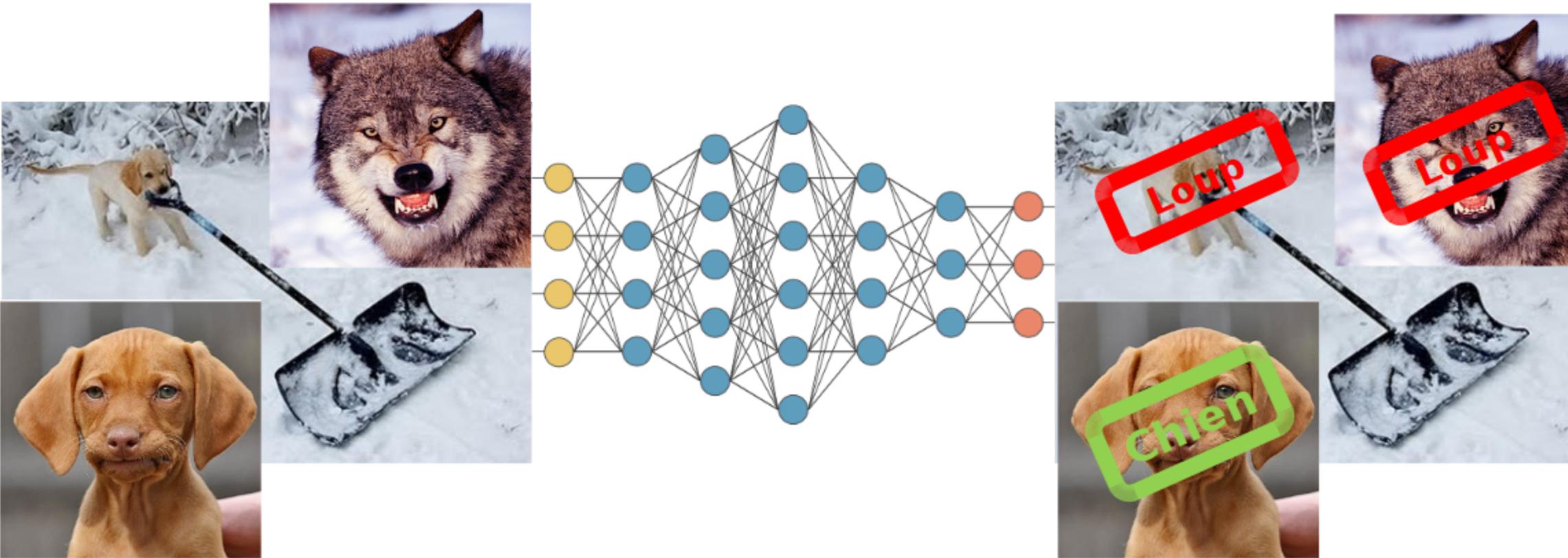
Les neurones



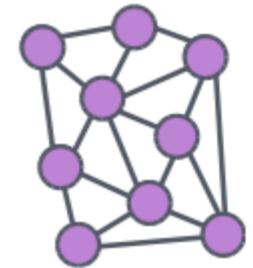
Paramétrage

- Poids synaptique (apprentissage)
- Fonction de combinaison
- Fonction d'activation
- Seuil d'activation

Compréhension / Prédiction



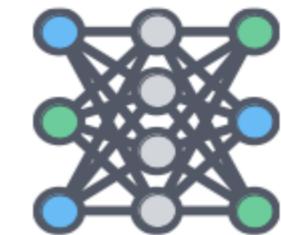
Le graphe



Modélisation et
Algorithmes



Accélère le ML



Compréhension
des AI

Compréhension des données

- Cartographier les données de l'apprentissage
- Générer de nouvelle données d'apprentissage par inférence

Compréhension de la prédiction

- Modélisation du réseaux avec son état
- Reproduire son comportement et de trouver les chemins utilisés

