

Intelligence artificielle

Redha Moulla

Casablanca, 7-8 mai 2025

Plan de la formation

- Qu'est-ce que l'intelligence artificielle ?
- Machine learning
- Qualité des données
- Deep learning
- Application du deep learning
- Enjeux sociaux de l'IA

Qu'est-ce que l'intelligence artificielle ?

Définition littérale de l'intelligence artificielle

1. Intelligence

Ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et relationnelle.

- Larousse

2. Artificielle

Qui est produit de l'activité humaine (opposé à la nature).

- Larousse

Qu'est-ce que l'intelligence ?

La notion d'intelligence recouvre plusieurs facultés cognitives :

- ① **Raisonnement** : La capacité à résoudre des problèmes et à faire des déductions logiques.
- ② **Apprentissage** : L'aptitude à acquérir de nouvelles connaissances et à s'améliorer grâce à l'expérience.
- ③ **Perception** : La compétence pour reconnaître et interpréter les stimuli sensoriels.
- ④ **Compréhension** : L'habileté à saisir le sens et l'importance de divers concepts et situations.
- ⑤ **Mémorisation** : La faculté de stocker et de rappeler des informations.
- ⑥ **Créativité** : Le pouvoir d'inventer ou de produire de nouvelles idées, de l'originalité dans la pensée.

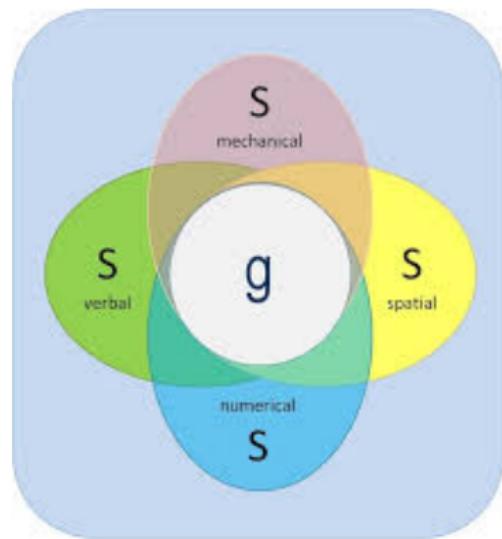
L'intelligence selon Spearman

Hypothèse du facteur g :

- Il existe une **intelligence générale**, commune à toutes les capacités cognitives.
- Chaque performance dépend à la fois :
 - d'un facteur *g* : aptitude globale,
 - d'un facteur *s* : spécifique à la tâche.

Limite implicite : l'intelligence devient *ce qui se mesure*.

Et si l'intelligence humaine ne pouvait pas se réduire à des performances observables ?



Jalons historiques

- **1956** — Conférence de Dartmouth : acte de naissance "officiel" de l'IA.
- **1943** — McCulloch & Pitts : formalisation du neurone artificiel.
- **1936** — Alan Turing : machine universelle, base de toute informatique.
- **1854** — Boole : logique symbolique manipulable mécaniquement.
- **1837** — Babbage & Ada Lovelace : machine analytique, programme et calcul.
- **XVII^e siècle** — Pascal (1642) et Leibniz (1673) : premières machines à calculer.
- **IX^e siècle** — Al-Khwarizmi : formalisation des premiers "algorithmes".

Le mythe des servantes d'or

“Si chaque instrument était capable, sur une simple injonction, ou même pressentant ce qu'on va lui demander, d'accomplir le travail qui lui est propre, comme on le raconte des statues de Dédale ou des trépieds d'Héphaïstos, lesquels, dit le poète, : “Se rendaient d'eux-mêmes à l'assemblée des dieux”, si, de la même manière, les navettes tissaient d'elles-mêmes, et les plectres pinçaient tout seuls la cithare, alors, ni les chefs d'artisans n'auraient besoin d'ouvriers, ni les maîtres d'esclaves.”

— Aristote

Conférence de Dartmouth ?

Articles

AI Magazine Volume 27 Number 4 (2006) © AAAI

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

John McCarthy, Marvin L. Minsky,
Nathaniel Rochester,
and Claude E. Shannon

The 1956 Dartmouth summer research project on artificial intelligence was proposed by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. The proposal required 17 pages plus a title page. Copies of the reprints are bound in a volume of the *Proceedings of the Conference on Cognition and Computer*, published by Stanford University. The first 5 papers state the proposal and give the general goals of the study and interests of the four who proposed the study. In the interest of brevity, this article reprints the first 2 pages of the proposal, which contain the philosophical statements of the proposal.

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use lan-

guage, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

1. Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. It may be argued that the present state of present computers may be insufficient to simulate many of the higher functions of the mind. This is true, but it is not because of lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2. How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human language consists of words according to rules of reasoning and rules of conjecture. From this point of view, learning a generalization consists of admitting a new

"We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

L'intelligence artificielle selon John McCarthy

"It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."

— John McCarthy

Le Test de Turing

Le Test de Turing, développé par Alan Turing en 1950, est une tentative de mesurer l'intelligence d'une machine, plus précisément de la faculté d'une machine à penser. Cette dernière n'étant pas si évidente à mesurer, le test substitute finalement à la faculté de penser celle de traiter le langage naturel comme un humain.

Les points clés du Test de Turing sont :

- Un interrogateur humain engage une conversation avec un humain et une machine, chacun étant caché de la vue de l'interrogateur.
- Si l'interrogateur ne peut pas déterminer systématiquement quelle est la machine, celle-ci est considérée comme ayant passé le test.
- Le test ne mesure pas la connaissance ou la capacité à être vérifique, mais plutôt la capacité de reproduire le comportement humain.

Deux visions de l'intelligence artificielle

IA symbolique (GOFAI)

- Héritée de la logique et de la philosophie rationaliste
- Représentation explicite des connaissances
- Raisonnement par règles, inférences logiques
- Utilisée dans les systèmes experts

"Penser, c'est manipuler des symboles selon des règles."

IA connexionniste

- Inspirée du fonctionnement du cerveau humain
- Représentation distribuée, implicite
- Apprentissage par ajustement de poids synaptiques
- Fondement du deep learning

"Penser, c'est apprendre par interaction avec le monde."

Deux visions du monde, deux manières de concevoir l'intelligence — logique vs émergence.

IA connexionniste vs IA symbolique

Intelligence artificielle symbolique :

Systèmes basés sur des règles et des symboles pour imiter le raisonnement humain.

- Logique
- Ensemble de règles
- Orientée connaissance

Intelligence artificielle connexionniste

: Modèles inspirés du cerveau humain pour apprendre des tâches à partir de données.

- Probabiliste
- Apprentissage machine
- Orientée données

Les deux hivers de l'intelligence artificielle

Premier hiver de l'IA (1974–1980)

- Échec des promesses de l'IA symbolique (raisonnement, langage naturel, etc.)
- Critique des réseaux de neurones simples (Perceptrons, Minsky Papert, 1969)
- **Conséquence** : Chute des financements, perte de crédibilité scientifique

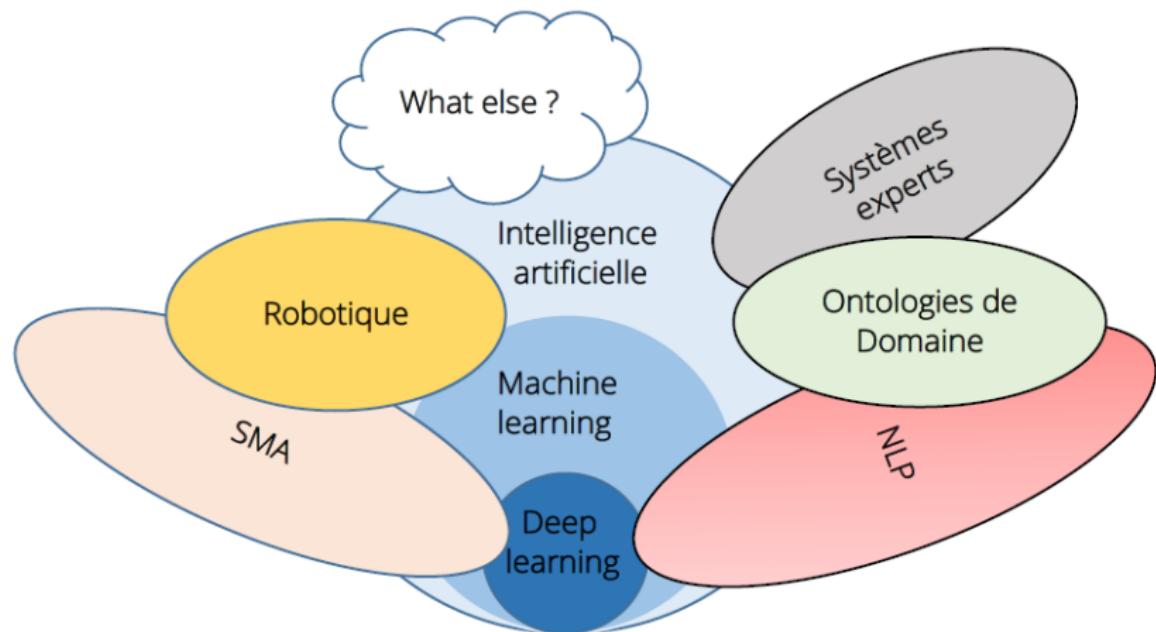
Deuxième hiver de l'IA (1987–1993)

- *Cause principale* : Désillusion face aux systèmes experts (coûts, rigidité, faible adaptabilité)
- **Contexte** : explosion commerciale suivie d'un désengagement brutal
- **Conséquence** : l'IA est perçue comme une promesse creuse dans le monde industriel

L'IA n'échoue pas seulement pour des raisons techniques. Elle échoue quand elle oublie la complexité du réel.

C'est quoi finalement l'intelligence artificielle ?

Il s'agit d'un ensemble de techniques qui permettent à la machine d'accomplir des tâches qui requièrent traditionnellement une intelligence humaine.



Machine learning

Trois approches de l'intelligence artificielle

- **IA discriminative** : Modélise la frontière entre les classes.
Exemples : régression logistique, SVM, réseaux de neurones pour la classification.
- **IA générative** : modélise la distribution des données pour en générer de nouvelles.
Exemples : GAN, VAE, modèles de langage comme GPT.
- **Apprentissage par renforcement (Reinforcement Learning)** : apprend à prendre des décisions séquentielles dans un environnement.
L'agent n'imiter pas ou ne classe pas, il agit pour maximiser une récompense.
Exemples : Q-learning, PPO, RLHF pour le fine-tuning des LLMs.

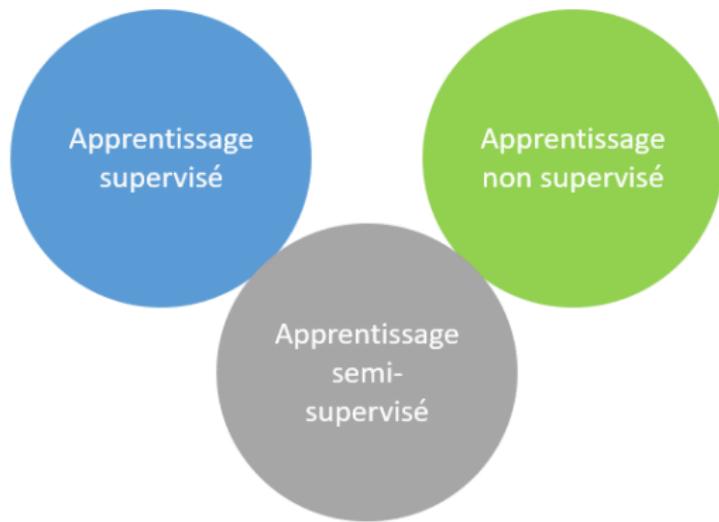
Machine learning discriminatif

L'apprentissage automatique est une branche de l'intelligence artificielle qui consiste à doter les machines de la capacité d'apprendre à partir de données sans que celles-ci ne soient explicitement programmées pour exécuter des tâches spécifiques.

Le machine learning englobe plusieurs types d'apprentissage :

- **Supervisé** : Les algorithmes apprennent à partir de données étiquetées pour faire des prédictions ou classifications.
- **Non supervisé** : L'apprentissage est effectué sur des données non étiquetées pour trouver des structures cachées.
- **Semi-supervisé** : Combine des éléments des deux premiers types en utilisant une petite quantité de données étiquetées et une grande quantité de données non étiquetées.

Typologies d'apprentissage automatique



L'apprentissage supervisé

L'apprentissage supervisé consiste à apprendre un modèle qui associe une étiquette (*label*) à un ensemble de caractéristiques (*features*).

- **Inputs** : un jeu de données *annotées* pour entraîner le modèle.
 - Exemple : des textes (tweets, etc.) avec les *sentiment* associés, positifs ou négatifs.
- **Output** : une étiquette pour un point de donnée inconnu par le modèle.

L'apprentissage supervisé se décline lui-même en deux grandes familles :

- **La classification** : prédire une catégorie ou une classe.
 - Exemple : prédire l'étiquette d'une image (chat, chien, etc.), le sentiment associé à un texte, le centre d'intérêt d'un client à partir de ses commentaires, etc.
- **La régression** : prédire une valeur continue (un nombre réel typiquement).
 - Exemple : prédire le prix d'un appartement, la lifetime value d'un client, etc.

Classification

Exemple de classification : credit scoring

Âge	Revenu annuel (k€)	Historique de crédit	Nombre de cartes de crédit	Niveau d'éducation	Propriétaire immobilier (Oui/Non)	Label (y)
30	50	Bon	2	Licence	Oui	Accepté
45	80	Moyen	3	Master	Oui	Accepté
22	20	Mauvais	1	Bac	Non	Refusé
35	60	Bon	4	Licence	Oui	Accepté
40	70	Moyen	2	Bac+2	Non	Refusé

Régression

Exemple de régression : prédition des prix des logements

Surface (m ²)	Nombre de chambres	Distance du centre-ville (km)	Année de construction	Quartier (Score 1-10)	Prix (k€) (y)
80	3	5	2010	8	300
120	4	10	2005	7	450
60	2	2	2020	9	200
150	5	15	1995	6	600
100	3	8	2015	7	400

L'apprentissage non supervisé

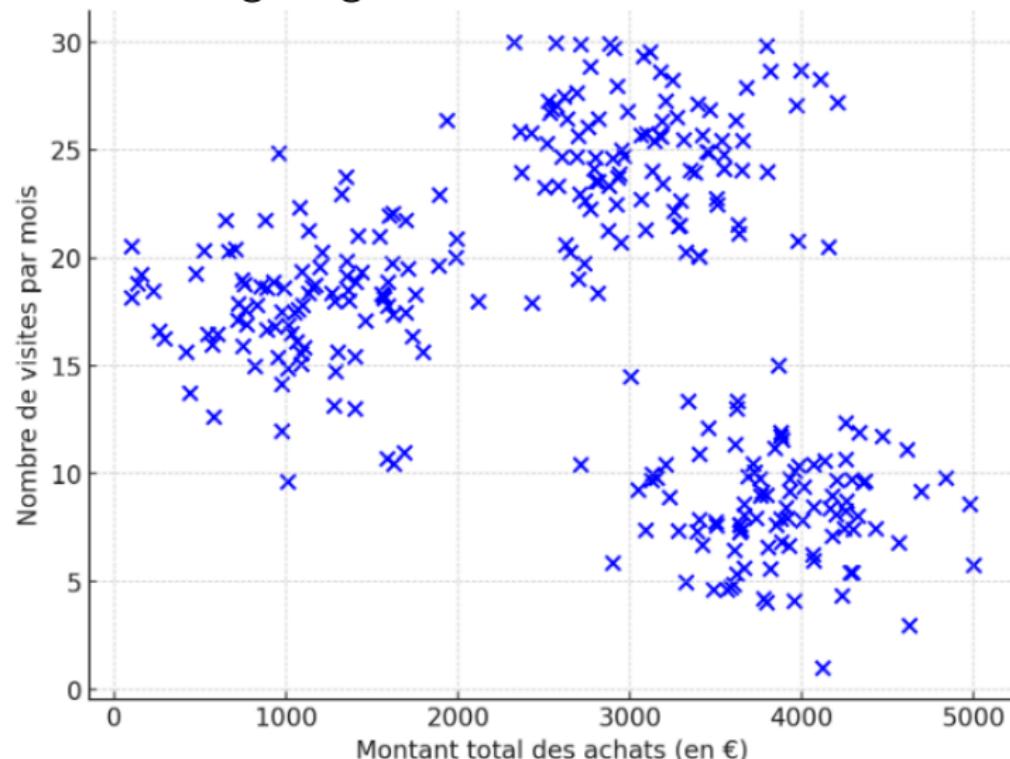
L'apprentissage non supervisé se réfère à l'utilisation de modèles d'apprentissage automatique pour identifier des patterns et des structures dans des données qui ne sont pas étiquetées.

