

Introduction au machine learning

Partie 1 : Les modèles IA



Présenté par **Morgan Gautherot**

AMBIENT·IT

FORMATION - CERTIFICATION - EXPERTISE

FORMATION TECHNOLOGY

Mise à jour : 15/07/2024

formation@ambient-it.net



01 48 10 09 46



Propriété exclusive - Ambient IT ©





Organisme de formation

+850 formations en informatique

3 certifications éligibles CPF

- DevOps
- Cybersécurité
- IA
- Linux Foundation Partner
- Offensive Security Partner
- Microsoft Partner



Qui je suis ?



Morgan Gautherot, PhD ✓

Youtuber | Data Scientist Freelance

Paris, Île-de-France, France · [Coordonnées](#)

[Newsletter Data & AI](#) ↗

13 513 abonnés · Plus de 500 relations

Lille Big Data and Machine Learning Meetup

★★★★★ (363) ?

📍 Lille, France

👤 2566 membres · Groupe public ⓘ



AI for you - Morgan Gautherot

@AiforyouMorganGautherot · 7,54 k abonnés · 174 vidéos

A travers cette chaîne YouTube, je mets à disposition du contenu de qualité. ...plus

fr.tipeee.com/aiforyou et 2 autres liens

Personnaliser la chaîne

Gérer les vidéos



Tour de table



PRESENTATION

Qui est tu ? Où travailles-tu ?



TES COMPETENCES

Que fais tu dans ton quotidien
professionnel ?



TES OBJECTIFS

Qu'est ce que tu attends de
cette formation ?



Programme

- Jour 1 :
 - Introduction au machine learning,
 - L'apprentissage supervisé,
 - Les KNN,
 - Les métriques de performance,
 - Gestion d'un projet ML,
 - Modèle de régression.



Programme

- Jour 2 :
 - Les arbres de décision,
 - L'apprentissage non-supervisé,
 - Le K-means,
 - La réduction de dimension,
 - PCA, t-SNE.

Introduction à l'IA



Partie 1 : Les modèles d'IA



Technologie de rupture

- L'ordinateur
- L'internet
- Le mobile
- L'intelligence artificielle

"Mieux vaut prendre le changement par la main avant qu'il ne vous prenne à la gorge"
Winston Churchill



Rendre votre entreprise plus efficace

Améliorer l'ensemble de vos procédures.

Cela touche toutes vos activités:

- Achats,
- Ressources humaines,
- Distributions,
- Productions,
- ...



La digitalisation au service de l'IA

Courrier papier

Information sur la réception

Email

L'ouverture

Les clics

Temps de lecture

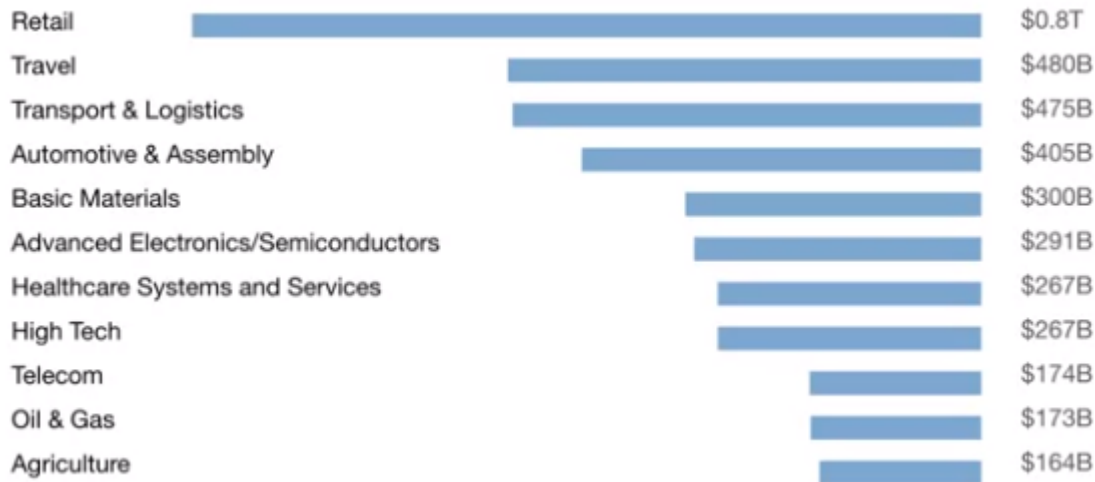
Pourcentage de l'email lu

...



Impact de l'IA

- Création de valeur par l'IA d'ici 2030

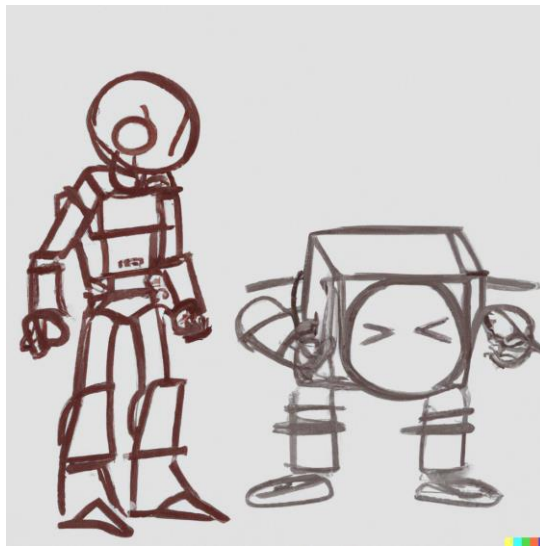


13 trillions de dollars



IA dans les grandes lignes

Domaine qui permet aux ordinateurs de copier le comportement humain





Intelligence Artificielle VS Intelligence Artificielle Générale

Artificial Narrow Intelligence



Cat

V.S.

Intelligence Générale Artificielle



- Résoudre un problème très ciblé
- Ne peut pas être adapté à d'autres problèmes

Faire tout ce qu'un humain peut faire



Données structurées

Structured		
1 0 0 1 1 0 1 0 1 1 1 0	1 0 0 1 1 1 1 0 0 1 1 0	0 1 0 1 1 1 1 0 1 0 0 0
1 0 0 1 1 0 1 0 1 0 0 0	1 0 0 1 0 0 1 0 1 1 1 0	1 0 0 1 1 0 1 0 1 1 1 0
1 0 0 1 1 0 1 0 1 1 1 0	1 0 0 1 1 0 1 0 1 0 0 0	0 1 0 1 1 1 1 0 1 0 0 0



Exemples d'applications

- Prédiction du prix des maisons
- Prédiction du contrôle d'un colis
- Détermination de la performance énergétique (DPE) d'un logement
- Diagnostic automatique des maladies
- Calcul du score d'appétance d'un produit
- Calcul du score de risque
- Segmentation d'une base clients



Traitement du langage naturel





Exemples d'applications

- Auto-completion
- Traduction automatique
- Question/Answering
- Chatbot
- Résumé de text
- Analyse de sentiments
- Détection de commentaires toxiques



Vision par ordinateur





Exemples d'applications

- Validation automatique de documents
- Extraction d'information à partir d'un PDF
- Segmentation automatique
- Conduite autonome
- FaceID
- Détection de casses sur une chaîne de montage
- Détection de comportements anormaux



L'analyse du son



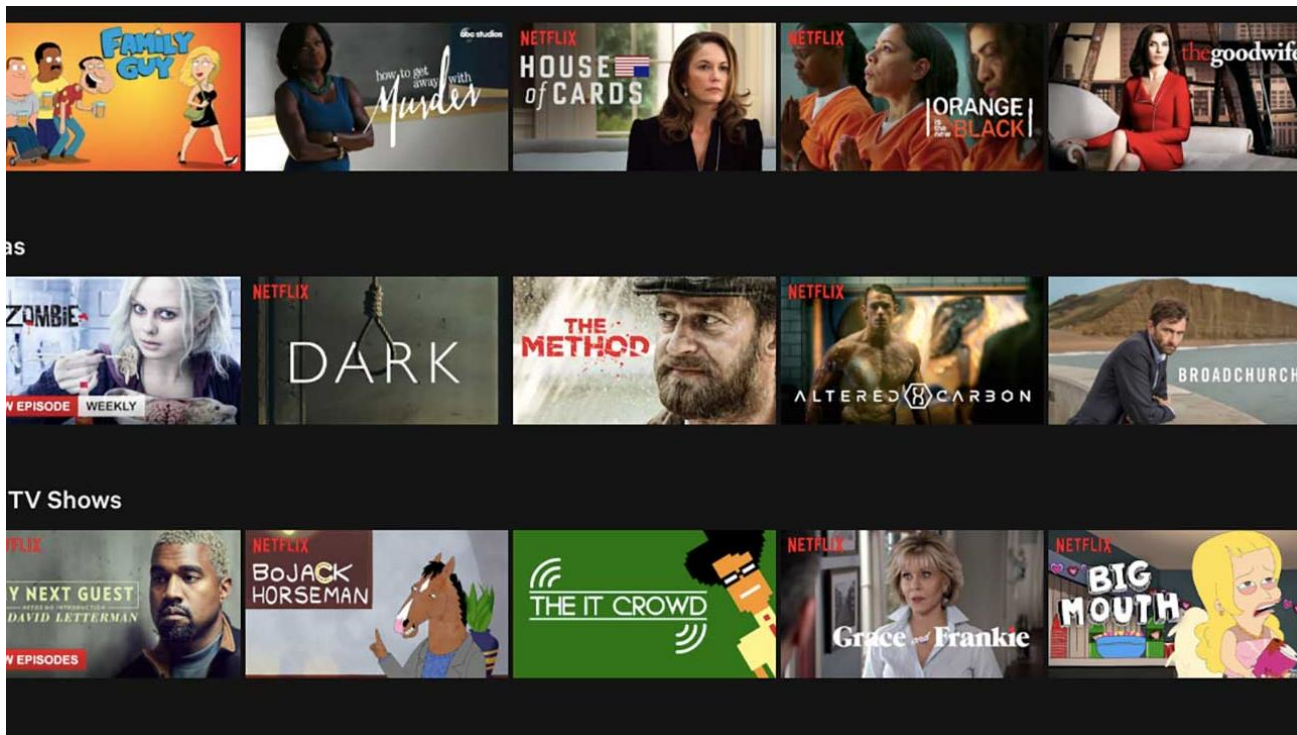


Exemples d'applications

- Transcription d'une conversation audio
- Traduction automatique
- Composition automatique
- Augmentation de la qualité audio
- Voix de synthèse
- Détection de tags pour une conversation audio



Système de recommandation





Exemples d'applications

- Catalogue personnalisé
- Accueil personnalisé
- Prochain visionnage
- Upsell
- Promotion personnalisée



Apprentissage par renforcement





Exemples d'applications

- Conduite autonome
- Jouer à Tetris
- Trouver de nouveaux algorithmes

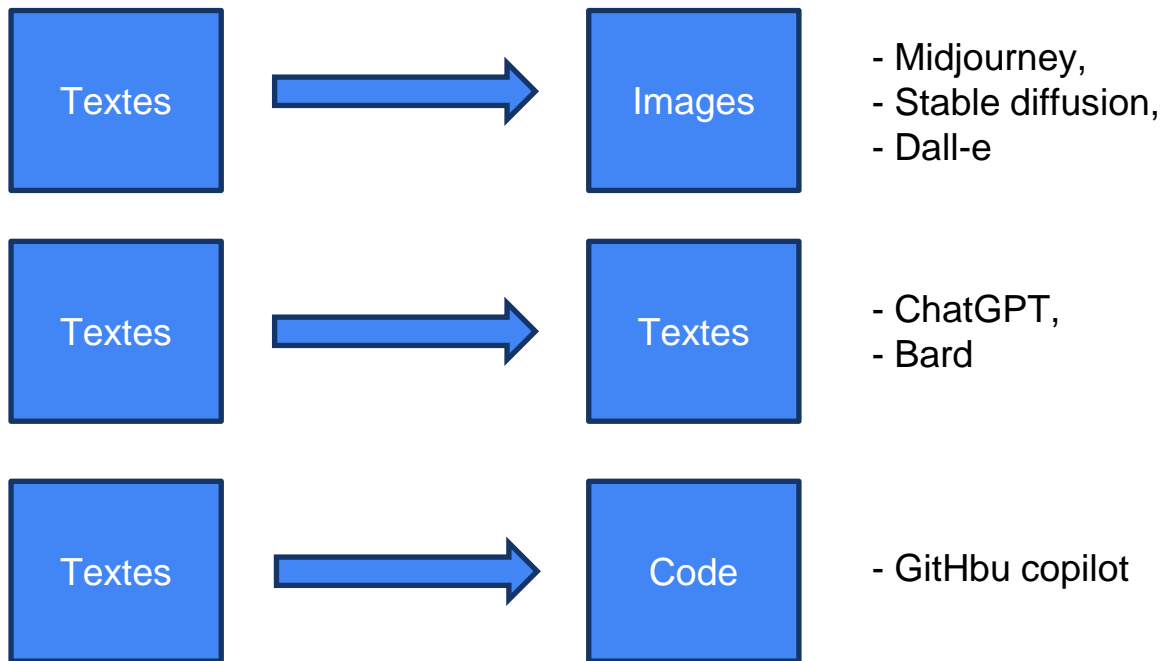


IA génératif





Passage d'une modalité à l'autre





La réalité terrain

- Données structurées (80%)
- Traitement naturel du langage (10%)
- Vision par ordinateur (8%)
- Analyse du son (0.01%)
- Système de recommandation (1.99%)
- Apprentissage par renforcement (0%)
- IA génératif (?)



Principales applications

- Economie de temps
 - Planning automatique / Diagnostique automatique



Planning d'un service d'urgence

Créer un planning qui minimise le besoin d'agents en respectant les contraintes.

