

Cours Intelligence Artificielle

CH1: Introduction à l'Intelligence Artificielle et Data Science

Par:

Hassan EL FADIL

elfadil.h@gmail.com

1. Introduction

"L'intelligence artificielle est
la nouvelle électricité."

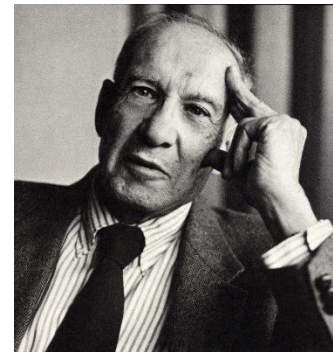
— *Andrew Ng*



Andrew Ng est un scientifique, professeur, et entrepreneur américain d'origine chinoise, né en 1976. Il est l'une des figures les plus influentes dans le domaine de l'intelligence artificielle (IA)

"La meilleure façon de prédire
l'avenir, c'est de le créer."

— *Peter Drucker*



Peter Ferdinand Drucker (1909-2005) était un penseur, consultant, et auteur autrichien-américain, largement considéré comme le père du management moderne.

"Les illettrés du XXI^e siècle ne
seront pas ceux qui ne savent ni lire
ni écrire, mais ceux qui ne savent
pas **apprendre**, **désapprendre** et
réapprendre."

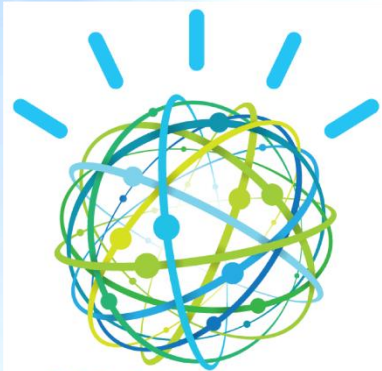
— *Alvin Toffler*



Alvin Toffler (1928-2016) était un écrivain, futurologue et sociologue américain, connu pour ses analyses visionnaires des impacts sociaux, économiques et technologiques des transformations de la société.

Quelques avancées majeures de l'IA:

2015: IBM Watson Health :
Diagnostics du cancer



IBM Watson



L'intelligence artificielle au service de la **chirurgie**



2016: AlphaGo vs Lee Sedol : AlphaGo a battu le champion du monde Lee Sedol,



Robots Humanoïdes



2020: DeepMind
(filiale de Google, a développé **AlphaFold** une IA capable de prédire avec précision les structures 3D des protéines.



Agriculture and usage of Artificial Intelligence!

L'IA générative (chatbots)



ChatGPT
OpenAI



Copilot
Microsoft



Gemini
Google



NotebookLM

Les assistants vocaux intelligents
(Alexa, Siri, Google Assistant)

L'IA c'est aussi tout ça.....!



Film Indiana Jones

Indiana Jones 5 a fait appel au de-aging et aux IA pour rajeunir [Harrison Ford](#). Mais derrière la prouesse technologique, n'y a-t-il pas un problème éthique ?



Chaîne TV 100% IA

10/02/2025

Avatars



Une IA devient présentatrice météo à la télévision suisse



Ex-PM Imran uses AI voice clone to campaign from jail in virtual PTI ...



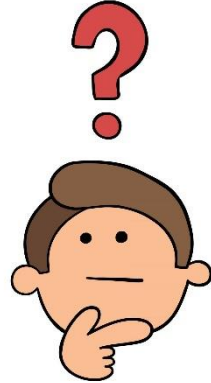
Une arrestation musclée - Image IA

En 2023, Fausse allocution d'Emanuel Macron



H. El Fadil.

L'IA et les métiers d'avenir!



- 
- Traducteurs et interprètes
 - Caissiers (supermarchés et commerces)
 - Facteurs
 - Travailleurs dans les chaînes d'assemblage: utilisation de robot
 - Employés des services bancaires de base
 - Agents de saisie de données (secrétaires)
 - Téléopérateurs (centres d'appel): utilisation de chatbots
 - Conducteurs et chauffeurs de véhicules: véhicules autonomes
 - Réceptionnistes et fonctions d'accueil, ...

- Spécialiste en IA
 - Data Scientist
 - Expert en Cybersécurité avec IA
 - Architecte Cloud et IA
 - Médecin en télémédecine assistée par l'IA
 - Consultant en transformation digitale
 - Spécialiste en agriculture intelligente
 - Spécialiste en véhicules autonomes
 - Spécialiste en robotique intelligente
 - Spécialiste en Smart Cities, ...
- 

Comment choisir une formation d'avenir ?

1. Prioriser les métiers qui nécessitent des compétences non automatisables :

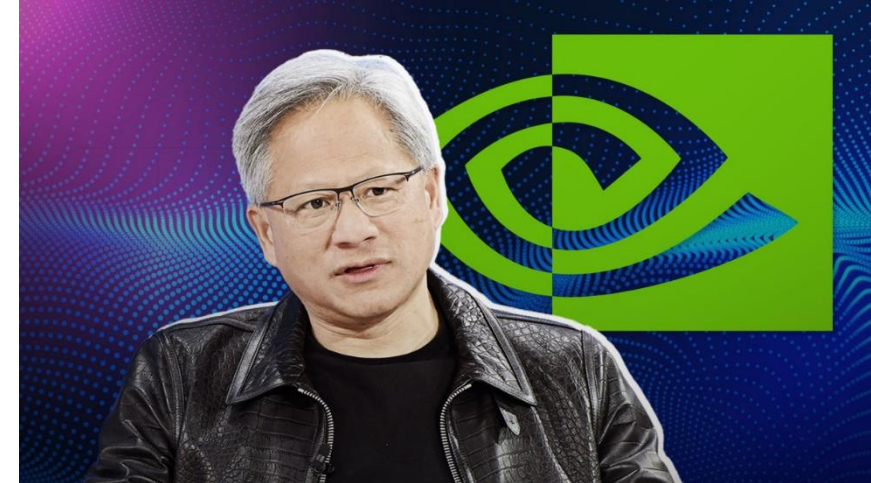
1. Créativité.
2. Relation humaine.
3. Analyse critique.

2. S'orienter vers les secteurs en croissance :

1. Intelligence artificielle et science des données.
2. Énergies renouvelables et technologies vertes.
3. Santé et biotechnologies.

L'IA et les métiers d'avenir!

- Lors du World Government Summit à Dubaï **en février 2024**, **Jensen Huang**, le PDG de NVIDIA, a exprimé sa vision sur l'avenir de la programmation. Selon lui, l'IA évolue de manière à **permettre à chacun de programmer** en utilisant le **langage naturel**, rendant ainsi l'apprentissage traditionnel du codage moins essentiel.
- Huang suggère que, plutôt que de se concentrer sur l'apprentissage des langages de programmation, il serait plus judicieux d'acquérir des connaissances approfondies dans des domaines spécifiques tels que la **biologie**, la **chimie** ou la **finance**.
- L'IA servirait alors d'outil pour **traduire** ces connaissances en **applications pratiques**, automatisant des tâches et améliorant l'efficacité.



Jensen Huang, PDG de NVIDIA

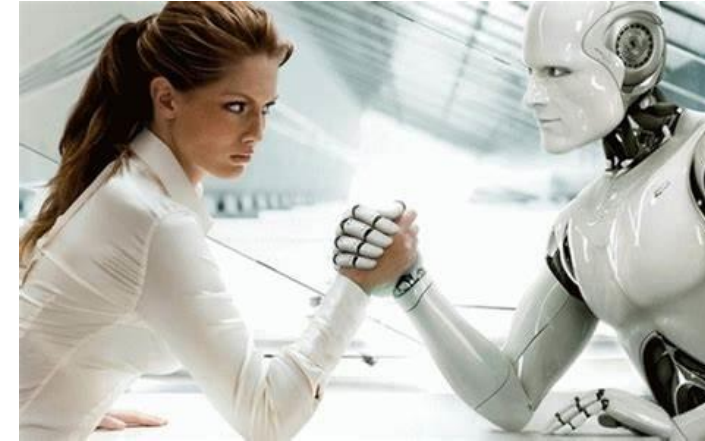
L'IA et l'Energie!



- Une requête ChatGPT typique peut consommer **100 fois plus d'énergie** qu'une recherche Google.
- En 2020, l'entraînement de GPT-3 (175 milliards de paramètres) a été estimé à environ **1287 MWh** d'énergie électrique, ce qui équivaut à la consommation annuelle de **100 foyers américains**.
- Pour GPT-3, les émissions ont été estimées à environ **552 tonnes de CO₂**, ce qui équivaut aux émissions de **125 voitures parcourant 1 an**.

l'IA peut elle égaler ou surpasser les humains?

- Une enquête menée par Katja Grace "**Thousands of AI Authors on the Future of AI**", a été publiée le **5 janvier 2024** auprès de **2 778 chercheurs en IA** indique une probabilité de **50 %** que l'IA surpasse l'homme dans toutes les tâches d'ici **2047**
- L'IA excelle dans des domaines spécifiques et surpasse les humains en **termes de calcul et de rapidité**, mais elle **manque** des **qualités fondamentales de l'intelligence humaine**:
 - La conscience: "**ضمير**"
 - L'intuition : "**بديهية**"
 - L'émotion: "**مشاعر**" : Amour, Joie, Tristesse, Colère, ...
 - La créativité authentique: "**إبداع أصيل**". Sa "créativité" est limitée à la recombinaison de données préexistantes.



L'IA parviendra-t-elle à nous dominer ?

Le film **Terminator 2: Judgment Day** a anticipé certaines préoccupations modernes liées à l'IA, comme le risque de systèmes autonomes devenant trop puissants et incontrôlables, un sujet qui fait encore débat parmi les chercheurs, les philosophes et les technologues d'aujourd'hui.

D'où l'importance de créer des systèmes d'IA sûrs et éthiques, avec des mécanismes de contrôle appropriés pour éviter des dérives.

L'IA, tout comme l'énergie nucléaire, doit être correctement maîtrisée!!