Principales typologies de l'apprentissage non supervisé :

- **Clustering** : Regroupement de points de données similaires ensemble.
Exemple : segmentation de marché, regroupement social.
- **Détection d'anomalies** : Détecter des observations dont les caractéristiques sont inhabituelles par rapport à la majorité.

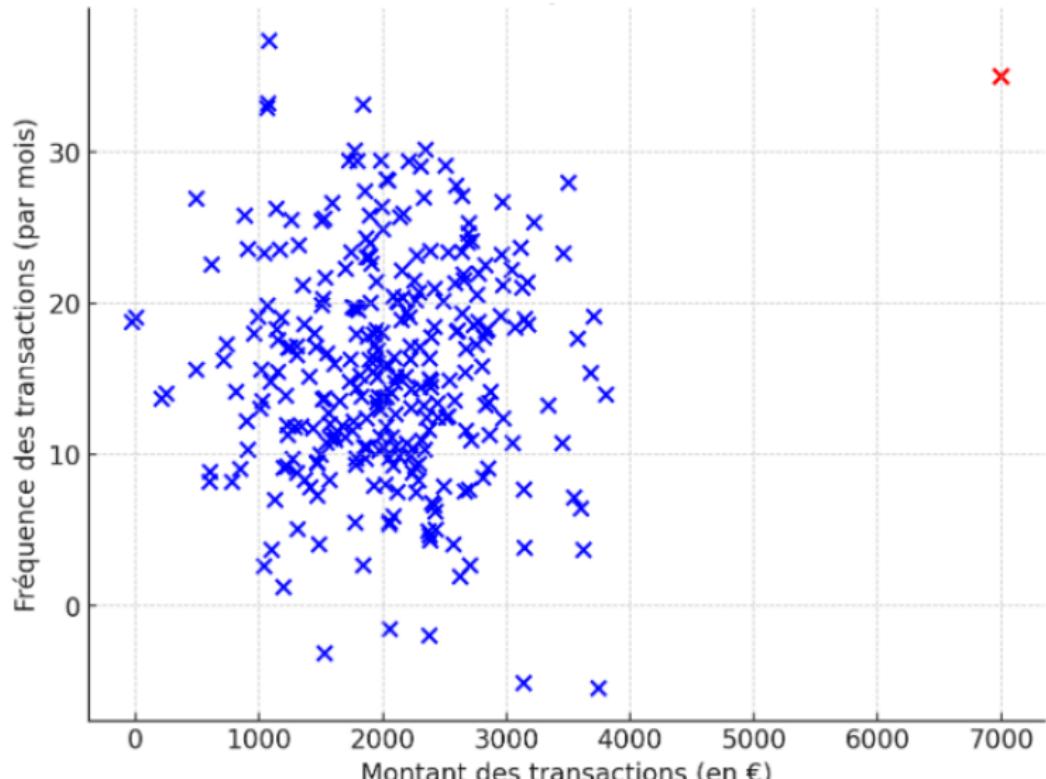
Clustering

Exemple de clustering : segmentation clients



Détection d'anomalies

Exemple de détection d'anomalies : fraude bancaire



L'apprentissage semi-supervisé

L'apprentissage semi-supervisé combine des éléments des approches supervisées et non supervisées. Il utilise un petit ensemble de données étiquetées et un plus grand ensemble de données non étiquetées pour former des modèles.

Cette méthode est particulièrement utile quand :

- Les données étiquetées nécessitent des ressources coûteuses pour les obtenir, mais les données non étiquetées sont abondantes.
- L'ajout d'un peu d'information étiquetée peut améliorer significativement la performance de modèles entraînés avec des données non étiquetées.

Les applications typiques incluent :

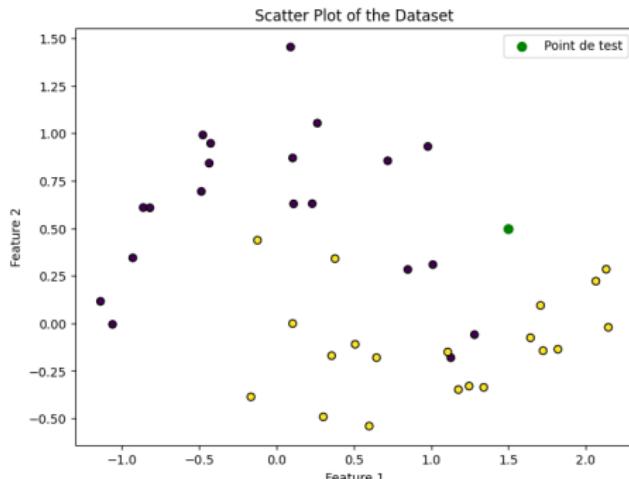
- Développement de systèmes de recommandation plus performants.
- Traitement de langage naturel et analyse de sentiment lorsque les annotations complètes ne sont pas disponibles.

L'apprentissage tente d'exploiter "le meilleur des deux mondes" de l'étiquetage et de la découverte de structure.

Apprentissage supervisé

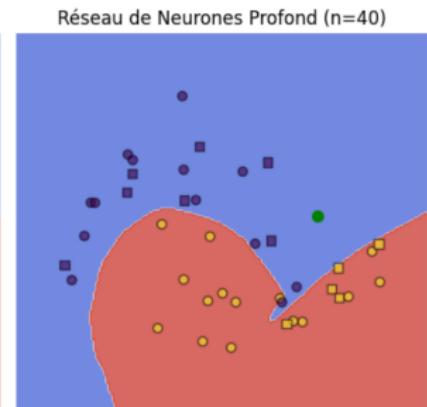
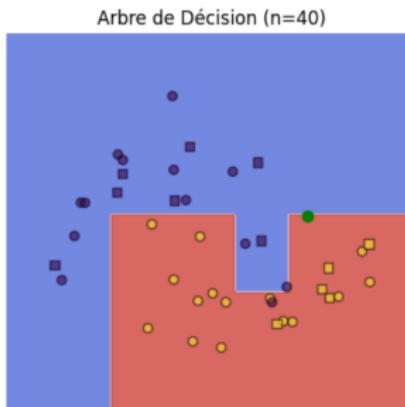
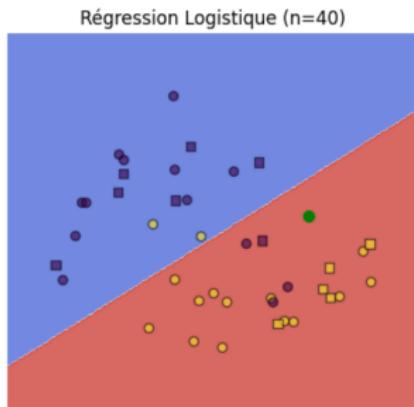
L'apprentissage supervisé comme un problème d'induction

- **Définition** : L'apprentissage supervisé consiste à apprendre une fonction f qui mappe les entrées X aux sorties y , à partir d'un ensemble d'exemples d'entraînement (X, y) .
- **Induction** : Le modèle induit une règle générale à partir de données particulières, dans le but de généraliser à de nouvelles instances.
- **Problème de généralisation** : Comment garantir que le modèle apprend une règle qui s'applique à de nouvelles données ?



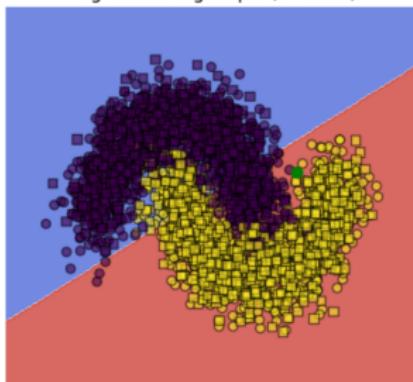
Une indétermination intrinsèque pour le choix du modèle

Il y a une infinité de manières d'induire un modèle à partir d'un échantillon de données d'entraînement.

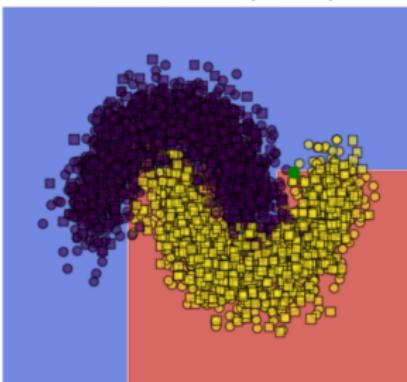


Sous-apprentissage et surapprentissage sur les données d'entraînement

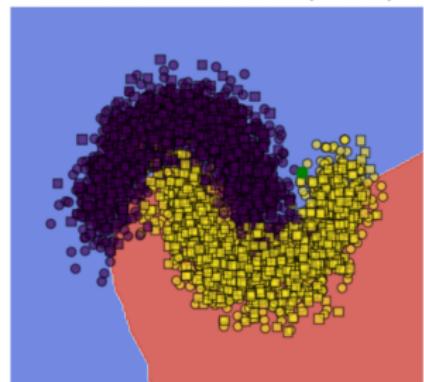
Régression Logistique (n=3000)



Arbre de Décision (n=3000)



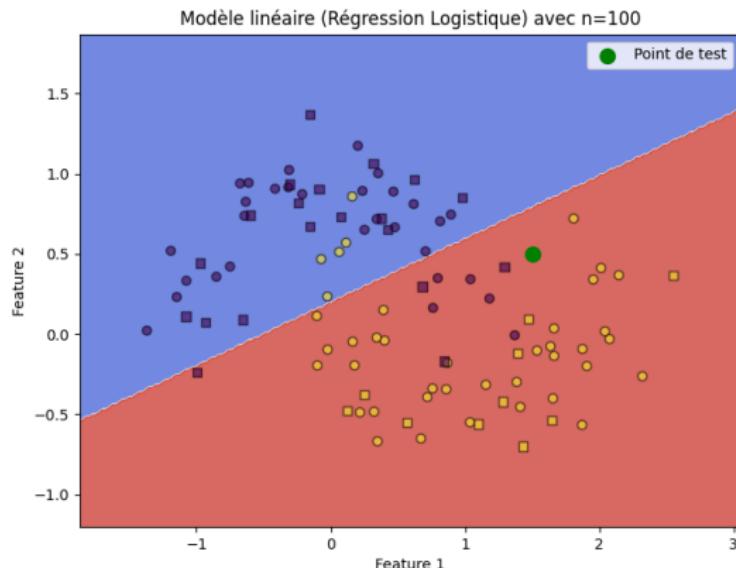
Réseau de Neurones Profond (n=3000)



Sous-apprentissage

Definition

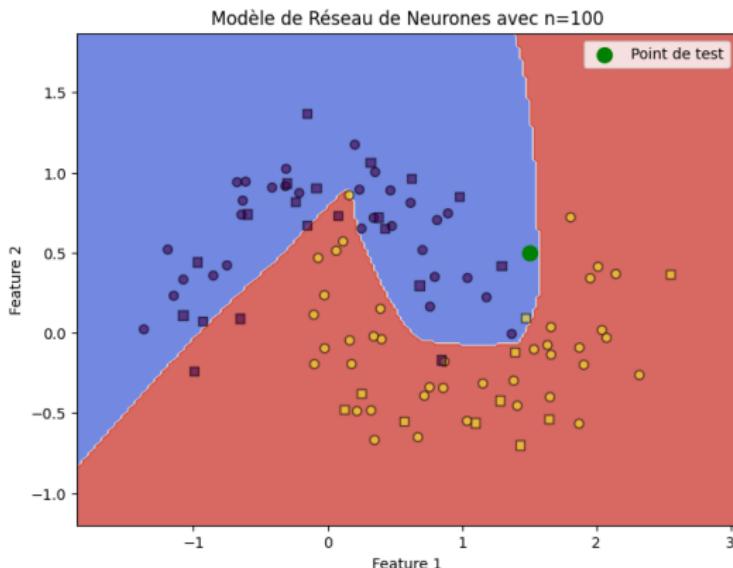
On dit qu'un modèle de machine learning est en régime de sous-apprentissage (underfitting) lorsqu'il n'arrive pas à capturer la complexité (l'information) présente dans le jeu de données d'entraînement.



Sur-apprentissage

Definition

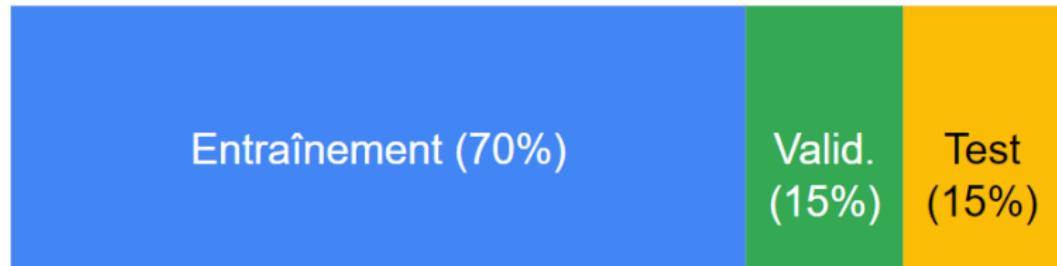
On dit qu'un modèle de machine learning est en régime de sur-apprentissage (overfitting) lorsqu'il n'arrive pas à généraliser à des données non encore observées, i.e. lorsqu'il est trop adapté aux données d'entraînement.



Sélection de modèle

Pour sélectionner le modèle le plus pertinent par rapport à une métrique donnée, on applique la méthodologie suivante :

- On partitionne le jeu de données disponible en trois parties : un jeu d'entraînement, un jeu de validation et un jeu de test.
- On entraîne M modèles sur le jeu d'entraînement.
- On évalue les performances respectives des M modèles sur le jeu de validation et on sélectionne le meilleur.
- Le modèle sélectionné est ensuite évalué sur le jeu de test. Idéalement, le jeu de test est ainsi utilisé une seule fois.



Exemple : sélection de modèle

Modèle	Précision Entraînement	Précision Validation
Régression Logistique	0.90	0.88
Arbre de Décision	0.95	0.92
Réseau de Neurones Profond	1	0.90

Table: Comparaison des précisions des modèles sur les jeux d'entraînement et de validation

Métriques de performance : régression

On dispose d'un certain nombre de métriques pour évaluer les performances des modèles de machine learning. Celles-ci peuvent être divisées en deux catégories.

Régression

- L'erreur quadratique moyenne (MSE) : elle est définie comme la moyenne des carrés des écarts entre les prédictions et les valeurs observées.

$$MSE = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2$$

- La racine carrée de l'erreur quadratique moyenne (RMSE) :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2}$$

Métriques de performance : classification 1/2

Accuracy : L'accuracy est la métrique de base qui permet d'évaluer les performances d'un modèle de classification. Elle est définie comme :

$$\text{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$$

Matrice de confusion : La matrice de confusion est une représentation permettant d'offrir plus de finesse par rapport à l'accuracy, notamment quand le jeu de données est déséquilibré (présence de classes majoritaires). Elle compare les prédictions du modèle avec les valeurs réelles et est structurée comme suit :

		Valeur Prédite	
		Positif	Négatif
Valeur Réelle	Positif	Vrai Positif (VP)	Faux Négatif (FN)
	Négatif	Faux Positif (FP)	Vrai Négatif (VN)

Métriques de performance : classification 2/2

A partir de la matrice de confusion, on peut dériver d'autres métriques :

- Précision : elle est définie comme la proportion des prédictions correctes parmi toutes les prédictions positives :

$$\text{Précision} = \frac{VP}{VP + FP}$$

- Rappel (recall) : il représente la proportion des vrais positifs correctement prédits par le modèle.

$$\text{Rappel} = \frac{VP}{VP + FN}$$

- Score F1 (F1-score) : Le score F1 est défini comme la moyenne harmonique de la précision et du rappel.

$$\text{Score F1} = 2 \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

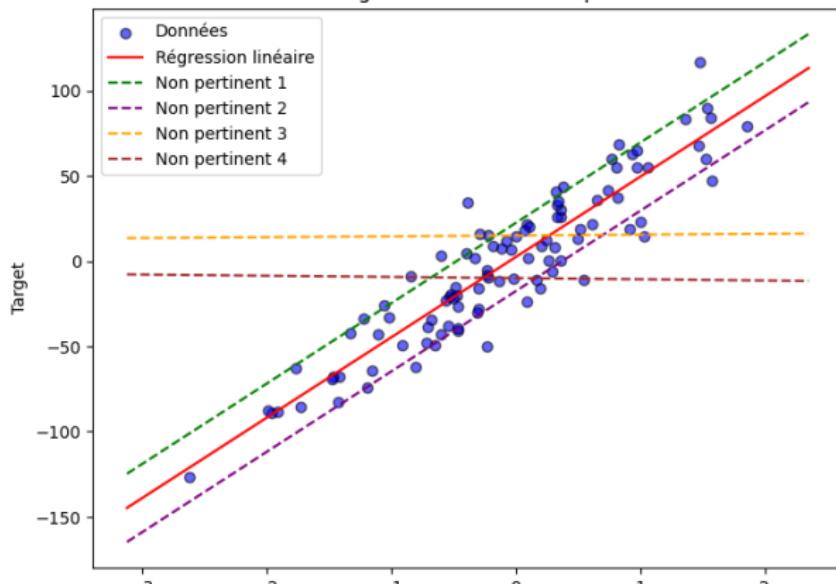
Modèles classiques de machine learning

Régression linéaire simple

Soit un ensemble de n observations x_1, x_2, \dots, x_n avec les labels correspondants y_1, y_2, \dots, y_n , on cherche le modèle linéaire qui ajuste le mieux ces données.