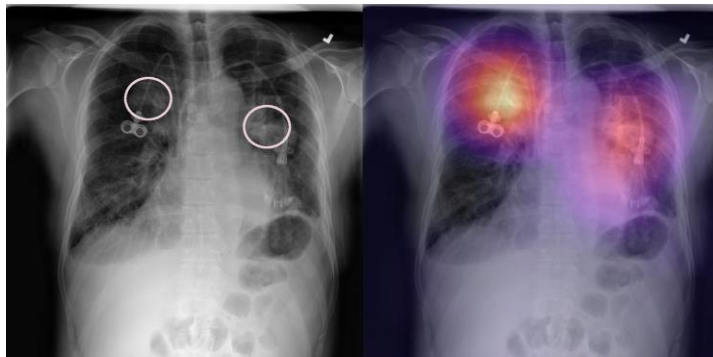
Contraintes:

- Minimum de 2 jours consécutif
- Maximum de 5 jours consécutif
- Avoir toujours au moins 3 agents en poste
- Poste de 8h ou 10h

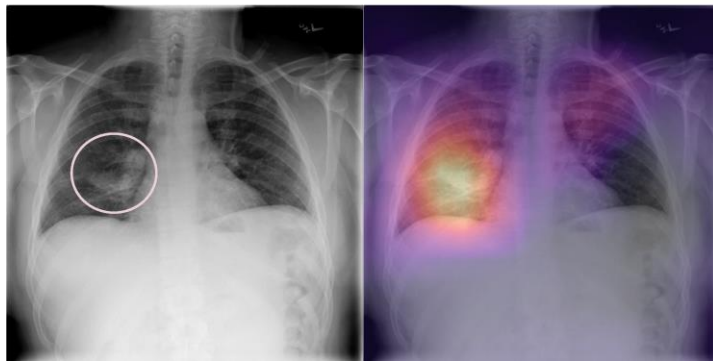


Diagnostic automatique

a



b





Diagnostique automatique

Pathology	Radiologists (95% CI)	Algorithm (95% CI)	Algorithm – Radiologists Difference (99.6% CI) ^a	Advantage
Atelectasis	0.808 (0.777 to 0.838)	0.862 (0.825 to 0.895)	0.053 (0.003 to 0.101)	Algorithm
Cardiomegaly	0.888 (0.863 to 0.910)	0.831 (0.790 to 0.870)	−0.057 (−0.113 to −0.007)	Radiologists
Consolidation	0.841 (0.815 to 0.870)	0.893 (0.859 to 0.924)	0.052 (−0.001 to 0.101)	No difference
Edema	0.910 (0.886 to 0.930)	0.924 (0.886 to 0.955)	0.015 (−0.038 to 0.060)	No difference
Effusion	0.900 (0.876 to 0.921)	0.901 (0.868 to 0.930)	0.000 (−0.042 to 0.040)	No difference
Emphysema	0.911 (0.866 to 0.947)	0.704 (0.567 to 0.833)	−0.208 (−0.508 to −0.003)	Radiologists
Fibrosis	0.897 (0.840 to 0.936)	0.806 (0.719 to 0.884)	−0.091 (−0.198 to 0.016)	No difference
Hernia	0.985 (0.974 to 0.991)	0.851 (0.785 to 0.909)	−0.133 (−0.236 to −0.055)	Radiologists
Infiltration	0.734 (0.688 to 0.779)	0.721 (0.651 to 0.786)	−0.013 (−0.107 to 0.067)	No difference
Mass	0.886 (0.856 to 0.913)	0.909 (0.864 to 0.948)	0.024 (−0.041 to 0.080)	No difference
Nodule	0.899 (0.869 to 0.924)	0.894 (0.853 to 0.930)	−0.005 (−0.058 to 0.044)	No difference
Pleural thickening	0.779 (0.740 to 0.809)	0.798 (0.744 to 0.849)	0.019 (−0.056 to 0.094)	No difference
Pneumonia	0.823 (0.779 to 0.856)	0.851 (0.781 to 0.911)	0.028 (−0.087 to 0.125)	No difference
Pneumothorax	0.940 (0.912 to 0.962)	0.944 (0.915 to 0.969)	0.004 (−0.040 to 0.051)	No difference

^aThe AUC difference was calculated as the AUC of the algorithm minus the AUC of the radiologists. To account for multiple hypothesis testing, the Bonferroni-

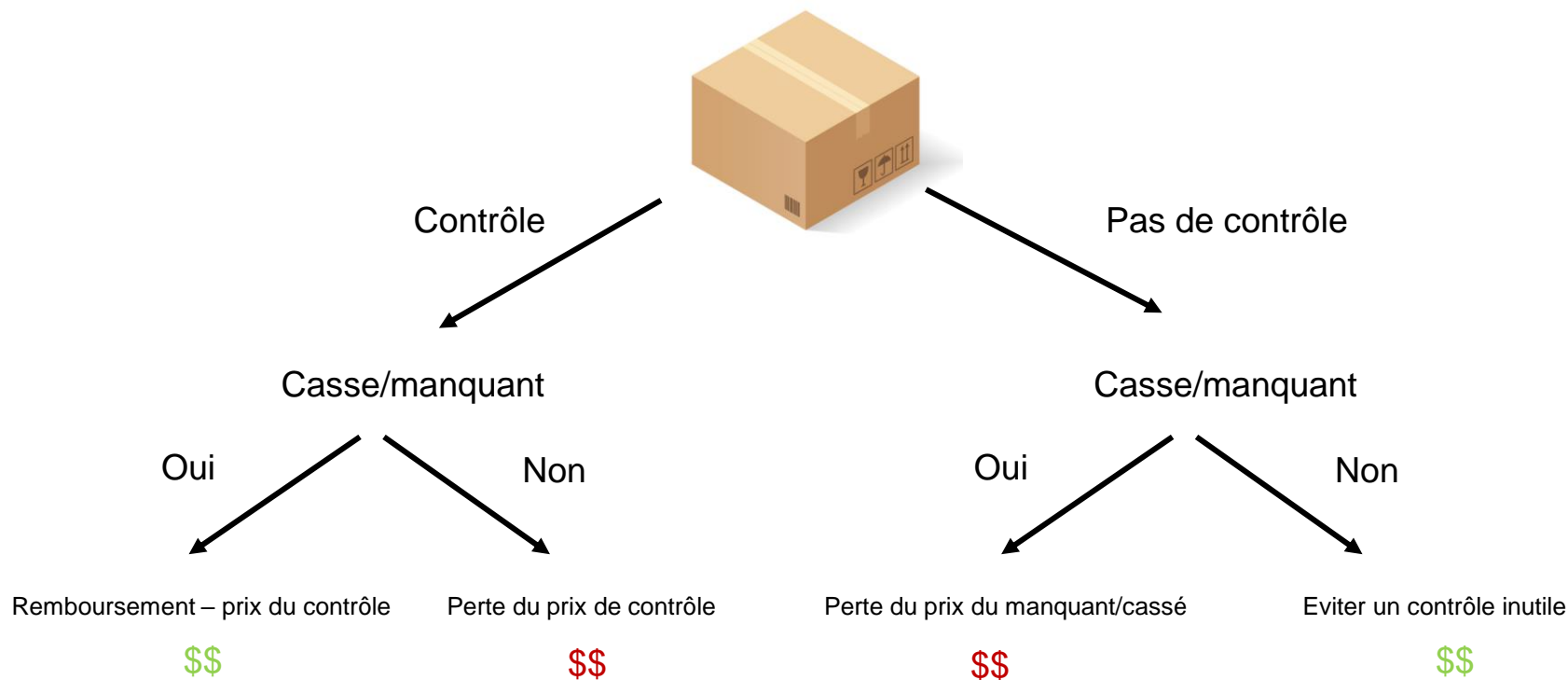


Principales applications

- Economie de temps
 - Planning automatique
- Economie argent
 - Scoring



Contrôle des colis





Résultat

- Processus existant:
 - 1€ investi en contrôle -> 0.3€ de gain
- Utilisation du machine learning:
 - 1€ investi en contrôle -> 1.5€ de gain
- + réduction du temps de contrôle

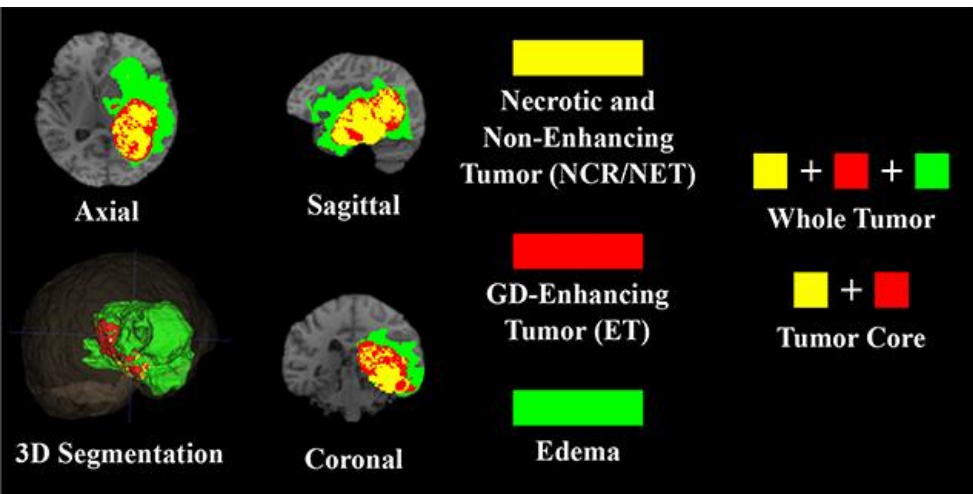


Principales applications

- Economie de temps
 - Planning automatique
- Economie argent
 - Scoring
- Augmentation de performances
 - Segmentation automatique

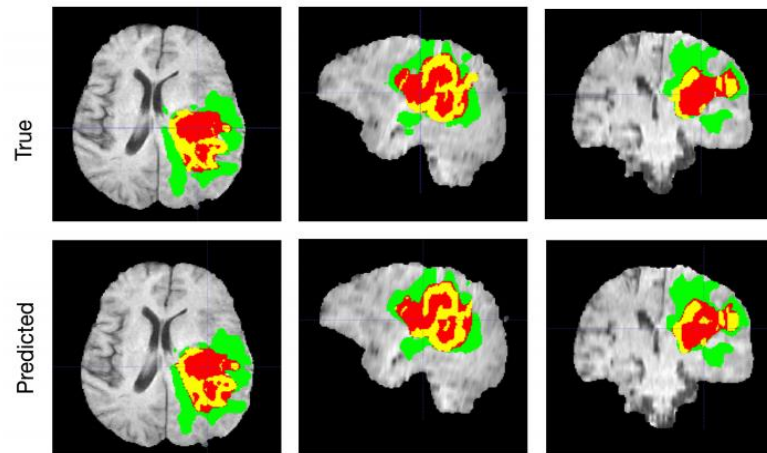


Segmentation de tumeur cérébrale



Segmentation manuelle de tumeur par des radiologues expert:

- La variabilité intra-opérateur était de $20\% \pm 15\%$.
- La variabilité inter-opérateur était de $28\% \pm 12\%$.



Expert system



Partie 1 : Les modèles IA



Definitions

- L'intelligence artificielle (IA) est un domaine de l'informatique dont l'objectif est de créer des programmes qui accomplissent des tâches normalement dévolues à l'intelligence humaine et de la simuler.

Intelligence artificielle





L'IA à travers le temps

1957



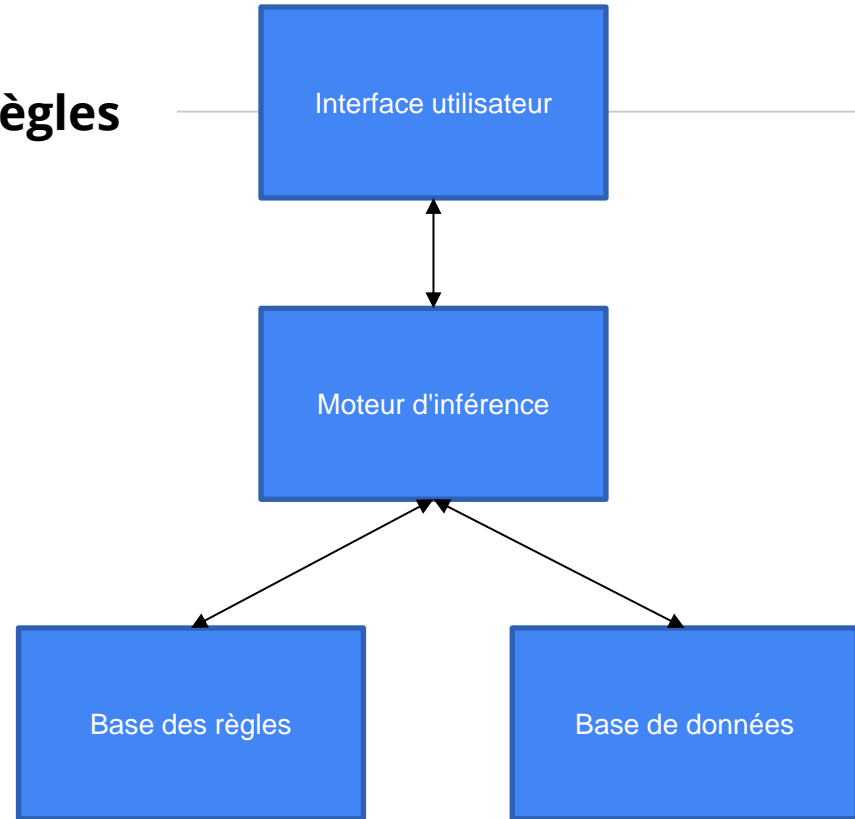
Système expert





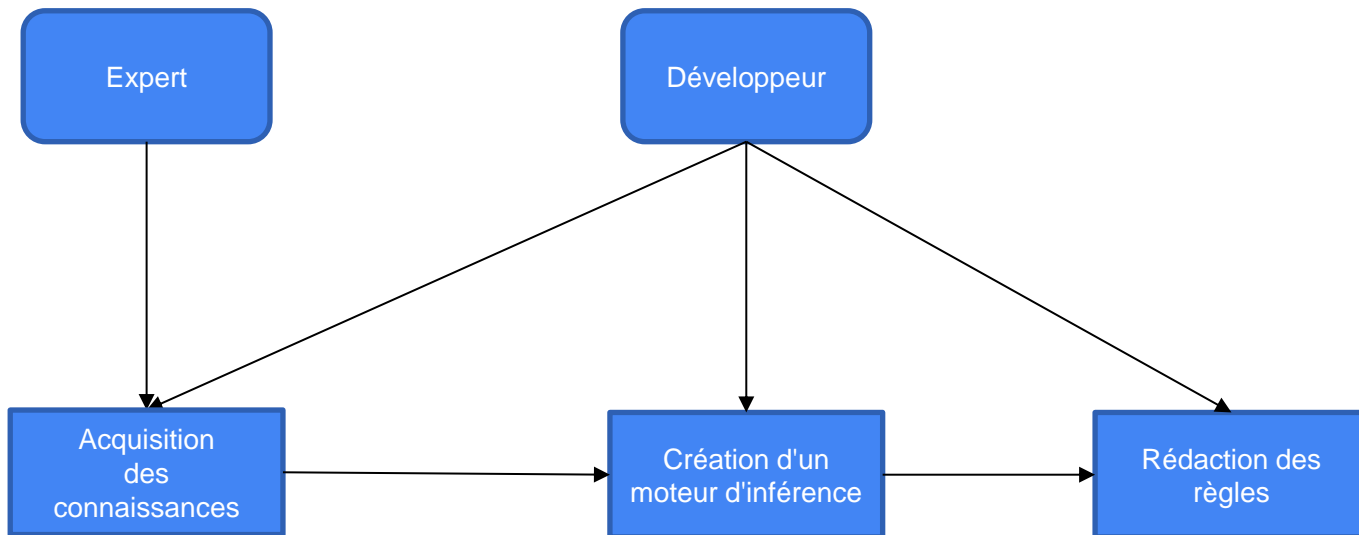
Système basé sur des règles

- Un système expert est un programme qui répond à des questions ou résoud des problèmes dans un domaine de connaissances donné, en utilisant des règles logiques dérivées des connaissances d'experts humains dans ce domaine.