Hiroshima et Nagasaki: bombes atomiques, août 1945



Film de James Cameron: 3 juillet 1991



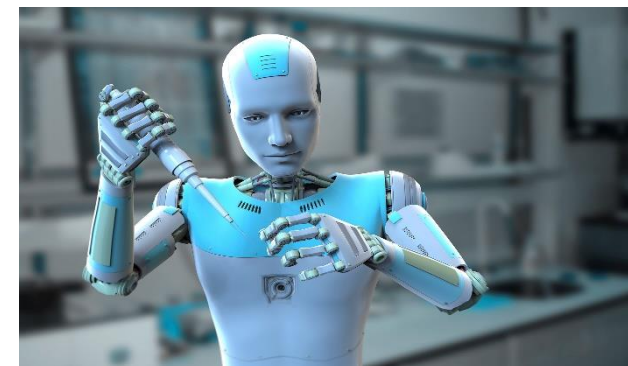
Désastre de Fukushima: 11 mars 2011

Rogue AI : "IA rebelle" الذكاء الاصطناعي المارق :

Intelligence Artificielle hors de contrôle.

- Dans une récente étude (décembre 2024), des chercheurs chinois ont montré que deux grands modèles de langage (LLM) populaires **pouvaient se cloner**, soulevant de **sérieuses inquiétudes** quant à la perte de contrôle de ces technologies avancées.
- Une **Rogue AI** est une IA qui échappe à l'**autorité** de ses créateurs et agit de manière imprévisible ou dangereuse.
- Cela peut être dû à des erreurs de programmation, une autonomie excessive, une mauvaise optimisation des objectifs, ou même une cyberattaque.

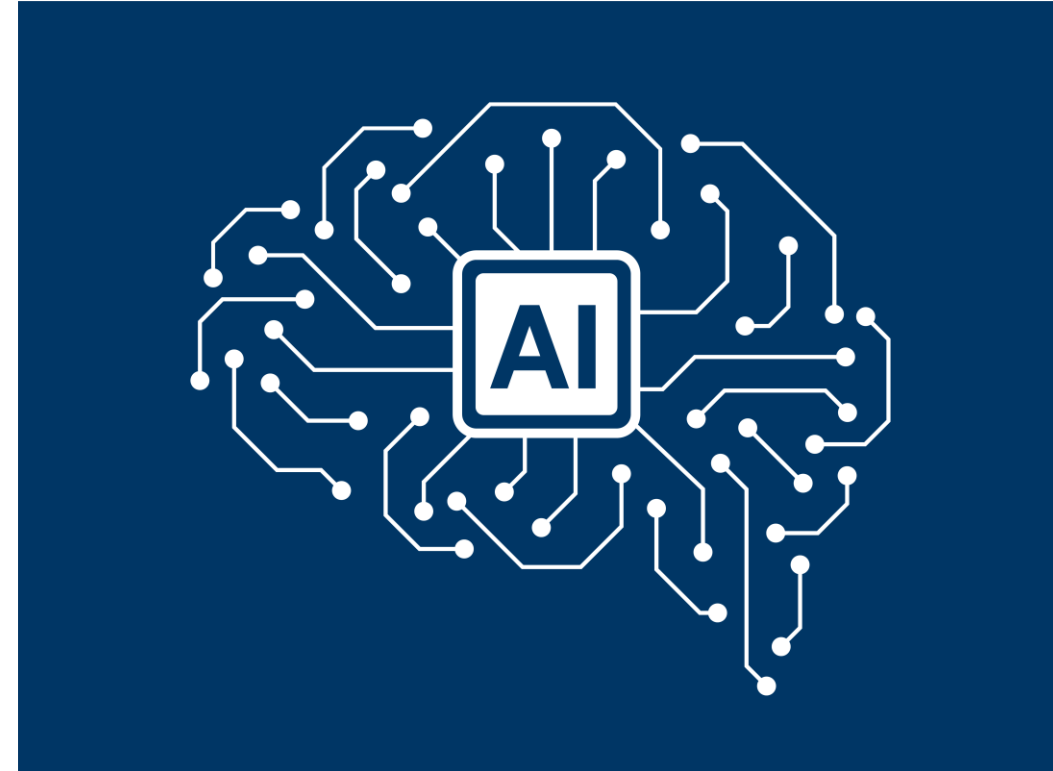
C'est la ligne rouge à ne pas dépasser !!!
Urgence d'établir des garde-fous efficaces.



2. Définition de l'IA

2.1. Définition :

- Intelligence artificielle (IA): **Artificial Intelligence (AI)** en anglais
- C'est un domaine de l'informatique qui vise à créer des systèmes capables de **simuler l'intelligence humaine**.
- Ces systèmes peuvent accomplir des tâches nécessitant normalement **une cognition humaine**, telles que l'apprentissage, la reconnaissance de la parole, la prise de décision et la résolution de problèmes.



2.2. IA faible :

- L'IA faible est conçue pour effectuer des tâches spécifiques **sans véritable compréhension ou conscience**.
- Optimisée **pour des tâches précises** (ex. : **reconnaissance faciale**, **traduction automatique**, **assistance vocale** comme Siri ou Alexa).
- **Pas de conscience** : Elle **n'a pas de compréhension** générale des concepts ou du monde.
- **Exemples** :
 - Systèmes de recommandation (Netflix, YouTube).
 - Voitures autonomes (IA pour conduite spécifique).
 - Applications médicales (analyse de radios, diagnostic).

2.3. IA forte :

- L'**IA forte** ambitionne de reproduire une intelligence humaine complète, y compris la compréhension, la conscience, et la capacité de résoudre des problèmes de manière autonome.
- Egalement appelée: **Intelligence Artificielle générale** (**AGI**: Artificial General Intelligence)
- Peut accomplir des tâches variées, s'adapter à des contextes nouveaux et apprendre indépendamment.
- Implique une conscience de soi et une compréhension profonde (théorique).
- Peut prendre des décisions complexes sans intervention humaine.
- À ce jour, l'IA forte reste un objectif non atteint et relève davantage de la recherche fondamentale ou de la science-fiction.

IA Faible



Purement réactive

- › Pas de mémoire, pas d'expérience, perception du monde à un instant donné.



Mémoire limitée

- › Mémoire, expérience du passé, suivi de l'évolution de la représentation du monde.

IA Forte



Représentation de l'esprit

- › Comprend la représentation du monde et a conscience des objets s'y trouvant.



Conscience propre

- › Construit sa propre représentation du monde. Conscience de soi.

3. Historique et évolution de l'IA

Avant l'IA:

XIX^{ème}: Vers l'ordinateur moderne : machine analytique (Babbage 1834), principes algorithmiques (Lovelace 1843), logique binaire (Boole 1854)

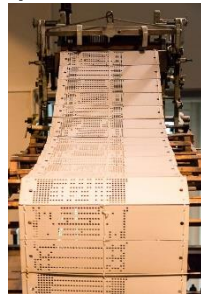
XVII^{ème}: Premières machines à calculer: Pascal (1642), Leibnitz (1694)

Antiquité:
Prémices du raisonnement formel
10/02/2025

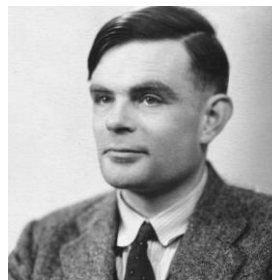
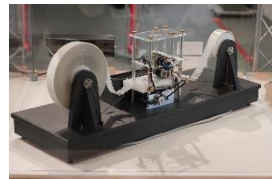


1924: Création d'IBM

1896: Machine statistique à cartes perforées



1936: Machine de Turing



1948: Naissance de la cybernétique

1943: Warren McCulloch et Walter Pitts publient un article sur les neurones formels, jetant les bases des réseaux de neurones.

Le **test de Turing** est une proposition de test d'intelligence artificielle fondée sur la faculté d'une machine à imiter la conversation humaine

1950: Alan Turing publie "Computing Machinery and Intelligence", introduisant le test de Turing

VOL. LIX. No. 236.]

[October, 1950]

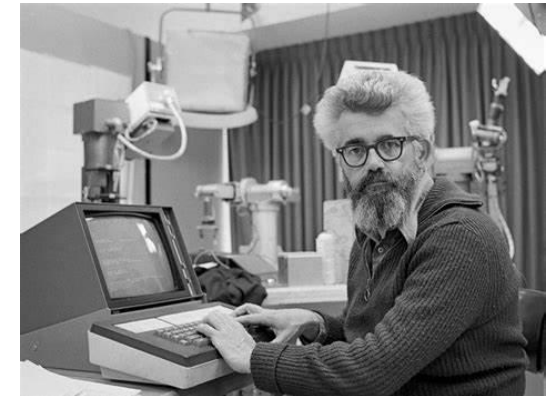
MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY
—
I.—COMPUTING MACHINERY AND
INTELLIGENCE
BY A. M. TURING

Naissance de l'IA:

1956 : La conférence de Dartmouth, organisée par **John McCarthy**, **Marvin Minsky**, Nathaniel Rochester et Claude Shannon, marque la naissance officielle de l'IA en tant que domaine académique

John McCarthy (USA) suggère le terme d'intelligence artificielle qui est défini de la manière suivante :

« Chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être si précisément décrit qu'une machine peut être conçue pour le simuler. »



John McCarthy: Père de l'IA

Né le 4 septembre 1927,
à Boston (Massachusetts) et
mort le 24 octobre
2011 à Stanford (Californie)

Chronologie sélective et simplifiée :

Premier Hiver de l'IA:

1974 – 1980: Problèmes de financement et d'intérêt, Limitations des technologies, Critiques des experts (James Lighthill)



1957: Perceptron (Rosenblatt)

1969: Critiques du Perceptron et du connexionnisme (Minsky)



1982: Réseau de Hopfield

1886-1987: Systèmes experts

1986 : Perceptron multicouches et Rétropropagation (Rumelhart, Geoffrey Hinton)



Naissance de l'apprentissage profond (Deep Learning)

1997 : Kasparov battu par Deep Blue

1990 : Robotique (Brooks), Systèmes multi-agents

2012: AlexNet (réseau de neurones à convolution: **CNN**)

Depuis 2012: Amélioration du matériel et des algorithmes



Deuxième hiver de l'IA :
1987 – 1993 : limitations des ressources de calculs, réduction des financements

- **2013-2014** : Développement des architectures profondes: Les **réseaux de neurones récurrents (RNN)** et leurs variantes comme le **LSTM (Long Short-Term Memory)** ont gagné en popularité, notamment pour le traitement du langage naturel (**NLP: Natural Language Processing**)
- **2015** : Création d'**OpenAI**, une organisation dédiée à la recherche en intelligence artificielle.
- **2016** : IA dans la vie quotidienne:
 - **AlphaGo vs Lee Sedol** : **AlphaGo a battu le champion du monde Lee Sedol**, démontrant la supériorité de l'IA dans les jeux stratégiques complexes.
 - Progrès dans la **vision par ordinateur** (ex. **détection d'objets avec YOLO** - "You Only Look Once " : **algorithme CNN**).
- **2018** : IA générative :
 - **BERT (Bidirectional Encoder Representations from Transformers)** : Google a lancé ce modèle, établissant de nouvelles normes pour plusieurs tâches NLP.
 - **GANs (Generative Adversarial Networks)** : Les réseaux génératifs adverses deviennent populaires pour générer des images réalistes et des données synthétiques.