Chatbot

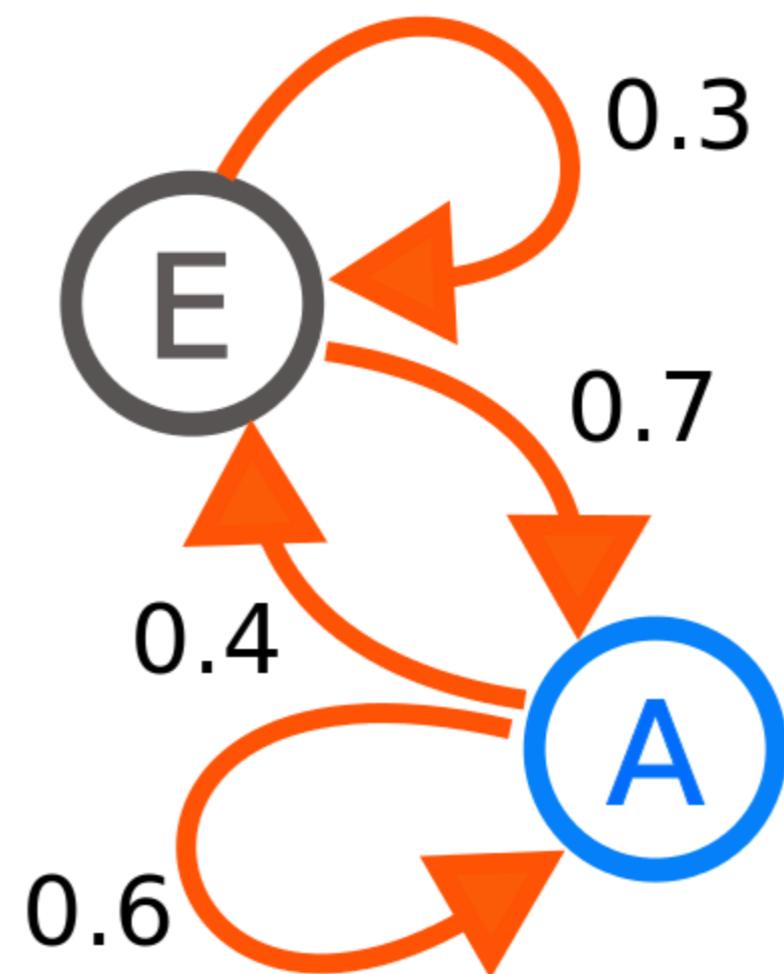
2016 - Microsoft Tay

Microsoft deletes 'teen girl' IA after it became a Hitler-loving sex robot within 24 hours.