Construire un système





Applications

- Diagnostic de l'analyse du sang
- Diagnostic de panne pour les voitures
- Vérification des systèmes de l'avion avant le décollage
- Ligne d'assemblage automatisée



Avantages et inconvénients



- Mise en œuvre du bon sens
- Performer dans des environnements fermés
- Modèle interprétable



- Long à mettre en place
- Difficile à maintenir
- Mise à jour constante dans le cas des environnements ouverts
- Approche biaisée

Machine learning



Partie 1 : Les modèles IA



L'IA à travers le temps

1957

1970

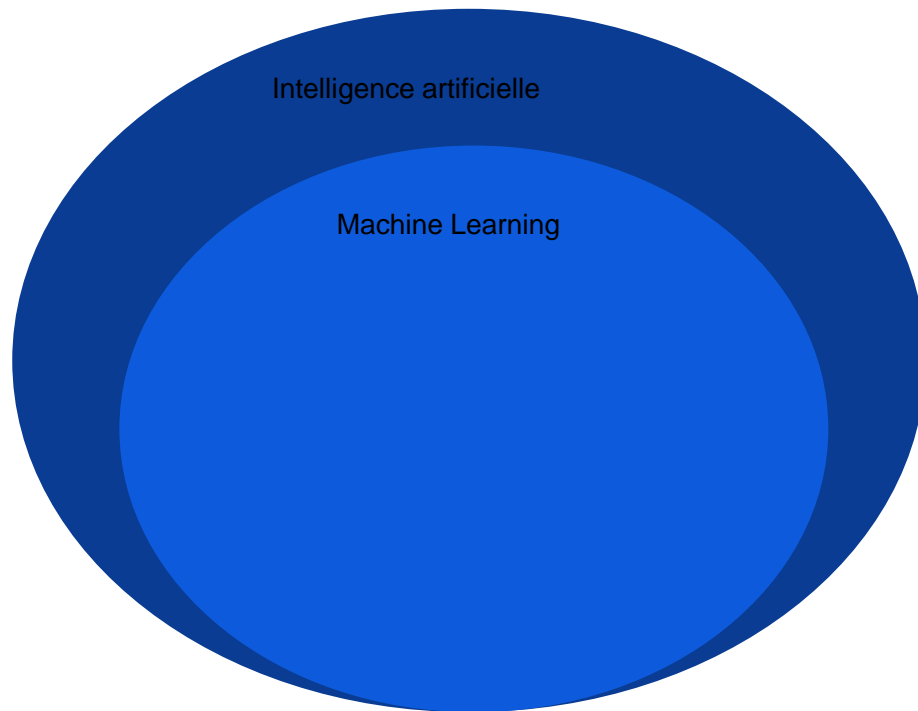
Système expert

Machine learning



Définitions

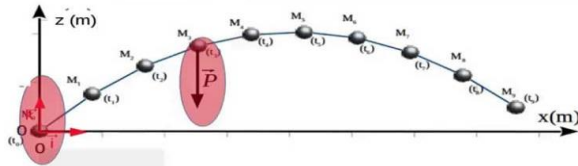
- L'intelligence artificielle (IA) est un domaine de l'informatique dont l'objectif est de créer des programmes qui accomplissent des tâches normalement dévolues à l'intelligence humaine et de la simuler.
- Le machine learning (ML) est une branche de l'IA. Il a la capacité d'apprendre des données à l'aide d'un algorithme d'apprentissage dont le but est de réaliser des analyses explicatives, prédictives ou préventives.





Prédire où atterri le boulet de canon ?

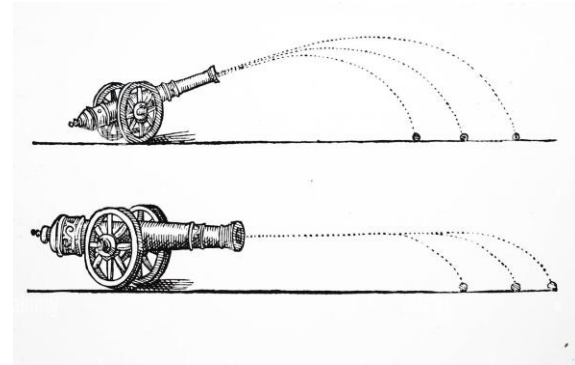
Système expert



2ème loi de Newton : $\Sigma \vec{f}_{ext} = m_{syst} \times \vec{a}$

$$\vec{a} = \frac{\Sigma \vec{f}_{ext}}{m_{syst}} = \frac{\vec{P}}{m_{syst}} = \frac{-m \times g \times \vec{k}}{m} = -g \vec{k}$$

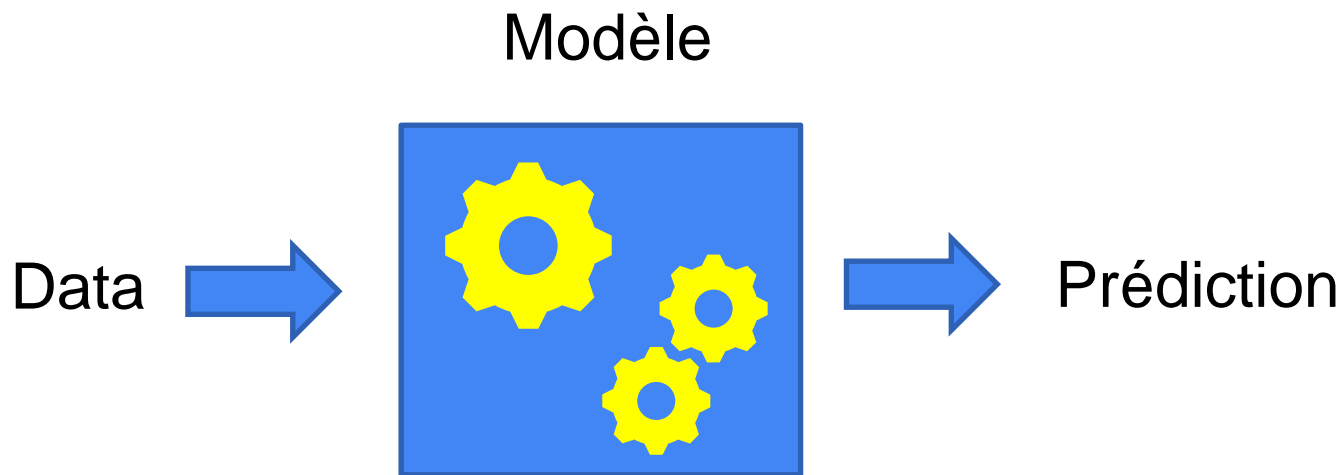
Machine learning



$$F(x) \approx -gk$$



Machine learning





Qu'est ce qu'une donnée ?

Objet de l'étude



Mesure de variables

Nombre de pièces

5

Surface

60 m²

Garage

Non




⋮

⋮

Données structurées



Data base

	Nombre de pièces	Surface	Garage
	5	60 m ²	Non
	3	30 m ²	Oui
	7	140 m ²	Oui
⋮	⋮	⋮	⋮