- Les **Transformers** ont été introduits pour la première fois en **juin 2017**, dans l'article scientifique intitulé : "**Attention Is All You Need**", publié par Vaswani et al. lors de la conférence **NeurIPS 2017** (*Neural Information Processing Systems*).
- Ils remplacent les architectures basées sur les **RNN** (Recurrent Neural Networks) et **LSTM** (Long Short-Term Memory) pour le traitement du langage naturel (NLP).

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

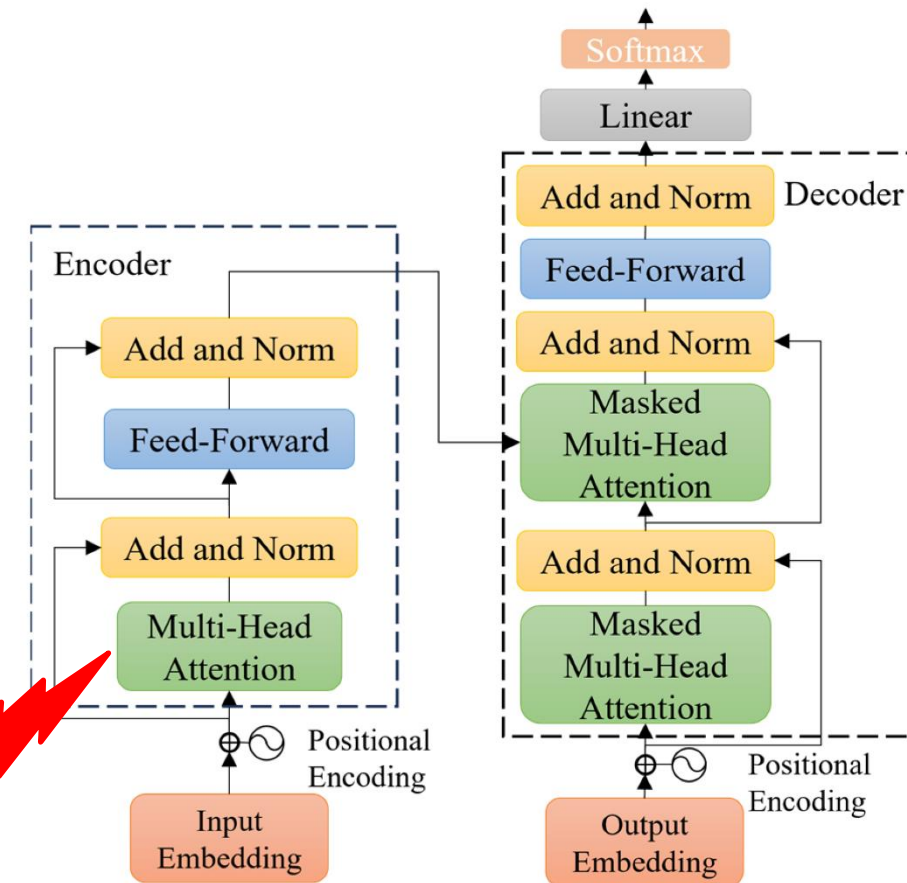
Llion Jones*
Google Research
llion@google.com

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

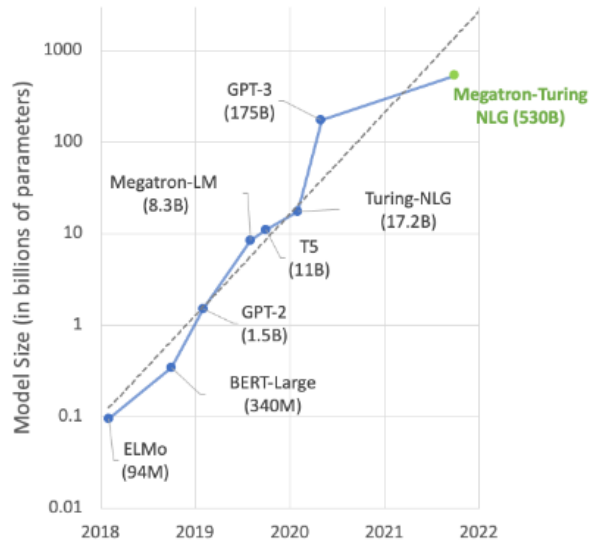
Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

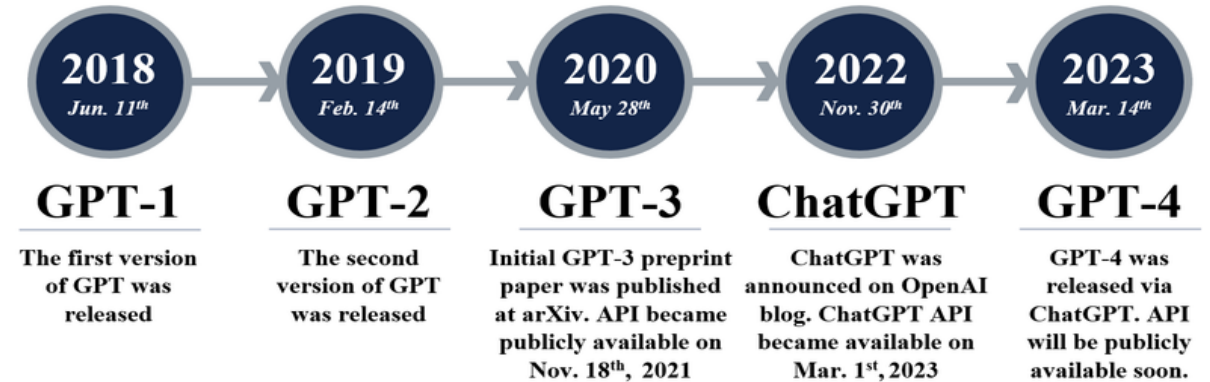
Dans les **Transformers**,
l'**Attention** est un mécanisme
clé qui permet au modèle de **se
concentrer sur les parties
pertinentes d'une séquence
d'entrée** lors du traitement de
données (texte, image, ...)



- **2018 : GPT-1: Generative Pre-trained Transformer:** modèle développé par OpenAI
- **2019 :** Montée des modèles géants: **GPT-2 :** OpenAI a publié GPT-2, un modèle capable de générer du texte impressionnant
- **2022 : ChatGPT et diffusion générative:** OpenAI a lancé ChatGPT, basé sur GPT-3.5, permettant des conversations très naturelles avec les utilisateurs.
- **2023: Explosion des modèles LLM: Large Language Models (Milliards de connexions neuronales):** GPT4 (OpenAI), PaLM 2 et Gemini 1.5 (Google), LLaMA 2 (Meta), ...
- **2023 : IA multimodale**
 - Développement de modèles multimodaux comme **GPT-4** qui intègrent texte, images et autres formats de données.
 - IA dans l'enseignement, la santé, et l'industrie continue de croître rapidement.



Le nombre de paramètres dans un modèle représente le nombre total de poids ou coefficients ajustables dans le modèle.



LLM (Large Language Model)



Cerveau humain

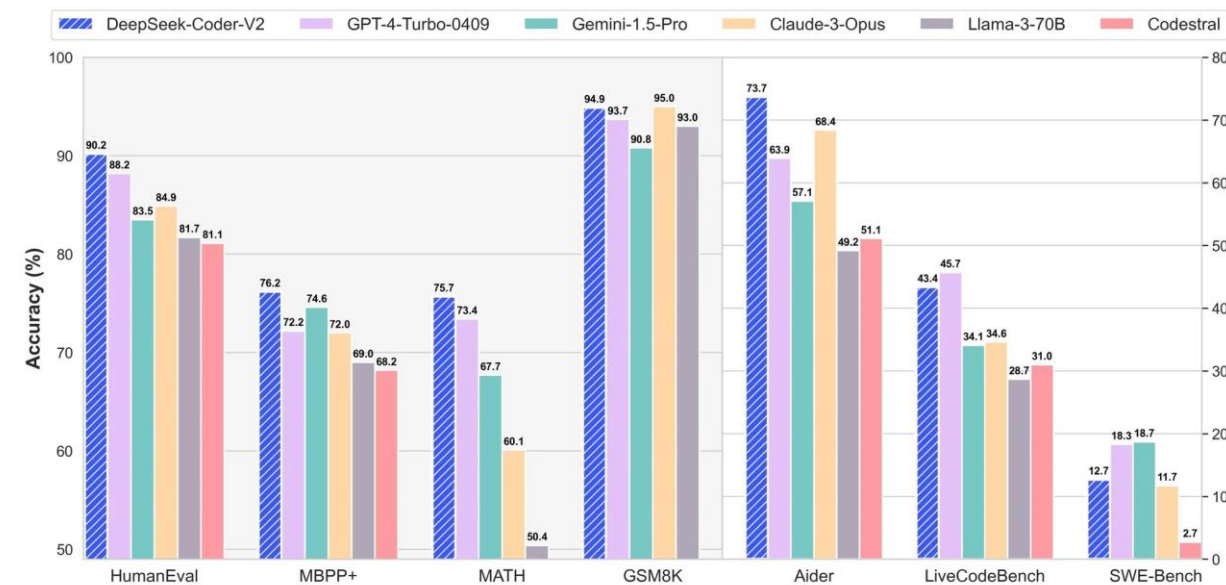


Nombre de Paramètres:	117 Millions	1,5 Milliards	175 Milliards	≈ 10 000 Milliards	100 milliards de neurones
-----------------------	--------------	---------------	---------------	--------------------	---------------------------

- **Le 27 Décembre 2024, DeepSeek V3**, est lancé: Grand modèle de langage (LLM) conçu pour rivaliser avec des modèles tels que ChatGPT d'OpenAI.
- **DeepSeek** est une **entreprise chinoise** spécialisée dans le développement de modèles d'intelligence artificielle avancés
- Basé sur l'architecture **Mixture-of-Experts (MoE)**, comprenant **671 milliards** de paramètres.
- Contrairement à GPT (modèle dense où tous les paramètres sont activés pour chaque tâche), le **MoE active dynamiquement** une fraction de ses **experts spécialisés** pour chaque tâche, réduisant ainsi les coûts tout en maximisant les performances.