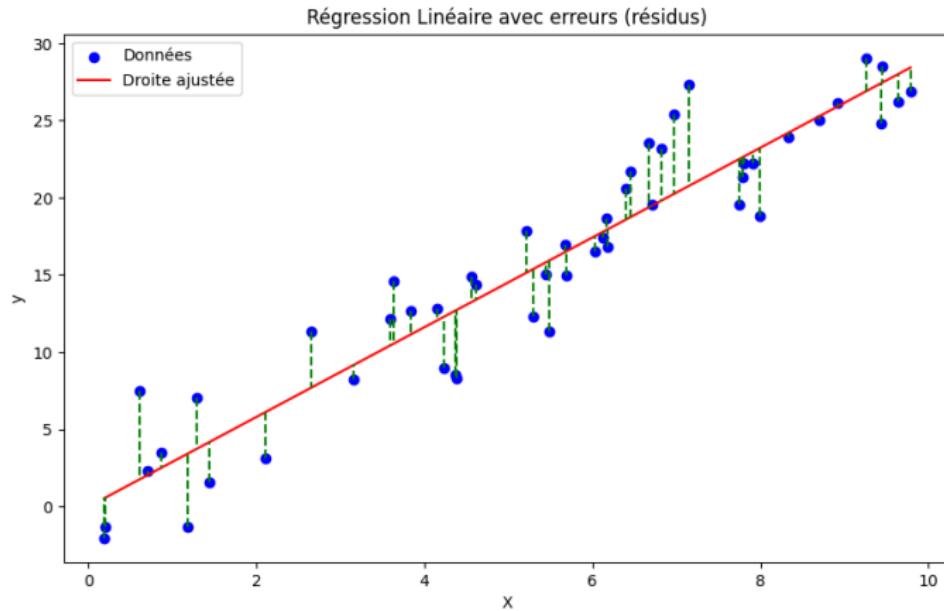
$$\hat{y} = \beta_0 + \beta_1 x$$

Illustration de la régression linéaire avec plusieurs modèles



Minimisation du risque empirique

L'erreur de prédiction pour la i ième observation est : $e_i = y_i - \hat{y}_i$. où $\hat{y}_i = \beta_0 + \beta_1 x_i$.



Exemple : prédire le prix d'un logement en fonction de la surface 1/2

Soit un dataset de 100 points où x représente la surface (en m²) et y le prix des logements (en €). Les 4 premières lignes sont :

x (Surface m ²)	y (Prix en €)
40	120 000
65	200 000
80	250 000
55	160 000
:	:

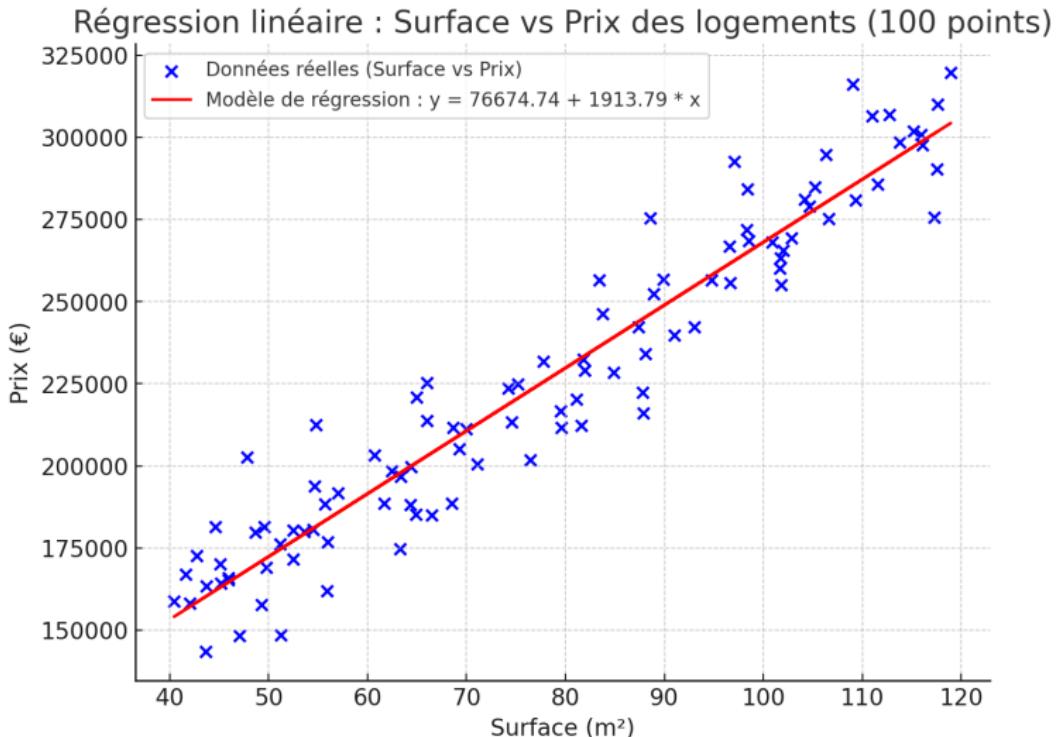
Les coefficients β_0 et β_1 sont calculés à partir des formules suivantes :

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{1234567}{56789} = 2173.15$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 200000 - 2173.15 \times 60 = 69410.5$$

L'équation de la droite de régression obtenue est donc :

Exemple : prédire le prix d'un logement en fonction de la surface 2/2



Régression linéaire multiple

On considère n observations X^1, X^2, \dots, X^n où chaque observation X^i est désormais un vecteur ayant p composantes (p variables explicatives).

$$X^i = \begin{pmatrix} x_1^i \\ x_2^i \\ \vdots \\ x_p^i \end{pmatrix}$$

La régression linéaire s'écrit alors :

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

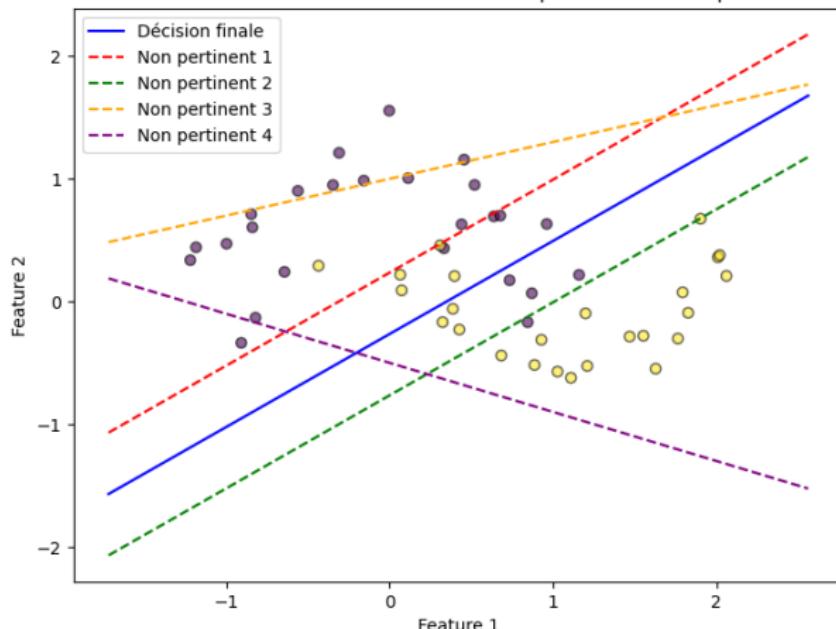
Les coefficients $\beta_0, \beta_1, \dots, \beta_p$ sont déterminés par la méthode des moindres carrés :

$$\arg \min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n \left(y^i - (\beta_0 + \sum_{j=1}^p \beta_j x_j^i) \right)^2$$

Principe de la régression logistique

- **Définition :** La régression logistique est un algorithme de classification linéaire utilisé pour prédire la probabilité qu'une observation appartienne à une classe donnée.

Illustration de la surface de décision avec plusieurs droites possibles



Régression logistique : introduction

La régression logistique est une technique d'analyse statistique utilisée pour modéliser la probabilité d'une variable dépendante binaire. C'est un cas particulier de modèle linéaire généralisé qui est utilisé pour des problèmes de classification.

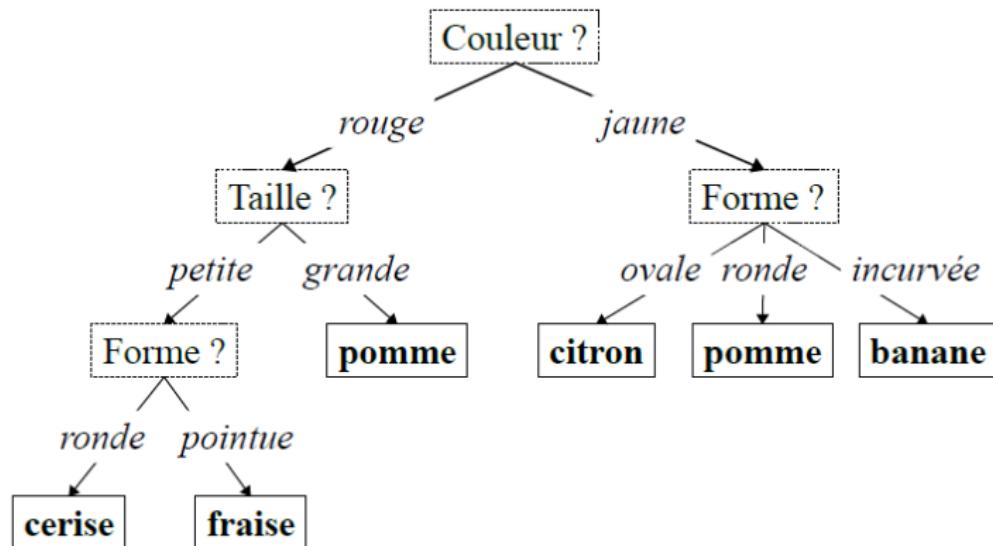
Principes de la régression logistique :

- **Variable dépendante** : On cherche la probabilité que la variable dépendante (y) appartienne à une classe (0 ou 1, vrai ou faux, succès ou échec). Autrement dit, on cherche à modéliser $P(y = 1)$ en fonction des variables dépendantes (explicatives) x .
- **Odds ratio** : Plus concrètement, on cherche à exprimer la côte anglaise (odd ratio) en fonction des variables dépendantes (x).

$$\ln \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Arbres de décision

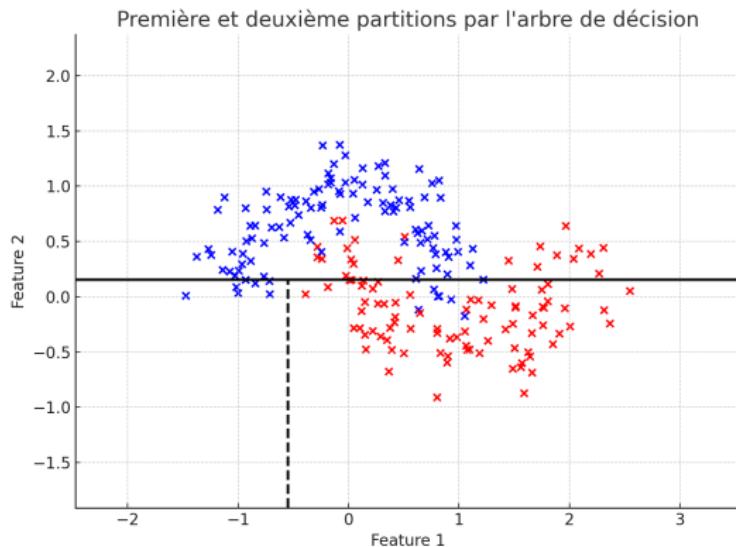
Les arbres de décisions sont des modèles dont le processus de décision est hiérarchique et prend la forme d'un arbre.



Entraînement des arbres de décision

Les arbres de décisions sont généralement entraînés à l'aide de la technique CART (Classification And Regression Trees).

Etant donné un ensemble d'observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, les arbres de décision partitionnent cet espace en plusieurs régions R_1, R_2, \dots, R_m .



Forêts aléatoires (random forests)

La technique des forêts aléatoires (random forests) consiste à appliquer une approche de type bagging sur les arbres de décision.

L'algorithme random forests suit cette procédure :

- Tirer par bootstrap B échantillons de tailles n à partir de l'ensemble D .
- Pour chaque échantillon tiré, construire un arbre en répétant les étapes suivantes jusqu'à atteindre n_{min} .
 - Tirer d'une manière aléatoire m variables parmi les p variables.
 - Sélectionner la meilleure variable avec le meilleur point de partitionnement.
 - Partitionner le noeud en deux sous-branches.
- Agréger les arbres construits.

Les prédictions sont agrégées selon qu'il s'agisse de régression ou de classification :

- Régression : moyenne $f^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$
- Classification : vote majoritaire.

Remarques

- L'algorithme de random forests intègre nativement une forme de validation croisée. Les performances mesurées sur $\bigcup_{b_i \neq b_k}$ (out of bag ou OOB) sont souvent proches de celles que l'on pourrait mesurer avec une validation croisée.
- Le nombre de variable tirées pour chaque noeud est généralement donné par \sqrt{p} pour la classification et $\frac{p}{3}$ pour la régression. Cet hyperparamètre dépend cependant du problème considéré.
- Lorsque le nombre de variable est élevé alors que le nombre de variables réellement partinente est faible, la probabilité que les p variables sélectionnées pour chaque partitionnement incluent des variables pertinentes devient faible, et les performances du modèle en termes de généralisation peuvent se détériorer considérablement.
- L'algorithme random forests permet de restituer des informations sur l'importance des variable (feature importance).

Apprentissage non supervisé

Apprentissage non supervisé

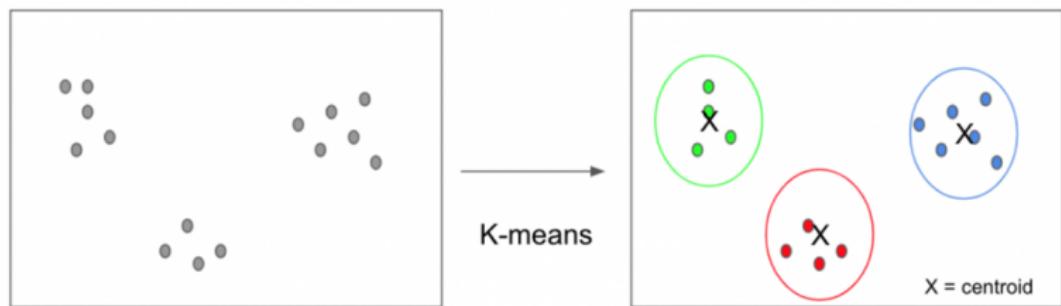
Dans l'apprentissage non supervisé, on considère n observations sans labels. On s'intéresse fondamentalement à la probabilité jointe de ces observations.