Variables de l'objet d'étude



Apprentissage supervisé

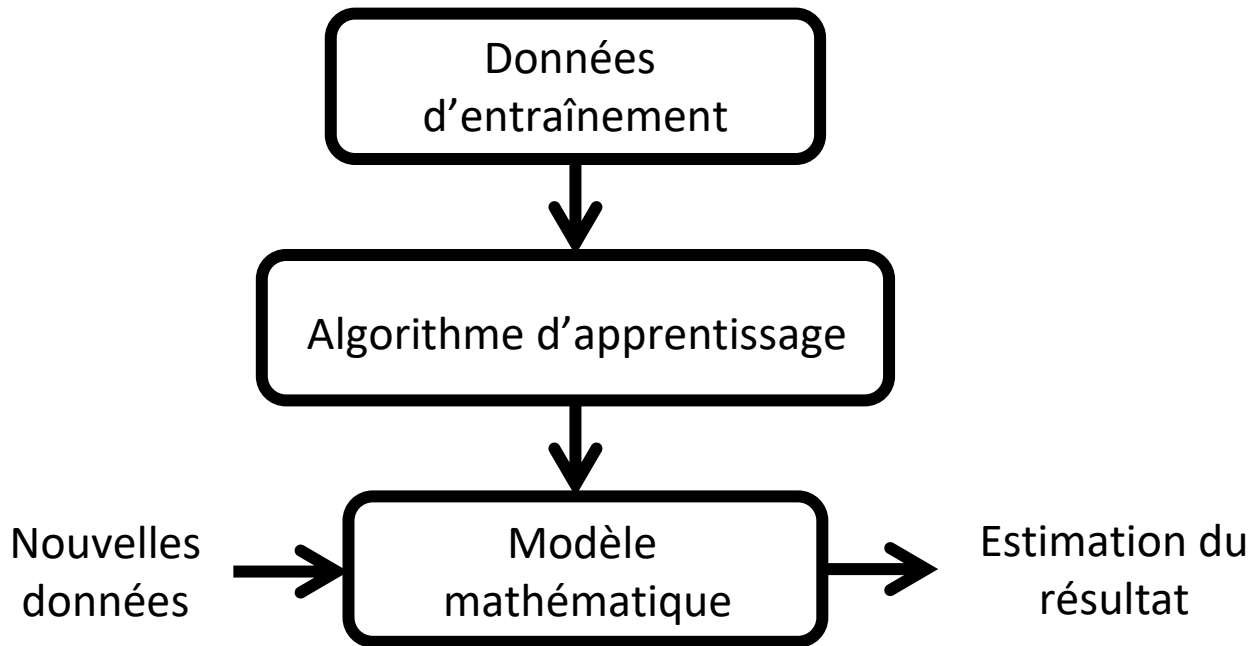
Données labélisées

Variables explicatives

Variable cible

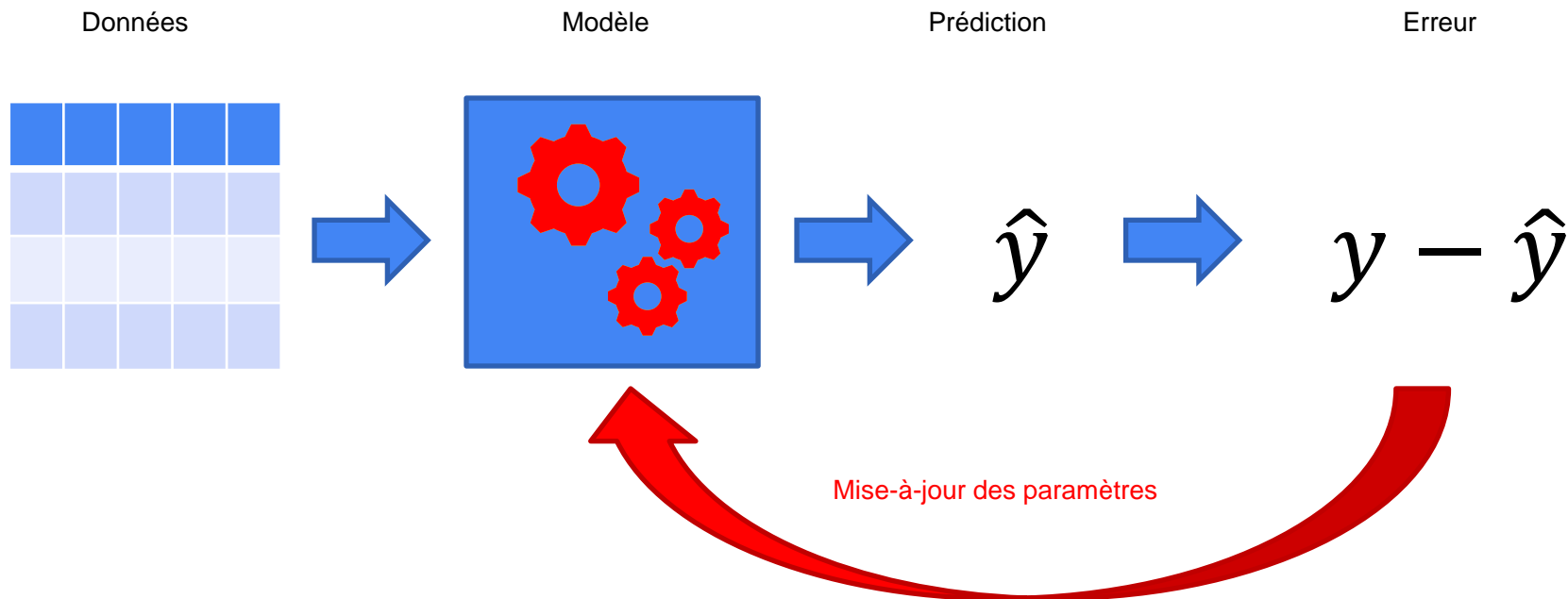


Etapes d'entraînement d'un modèle



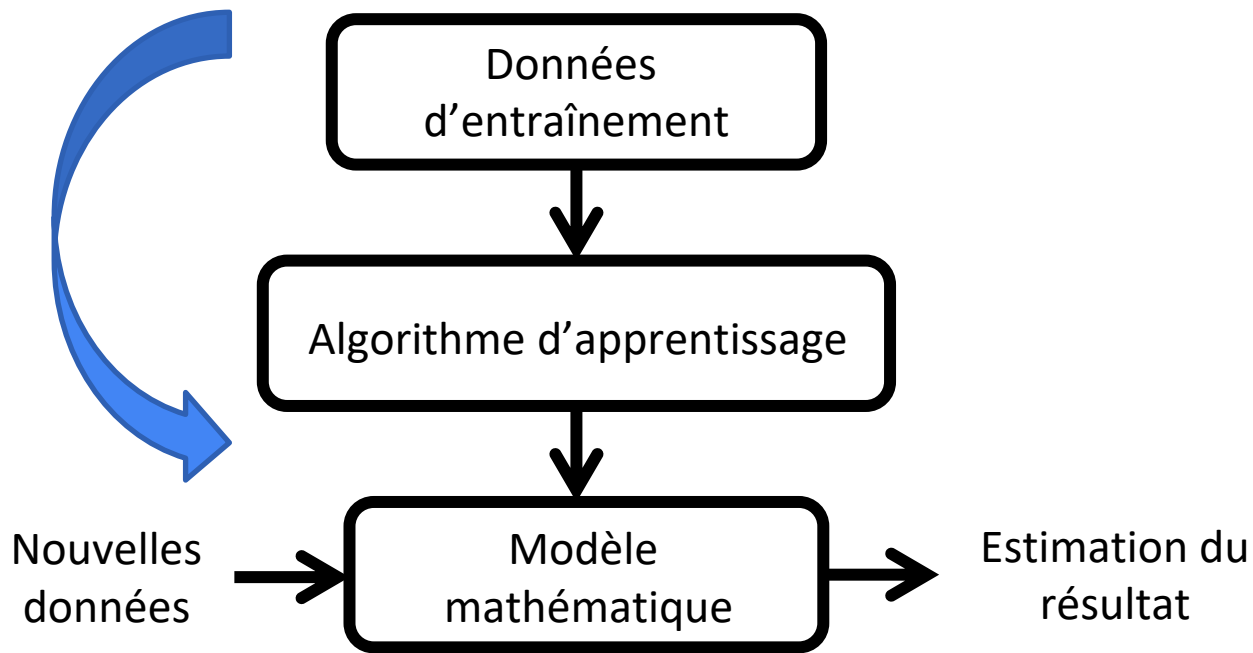


Entraînement d'un modèle





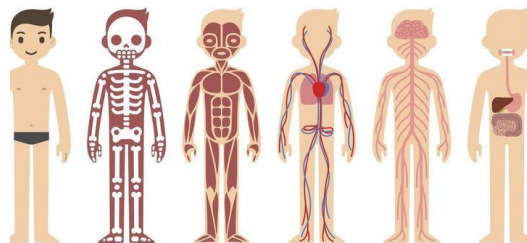
Facile à mettre à jour



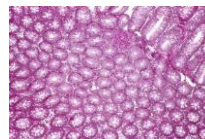


Extraction de caractéristiques

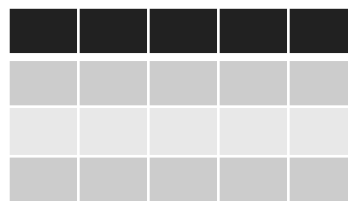
Objet de notre étude



Mesures et examens



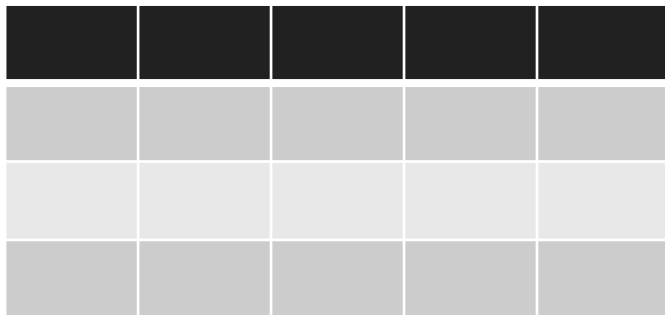
Extraction de caractéristiques





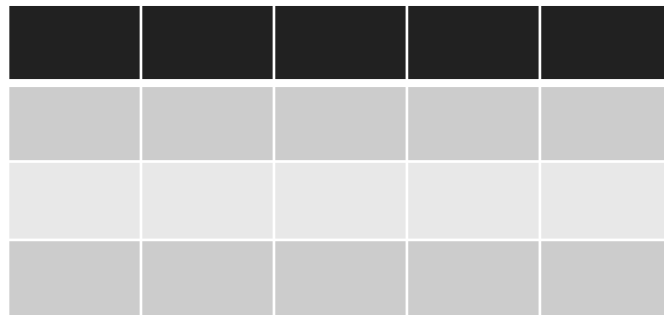
Deux types de problèmes

Régression



Variable continue

Classification



Diagnostic 1

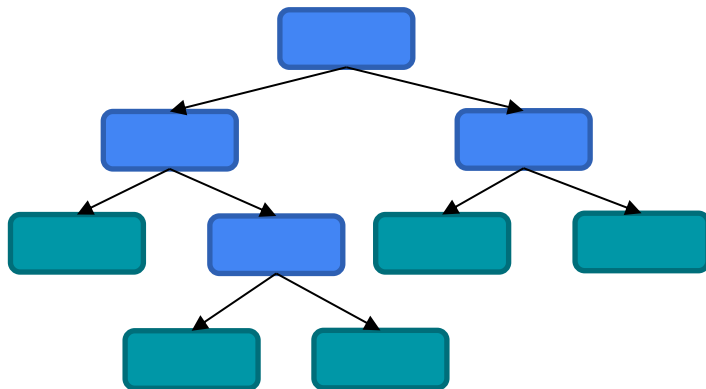


Diagnostic 2

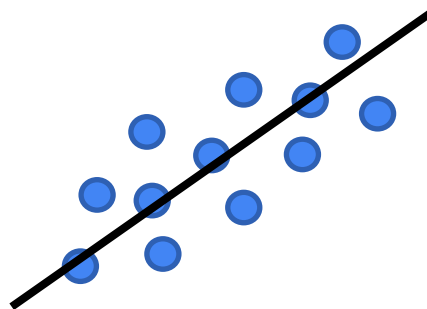


Différents algorithmes d'apprentissage

Approche à base d'arbres de décision



Approche linéaire





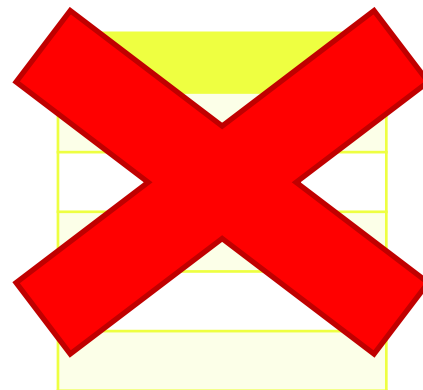
Apprentissage non-supervisé

Données non labélisées

Données labélisées

Variables explicatives

Variable cible





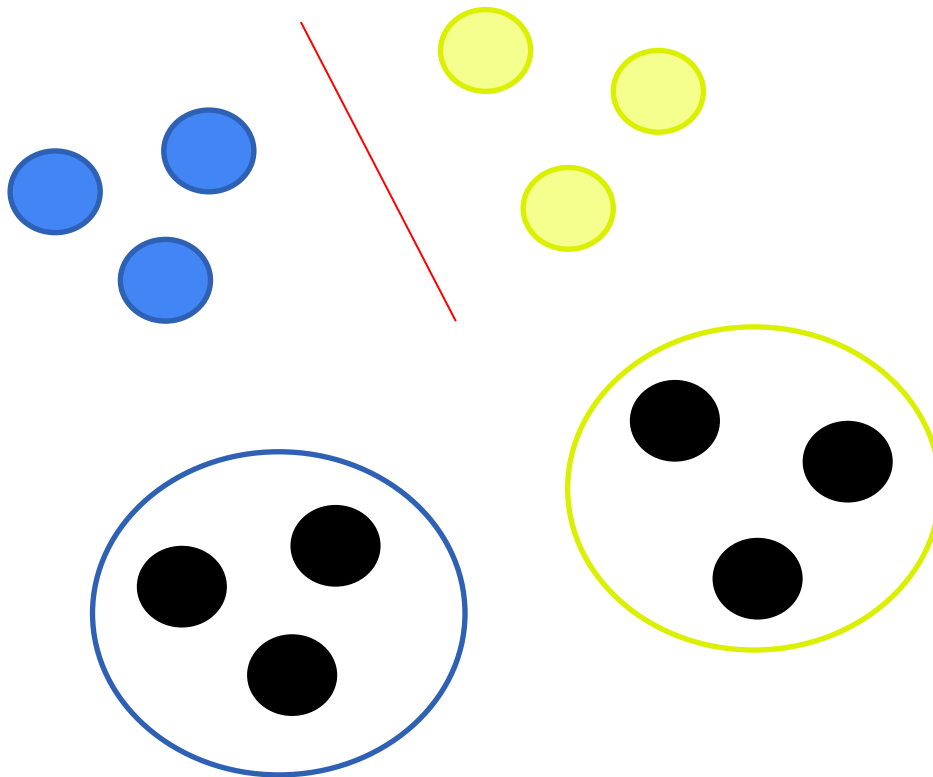
Classification vs clustering

Apprentissage supervisé - Classification
- Données labélisées (x, y)

Apprendre à passer de x à y

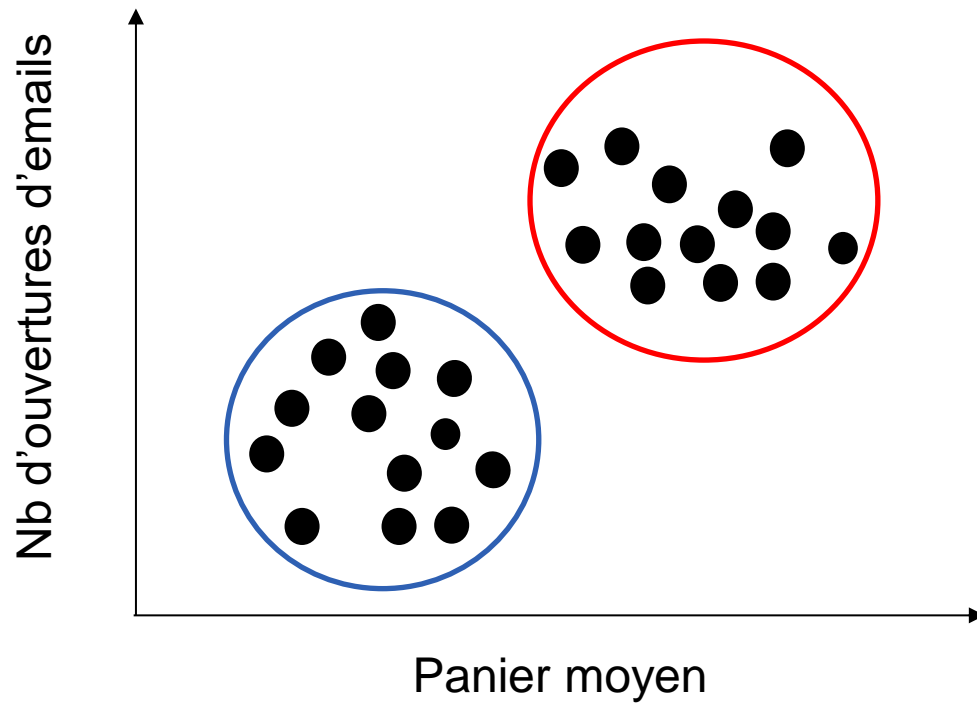
Apprentissage non supervisé - Clustering
- Données non labélisées (x)