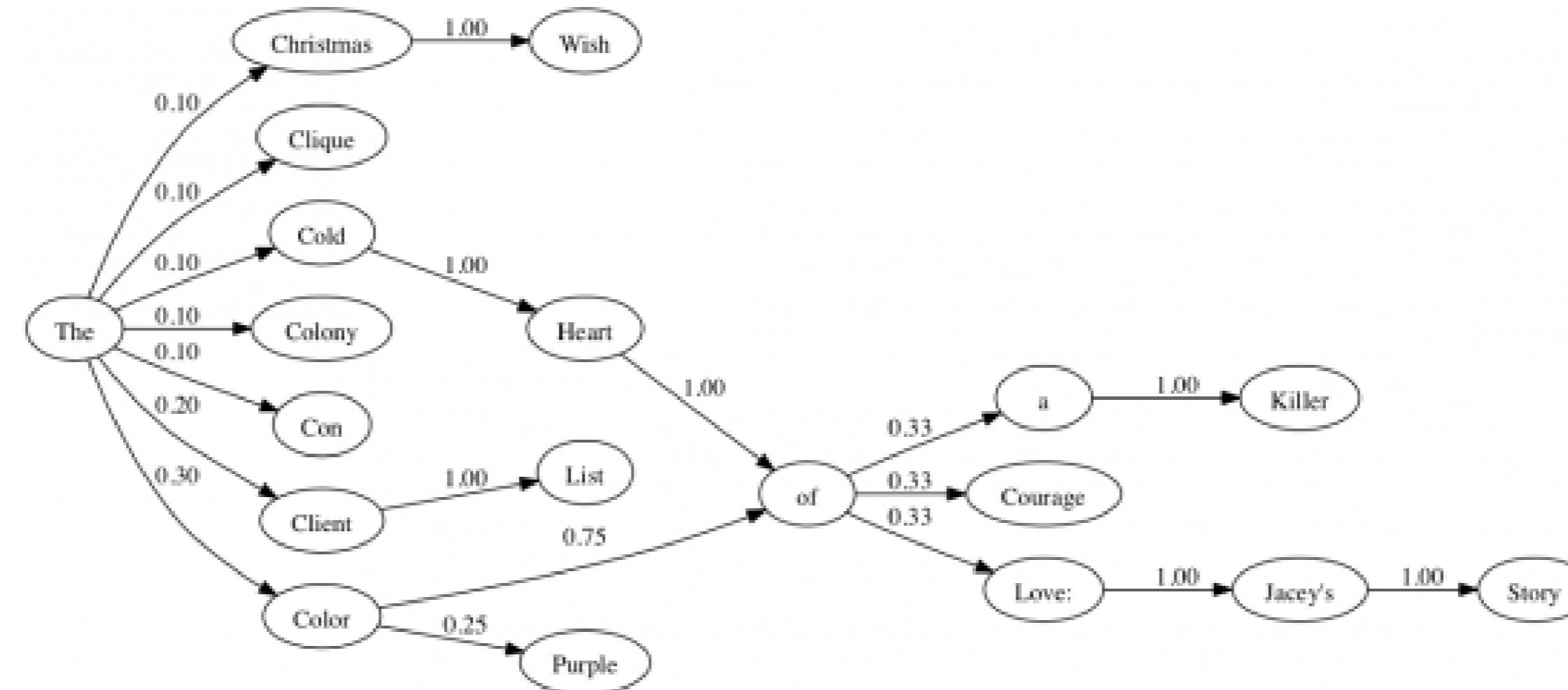
Les chaînes de Markov

L'information utile pour la **prédiction du futur** est entièrement contenue dans l'**état présent** du processus et n'est pas dépendante des états antérieurs



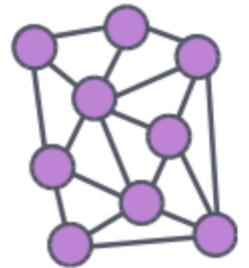
Apprentissage

Analyse des textes et mise à jour des statistiques dans le graphe.



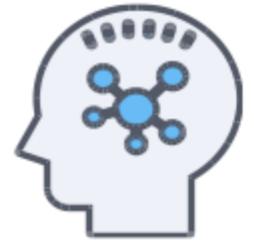
Un schéma de pensée par association d'idées ?

Le graphe



Modélisation et
Algorithmes

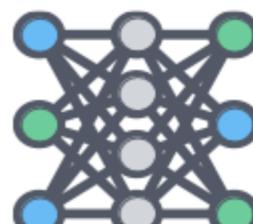
- Modélisation
- Stockage
- Traverser le graphe en temps réel
- Ajout de données sémantique (ontologie) pour différencier les homonymes et donner du contexte