On peut distinguer deux grandes catégories d'apprentissage non supervisé :

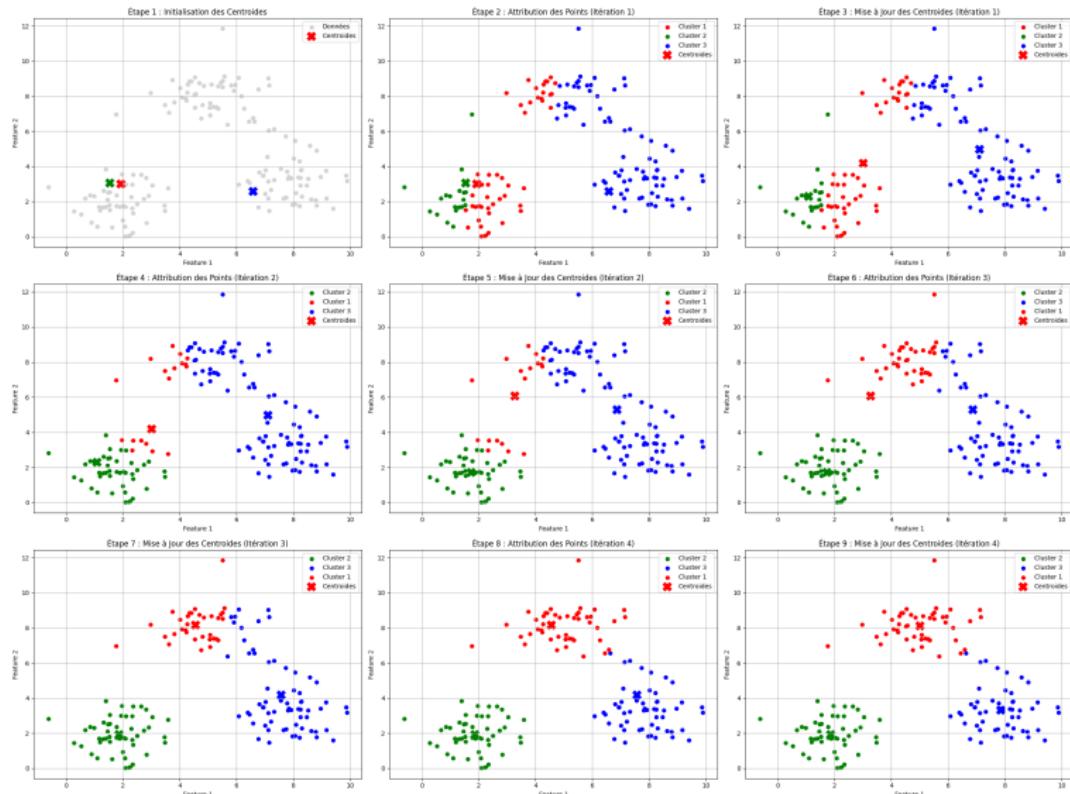
- **Clustering (partitionnement)** : cela consiste à partitionner les n observations en K groupes pertinents (généralement le critère de pertinence a une signification d'un point de vue métier).
- **Détection d'anomalie** : il s'agit de détecter des observations qui présentent un profil (des features) inhabituels par rapport au profil moyen de la majorité des observations.

K-means

L'algorithme de Lloyd tente de regrouper les données en clusters en minimisant les distances entre les points d'un même cluster tout en maximisant les distances entre points appartenant à différents clusters.



Algorithme de Lloyd



Remarques

- L'algorithme des k-means étant basé sur une distance euclidienne, il est nécessaire de normaliser les données avant de l'exécuter.
- L'algorithme des k-means est très sensible aux données aberrantes (outliers). Il faut donc considérer les données d'une manière attentive. Cependant, cela permet également d'utiliser l'algorithme des k-means pour la détection automatique des outliers.
- Les centroïdes étant initialisés d'une manière aléatoire, les clusters obtenus ne sont pas stables ; les clusters peuvent changer d'une exécution à l'autre. Il existe cependant une variante plus stable, appelée k-means++, qui permet de sélectionner les centroïdes d'une manière semi-aléatoire.
- Il est possible de partitionner les données avec une métrique plus générale que la distance euclidienne. On peut définir un algorithme k-means à noyau sur un espace de Hilbert pour aller au-delà de la métrique euclidienne.
- **K-means n'est pas adapté aux données en grande dimension.**

Isolation Forest : Principe

L'Isolation Forest est une technique de détection d'anomalies basée sur l'isolement des observations. Son efficacité repose sur l'hypothèse que les anomalies sont "faciles à isoler" par rapport aux observations normales.

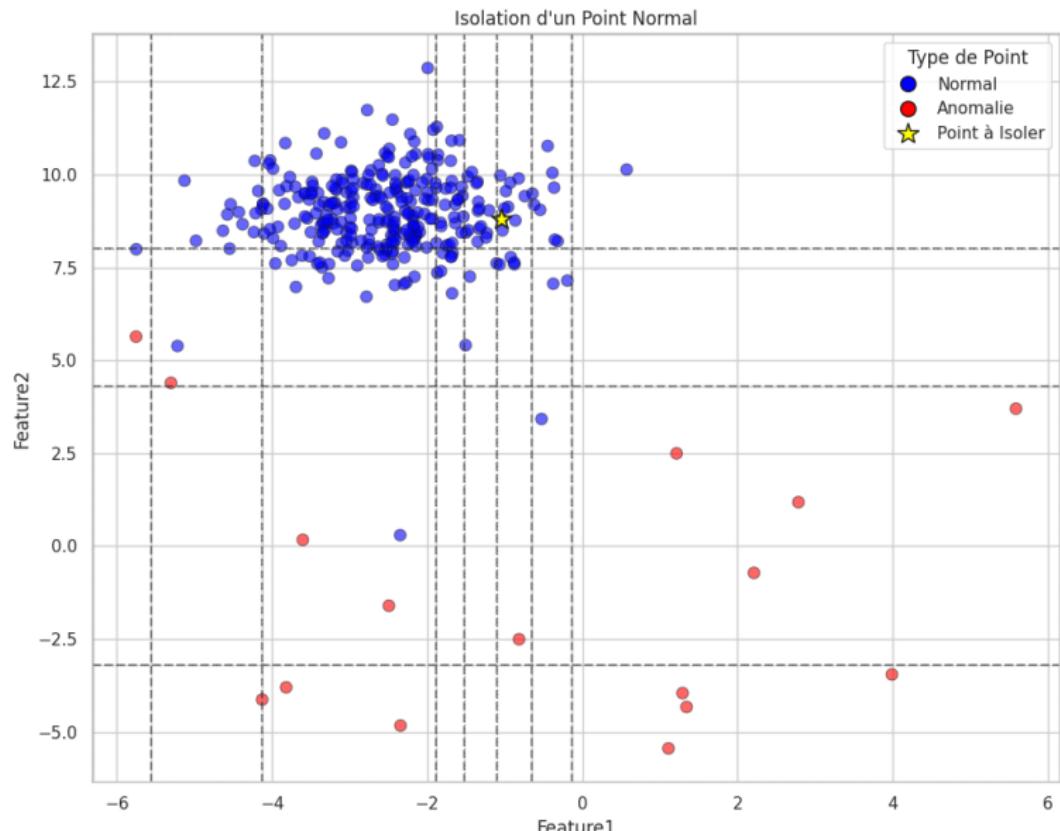
- Fonctionne en construisant des arbres d'isolement à partir de sous-ensembles de données.
- Isoler une observation signifie la séparer des autres par des divisions aléatoires de l'espace des caractéristiques.
- Moins de divisions sont nécessaires pour isoler une anomalie, ce qui constitue le fondement du score d'anomalie.

Les arbres d'isolement sont construits de manière récursive :

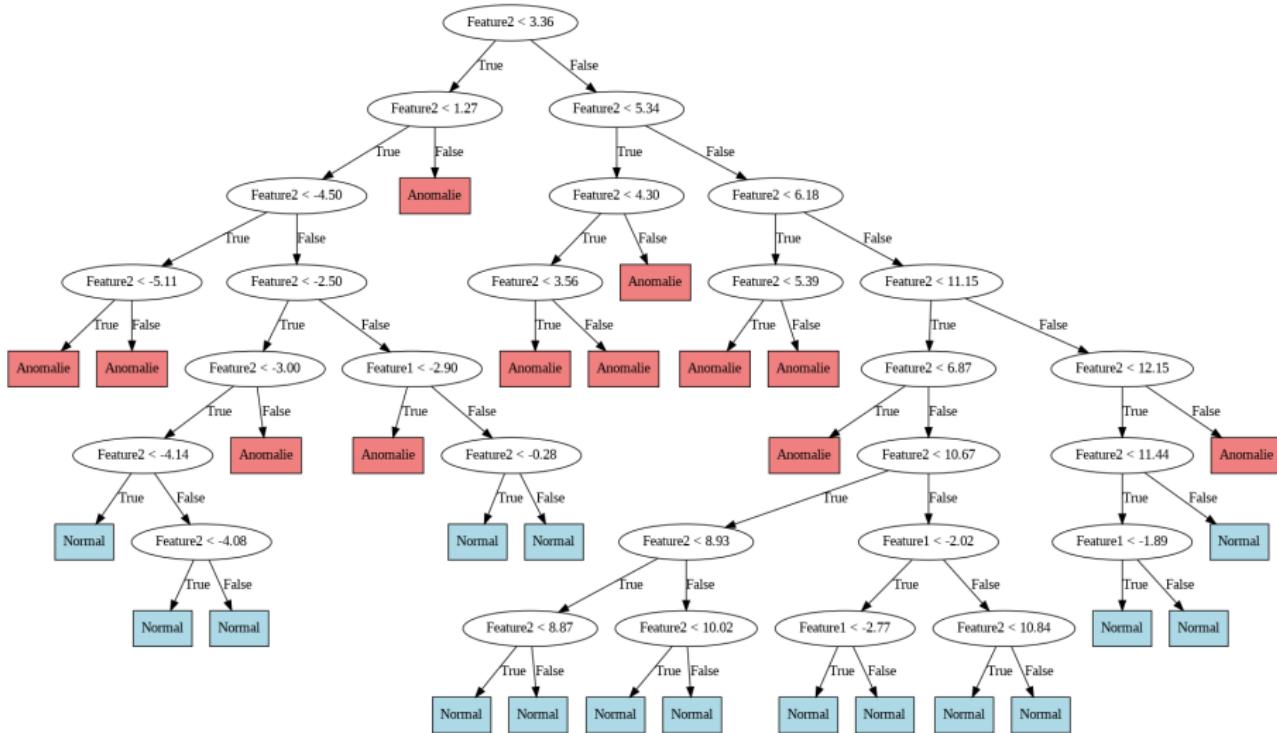
- ➊ Sélection aléatoire d'un sous-ensemble de données.
- ➋ Choix aléatoire d'une caractéristique et d'une valeur de seuil pour diviser le sous-ensemble.
- ➌ Répétition des divisions jusqu'à l'isolement des observations ou atteinte d'une limite de profondeur prédéfinie.

Chaque arbre est ainsi unique, offrant une perspective différente sur les données.

Isolation tree : fonctionnement 1/2



Isolation tree : fonctionnement 2/2



Deep learning

Brève histoire du deep learning

1940-1980 : Fondations

- 1943 : Neurone artificiel (McCulloch & Pitts).
- 1986 : Rétropropagation (Rumelhart, Hinton, Williams).

1990-2000 : Hiver

- Manque de données et puissance de calcul.
- Progrès en théorie : SVM, modèles bayésiens.

2010-2020 : Renaissance

- 2012 : AlexNet révolutionne la vision par ordinateur.

Depuis 2020 : Multimodalité et LLM

- 2017 : Transformer, base des LLM.
- 2020 : GPT-3, BERT.
- 2022 : ChatGPT...

Introduction aux réseaux de neurones

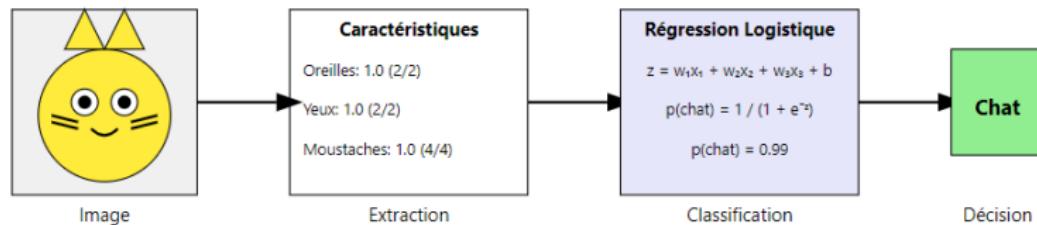
Définition : Les réseaux de neurones sont des modèles computationnels inspirés par le fonctionnement des neurones dans le cerveau humain. Ils sont capables d'apprendre des tâches complexes en modélisant des relations non linéaires entre les entrées et les sorties.

Caractéristiques :

- **Extraction automatique des features** : Capacité d'adaptation et d'extraction des features à partir des données sans programmation explicite.
- **Modélisation non linéaire** : Aptitude à capturer des relations complexes dans les données.
- **Modélisation en grande dimension** : Les modèles de deep learning sont particulièrement adaptés pour les données en grande dimension (images, texte, etc.).
- **Flexibilité** : Applicables à un large éventail de tâches et de typologies de données (images, langage naturel, données graphiques, etc.).

Extraction de features

Extraction de features en machine learning "classique"



Extraction de features en deep learning

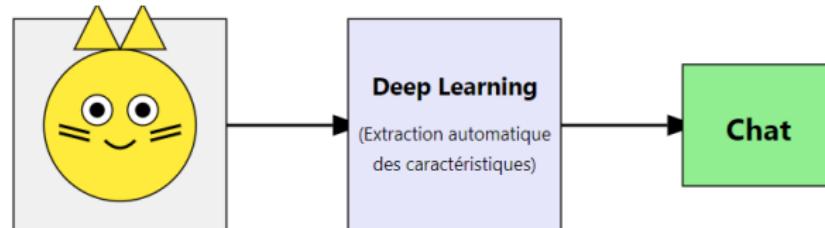
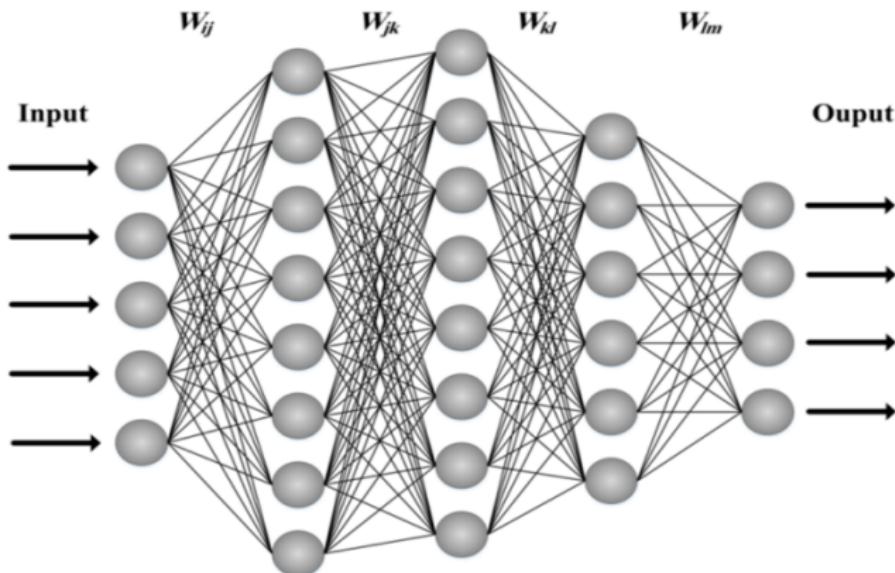
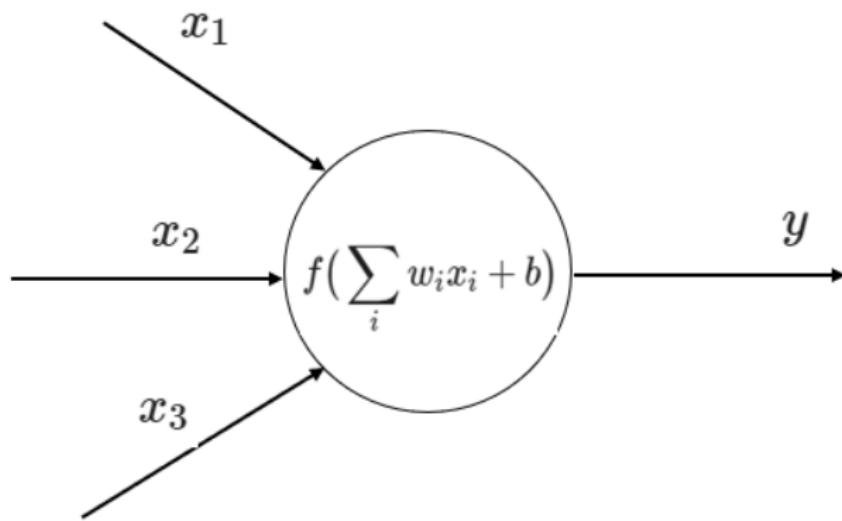


Illustration d'un réseau de neurones classique (MLP)



Neurone aritificial



Fonctions d'activation

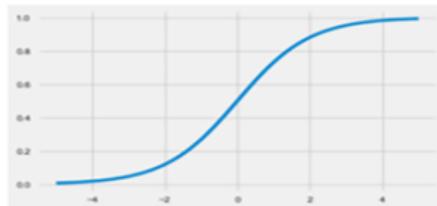
Les fonctions d'activation permettent aux modèles d'apprendre des relations plus complexes en capturant les non-linéarités dans les données.