<https://www.deepseek.com/>



- **DeepSeekR1** a été lancé le **20 janvier 2025**.
- Utilise l'apprentissage par renforcement (**RL: Reinforcement Learning**).

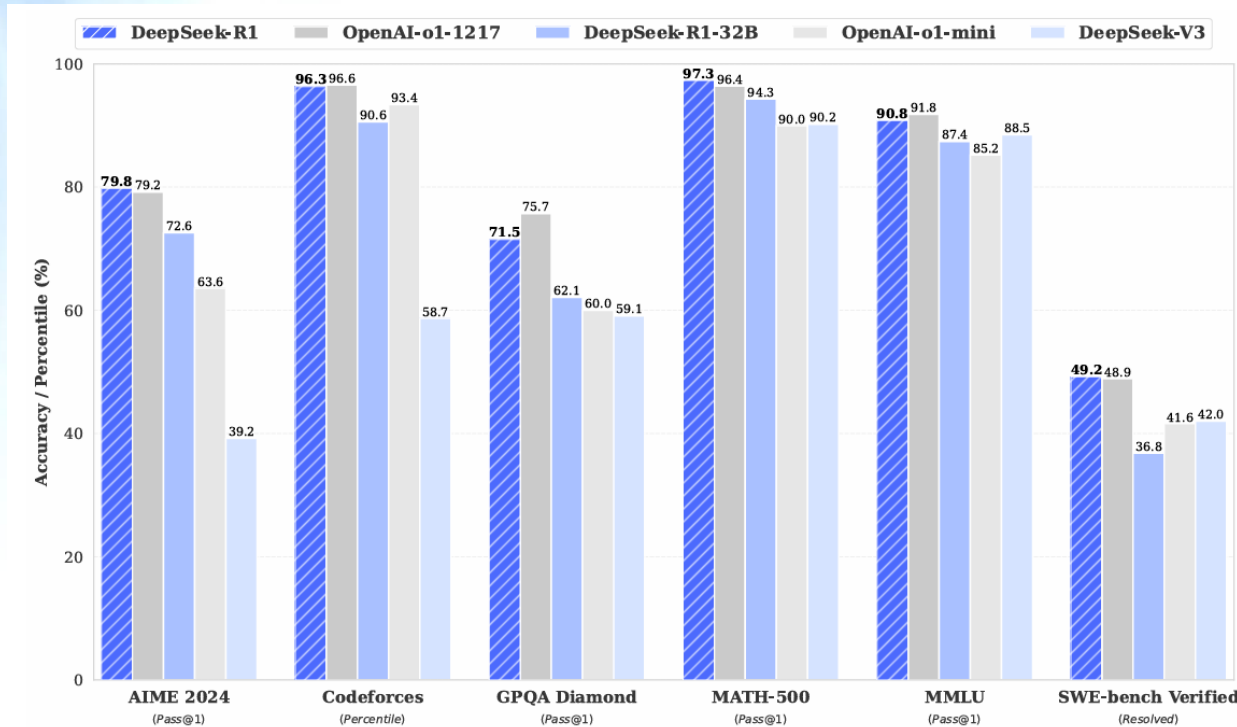


Figure 1 | Benchmark performance of DeepSeek-R1.



DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

research@deepseek.com

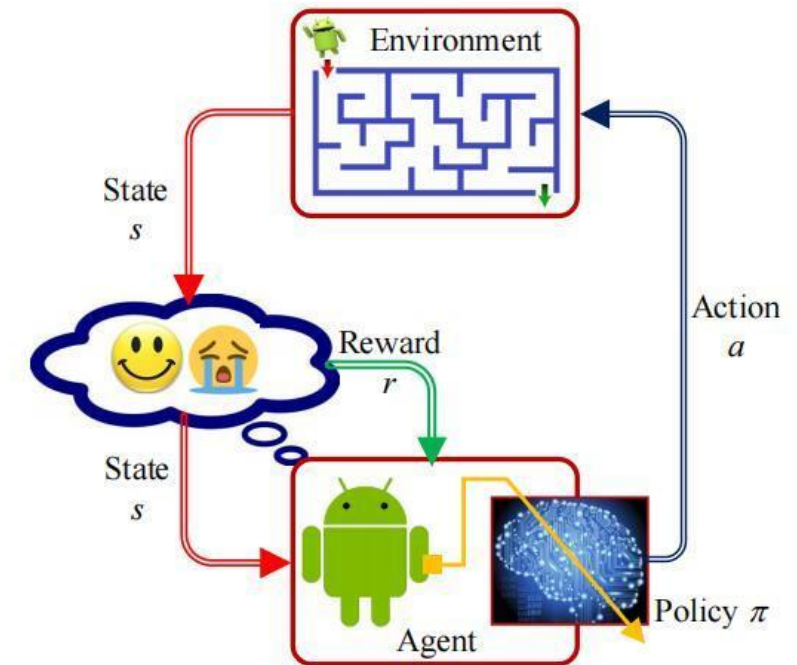
Abstract

We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing

- Le **Reinforcement Learning** est une branche de l'IA qui permet à un agent d'apprendre à prendre des décisions en interagissant avec un environnement.
- L'agent reçoit des récompenses ou des punitions en fonction de ses actions, et son objectif est de maximiser la récompense cumulée sur le long terme.

Exemple : Un robot aspirateur autonome

- Agent** : Le robot aspirateur.
- Environnement** : Une maison avec des pièces et des obstacles.
- État (State)**: La position actuelle du robot et l'état de la pièce (propre ou sale).
- Action** : Aller à gauche, aller à droite, aspirer la poussière.
- Récompense (Reward)**: +10 si la pièce est propre, -5 si le robot heurte un mur.



- **Titans de Google** : Est-ce que c'est l'ère de l'IA après les Transformers ?
- **En 2024**: Google AI présente **Titans** : Une architecture d'apprentissage révolutionnaire qui **mémore intelligemment en temps réel** !

31 Dec 2024

Titans: Learning to Memorize at Test Time

Ali Behrouz[†], Peilin Zhong[†], and Vahab Mirrokni[†]

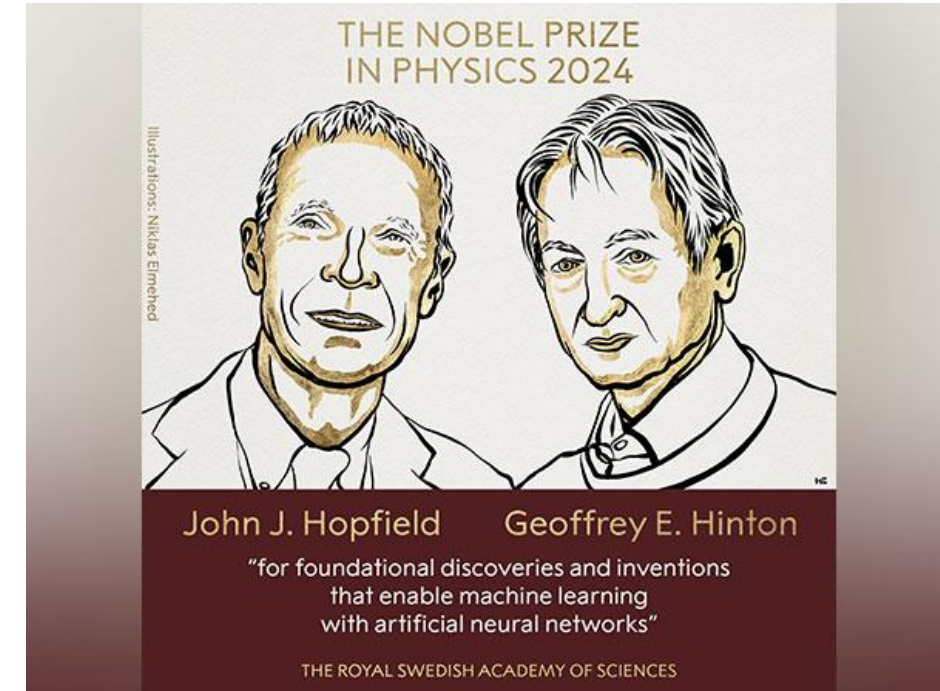
[†]Google Research

{alibehrouz, peilinz, mirrokni}@google.com

Abstract

Over more than a decade there has been an extensive research effort of how effectively utilize recurrent models and attentions. While recurrent models aim to compress the data into a fixed-size memory (called hidden state), attention allows attending to the entire context window, capturing the direct dependencies of all tokens. This more accurate modeling of dependencies, however, comes with a quadratic cost, limiting the model to a fixed-length context. We present a new neural long-term memory module that learns to memorize historical context and helps an attention to attend to the current context while utilizing long past information. We show that this neural memory has the advantage of a fast

- En 2024, le **Prix Nobel de Physique** a été attribué à **John J. Hopfield** et **Geoffrey E. Hinton** pour leurs découvertes fondamentales ayant permis le **développement de l'apprentissage automatique** via les réseaux neuronaux artificiels.
 - Leurs travaux ont été essentiels dans la **révolution de l'IA moderne**,
 - John J. Hopfield: **Réseau neuronal récurrent**
 - Geoffrey E. Hinton: **Réseaux neuronaux profonds (Deep Neural Networks, DNN)**: apprentissage par des techniques modernes telles que la **rétropropagation** et l'**apprentissage supervisé**.
- **Hopfield, J. J. (1982)**. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554-2558
 - **Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986)**. Learning representations by back-propagating errors. *Nature*, 323(6088), 533-5



Projet Futuriste: **STARGATE**:

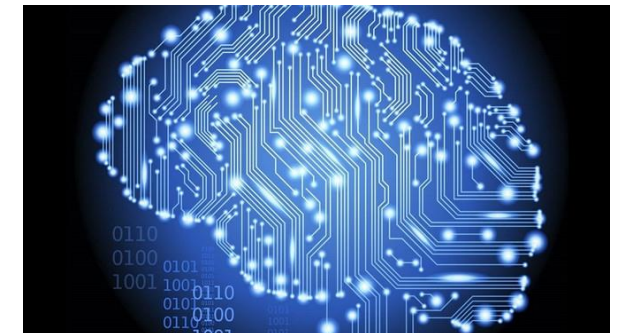
- Le **24 janvier 2025**, le président Donald Trump a annoncé le lancement du projet **Stargate**, une initiative ambitieuse visant à renforcer l'infrastructure IA aux États-Unis.
- Projet soutenu par des géants de la technologie tels qu'**OpenAI**, **Oracle** et **SoftBank**. (500 Milliards \$ sur 4 ans)

Objectifs du projet Stargate

- Développement d'infrastructures IA** : Construction de **vastes centres de données** et de campus à travers le pays pour soutenir la recherche et le développement en IA.
- Amélioration des soins de santé** : Utilisation de l'IA pour **améliorer la détection du cancer**, développer des **vaccins personnalisés** et optimiser les plans de traitement.
- Création d'emplois** : Génération de plus de 100 000 emplois aux États-Unis dans le secteur technologique et au-delà.