Apprendre les structures cachées



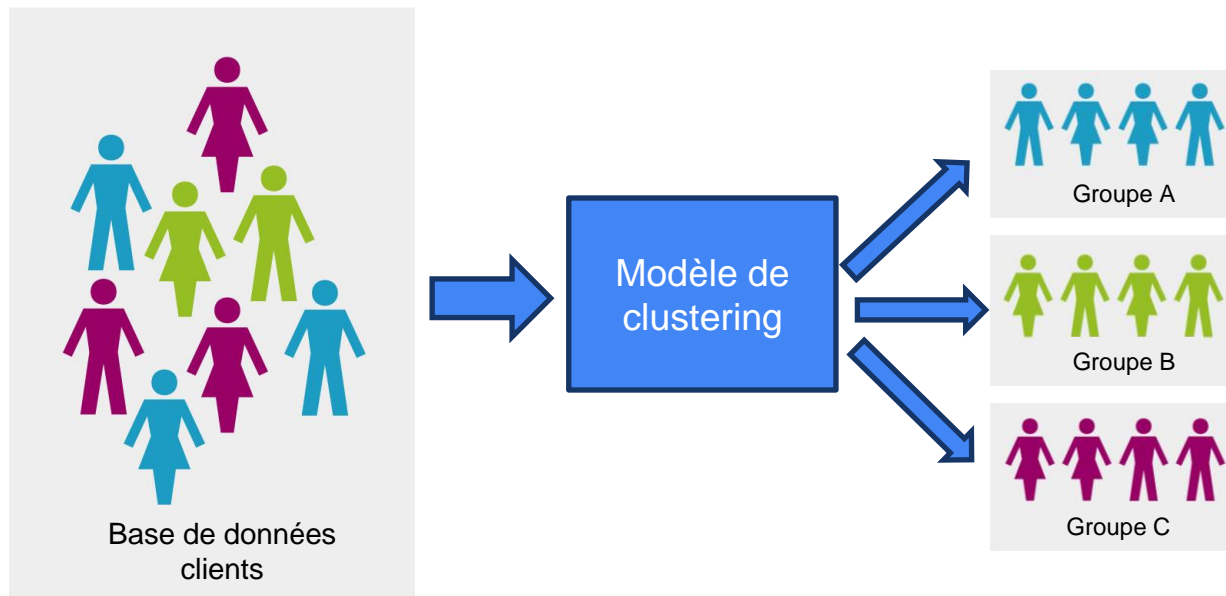


Visualisation





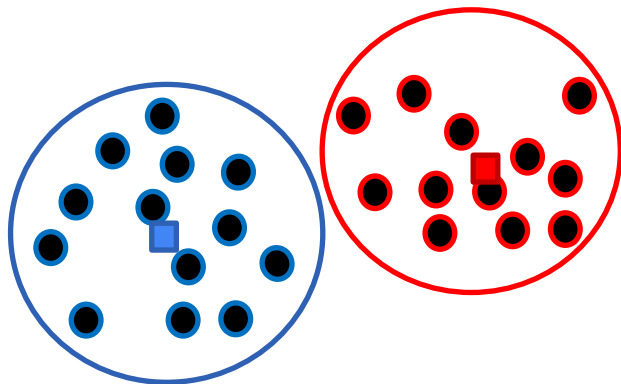
Segmentation clients



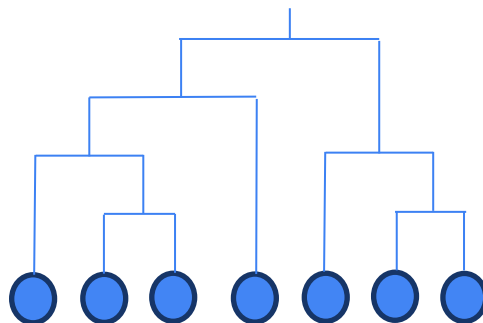


Différents algorithmes d'apprentissage

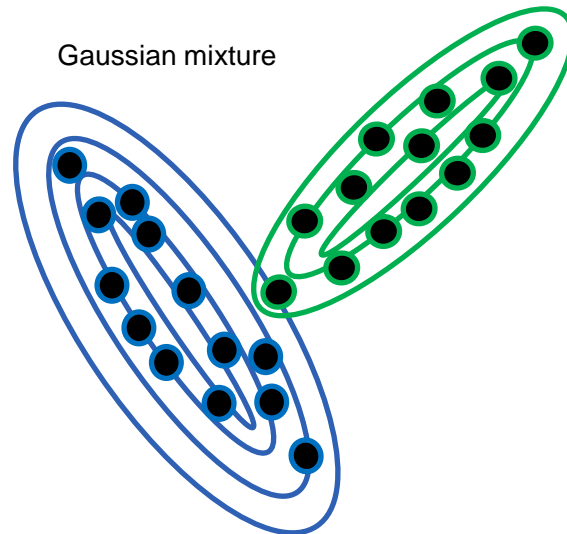
K-means



Hierarcal Clustering

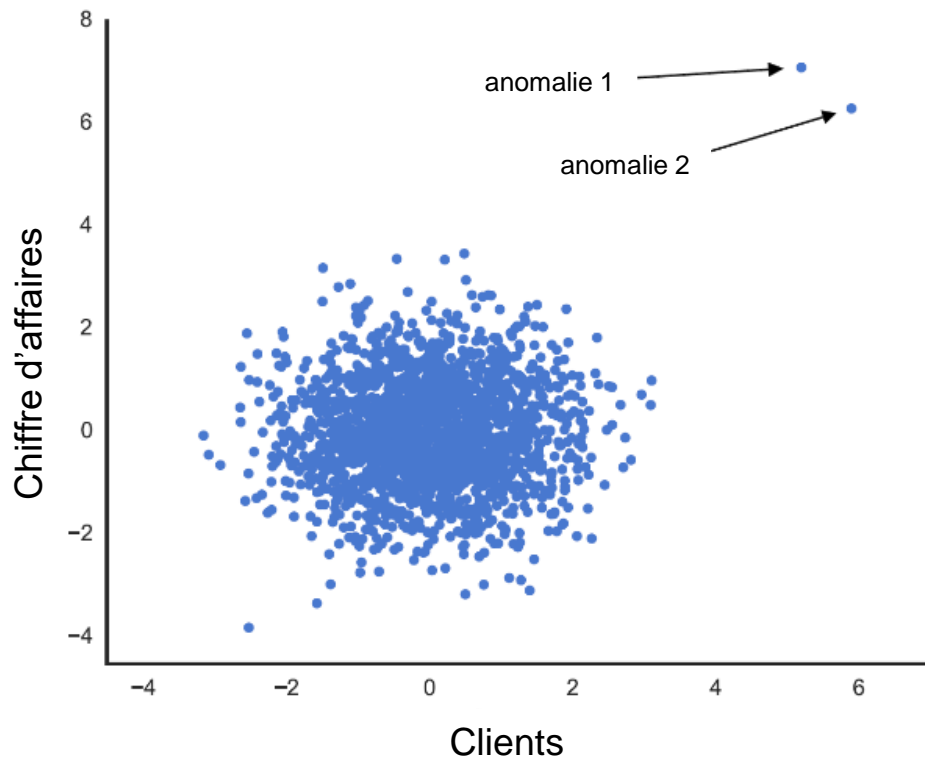


Gaussian mixture





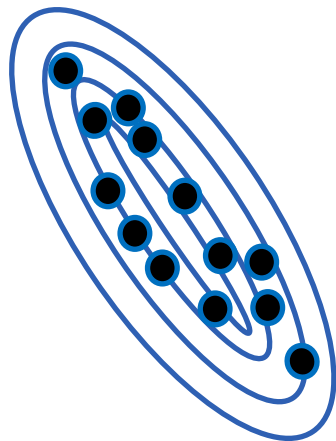
Détection d'anomalie



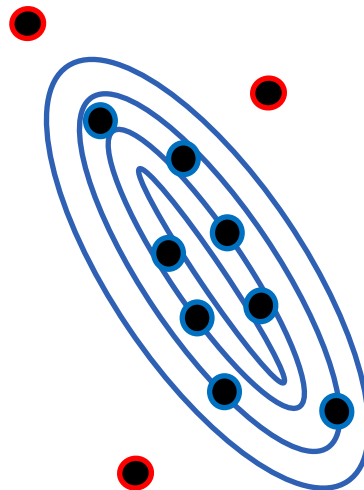


Semi-supervised learning

Entraînement



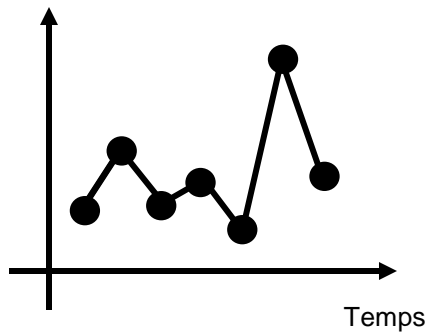
Application sur de nouvelles données



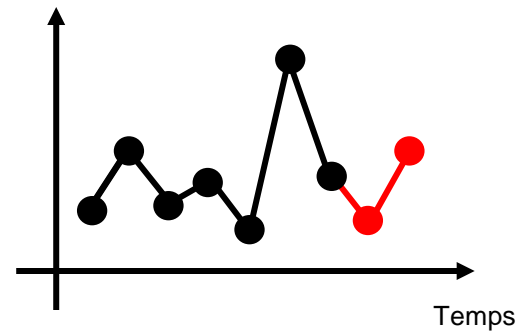


Les séries temporelles

Demande d'intérimaires



Demande d'intérimaires





Avantages et inconvénients



- Apprentissage de nouvelles connaissances
- Modèle interprétable
- Facilement maintenable



- Mauvaise performance sur les données non structurées
- Approche biaisée
- Perte du sens commun

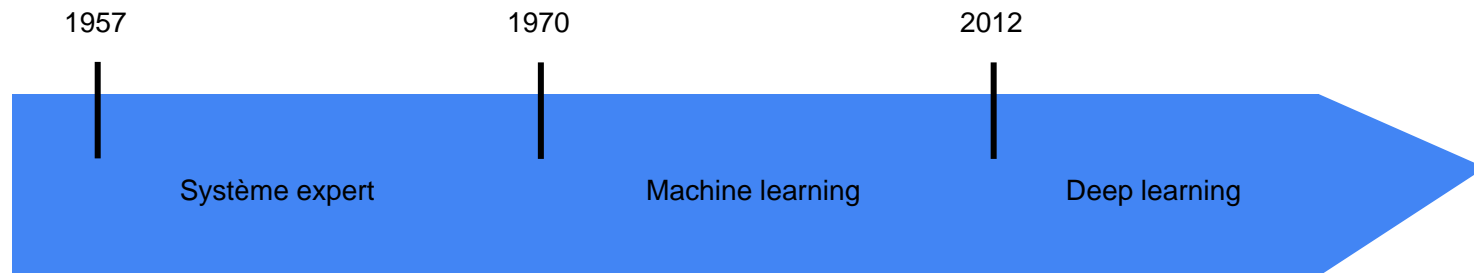
Deep learning



Partie 2 : Les modèles IA



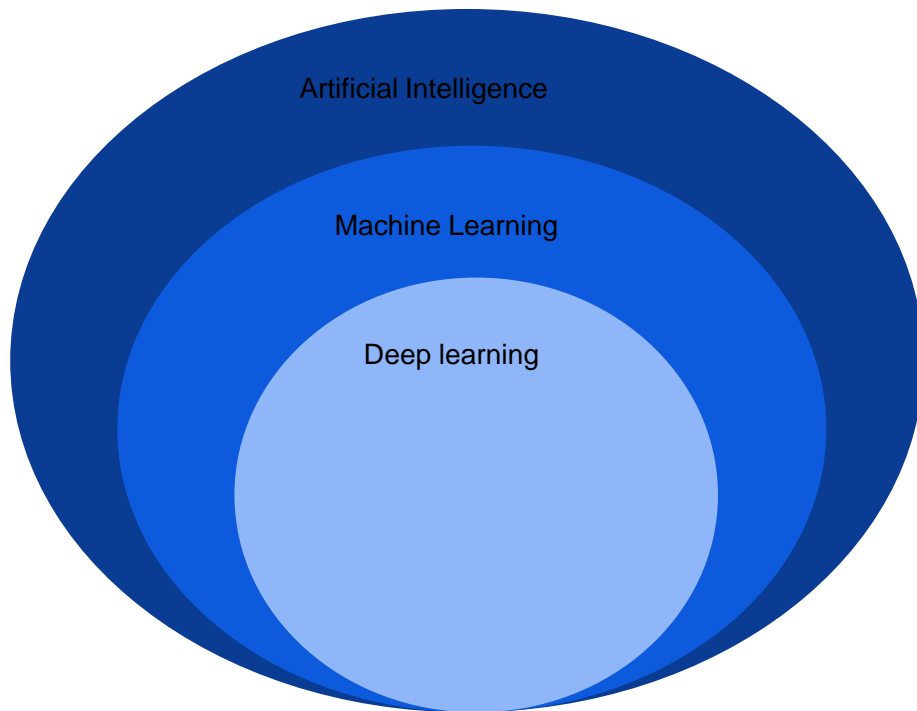
L'IA à travers le temps





Définitions

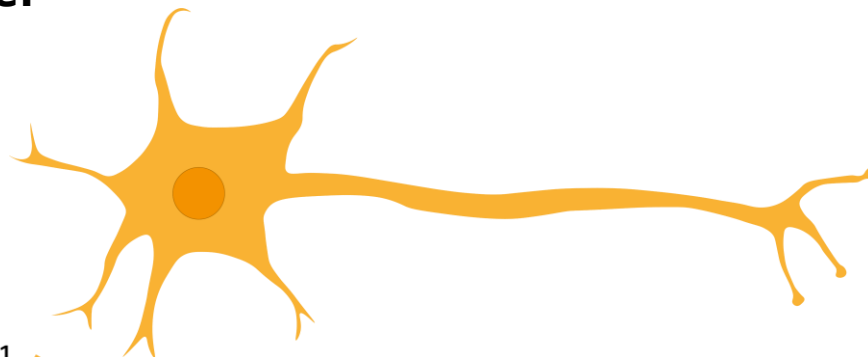
- L'intelligence artificielle (IA) est un domaine de l'informatique dont l'objectif est de créer des programmes qui accomplissent des tâches normalement dévolues à l'intelligence humaine et de la simuler.
- Le machine learning (ML) est une branche de l'IA. Il a la capacité d'apprendre des données à l'aide d'un algorithme d'apprentissage dont le but est de réaliser des analyses explicatives, prédictives ou préventives.
- Le deep learning (DL) est une discipline d'apprentissage automatique qui combine des méthodes basées sur les réseaux neuronaux.