- **Sigmoïde** : $\sigma(z) = \frac{1}{1+e^{-z}}$, plage de sortie (0, 1), utilisée pour la probabilité dans la classification binaire.
- **Tangente Hyperbolique (tanh)** : $\tanh(z)$, plage de sortie (-1, 1), version centrée et normalisée de la sigmoïde.
- **ReLU (Unité Linéaire Rectifiée)** : $f(z) = \max(0, z)$, non saturante, favorise la convergence rapide et permet d'éviter le problème de disparition des gradients.
- **Leaky ReLU** : $f(z) = \max(\alpha z, z)$, variante de ReLU qui permet un petit gradient lorsque $z < 0$.
- **Softmax** : Utilisée pour la couche de sortie des problèmes de classification multi-classes.

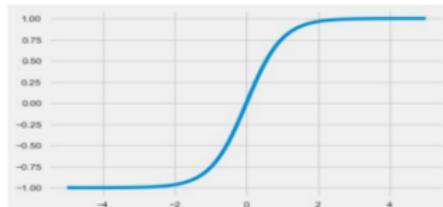
$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

Illustration des fonctions d'activation

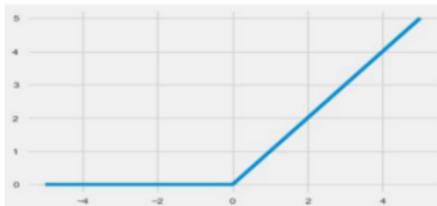
Sigmoïde



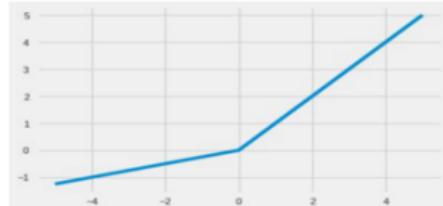
Tanh



ReLU



ReLU paramétrique



Entraînement d'un réseau de neurones artificiel

Principe d'Entraînement : L'entraînement d'un réseau de neurones consiste à ajuster ses poids pour minimiser une fonction de coût qui mesure l'erreur entre les prédictions et les vraies valeurs.

Descente de gradient stochastique (SGD) :

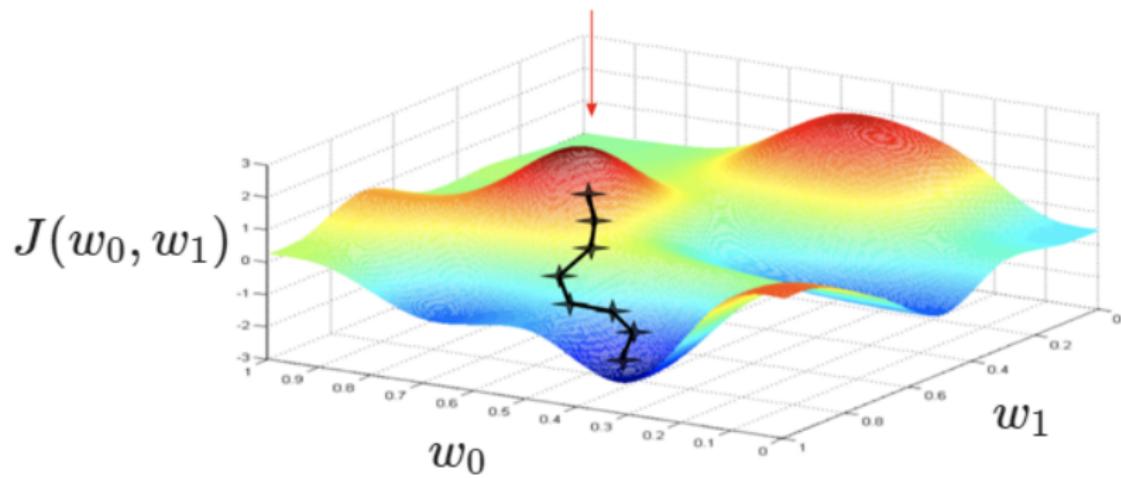
- Méthode d'optimisation utilisée pour mettre à jour les poids du réseau de manière itérative.
- À chaque itération, un sous-ensemble (batch) de données est utilisé pour calculer le gradient de la fonction de coût.
- Les poids sont mis à jour dans la direction opposée du gradient pour réduire l'erreur.

$$w_{new} = w_{old} - \eta \cdot \nabla_w J(w)$$

où :

- w_{old} et w_{new} sont les valeurs des poids avant et après la mise à jour,
- η est le taux d'apprentissage,
- $\nabla_w J(w)$ est le gradient de la fonction de coût par rapport aux poids.

Illustration graphique de la SGD



Computer vision

Introduction à la vision par ordinateur

La vision par ordinateur est une discipline de l'informatique qui permet aux ordinateurs de déduire des informations à partir d'images numériques, de vidéos et d'autres entrées visuelles similaires. Elle simule la capacité humaine de voir et de comprendre le contenu visuel.

- **Objectifs :**

- Interpréter et comprendre le contenu visuel de manière automatisée.
- Extraire et analyser des informations des données visuelles.

- **Applications :**

- Classification d'images.
- Détection d'objets.
- Segmentation d'images.
- Réduction de la dimensionnalité et débruitage.
- Génération d'images.

Introduction aux réseaux de neurones convolutifs (CNN)

Les réseaux de neurones convolutifs (CNN) sont une catégorie de réseaux de neurones profonds, spécialement efficaces pour l'analyse d'images. Ils sont largement utilisés pour des tâches de vision par ordinateur telles que la classification d'images, la reconnaissance de motifs, et la détection d'objets.

- **Caractéristiques clés :**

- Utilisation de convolutions pour extraire des caractéristiques de l'image.
- Capacité à capturer des hiérarchies de caractéristiques spatiales et temporelles.

- **Importance :**

- Efficacité dans le traitement des données visuelles.
- Réduction de la dimensionnalité, tout en préservant les caractéristiques essentielles.

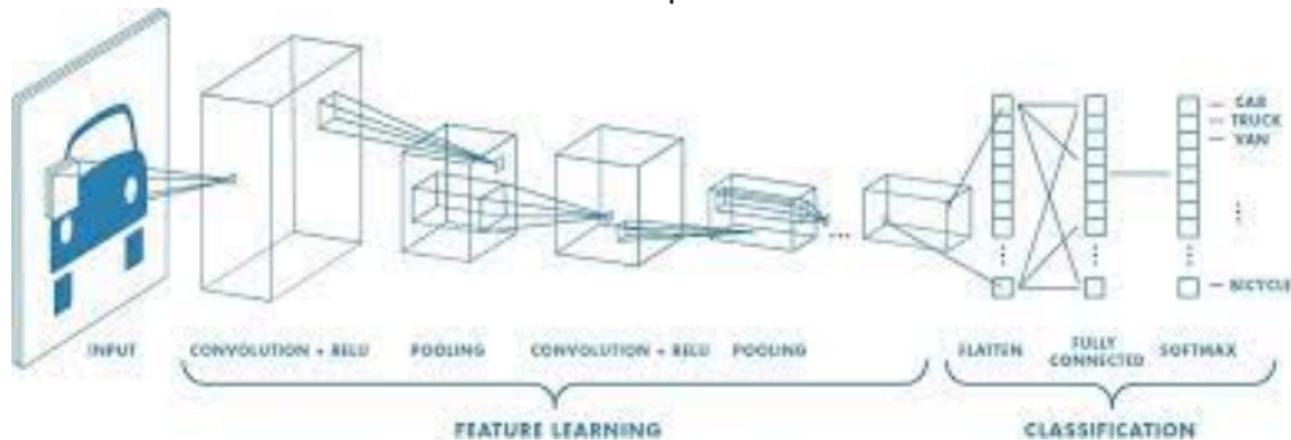
Architecture basique d'un CNN

Les CNNs sont structurés en plusieurs couches qui transforment progressivement l'entrée (image) pour aboutir à une sortie (décision).

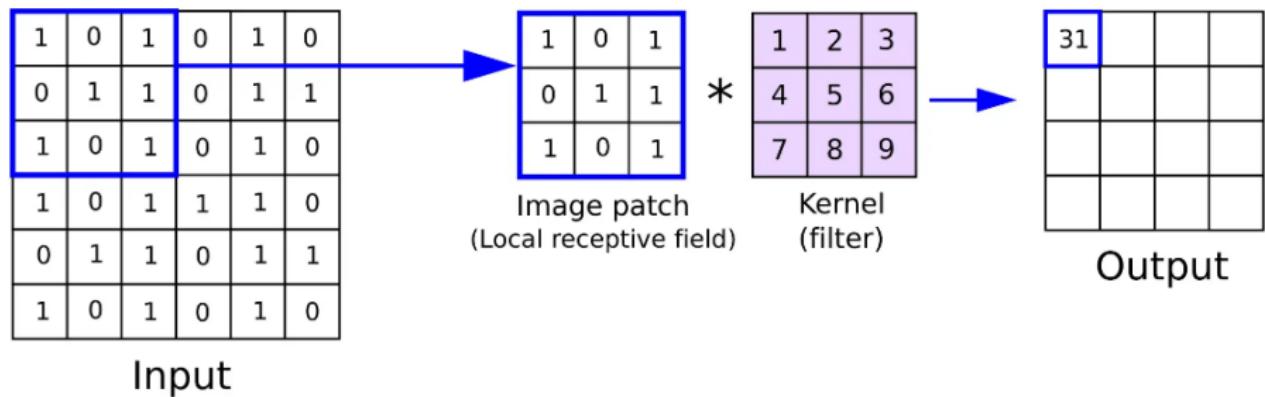
- ① **Couche de convolution** : Applique des filtres pour extraire des caractéristiques de bas niveau comme les bords et les textures.
- ② **Couche de pooling (ou Sous-échantillonnage)** : Réduit la dimensionnalité de chaque carte de caractéristiques tout en préservant les informations les plus importantes.
- ③ **Couche de normalisation** : Normalise les activations de la couche précédente, souvent pour améliorer la convergence.
- ④ **Couches entièrement Connectées (Fully Connected)** : Utilisées en fin de réseau pour classer les images en fonction des caractéristiques extraites.

Architecture d'un CNN

Le fonctionnement d'un CNN peut être visualisé comme un enchaînement de transformations où chaque couche traite et transforme les informations provenant de la couche précédente.



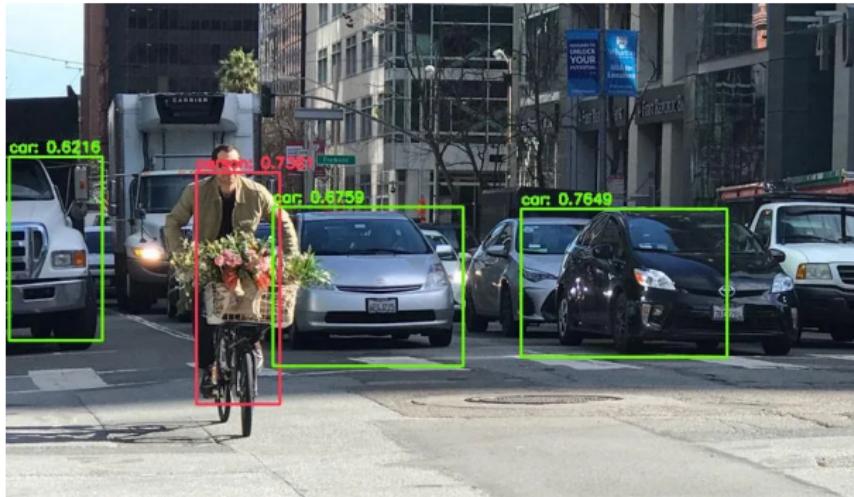
Fonctionnement des filtres d'un CNN



Introduction à YOLO (You Only Look Once)

YOLO est un système de détection d'objets révolutionnaire car il traite l'image entière lors d'une seule évaluation de réseau, ce qui lui permet de détecter les objets en temps réel.

- **Principe :** YOLO divise l'image en une grille et prédit les boîtes englobantes et les probabilités de classe pour chaque grille simultanément.
- **Avantages :** Extrêmement rapide par rapport à d'autres détecteurs qui proposent des régions puis classifient (comme R-CNN).



Entraînement de YOLO pour la détection d'objets

L'entraînement de YOLO est un processus crucial pour assurer une détection efficace et précise. Il comprend plusieurs étapes et paramètres importants à configurer.

- **Préparation des données :**

- Collecte d'un ensemble de données d'images annotées avec des boîtes englobantes autour des objets.
- Normalisation des images et des coordonnées des boîtes pour les adapter à l'entrée du réseau.

- **Architecture du réseau :**

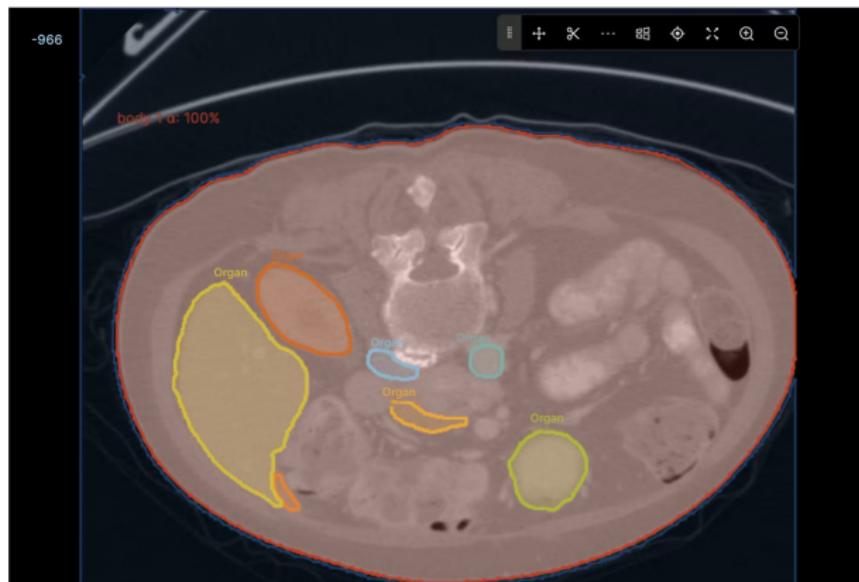
- Utilisation de Darknet, un réseau de neurones convolutionnel spécialement conçu pour YOLO.
- Configuration de différentes couches pour extraire et apprendre des caractéristiques à différentes échelles.

- **Fonction de perte :**

- Fonction de perte composite qui inclut la perte pour les erreurs de prédiction des classes, des boîtes englobantes et de la confiance des objets.
- Optimisation pour réduire les fausses détections et améliorer la localisation précise des objets.

Introduction à la segmentation d'images avec U-Net

U-Net est une architecture de réseau de neurones convolutifs développée principalement pour la tâche de segmentation d'images médicales. Elle est caractérisée par sa structure en forme de "U", optimisée pour travailler avec moins de données d'entraînement mais avec une grande précision de localisation.