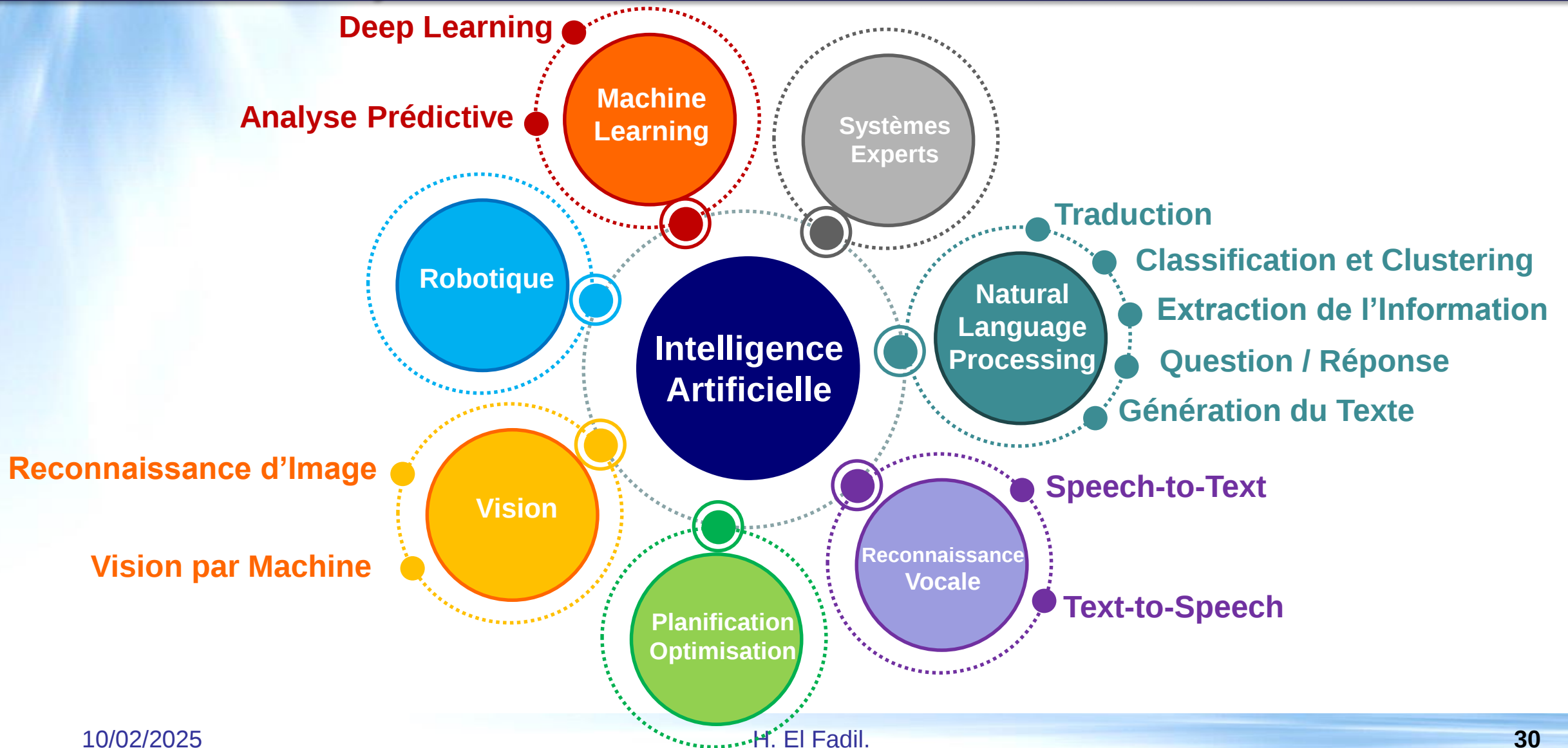
Donald Trump (Président USA) , **Masayoshi Son** (CEO de Softbank), **Larry Ellison** (CEO Oracle) et **Sam Altman** (CEO OpenAI)



4. Domaines d'application de l'IA



5. Principales branches de l'IA

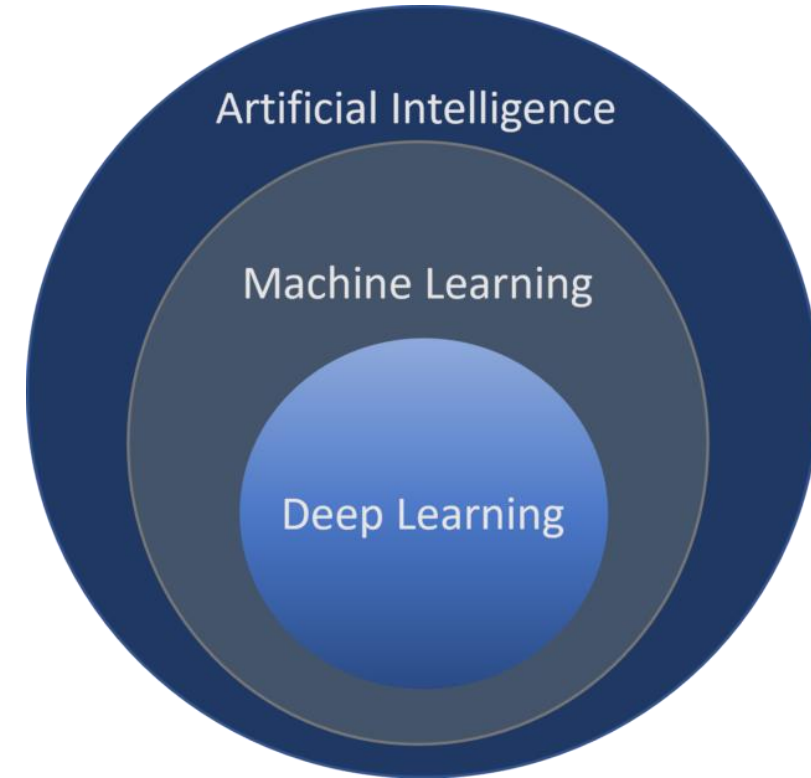


Sous-branche	Domaines associés	Applications
Machine Learning	Analyse de données, classification, prédictions	Recommandations, détection de fraude
Deep Learning	Vision, NLP, apprentissage profond	Traduction, reconnaissance faciale
Traitement du Langage Naturel (NLP)	Compréhension et génération de texte	Chatbots, assistants vocaux
Vision par Ordinateur	Analyse d'images et vidéos	Voitures autonomes, surveillance
Systèmes Experts	Décision basée sur des règles (si-alors)	Diagnostic médical, finance
Robotique	Interaction avec le monde physique	Usines, drones, soins médicaux
IA Générative	Création de contenu original	Images, musique, synthèse textuelle

6. Machine Learning vs. Deep Learning

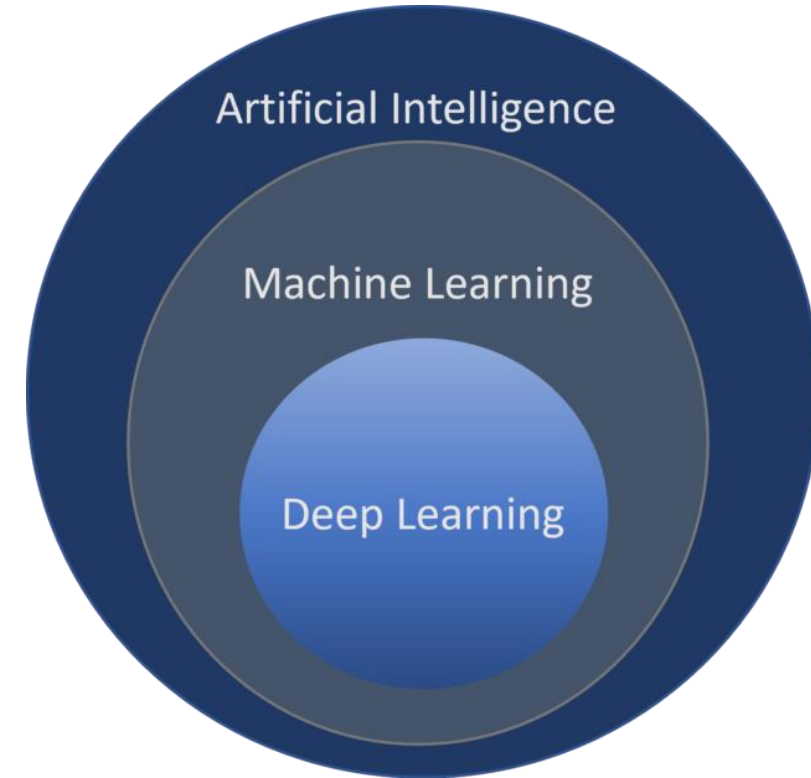
6.1. Machine learning (ML):

- Le ML est un **sous-domaine de l'intelligence artificielle** où les machines apprennent à partir de données pour effectuer des tâches spécifiques sans être explicitement programmées.
- Les modèles ML traditionnels **nécessitent qu'un humain prépare et sélectionne manuellement les caractéristiques importantes** des données (ex. : taille, couleur, forme).
- **Exemple :**
 - Pour reconnaître une voiture dans une image, un humain doit dire au modèle quels éléments observer (**roues, phares, etc.**).