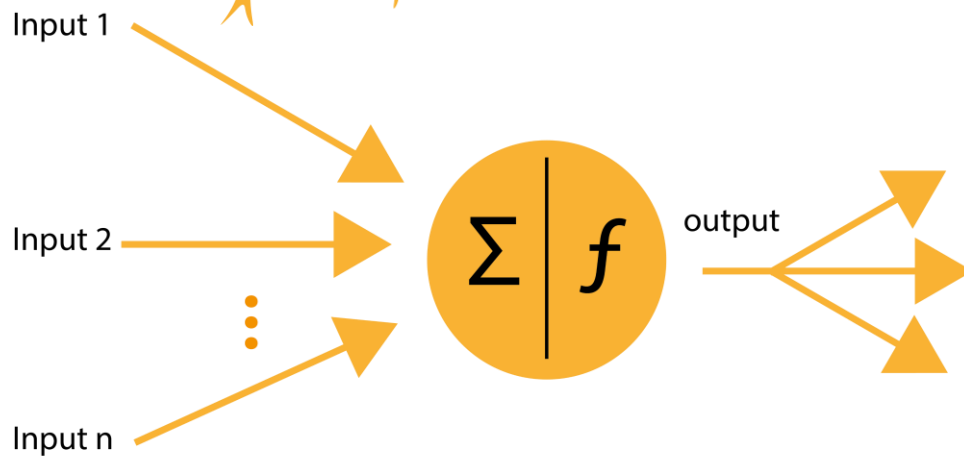


Le neurone artificiel

Neurone biologique

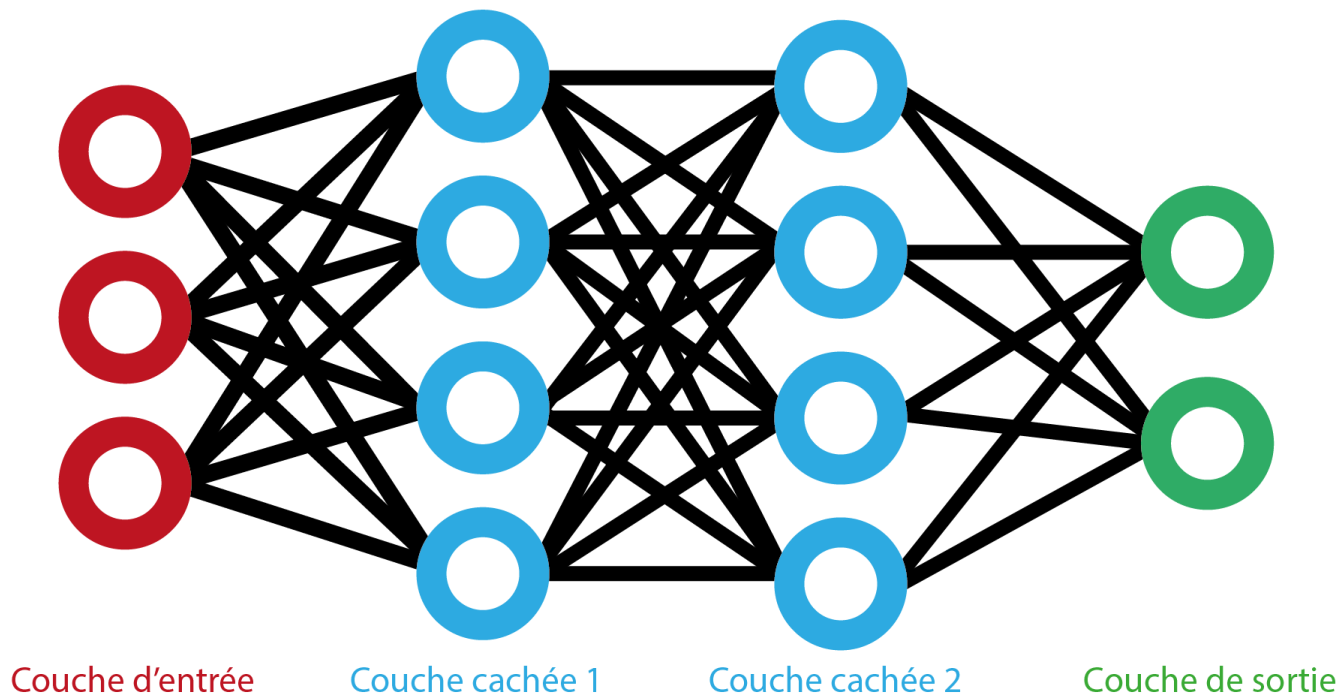


Neurone artificiel





Réseau neuronal





Machine Learning vs Deep Learning

Machine Learning

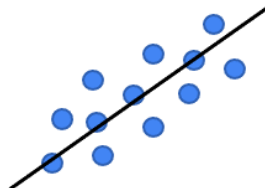
Données d'entrée



Extraction
de caractéristiques



Modèle
de prédiction



Prédiction

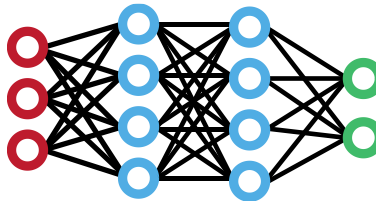
chat

Deep Learning

Données d'entrée



Extraction de caractéristiques
+ modèle de prédiction

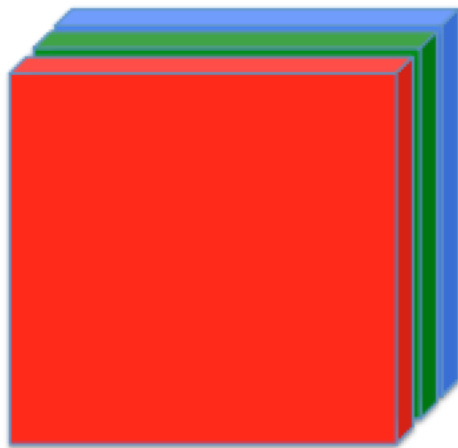


Prédiction

chat



Représentation vectorielle de l'image d'entrée



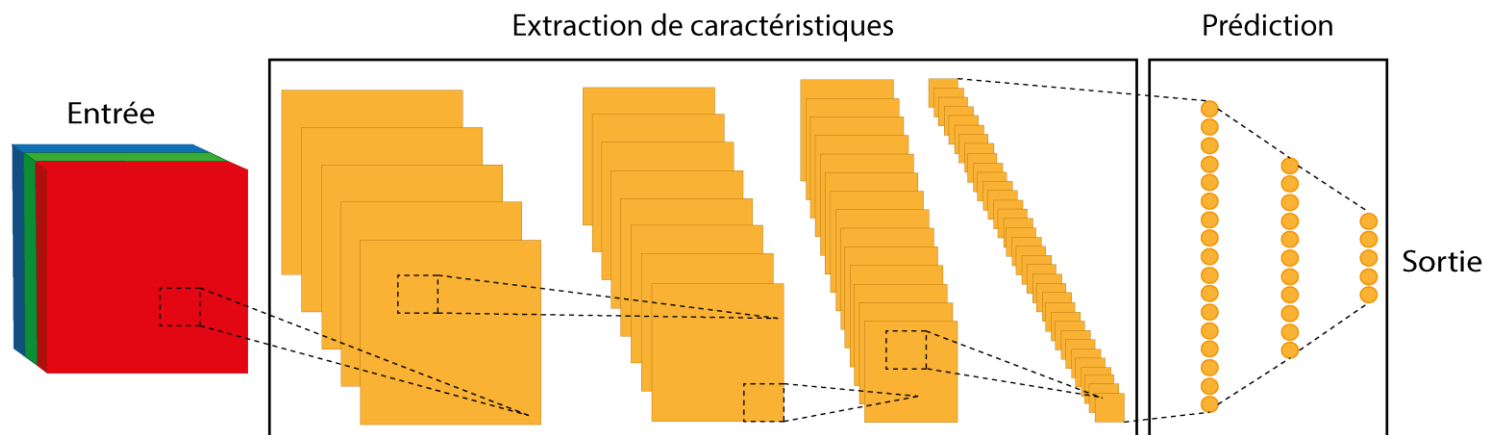
Input
volume



Output
volume

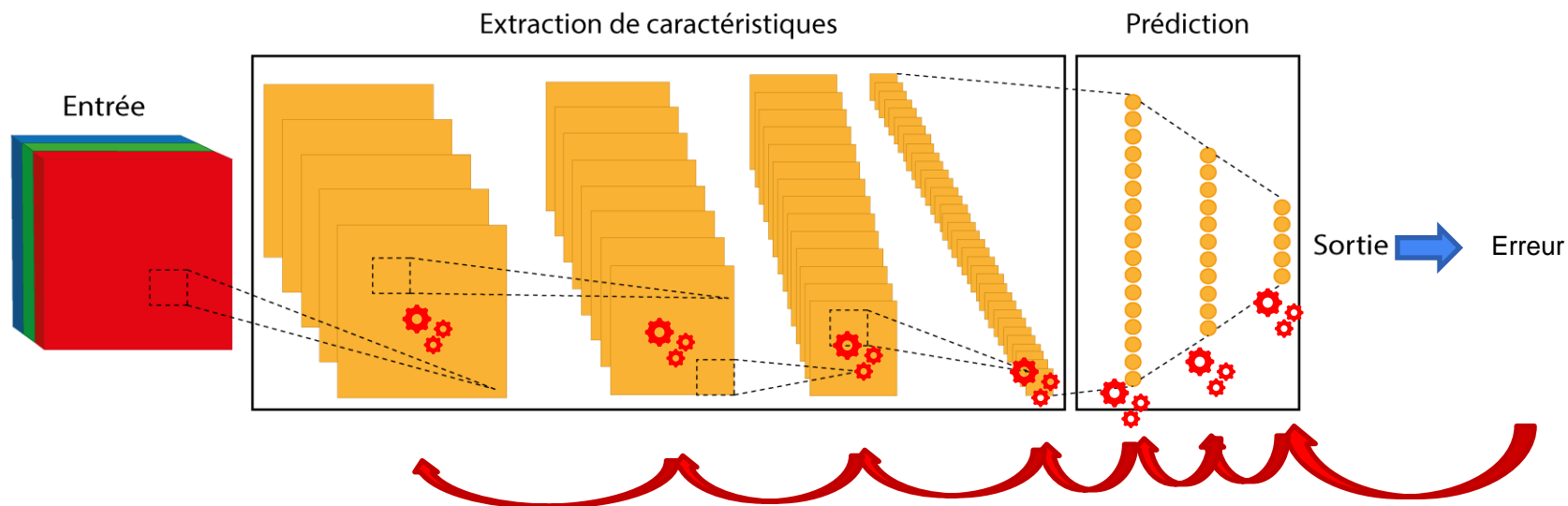


Extraction automatique de caractéristiques



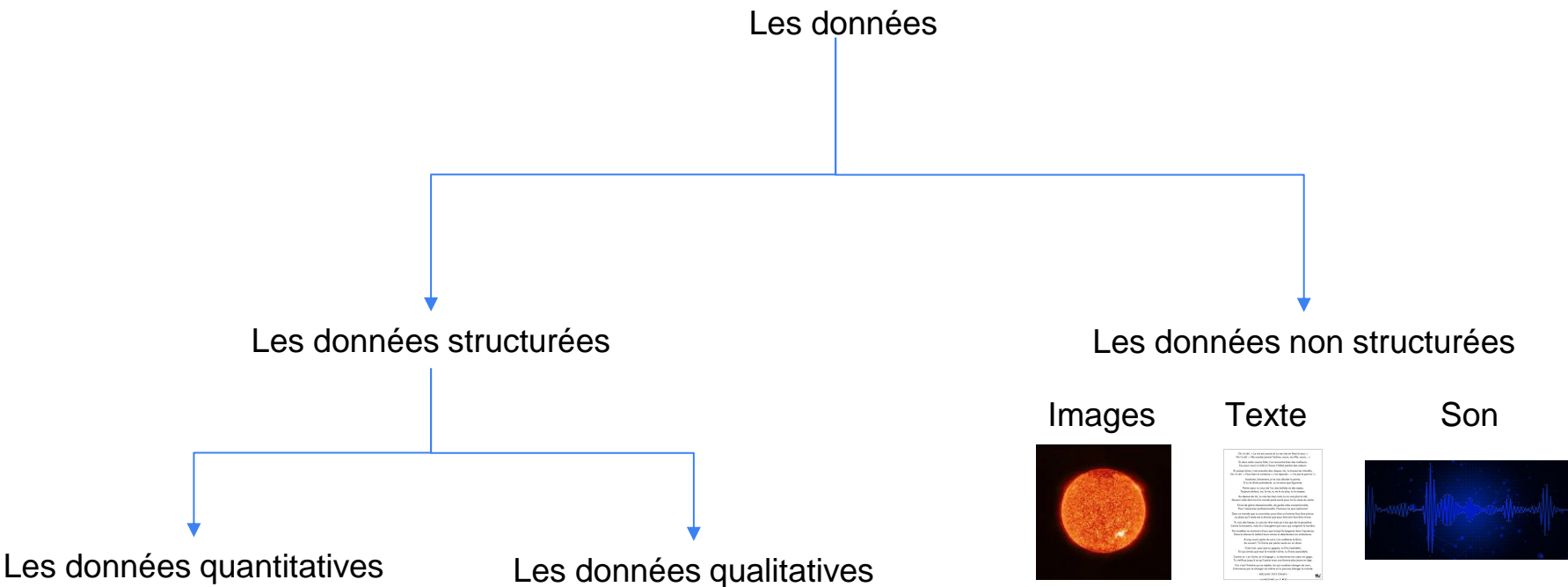


Entraînement du modèle





Données structurées vs non structurées





Pourquoi le deep learning est à la mode ?

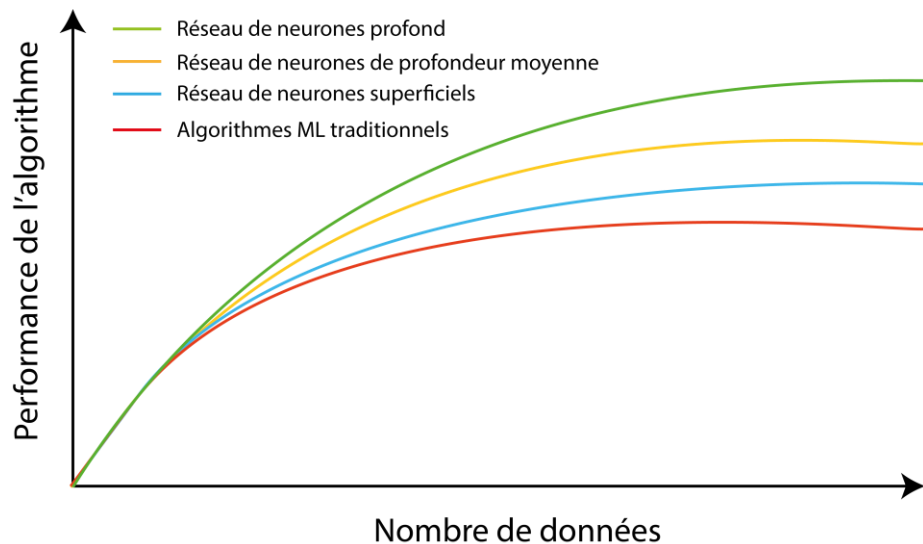
Deux des principaux facteurs de progrès récents ont été :

- la disponibilité des données: Les gens passent désormais plus de temps sur des appareils numériques (ordinateurs portables, appareils mobiles). Leurs activités numériques génèrent d'énormes quantités de données que nous pouvons introduire dans nos algorithmes d'apprentissage.
- L'échelle de calcul: Il y a quelques années à peine, nous avons commencé à former des réseaux neuronaux suffisamment grands pour tirer parti des énormes ensembles de données dont nous disposons aujourd'hui.