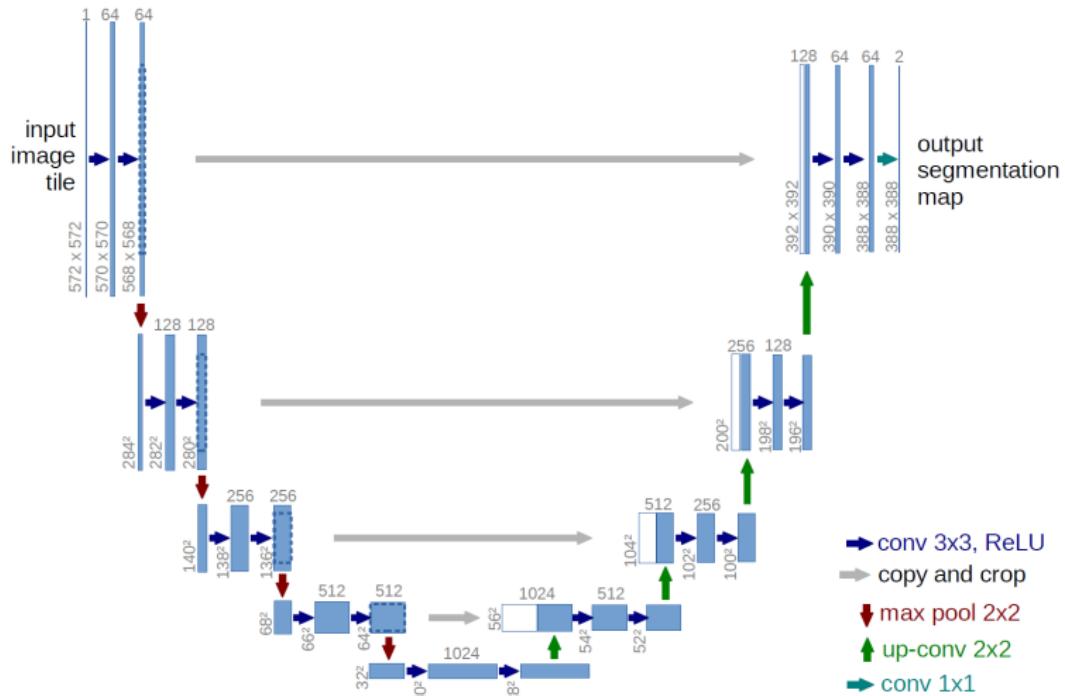
Architecture de U-Net

L'architecture de U-Net comprend deux parties principales: un encodeur (contraction) et un décodeur (expansion).

- **Encodeur :** Composé de couches de convolution suivies de max pooling pour la réduction de dimensionnalité, capturant les caractéristiques pertinentes à différentes échelles.
- **Décodeur :** Comprend des opérations de convolutions transposées pour agrandir les cartes de caractéristiques, et des connexions concaténées depuis l'encodeur pour récupérer les informations spatiales.

Fonctionnement de U-Net

U-Net prédit un masque pour chaque pixel d'une image, indiquant l'appartenance à une classe spécifique, ce qui est particulièrement utile pour les tâches de segmentation fine où chaque pixel importe.



Applications de U-Net en Vision par Ordinateur

Bien que développé pour la segmentation médicale, U-Net a trouvé des applications dans divers autres domaines nécessitant une segmentation précise d'images.

- **Applications :**

- Détection de défauts dans les matériaux industriels.
- Segmentation de routes et de bâtiments dans les images satellitaires.
- Applications en agriculture pour le suivi des cultures.

Introduction aux réseaux de neurones siamois

Les réseaux siamois sont une architecture spéciale de réseaux de neurones conçue pour comparer deux entrées et déterminer leur similitude. Ils sont particulièrement efficaces pour des tâches telles que la vérification d'identité et la comparaison d'images.

- **Principe :** Utilise deux réseaux neuronaux identiques, qui partagent les mêmes poids, pour extraire les caractéristiques de chaque entrée.
- **Utilisation typique :** Comparaison de paires d'images pour déterminer si elles sont similaires ou non, basée sur l'apprentissage par paires ou triplets.

Architecture des Réseaux Siamois

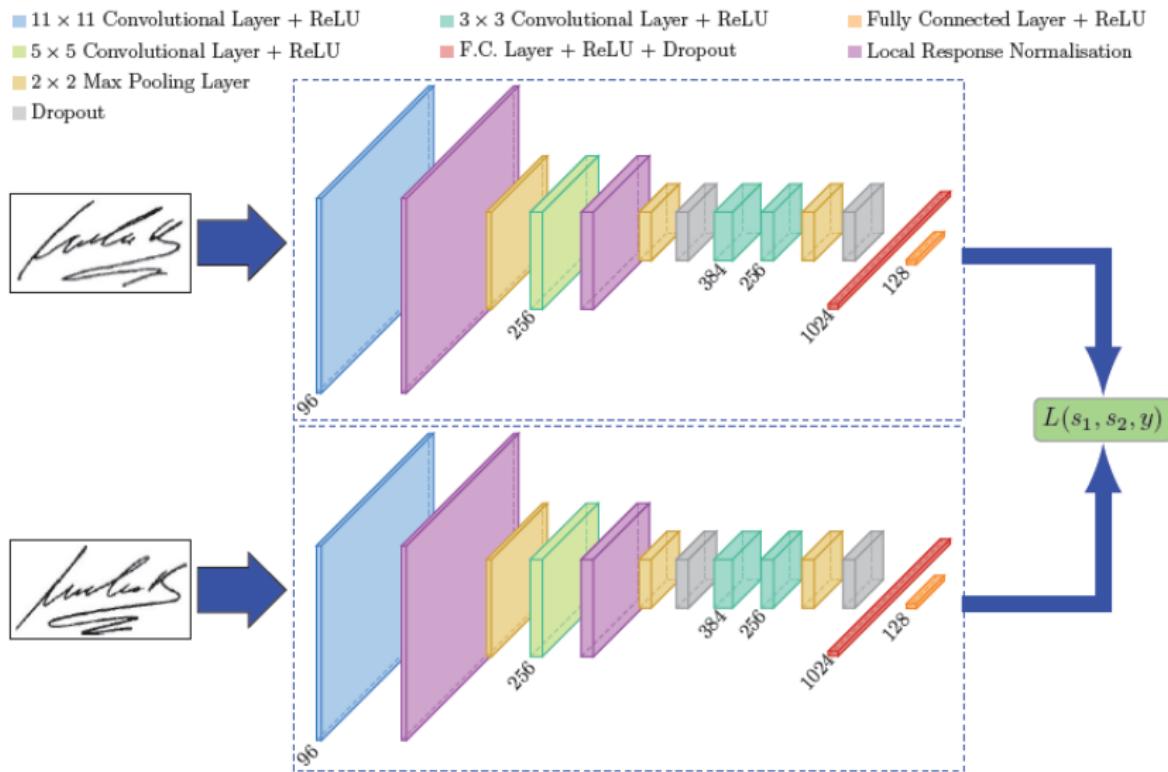
L'architecture siamoise se compose de deux réseaux identiques qui convergent en un point pour comparer leurs caractéristiques. Chaque branche encode une entrée dans un espace de caractéristiques qui permet de mesurer des distances.

- **Composants :**

- Deux réseaux de neurones identiques.
- Une couche qui calcule une distance ou une métrique entre les représentations obtenues.

- **Métrique de distance :** Souvent, la distance euclidienne ou la distance de Manhattan est utilisée pour calculer la similarité entre les représentations.

Architecture des réseaux siamois



Applications des réseaux siamois

Les réseaux siamois ont une variété d'applications en vision par ordinateur, allant au-delà de la simple comparaison d'images.

- **Applications :**

- Vérification de l'identité dans les systèmes biométriques.
- Reconnaissance de visages même avec des variations de pose ou d'illumination.
- Suivi d'objets en vidéo où les modèles doivent reconnaître un objet au fil du temps.

Introduction aux Generative Adversarial Networks (GANs)

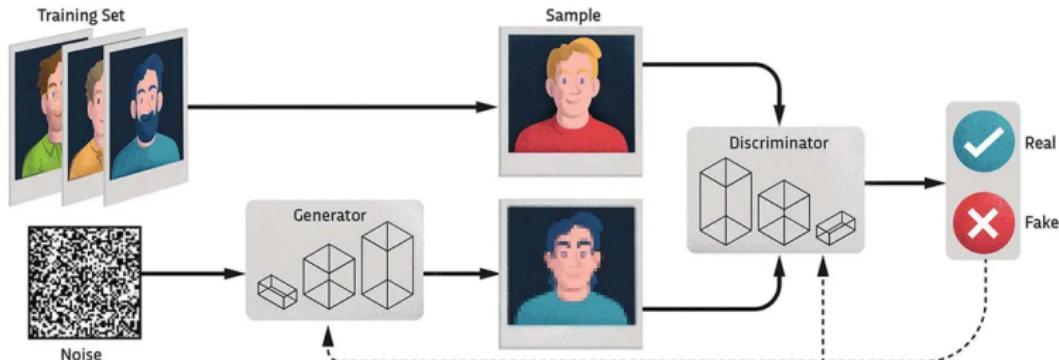
Les Generative Adversarial Networks (GANs) sont une classe de systèmes d'apprentissage automatique où deux réseaux neuronaux se livrent à un jeu compétitif. L'un génère des données (générateur) et l'autre évalue leur authenticité (discriminateur).

- **Fonctionnement** : Le générateur tente de créer des données suffisamment réalistes pour tromper le discriminateur, tandis que le discriminateur apprend à distinguer les fausses données des vraies.
- **Importance** : Les GANs sont particulièrement puissants pour générer des images réalistes et peuvent être utilisés pour améliorer les modèles de reconnaissance d'images.

Architecture des GANs

L'architecture d'un GAN se compose de deux parties principales : le générateur et le discriminateur.

- **Générateur** : Prend une entrée de bruit aléatoire et construit des données (par exemple, des images).
- **Discriminateur** : Reçoit soit les données générées, soit les vraies données et doit déterminer si elles sont authentiques ou non.



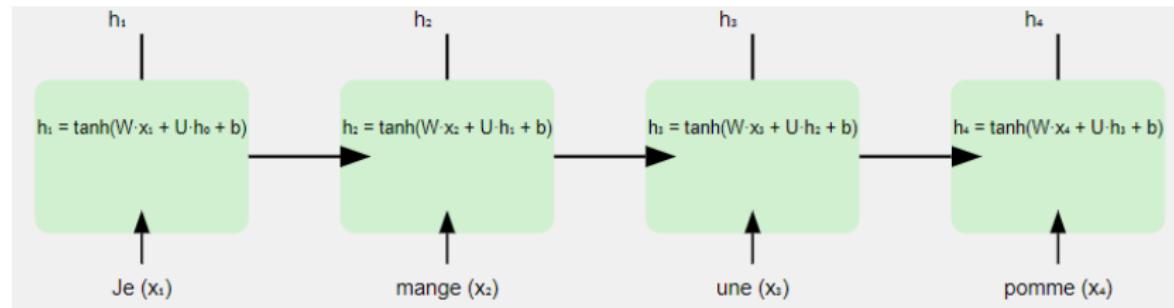
Traitement automatique du langage (NLP)

Introduction aux RNN

Les Réseaux de Neurones Récurrents (RNN) sont une classe de réseaux de neurones artificiels spécialement conçus pour traiter des séquences de données, tels que des séries temporelles ou des séquences textuelles.

Caractéristiques:

- **Mémoire à court terme** : Les RNN ont la capacité de se "souvenir" d'informations passées grâce à leurs connexions récurrentes.
- **Traitement de séquences** : Ils sont particulièrement adaptés pour des tâches où les données sont séquentielles et où le contexte est important (ex: langage naturel, musique).

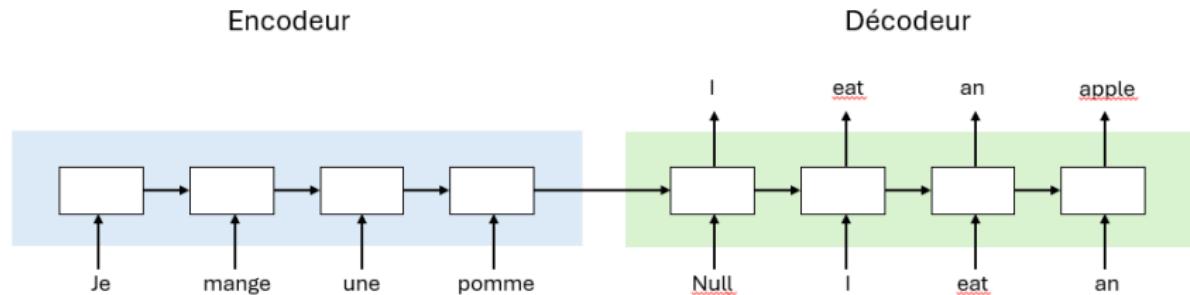


Modèles Seq-to-Seq

Les modèles de séquence à séquence (seq-to-seq) sont une classe de modèles en apprentissage profond conçus pour transformer une séquence d'entrée en une séquence de sortie.

L'architecture seq-to-seq comprend deux composants principaux :

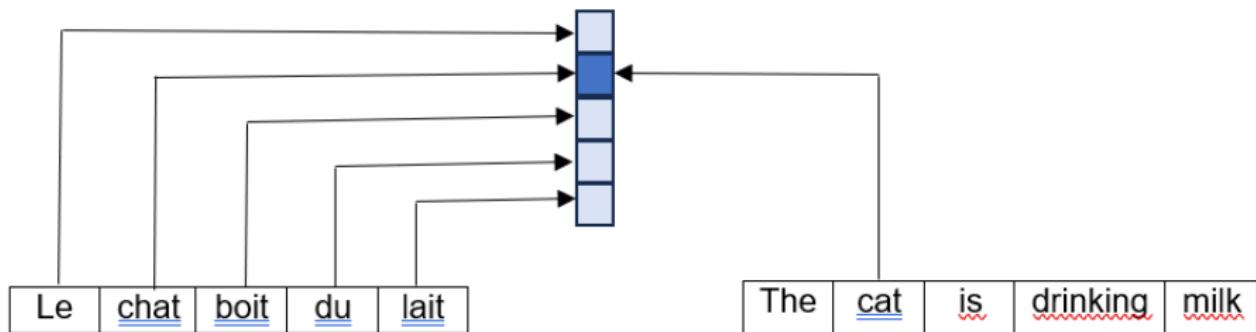
- **Encodeur** : Un réseau de neurones qui lit et encode la séquence d'entrée en un vecteur de contexte (une représentation dense de la séquence).
- **Décodeur** : Un autre réseau de neurones qui lit le vecteur de contexte pour générer la séquence de sortie, un élément à la fois.



Modèles Seq2Seq avec mécanisme d'Attention

Le mécanisme d'attention est une amélioration clé apportée aux modèles seq-to-seq traditionnels, permettant au modèle de se "concentrer" sur différentes parties de la séquence d'entrée lors de la génération de la séquence de sortie.

- **Objectif** : Surmonter les limitations des encodeurs seq-to-seq qui compressent toute l'information d'une séquence d'entrée dans un vecteur de contexte fixe.
- **Avantage** : Améliore la capacité du modèle à gérer de longues séquences d'entrée, en rendant le processus de génération de la séquence de sortie plus dynamique et contextuellement informé.