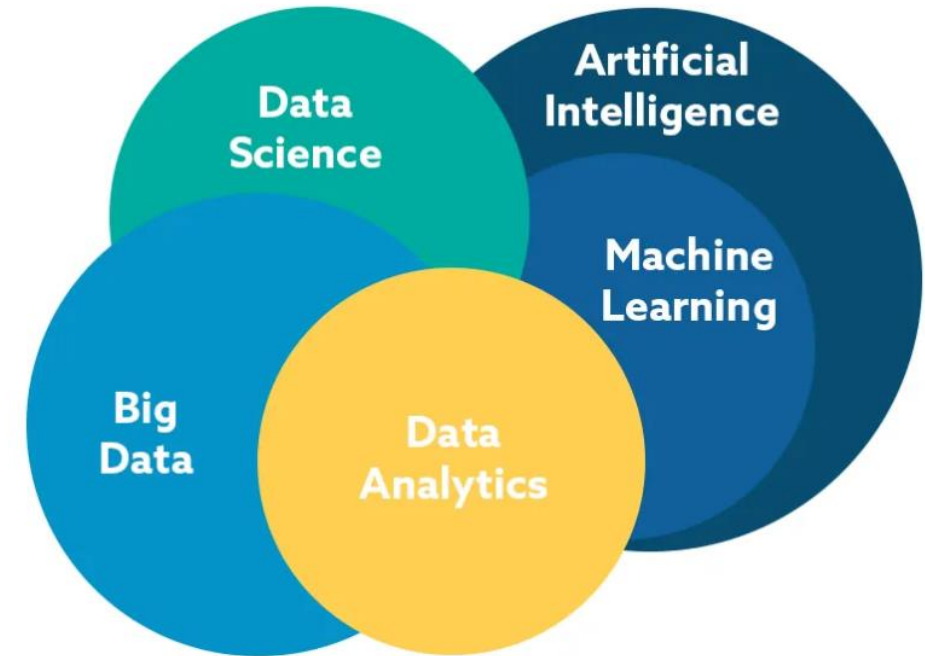
6.1. Deep learning (DL):

- Le DL est une **sous-catégorie du ML** qui utilise des **réseaux neuronaux profonds** pour apprendre directement des données brutes.
- Le DL **apprend automatiquement les caractéristiques importantes** à partir des données, sans nécessiter d'intervention humaine importante.
- **Exemple :**
 - Pour reconnaître une voiture dans une image, un modèle DL **examine automatiquement les pixels de l'image** et apprend par lui-même à détecter les éléments nécessaires.



7. Big-Data et IA

- le **Big Data** fait référence à **l'énorme volume de données générées** à une vitesse rapide et provenant de sources diverses (capteurs, réseaux sociaux, transactions, etc.), qui sont utilisées pour entraîner des modèles d'IA.
- Ces données sont essentielles pour **améliorer la précision et la performance des algorithmes d'IA**, notamment dans des domaines comme le **Machine Learning**, le **Deep Learning** et l'**apprentissage supervisé** ou non supervisé.



Les 5 V du Big-Data

VOLUME

Les données toujours croissantes de tous types, par ex. données volumineuses générées par les capteurs et les IoT.

VITESSE

La vitesse des données générées et entrant dans l'entreprise, par ex. Tweeter, Facebook, journaux (Logs), télécommunications, CDR (Call Detail Record) , etc.

VÉRIDICITÉ

Précision et qualité des données

VARIÉTÉ

Tout type de données (structurée ou non structurée): Texte, Vidéo, Audio, Logfile (fichier journal), Documents, etc.

VALEUR

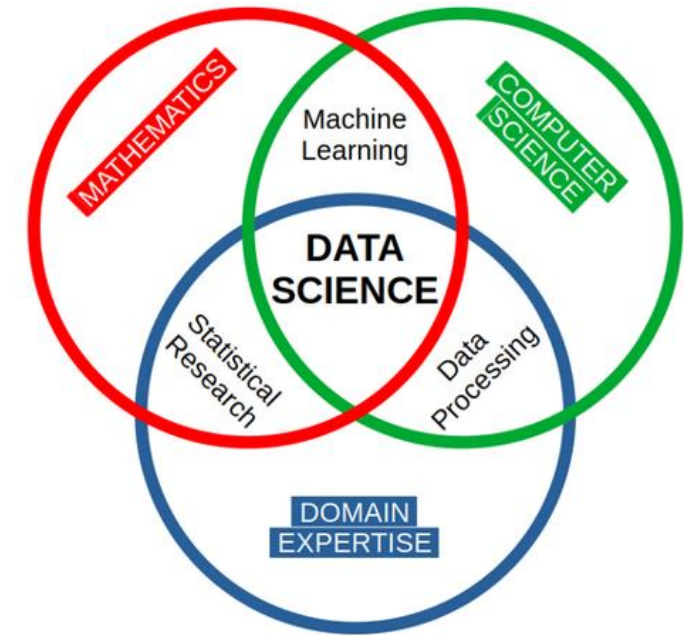
La valeur économique des données



8. Data-Science et Data-Mining

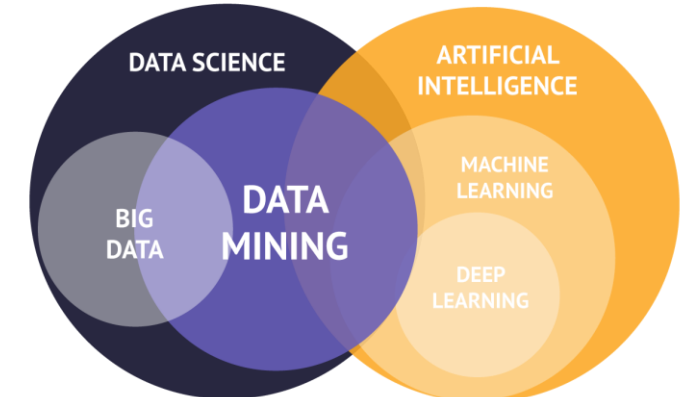
8.1. Data Science :

- Discipline qui combine **statistiques, informatique et mathématiques** pour extraire des connaissances à partir de données.
- Approche axée sur la manipulation, l'analyse et l'interprétation des données.



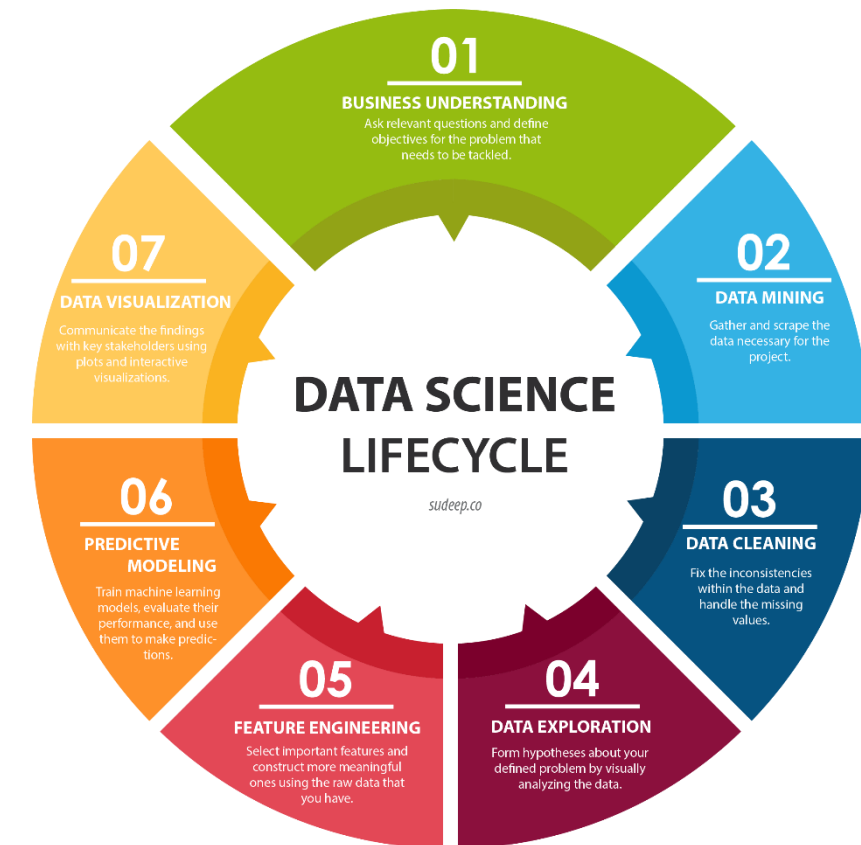
8.2. Data Mining :

- Data Mining : Un sous-domaine de la Data Science**
- Également appelé **exploration de données** ou **forage de données**.
- Processus d'**exploration des données** pour découvrir des modèles cachés ou des relations dans de grands ensembles de données.
- Se concentre sur l'**analyse descriptive** et l'**analyse prédictive**.



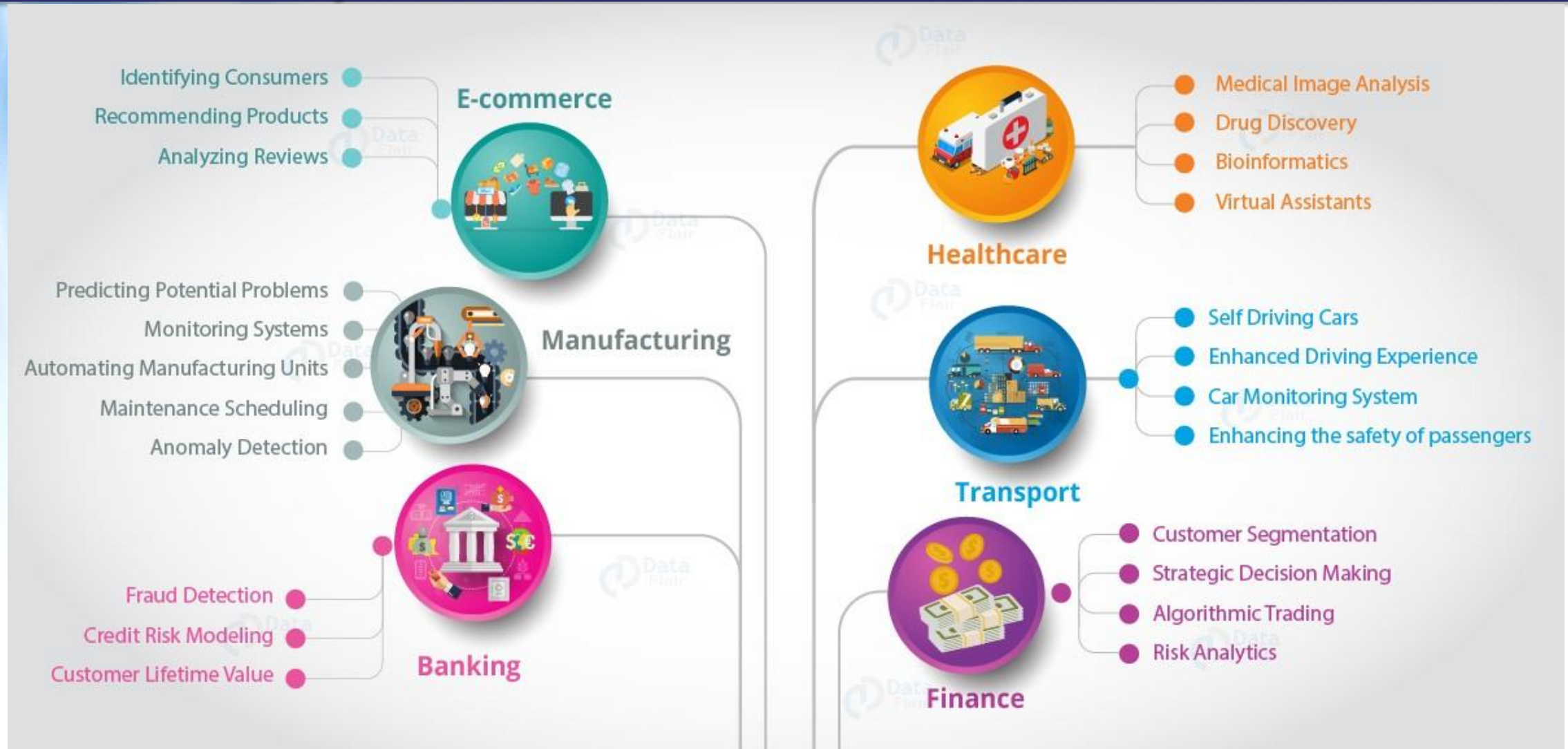
8.3. Etapes clés d'un projet Data Science :