Graphe de
connaissance



Accélère le ML



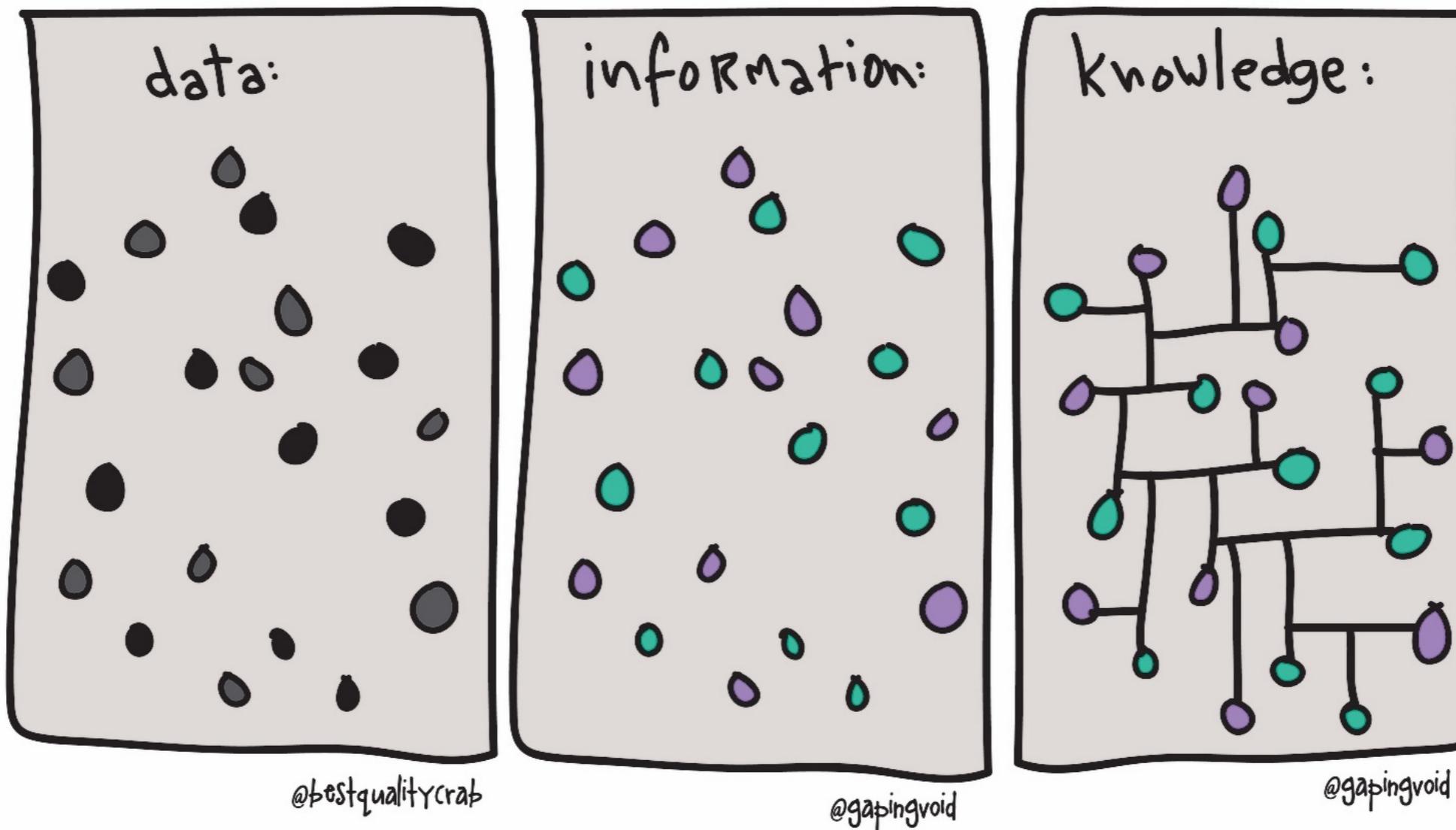
Compréhension
des AI



Les graphes sont partout en IA

Graphe de connaissance

Base de connaissance != Graphe de connaissance



Les algorithmes de graphe

Pathfinding & Search



- Parallel Breadth First Search & DFS
- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- Minimum Spanning Tree

- A* Shortest Path
- Yen's K Shortest Path
- K-Spanning Tree (MST)

Centrality / Importance



- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- PageRank

- Harmonic Closeness Centrality
- Dangalchev Closeness Centrality
- Wasserman & Faust Closeness Centrality
- Approximate Betweenness Centrality
- Personalized PageRank
- Balanced Triad (identification)
- Louvain - Multi-Step

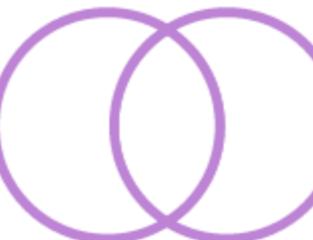
Community Detection



- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation
- Louvain Modularity - 1 Step