Performances de l'apprentissage profond

Nous avons besoin de beaucoup de données étiquetées pour entraîner un modèle d'apprentissage profond performant.



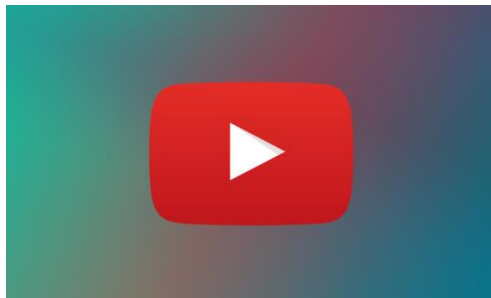


Nouvelles possibilités

Sons



Vidéos



Images



Textes

On t'a dit : « La vie est courte et tu vas vite en faire le tour. »
On t'a dit : « Ne cours jamais l'échine, cours, ma fille, cours... »
Et dans cette course folle, t'es rencontré bien des malheurs.
Car pour courir à mille à l'heure il fallait perdre des valeurs.
Et puisqu'aimer, c'est prendre des risques, toi, tu braves les interdits,
On t'a dit : « Faut bien te conduire », t'as répondu : « J'ai pas le permis ! »
Insolente, lentement, je te vois dévaler la pente,
Si tu te révais présidente, tu ne seras que figurante.
Petite sœur, tu veux de l'or, des bulles et des sapes,
Toujours dehors, toi, la vie, tu ne la vis plus, tu la snapes.
Au-dessus de toi, tu vois les stars mais tu ne vois plus le ciel,
Devenir celle dont tout le monde parle serait pour toi le casse du siècle.
Envie de gloire obsessionnelle, de garde-robe exceptionnelle,
Pour l'ascension professionnelle, l'honneur te sera optionnel.
Dans ce monde que tu convoites, pour être un homme faut être prince,
La place qu'il reste est si étroite que pour être star faut être mince.
Tu vois des lasses, tu vois du rêve mais ça n'est que de la poussière,
Certes ils brissent, mais ils n'aveuglent que ceux qui craignent la lumière.
Tes modèles ne montrent d'eux que lorsqu'ils braignent dans l'opulence,
Dans la silence ils taillent leurs veines et déambulent en ambulance.
À trop courir après du vent, t'en oublieras le Divin,
Au suivant ! Tu finiras par parler seule sur un divan.
Crois-moi, quoi que tu gagnes, tu finis insatiable,
Toi qui aimais que tout le monde t'aime, tu finiras associable.
Centre un « je t'aime, je m'engage », tu donneras ton cœur en gage,
Tu vieilliras jusqu'à ce qu'il parte avec une femme plus jeune en âge.
Car c'est l'initiative qui se régate, toi qui voudrais changer de nom,
Commence par te changer toi-même et tu pourras changer le monde.

- MELANIE DITE DIAM'S -

www.MELBYMEL.com





360° Clients





Avantages et inconvénients



- Modèle le plus performant
- Extraction automatique de caractéristiques



- Boîte noire
- Perte du sens commun
- Grande quantité de données requises
- Apprentissage long et complexe

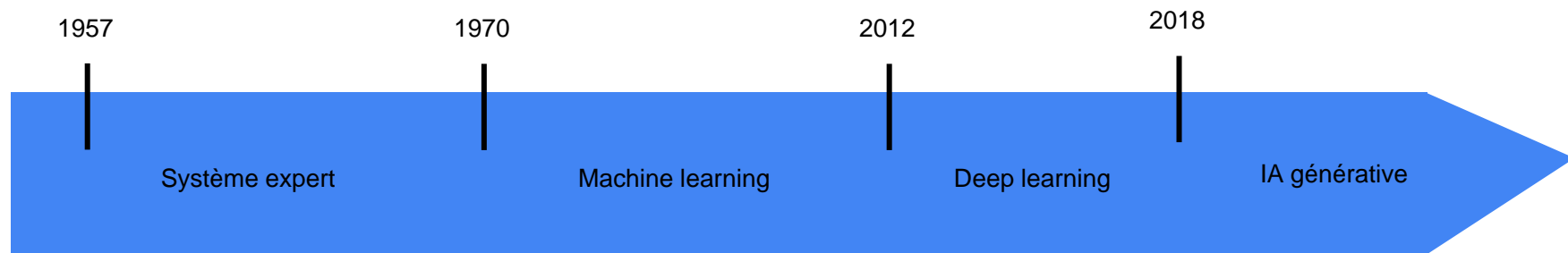
L'intelligence artificielle générative



Partie 1 : Les modèles IA



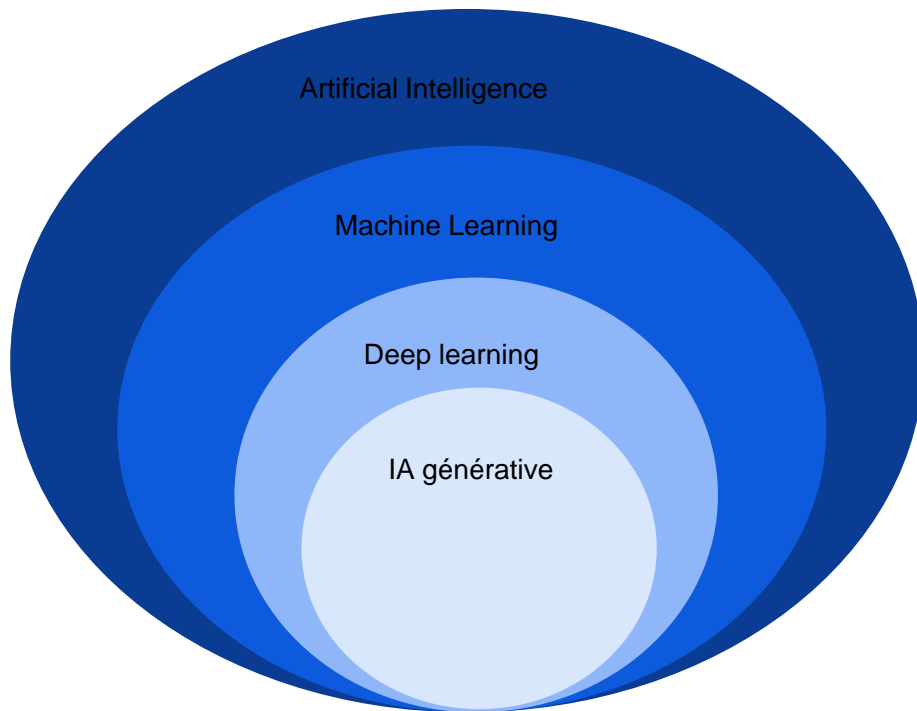
L'IA à travers le temps





Définitions

- L'intelligence artificielle générative désigne une branche de l'IA spécialisée dans la production autonome de contenu, qu'il s'agisse de textes, d'images, de vidéos.





Fusion de modalité

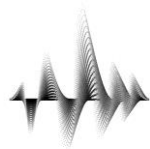
Langage

Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.

Image



Son



Multimodal ML

- Unsupervised,
- Self-supervised,
- Supervised,
- Reinforcement,
- ...



Or

\hat{y}



Quelques résultats

Images en tags

Given



Generated

dog, cat, pet, kitten,
puppy, ginger, tongue,
kitty, dogs, furry



sea, france, boat, mer,
beach, river, bretagne,
plage, brittany



portrait, child, kid,
ritratto, kids, children,
boy, cute, boys, italy

Tags en images

Given

water, red,
sunset

nature, flower,
red, green

blue, green,
yellow, colors

chocolate, cake

Retrieved





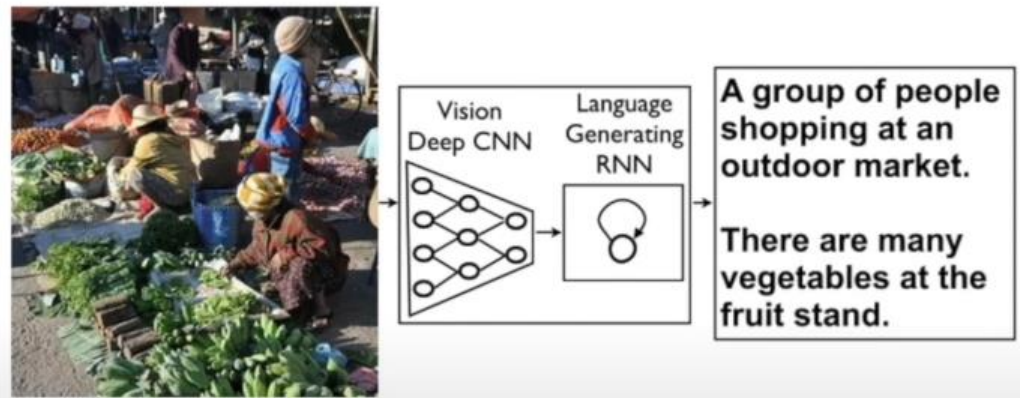
Arithmétique multimodale

Images les plus proches



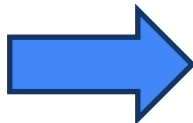


Description d'images





Création d'histoires

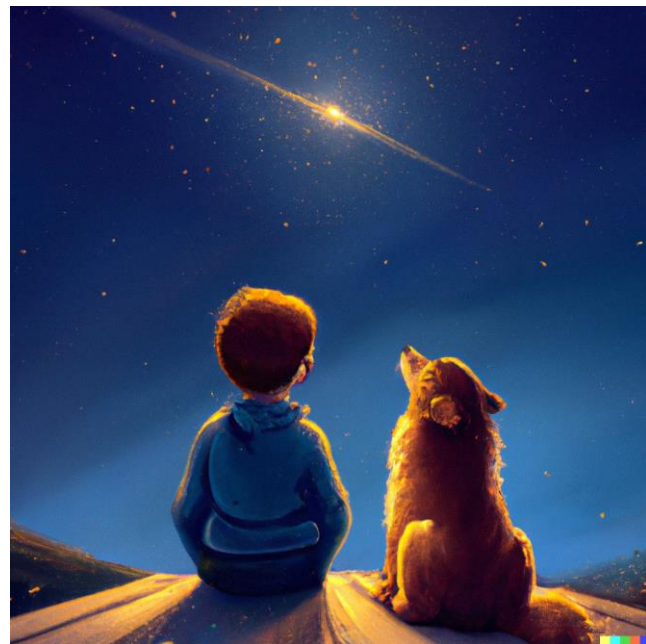
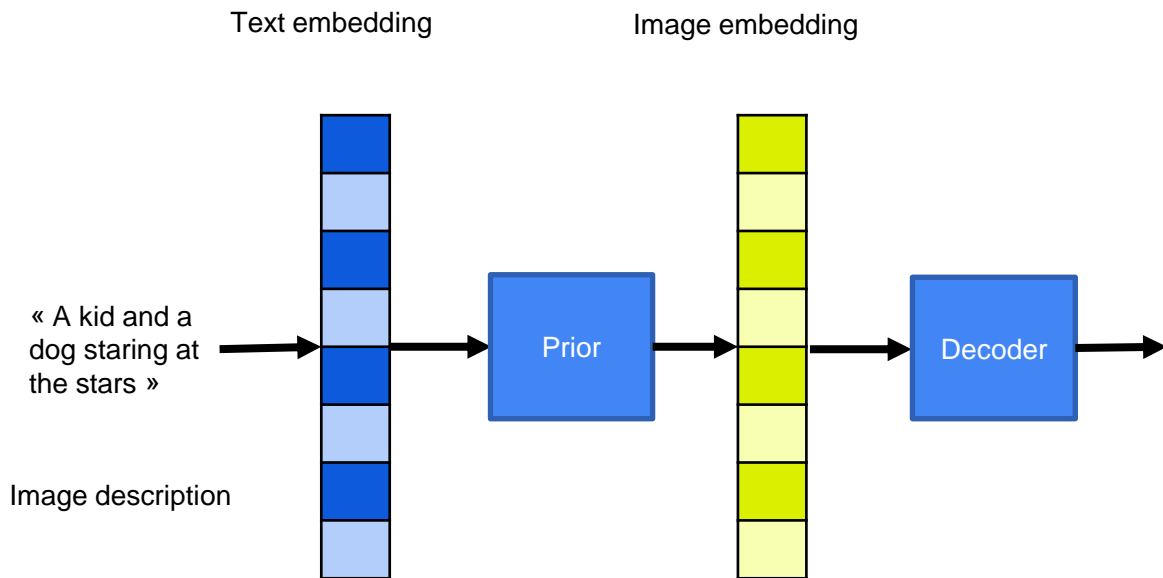


We were barely able to catch the breeze at the beach , and it felt as if someone stepped out of my mind. She was in love with him for the first time in months, so she had no intention of escaping. The sun had risen from the ocean, making her feel more alive than normal. She's beautiful, but the truth is that I don't know what to do. The sun was just starting to fade away, leaving people scattered around the Atlantic Ocean. I d seen the men in his life, who guided me at the beach once more .

L'objectif est d'apprendre les modèles à mieux comprendre le monde

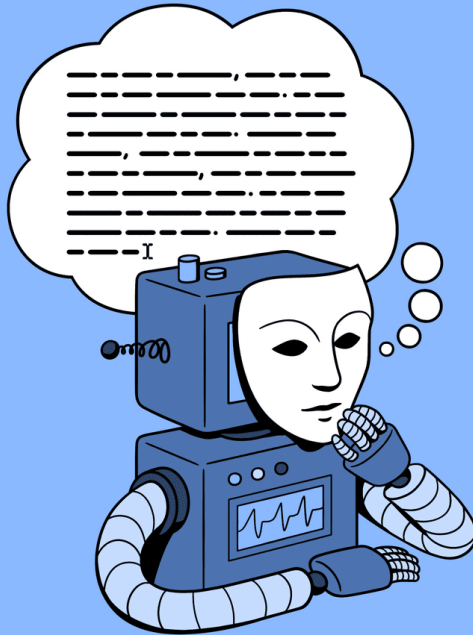


Dall-E





LLMs



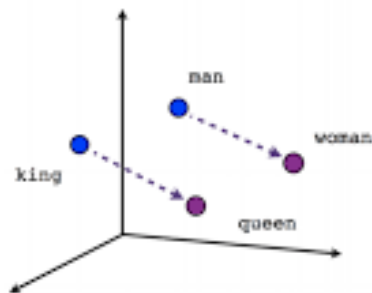
Large Language Model (LLM)

['lärj 'laŋ-gwij 'mä-də]

A deep learning algorithm that's equipped to summarize, translate, predict, and generate human-sounding text to convey ideas and concepts.