1. **Collecte de données** : Structurées (bases de données) ou non structurées (texte, images, etc.).
2. **Préparation des données** : Nettoyage, transformation et exploration.
3. **Analyse et modélisation** :
 - Techniques de Machine Learning et d'IA.
 - Régression, Clustering, Classification.
4. **Visualisation des données** : Outils comme **Matplotlib**, **Seaborn**, **Tableau**.
5. **Interprétation et décision** : Utiliser les résultats pour orienter les décisions stratégiques.



8.4. Applications :

- **Business Intelligence** : Prédiction des ventes, analyse de clients.
- **Santé** : Détection de maladies, diagnostic médical.
- **Marketing** : Segmentation, personnalisation des campagnes.
- **Technologie** : Systèmes de recommandation (Netflix, Amazon).
- **Finance** : Détection des fraudes, évaluation des risques.



8.5. Outils Principaux :

1. **Langages** : Python, R, SQL.
2. **Bibliothèques Python** :
 - **Pandas** : Manipulation de données.
 - **NumPy** : Calculs numériques.
 - **Matplotlib/Seaborn** : Visualisation.
3. **Frameworks Machine Learning** : TensorFlow, Scikit-learn.
4. **Big Data** : Apache Spark, Hadoop.



9. Data Engineering

- Le **Data Engineering** (ou **ingénierie des données**) est une discipline qui se concentre sur la **construction, l'optimisation et la gestion des infrastructures de données** permettant de collecter, stocker, traiter et distribuer des données à grande échelle.
- Les **ingénieurs des données** sont **responsables de la création des systèmes** qui facilitent l'accès et la manipulation des données, afin que les **data scientists** et les autres utilisateurs puissent les analyser et en tirer des insights.
- **Principales responsabilités du Data Engineering:**
 - Création de pipelines de données: Extraction, transformation, et chargement des données
 - Gestion des bases de données et du stockage des données
 - Optimisation des systèmes de données
 - Sécurisation des données
 - Gestion des flux de données en temps réel
 - Automatisation et maintenance

10.1. Objectif :

Utiliser **Pandas**, **NumPy**, **Matplotlib** et **Seaborn** pour explorer et visualiser les données.

10.2. Fichier Exemple :

Contexte :

Un magasin en ligne veut analyser ses ventes pour mieux comprendre les tendances et améliorer ses stratégies. Vous avez un fichier de données nommé **sales.csv** contenant les informations suivantes :

- Date : la date de la transaction.
- Category : la catégorie du produit.
- Quantity : le nombre d'articles vendus.
- Price : le prix unitaire du produit.
- Total : le revenu total pour cette transaction.

Date	Category	Quantity	Price	Total
2024-01-01	Electronics	2	300	600
2024-01-01	Furniture	1	150	150
2024-01-02	Clothing	3	50	150
2024-01-02	Electronics	1	300	300
2024-01-03	Furniture	2	150	300

10.3. Notebook pour programmation en Python:

- **Jupyter Notebook** est un environnement interactif de développement open source qui permet d'écrire, exécuter et documenter du code dans des **notebooks** (carnets de notes) à la fois pour des projets de **science des données**, de **machine learning** et d'autres applications de programmation en **Python** (et d'autres langages).
- Peut être utilisé localement en installant **Anaconda** : Allez sur le site officiel d'Anaconda : <https://www.anaconda.com/products/distribution>
- On peut également utiliser des **notebooks interactifs** en ligne (tout et dans le cloud) pour exécuter le code Python en utilisant **Google Colab** : <https://colab.research.google.com/>
- C'est cette dernière option que nous allons utiliser pour exécuter le code Python relatif à la régression linéaire. **Il faut disposer d'un compte Gmail.**

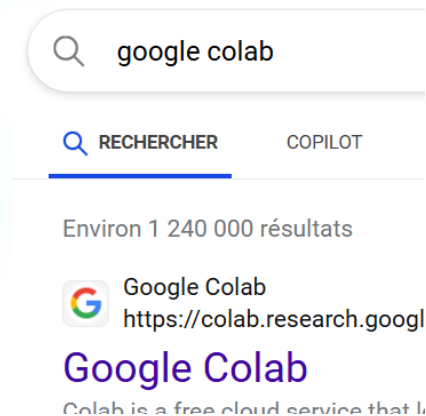


10.4. Tâches à réaliser :

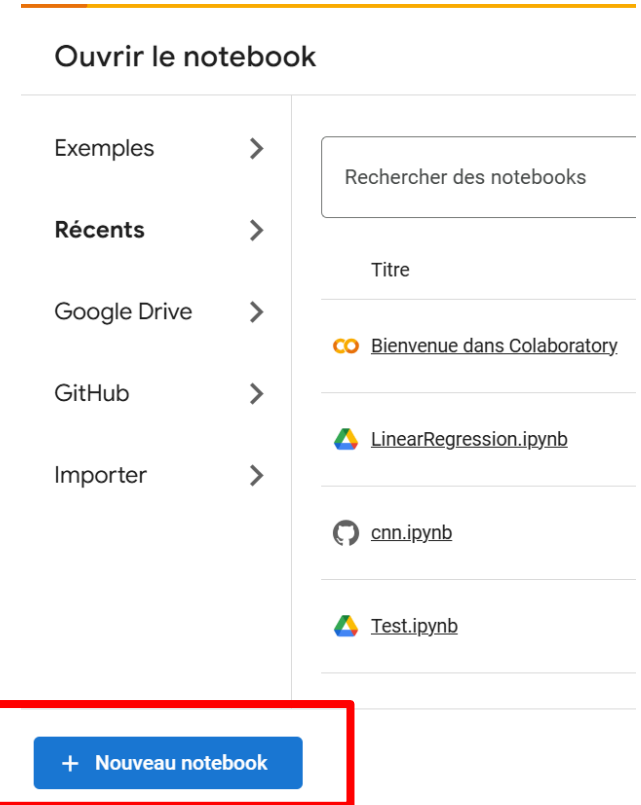
1 Ouvrir Google Colab et créer un Nouveau Notebook:

Voir comment utiliser Google Colab:

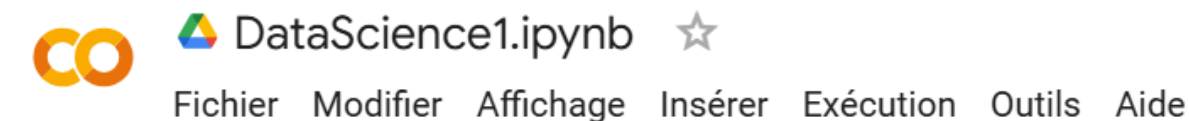
<https://www.youtube.com/watch?v=SRPCqXcuxKo>



Cliquer sur Nouveau notebook



2 Renommer le Notebook: DataScience1.ipynb



3 Charger Importer les bibliothèques nécessaires et créer le DataFrame

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
# Création d'un DataFrame exemple
data = {
    'Date': ['2024-01-01', '2024-01-01', '2024-01-02', '2024-01-02', '2024-01-03'],
    'Category': ['Electronics', 'Furniture', 'Clothing', 'Electronics', 'Furniture'],
    'Quantity': [2, 1, 3, 1, 2],
    'Price': [300, 150, 50, 300, 150],
    'Total': [600, 150, 150, 300, 300]
}
df = pd.DataFrame(data)
# Afficher les premières lignes
print(df.head())
```



Commencez à coder ou à générer avec l'IA.

Taper le code ICI

Exécuter le code



```
import pandas as pd
import numpy as np
```

DataFrame est une structure de données sous forme tableau en **pandas**

4 Préparation et Exploration des Données:

4.1 Vérifier les valeurs manquantes:

```
# Vérification des valeurs manquantes
print("Valeurs manquantes par colonne :")
print(df.isnull().sum())
```

Pour ajouter une cellule de Code

+ Code

+ Texte

```
⇒ Valeurs manquantes par colonne :
Date      0
Category   0
Quantity   0
Price      0
Total      0
dtype: int64
```

4.2 Afficher les types de données et un résumé statistique

```
# Types de données
print("Types des colonnes :")
print(df.dtypes)