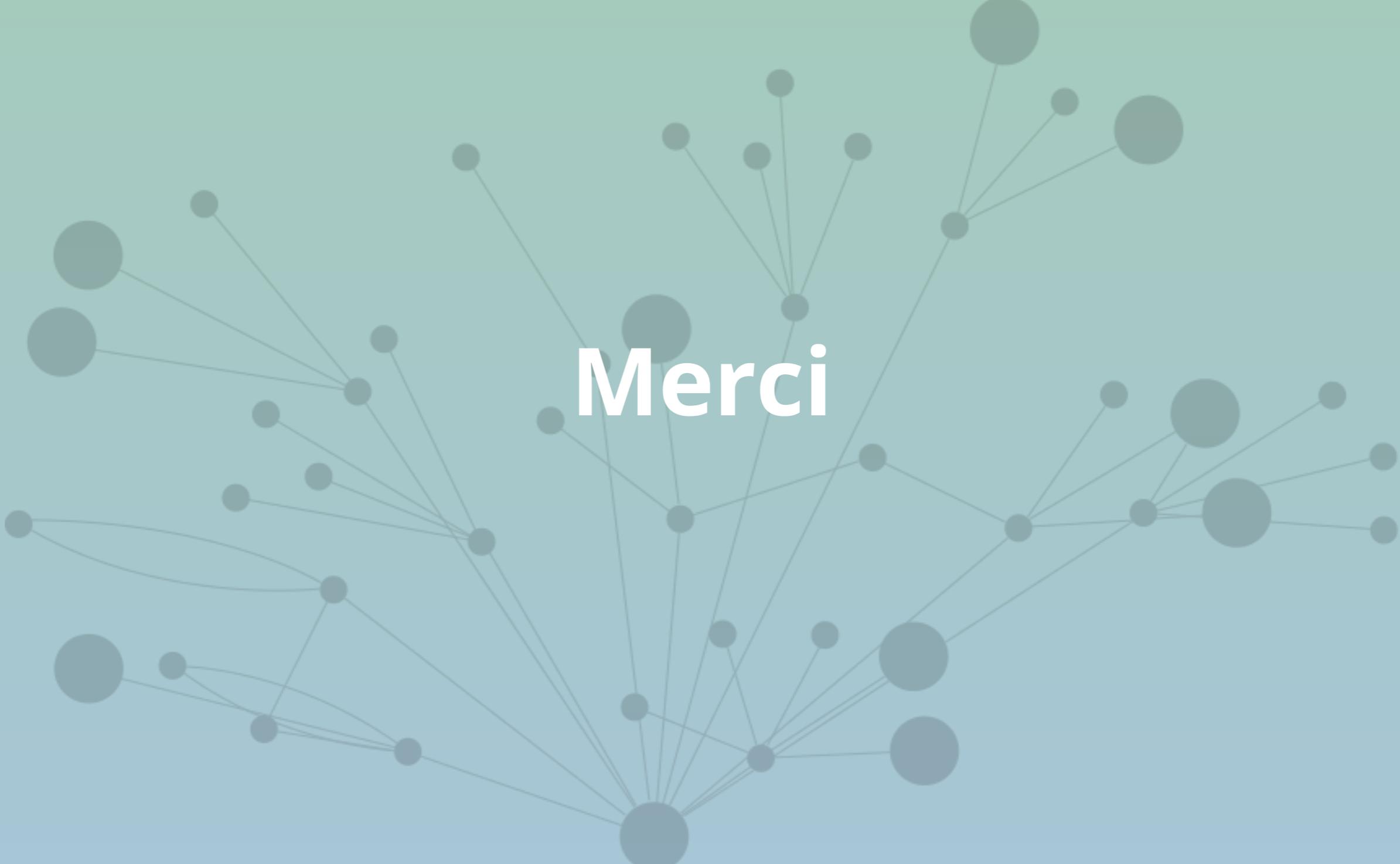
- **Euclidean Distance**
- **Cosine Similarity**
- **Jaccard Similarity**
- **Random Walk**
- **One Hot Encoding**

Similarity & ML Workflow



Le plus grand graphe du monde





Merci

Des questions ?