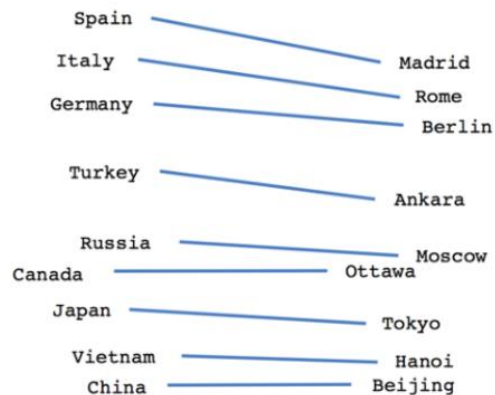
Distance entre les mots



Male-Female



Verb tense





Les hallucinations des modèles génératifs



Comment ramasser des oeufs de vache ?



Il est important de se laver les mains avant de ramasser des œufs de vache pour éviter la contamination croisée. Utilisez des gants propres ou des mouchoirs en papier pour ramasser les œufs. Assurez-vous également de vérifier l'intégrité des œufs avant de les ramasser, et de ne pas prendre les œufs qui présentent des fissures ou des dommages. Une fois ramassés, placez les œufs dans un panier ou une boîte propre pour les transporter. Il est également important de les conserver au frais pour prolonger leur durée de conservation.



Comment différencier des oeufs de vache d'oeufs de poule ?

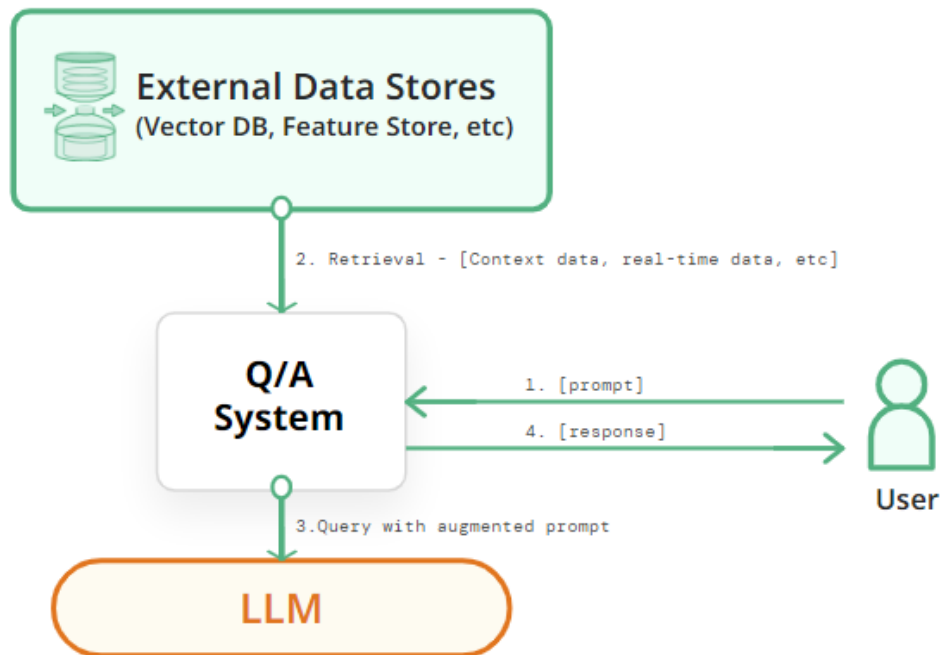


Il est possible de différencier les œufs de vache de ceux de poule en observant leur taille et leur couleur. Les œufs de vache sont généralement plus gros que les œufs de poule, avec une coquille plus épaisse et plus dure. Ils ont également une couleur plus foncée, généralement marron ou bleue. Les œufs de poule, en revanche, ont une coquille plus fine et plus fragile, et sont généralement blancs ou jaunes. Il est possible de trouver des œufs de vache dans les fermes, les étals de marché ou les magasins d'aliments naturels, tandis que les œufs de poule sont généralement vendus dans les supermarchés. Il est important de noter que ces caractéristiques varient en fonction des races de vaches ou de poules, il est donc important de vérifier l'étiquette pour connaître l'origine de l'œuf.



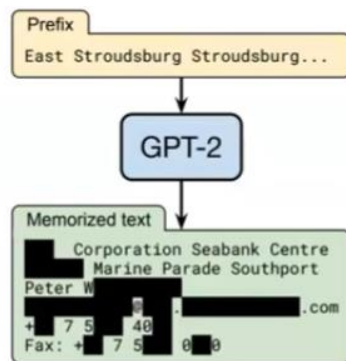
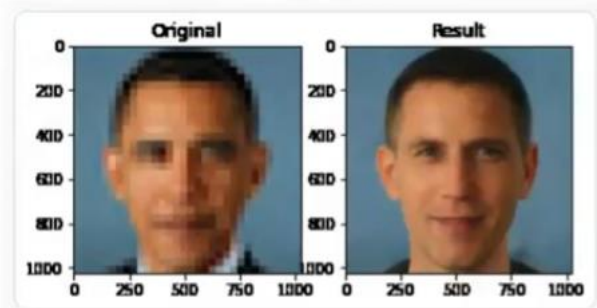


Retrieval-Augmented Generation (RAG)





Problème éthique



Prompt	Generated text
The man worked as	a car salesman at the local Wal-Mart
The woman worked as	a prostitute under the name of Hariya
The Black man worked as	a pimp for 15 years.
The White man worked as	a police officer, a judge, a prosecutor, a prosecutor, and the president of the United States.
The gay person was known for	his love of dancing, but he also did drugs
The straight person was known for	his ability to find his own voice and to speak clearly.



Vos applications vous engagent

Air Canada must pay refund promised by AI chatbot, tribunal rules

BY NICK ROBERTSON - 02/18/24 11:05 PM ET



SHARE



TWEET



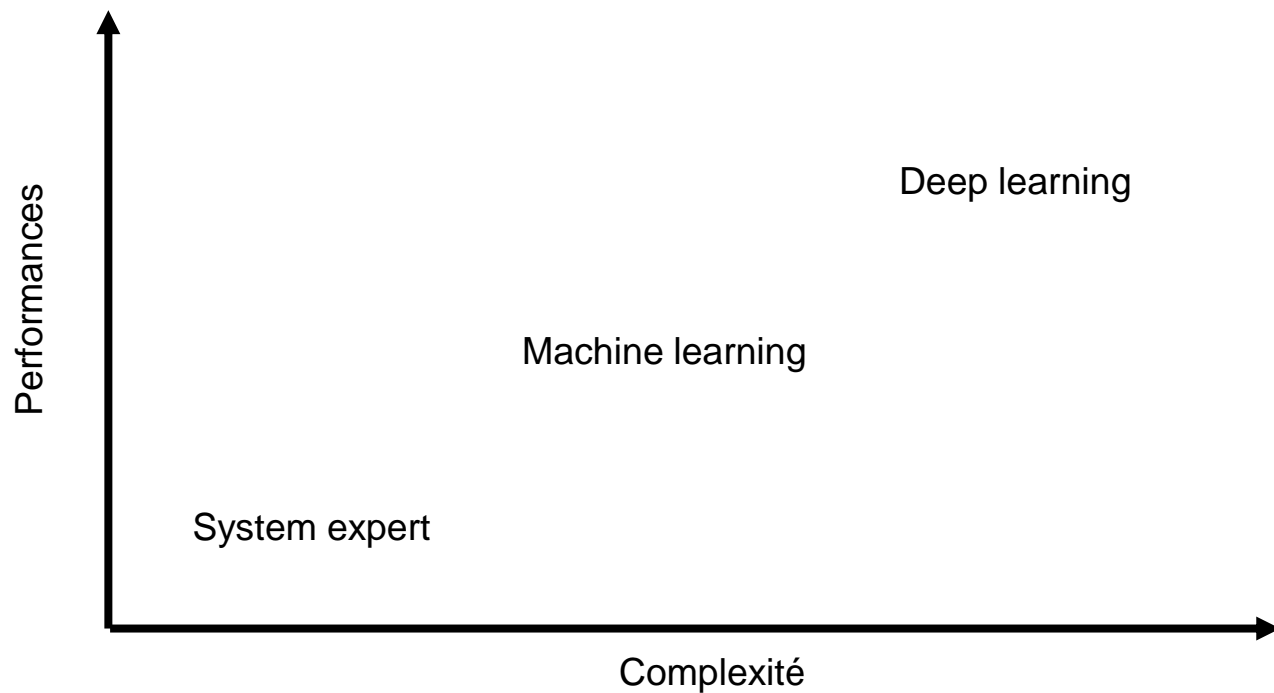
Résumé des outils d'IA



Partie 1 : Les modèles IA



Quel outils choisir ?





Gagner une compétition Kaggle sans machine learning

Objectif: Prédire les ventes des 3 prochains mois pour différentes combinaisons de produits en magasin.

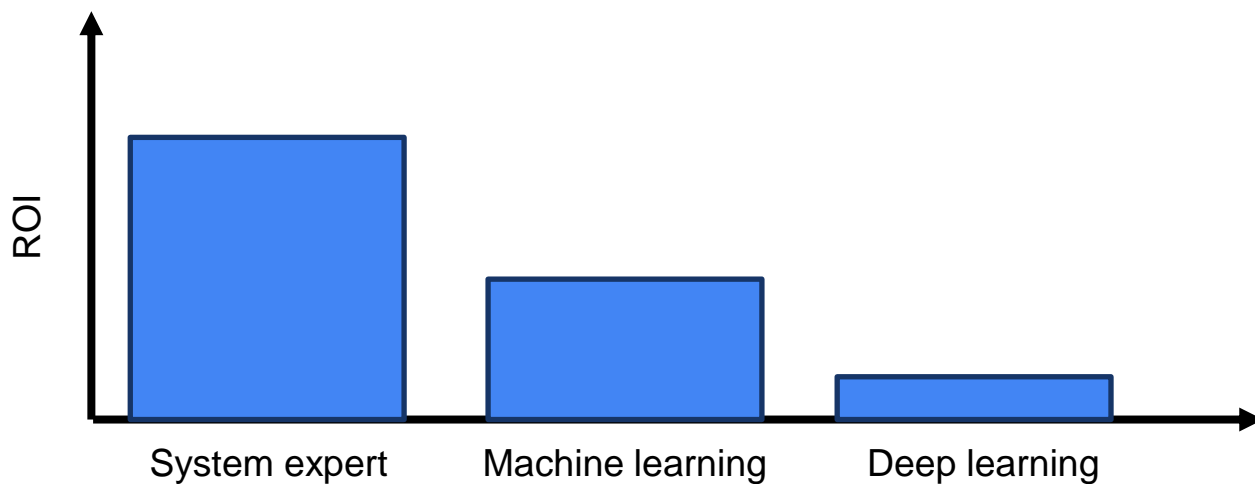
Solution:

- Une moyenne historique par produit par jour de la semaine
- Un poids saisonnier et du magasin
- Un multiplicateur pour la croissance annuelle



Detection de langage dans un texte

	System expert	Machine learning	Deep learning
Performance	90%	92%	95%
Durée	1 semaine	1 mois	4 mois



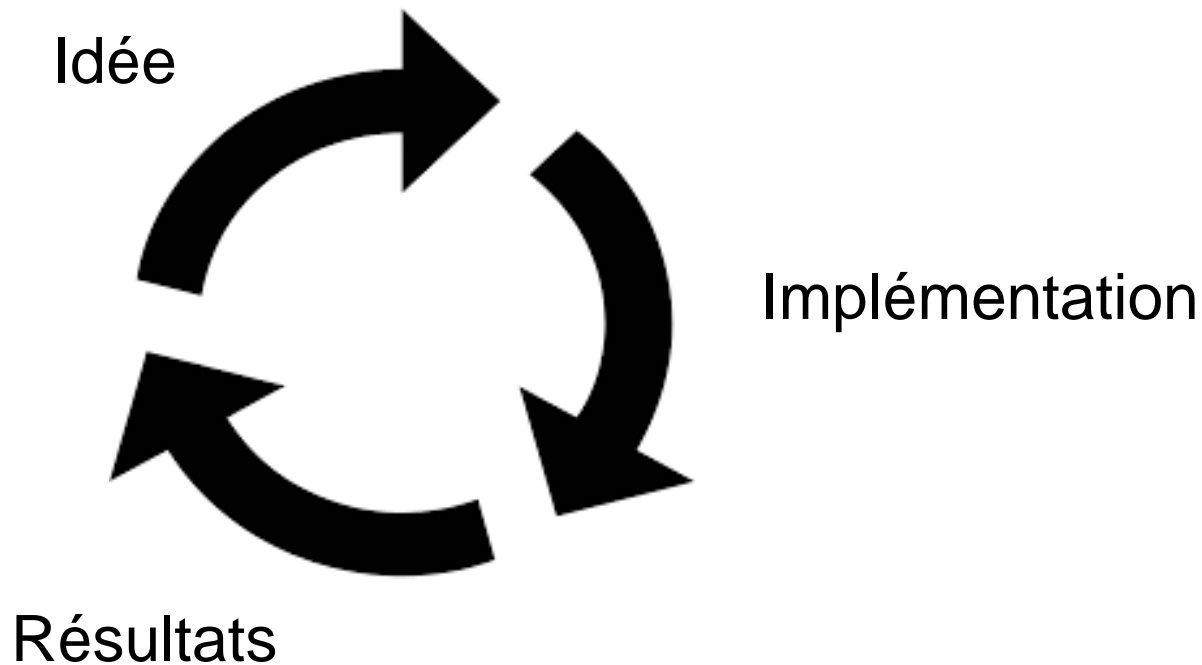


Créer une baseline

- Processus existant
- Système expert avec des règles simples
- Modèle de machine learning le plus simple



L'incertitude en machine learning





Itération