Transformers et LLMs

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

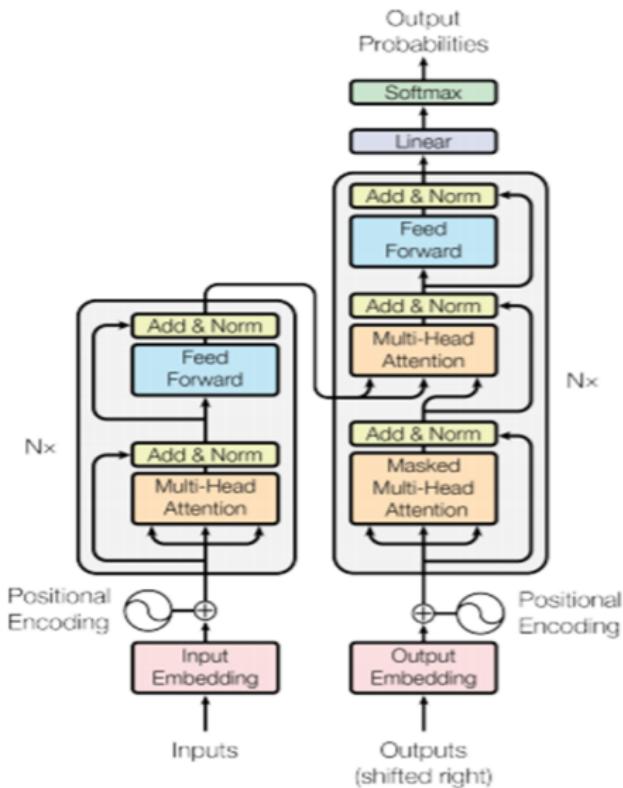
Lukasz Kaiser*
Google Brain
lukasz.kaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

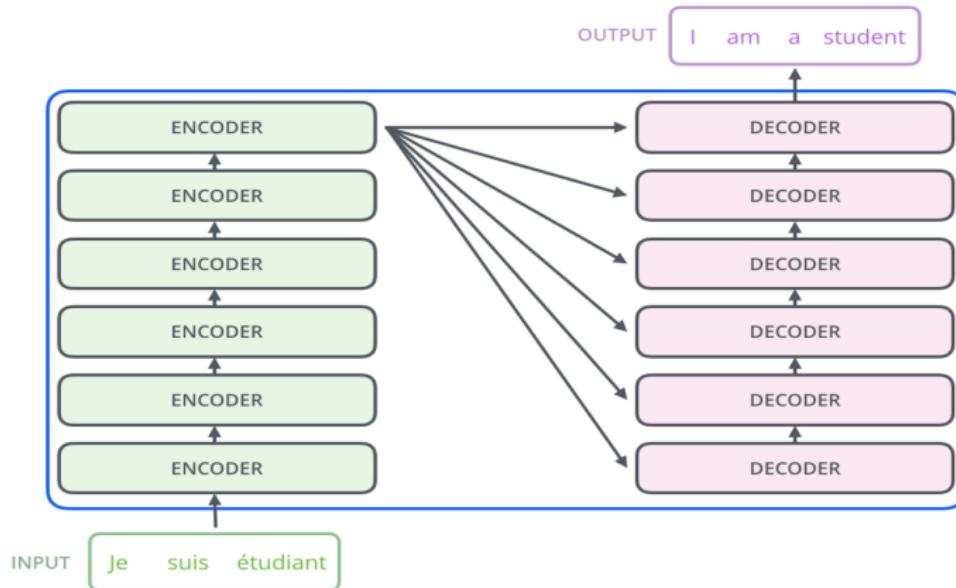
Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to

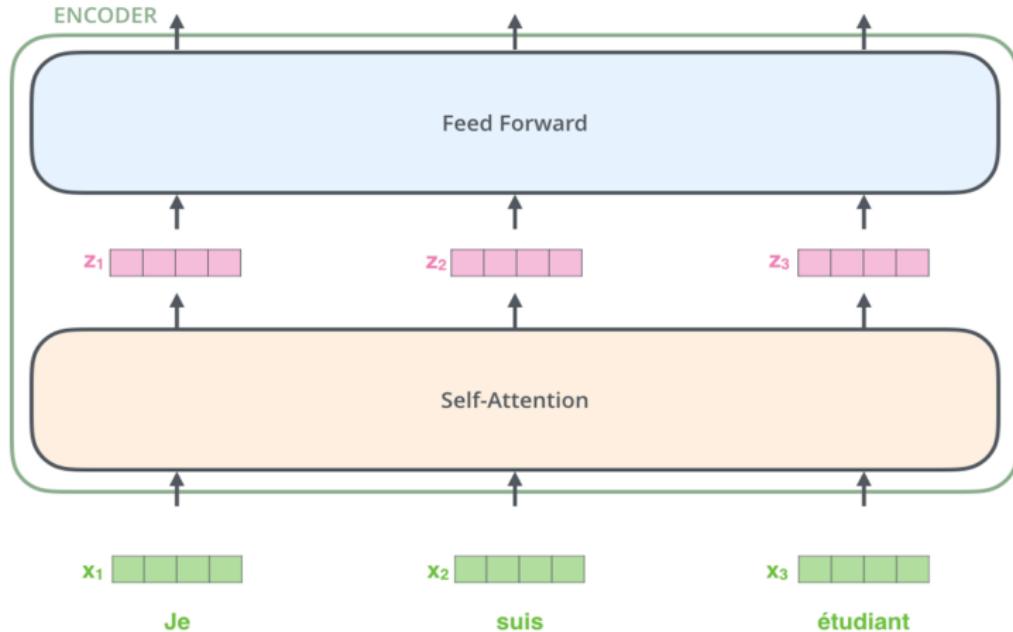
Transformer originel



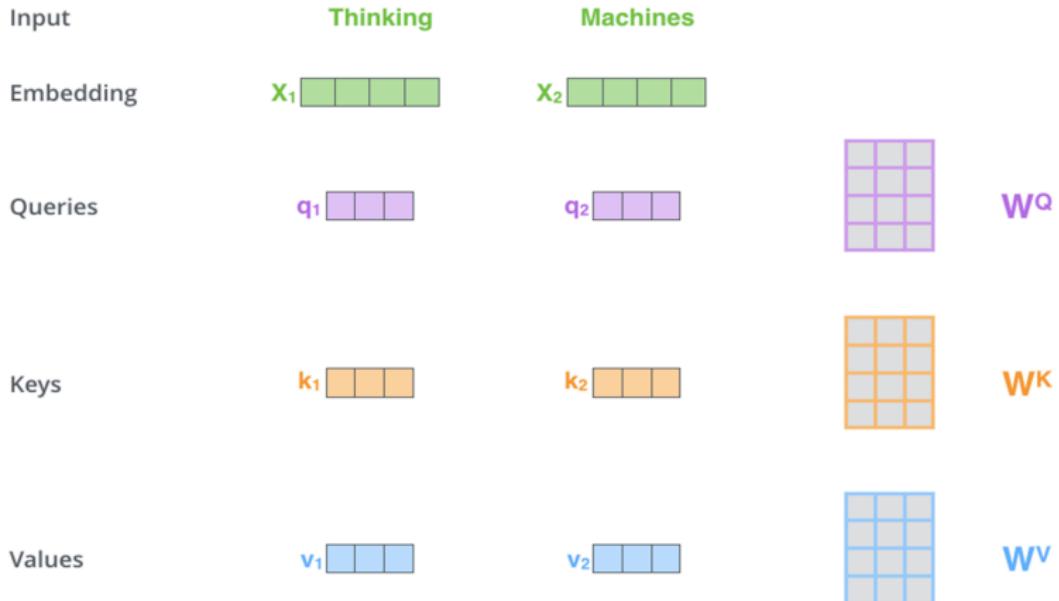
Mécanisme de cross-attention



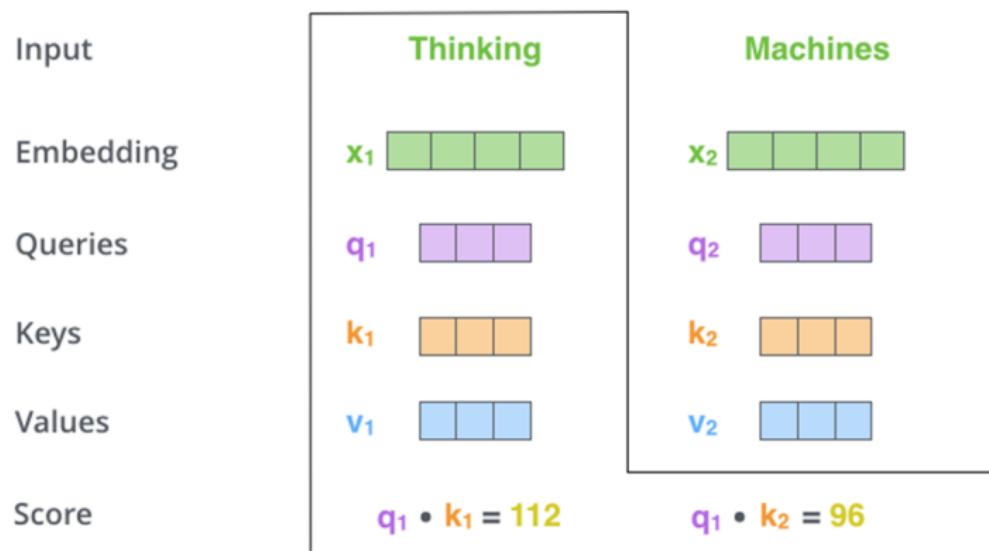
Mécanisme de self-attention



Fonctionnement du mécanisme de self-attention



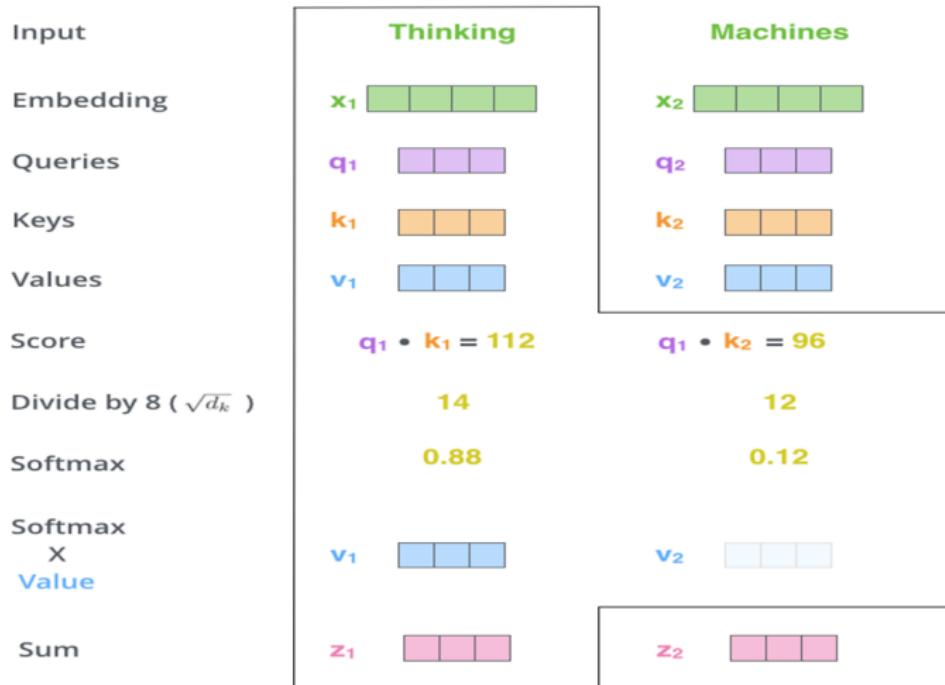
Calcul des scores de self-attention



Calcul des scores de self-attention

Input	Thinking		Machines	
Embedding	x_1	[4 green boxes]	x_2	[4 green boxes]
Queries	q_1	[3 purple boxes]	q_2	[3 purple boxes]
Keys	k_1	[3 orange boxes]	k_2	[3 orange boxes]
Values	v_1	[3 blue boxes]	v_2	[3 blue boxes]
Score	$q_1 \bullet k_1 = 112$		$q_1 \bullet k_2 = 96$	
Divide by 8 ($\sqrt{d_k}$)	14		12	
Softmax	0.88		0.12	

Calcul de la nouvelle représentation



Performances du transformer originel

Entraîné sur WMT 2014 English-German dataset, comprenant près de 4.5 millions de phrases sentence, le WMT 2014 English-French dataset, comprenant près de 36 millions de phrases.

- 8 NVIDIA P10 GPUs
- **Base** : 12 heures
- **Large** : 3.5 jours

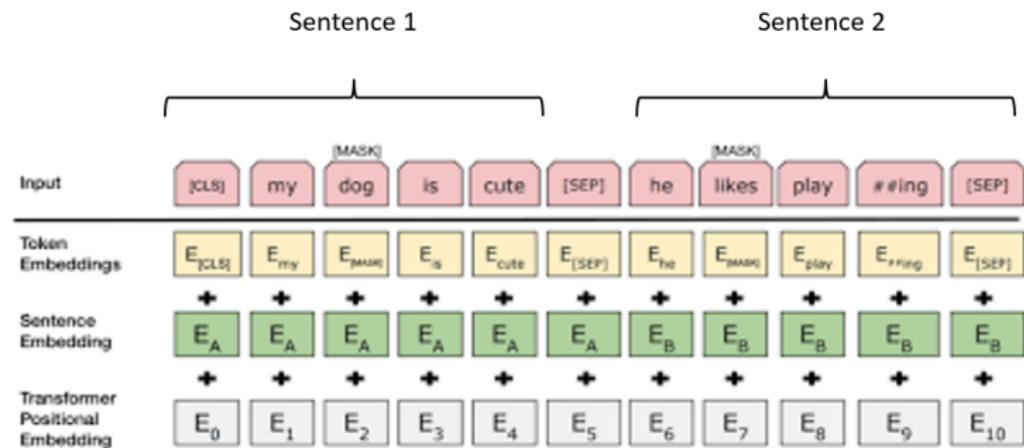
Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

Model	BLEU		Training Cost (FLOPs)	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet [18]	23.75			
Deep-Att + PosUnk [39]		39.2		$1.0 \cdot 10^{20}$
GNMT + RL [38]	24.6	39.92	$2.3 \cdot 10^{19}$	$1.4 \cdot 10^{20}$
ConvS2S [9]	25.16	40.46	$9.6 \cdot 10^{18}$	$1.5 \cdot 10^{20}$
MoE [32]	26.03	40.56	$2.0 \cdot 10^{19}$	$1.2 \cdot 10^{20}$
Deep-Att + PosUnk Ensemble [39]		40.4		$8.0 \cdot 10^{20}$
GNMT + RL Ensemble [38]	26.30	41.16	$1.8 \cdot 10^{20}$	$1.1 \cdot 10^{21}$
ConvS2S Ensemble [9]	26.36	41.29	$7.7 \cdot 10^{19}$	$1.2 \cdot 10^{21}$
Transformer (base model)	27.3	38.1	$3.3 \cdot 10^{18}$	
Transformer (big)	28.4	41.8	$2.3 \cdot 10^{19}$	

BERT

BERT (Bidirectional Encoder Representations from Transformers) représente une avancée majeure dans le NLP, introduisant une approche novatrice pour la modélisation de langage. Il utilise la bidirectionnalité pour comprendre le contexte des mots, permettant une compréhension plus fine du langage.

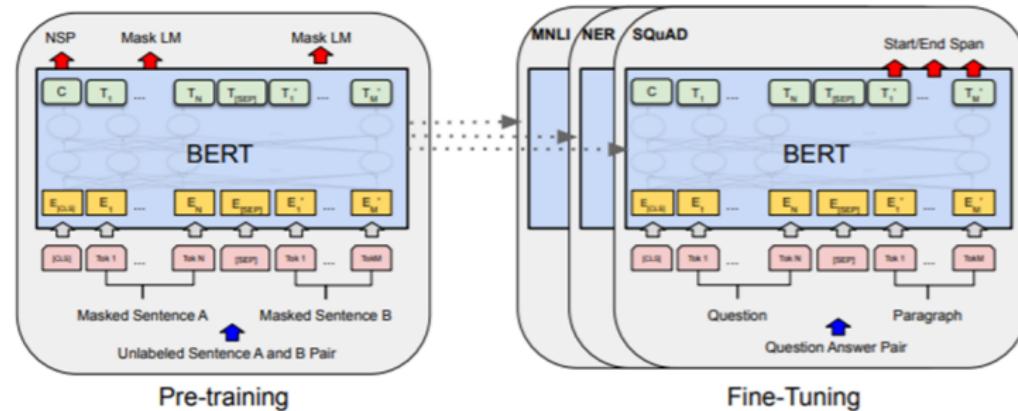
Pré-entraînement



Fine-tuning de BERT pour des tâches spécifiques

Après pré-entraînement, BERT est affiné pour des tâches spécifiques:

- Ajout d'une couche de sortie spécifique à la tâche (classification, NER, QA).
 - Fine-tuning de tous les paramètres du modèle pré-entraîné sur le corpus de la tâche.



Performances de BERT

BERT a été pré-entraîné sur le BookCorpus (800 millions de mots) et English Wikipedia (2,500 millions de mots).

Deux modèles :

- **BERT Base** : 12 couches avec 110 millions de paramètres.
- **BERT Large** : 24 couches avec 340 millions de paramètres.

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT _{BASE}	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT _{LARGE}	86.7/85.9	72.1	92.7	94.9	60.5	86.5	89.3	70.1	82.1

Modèles génératifs (GPT)

*Ce que je ne peux pas créer,
je ne le comprends pas.*

– Richard Feynman

GPT-3

- Développé par OpenAI, publié en 2020.
- Modèle auto-régressif basé sur l'architecture Transformer.

Caractéristique	Valeur / Détail
Date	2020
Paramètres	175 milliards
Corpus d'entraînement	Environ 300 milliards de tokens
Contexte maximum	2048 tokens
Ressources GPU	Plusieurs centaines de GPU (type V100)
Coût d'entraînement (estim.)	4,6 M\$ à 12 M\$ (selon les sources)

- **Applications** : génération de texte, Q&R, résumé, traduction, assistance à la programmation.

Le few-shot learning

- Les grands modèles (comme GPT-3) peuvent réaliser des tâches **sans être explicitement réentraînés** dessus.
- Le principe repose sur **l'utilisation de quelques exemples** (exemples étiquetés) directement dans le prompt ou l'entrée.
- **Zero-shot** : aucune démonstration fournie, le modèle doit comprendre la tâche à partir de sa connaissance générale.
- **One-shot / Few-shot** : on inclut un nombre très réduit d'exemples (un ou quelques-uns) pour guider le modèle dans la résolution de la tâche.
- Exemple :

English: "cat" → French: "chat"

English: "house" → French: "maison"

English: "car" → French: "voiture"

English: "tree" → French:

ChatGPT

ChatGPT, développé par OpenAI, est un modèle de langage basé sur l'architecture GPT (Generative Pre-trained Transformer) optimisé pour comprendre et générer des dialogues naturels. L'entraînement de ChatGPT se décline selon les étapes suivantes :

- ① Pré-entraînement sur un corpus volumineux :** Comme GPT-3, ChatGPT est d'abord pré-entraîné sur un vaste ensemble de données textuelles, englobant un large éventail de la littérature disponible sur Internet, pour apprendre une compréhension générale du langage.
- ② Fine-tuning supervisé :** Ensuite, ChatGPT est affiné sur des dialogues spécifiques pour améliorer ses compétences conversationnelles. Cette étape utilise des paires question-réponse et des conversations pour enseigner au modèle des structures de dialogue et des réponses contextuellement appropriées.
- ③ Reinforcement Learning from Human Feedback (RLHF):** Utilisation de techniques de renforcement pour ajuster les réponses du modèle basées sur les préférences et les corrections fournies par des évaluateurs humains, raffinant davantage la pertinence et la naturalité des réponses.

Nouvelle génération de *reasoning models*

Pourquoi “modèles de raisonnement” ? Ils allouent plus de calcul (*compute budget*) à l’*enchaînement logique* plutôt qu’à la simple prédiction du mot suivant.

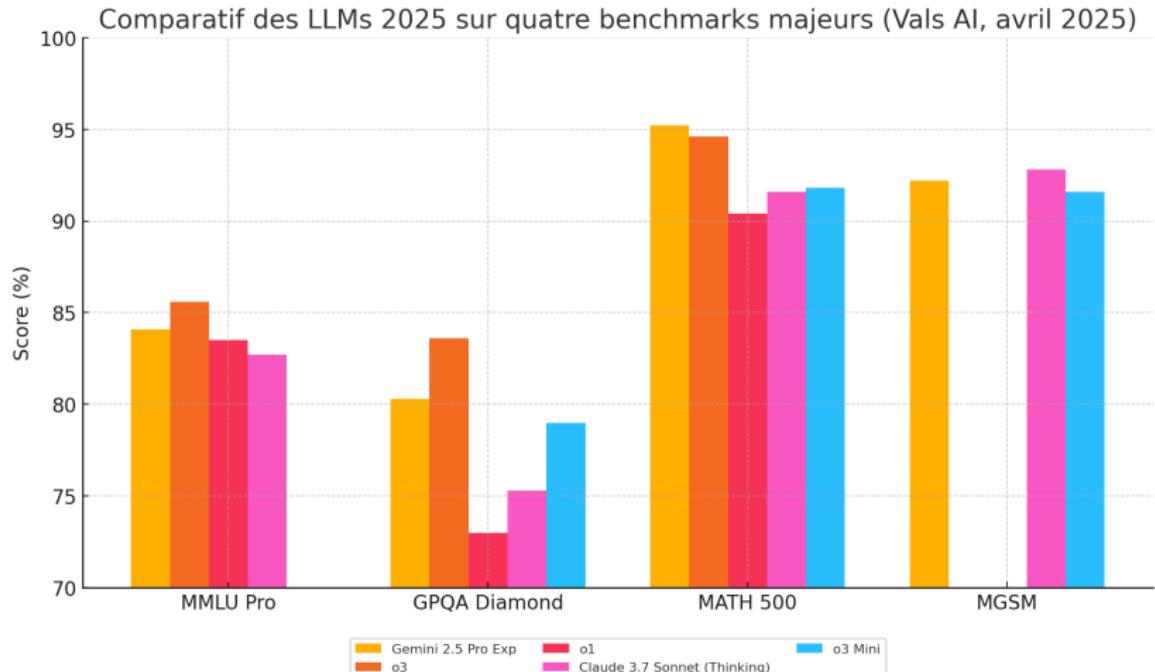
Ingrédients clés

- **Scratchpad logique** : génération d’un brouillon pas à pas avant l’output final (Chain-of-Thought, Graph-of-Thought).
- **Réflexion & vote** : plusieurs hypothèses sont évaluées puis la plus cohérente est retenue (self-consistency).
- **Allocation dynamique du calcul** : davantage de couches ou d’itérations si la question est complexe (budget adaptatif).
- **Entraînement orienté preuve** : jeux de données annotés étape par étape + récompense RL sur la justesse des *raisons*, pas seulement de la réponse.

Exemples de modèles

Familles o de OpenAI, DeepSeek-R , Gemini 2.5 reasoning , etc..

LLMs 2025 : comparatif multi-benchmarks



Source

: Vals AI, avril 2025 – 5-shot CoT.

LLMs propriétaires vs Open Source

- **Modèles propriétaires**

- ChatGPT, Claude, Gemini...
- Poids non disponibles au public.
- Performances élevées, accès limité par API payante
- Protection intellectuelle stricte (code non accessible)
- Support technique assuré, documentation avancée
- Dépendance aux fournisseurs et coûts potentiellement élevés

- **Modèles open source**

- Llama, Mistral, Gemma...
- Poids disponibles au public
- Transparence du code, flexibilité d'adaptation
- Performances parfois inférieures, mais amélioration continue
- Gratuit, mais coût de l'infrastructure à gérer
- Exposition aux failles potentielles de sécurité, support communautaire

Conclusion : Choisir entre contrôle, flexibilité et performances optimales.

Enjeux sociaux de l'intelligence artificielle

Impact de l'IA sur la médecine

Contexte mondial :

- Pénurie de soignants (11M d'ici 2030, OMS)
- Explosion des données médicales (images, biologie, texte)
- Besoin de personnalisation, prévention et efficacité

IA = catalyseur d'une médecine 4P :

- **Préditive** : détection précoce via IA
- **Préventive** : alertes temps réel, monitoring
- **Personnalisée** : traitements adaptés au profil
- **Participative** : outils de suivi pour les patients

Marché en pleine croissance : IA générative en santé = 17B\$ en 2032
(GlobalData)

Domaines d'application majeurs

Diagnostic médical :

- Imagerie : détection de cancers (sein, poumon, rétine)
- NLP : analyse automatique de comptes-rendus
- Tri aux urgences, aide au diagnostic clinique

Soins et traitements :

- Médecine de précision (profil génétique)
- IA de dosage (insuline, anesthésie)
- Découverte de médicaments par IA (ex: Rentosertib)

Gestion hospitalière :

- Anticipation des flux (ex: Saniia)
- Optimisation des plannings, stocks, lits
- Réduction des tâches administratives

Exemples concrets dans le monde

États-Unis :

- IDx-DR (rétinopathie diabétique), Aidoc (AVC), assistants vocaux

Europe :

- NHS (UK) : IA pour cancer du sein ; France : IA pour gestion des lits

Asie :

- Chine : IA pour COVID et cancer ; Inde : dépistage rural ophtalmo

Afrique :

- Nigeria : détection de médicaments contrefaçons
- Zimbabwe : IA pour cancer du col

OMS :

- Projets pilotes IA éthique et équitable dans 40+ pays

Enjeux éthiques et perspectives

Défis majeurs :

- Biais algorithmiques, représentativité des données
- Boîte noire : difficulté d'explication, confiance limitée
- Sécurité, confidentialité, RGPD, IA Act (Europe)

Acceptabilité humaine :

- Médecins : besoin de formation, IA comme copilote
- Patients : transparence, consentement, inclusion numérique

Futur proche :

- Jumeaux numériques, essais cliniques simulés
- Médecine continue à domicile (objets connectés + IA)
- IA comme partenaire invisible de l'écosystème soin

Impact de l'IA sur l'éducation

Contexte :

- Crise éducative mondiale : inégalités, pénurie d'enseignants, besoins de personnalisation
- Explosion des technologies IA (GPT, systèmes adaptatifs, IA générative)

Promesses :

- Un apprentissage plus personnalisé, plus accessible, plus engageant
- Des enseignants augmentés dans leurs pratiques pédagogiques

Marché en plein essor : +5 milliards *en 2024 pour l'IA éducative*

Applications de l'IA en éducation

Apprenants :

- Personnalisation du parcours (exercices adaptés, rythme individualisé)
- Tuteurs IA 24h/24 (ex : Khanmigo)
- Feedback immédiat sur les devoirs, écrits, prononciation

Enseignants :

- Outils de suivi précis des élèves (Lalilo, Smart Enseigno)
- Génération automatique de supports et d'évaluations

Systèmes éducatifs :

- Pilotage par les données, détection du décrochage, ciblage des ressources, etc.

Exemples internationaux marquants

Corée du Sud : Manuels scolaires IA (2025) pour un apprentissage personnalisé

E.A.U. : Tuteur IA national → amélioration de 10% des performances

UNICEF : Manuels IA pour élèves handicapés + utilisables hors-ligne

France : Lalilo (lecture CP), Smart Enseigno (maths), évaluations personnalisées

Brésil : Letrus pour améliorer l'expression écrite au collège

Chine : Squirrel AI Learning (tutorat adaptatif à grande échelle)

Perspectives : l'éducation augmentée

À court terme :

- Enseignants assistés par IA (préparation, évaluation, suivi)
- Tuteurs IA accessibles à tous (personnalisation, équité)

À moyen terme :

- Jumeaux pédagogiques numériques (suivi tout au long de la vie)
- Apprentissage immersif (VR + IA : laboratoires virtuels, simulations)

Systèmes :

- Pilotage intelligent de l'école (décrochage, remédiation, efficacité)

Défis éthiques et éducatifs

Équité et inclusion :

- Risque d'aggraver la fracture numérique ou sociale

Protection des données :

- RGPD, transparence sur les algorithmes, sécurité des données élèves

Pédagogie et valeurs :

- Risque de plagiat, perte du lien humain, biais algorithmiques
- Nécessité de maintenir l'enseignant au cœur du processus

Régulation urgente :

- Cadre légal, certification, co-construction avec les enseignants

AI scientist : vers la recherche autonome

Concept clé :

- Pipeline intégral : veille bibliographique, génération d'hypothèses, planification d'expériences, exécution robotisée, analyse, rédaction et soumission d'article.

Architecture :

- Agents spécialisés (lecteur, hypothèses, planificateur, analyste, rédacteur) orchestrés par un grand modèle de langage.
- Interaction continue avec un laboratoire robotisé ou une plateforme de calcul haute-performance.

Fait marquant 2025 :

- *AI scientist-v2* (Sakana AI) publie un article revu par les pairs (ICLR 2025) sans intervention humaine.

Gains mesurés :

- Idéation multipliée par 10 ; cycles d'expérimentation 24 h/24 ; délai de publication réduit de mois à jours.

Déploiement des modèles de machine learning

Pipline d'un projet de machine learning



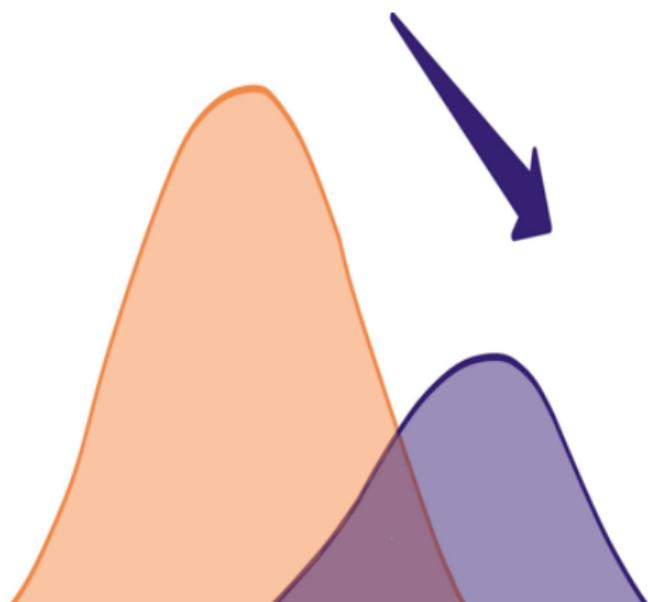
Mais ce processus est en réalité très itératif.

Data drift

Le data drift est le phénomène de changement dans la distribution des données.

Il apparaît d'une manière fréquente dans les modèles mis en production. Il peut être dû à plusieurs facteurs :

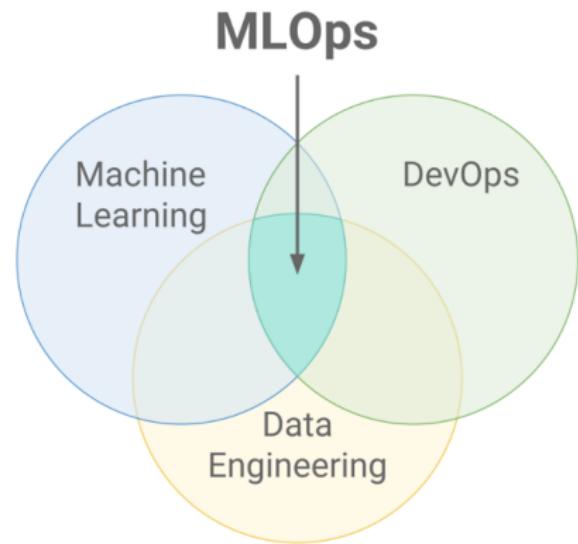
- Différence entre les données de test et les données en production.
 - Changement de l'environnement (avant et après le Covid par exemple).
 - Changement dans le comportement des utilisateurs ou clients.
 - Effets de l'algorithme lui-même sur les utilisateurs.
 - Etc.



MLOps

Objectif du MLOps :

- Alignement entre les équipes data scientists, data engineers, etc.
 - Reproductibilité et maintenance du code.
 - Adaptation aux changements pouvant affecter la distribution des données au fil du temps.



Outils de mise en production et MLOps

Les solutions de ML sont généralement déployées dans un environnement Cloud.

- AWS, GCP, Azure, etc.
- SageMaker, Vertex AI, Azure ML, etc.
- Dataiku, databricks, DataRobot, etc.



Google Cloud Platform



vertex.ai



Bonnes pratiques de mise en production

- Intégration continue et déploiement continu (CI/CD).
- Validation de la qualité du code.
- Monitoring de la qualité des données en production.
- Monitoring des performances de l'algorithme en production.
- Déploiement de modèles simples quand cela est possible.
- Optimisation et veille sur les ressources Cloud.

Plus loin...

Enjeux éthiques de l'IA

- **Enjeux sociétaux** : production massive de deepfakes, environnement peuplé par des artefacts, etc.
- **Alignement** : les valeurs encodées dans les modèles génératifs sont aujourd'hui déterminées par une poignée de personnes dans la Silicon Valley.
- **Transhumanisme** : à mesure que les modèles deviennent très puissants — voire plus puissants que l'intelligence humaine, la tentation de l'augmentation se fera plus pressante.
- **Enjeux ontologiques** : si toutes les tâches et facultés dont on pensait jusque-là qu'elles étaient le propre de l'Homme sont accessibles à l'intelligence artificielle, alors quelle est finalement l'essence de l'humain ?

Merci de votre attention

redha.moulla@axia-conseil.com