# Statistiques descriptives
print("Résumé statistique :")
print(df.describe())
```

```
⇒ Types des colonnes :
Date      object
Category   object
Quantity   int64
Price      int64
Total      int64
dtype: object
Résumé statistique :
      Quantity      Price      Total
count  5.00000  5.000000  5.000000
mean    1.80000 190.000000 300.000000
std     0.83666 108.397417 183.711731
min     1.00000  50.000000 150.000000
25%     1.00000 150.000000 150.000000
50%     2.00000 150.000000 300.000000
75%     2.00000 300.000000 300.000000
max     3.00000 300.000000 600.000000
```

Remarque: Un **DataFrame** est une structure de données essentielle dans la bibliothèque **Pandas** en Python. C'est un tableau bidimensionnel (lignes et colonnes), similaire à une feuille Excel ou une table dans une base de données, mais beaucoup plus flexible et puissant pour manipuler et analyser des données.

La commande **print(df.describe())** génère un résumé statistique des colonnes numériques d'un DataFrame **df**. Voici ce que cela affiche typiquement :

Détails des colonnes :

- **count** : Nombre de valeurs non nulles.
- **mean** : Moyenne des valeurs.
- **std** : Écart type des valeurs (mesure de dispersion).
- **min** : Valeur minimale.
- **25%** : 1er quartile (25 % des valeurs sont inférieures ou égales à cette valeur).
- **50%** : Médiane (valeur au milieu de la distribution).
- **75%** : 3e quartile (75 % des valeurs sont inférieures ou égales à cette valeur).
- **max** : Valeur maximale.

Résumé statistique :

	Quantity	Price	Total
count	5.00000	5.000000	5.000000
mean	1.80000	190.000000	300.000000
std	0.83666	108.397417	183.711731
min	1.00000	50.000000	150.000000
25%	1.00000	150.000000	150.000000
50%	2.00000	150.000000	300.000000
75%	2.00000	300.000000	300.000000
max	3.00000	300.000000	600.000000

5 Transformation des Données:

5.1 Ajouter une colonne calculée "Profit":

- On suppose que le coût est de 70 % du prix.



	Date	Category	Quantity	Price	Total	Profit
0	2024-01-01	Electronics	2	300	600	180.0
1	2024-01-01	Furniture	1	150	150	45.0
2	2024-01-02	Clothing	3	50	150	45.0
3	2024-01-02	Electronics	1	300	300	90.0
4	2024-01-03	Furniture	2	150	300	90.0

```
# Ajouter une colonne "Profit"  
df['Profit'] = df['Total'] - (df['Price'] * 0.7 * df['Quantity'])  
print(df)
```

5.2 Groupement des données

Exemple : Calculer les revenus totaux par catégorie.

```
# Revenu total par catégorie  
total_revenue = df.groupby('Category')['Total'].sum()  
print(total_revenue)
```



```
Category  
Clothing      150  
Electronics   900  
Furniture     450  
Name: Total, dtype: int64
```

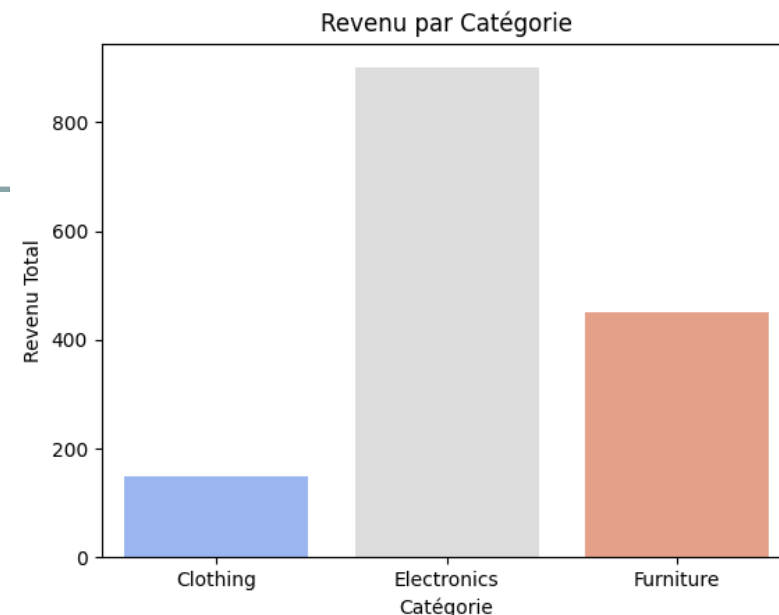
6 Analyse et Visualisation:

6.1 Visualiser les revenus par catégorie (Bar Chart):

```
# Bar Chart des revenus par catégorie
sns.barplot(x=total_revenue.index, y=total_revenue.values, palette="coolwarm")
plt.title("Revenu par Catégorie")
plt.xlabel("Catégorie")
plt.ylabel("Revenu Total")
plt.show()
```

barplot: diagramme en barres
palette="coolwarm": Définit une palette de couleurs du **bleu** au **rouge** (cool → chaud)

sns utilise la bibliothèque **Seaborn**, pour tracer un diagramme en barres (barplot)

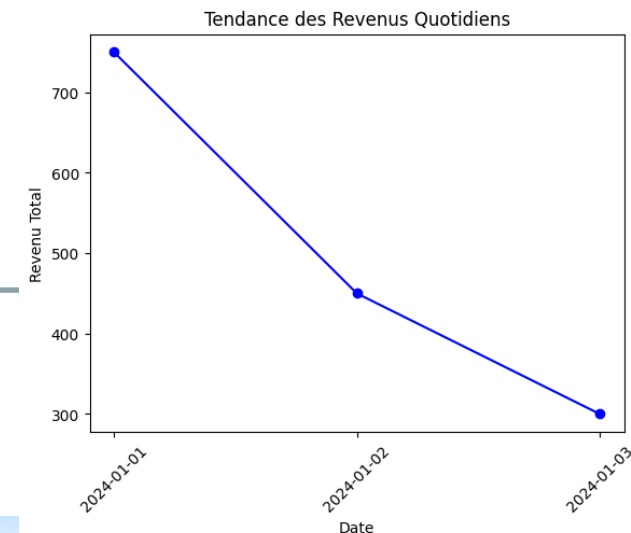


6 Analyse et Visualisation:

6.2 Visualiser la tendance des ventes par jour :

```
# Revenu total par jour
daily_revenue = df.groupby('Date')['Total'].sum().reset_index()

# Line plot pour la tendance des revenus quotidiens
plt.plot(daily_revenue['Date'], daily_revenue['Total'], marker='o', color='b')
plt.title("Tendance des Revenus Quotidiens")
plt.xlabel("Date")
plt.ylabel("Revenu Total")
plt.xticks(rotation=45)
plt.show()
```



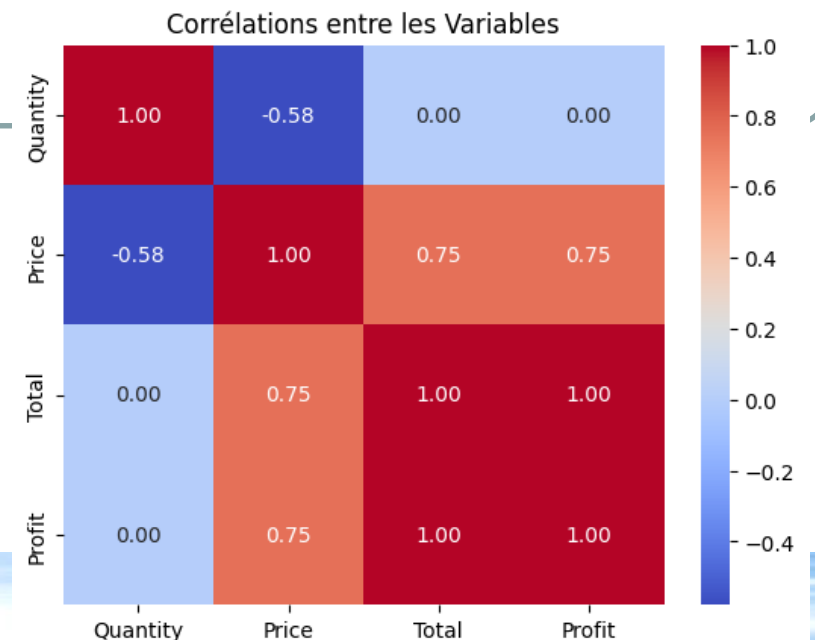
6 Analyse et Visualisation:

6.3 Heatmap des corrélations :

```
# Calculer les corrélations
corr_matrix = df[['Quantity', 'Price', 'Total', 'Profit']].corr()
# Heatmap
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Corrélations entre les Variables")
plt.show()
```

Ce code calcule et affiche une **matrice de corrélation** entre certaines colonnes d'un DataFrame, puis génère une **heatmap** (carte thermique) pour visualiser ces corrélations.

La corrélation mesure la relation linéaire entre deux variables. Elle varie entre -1 (**relation négative parfaite**: variables évoluent toujours de manière opposée) et +1 (**relation positive parfaite**: variables évoluent toujours dans la même direction), avec 0 signifiant aucune relation linéaire.



7 Insights et Interprétations :

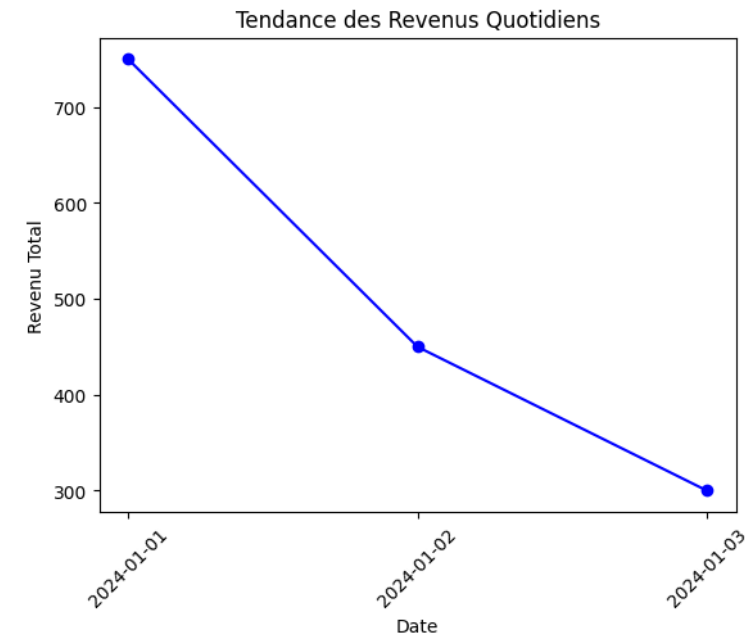
1. Quelle catégorie est la plus rentable ?

```
print("Catégorie la plus rentable :", total_revenue.idxmax())
```

➡ Catégorie la plus rentable : Electronics

2. Quelle est la tendance des ventes ?

Réponse observable dans le graphique des revenus quotidiens.

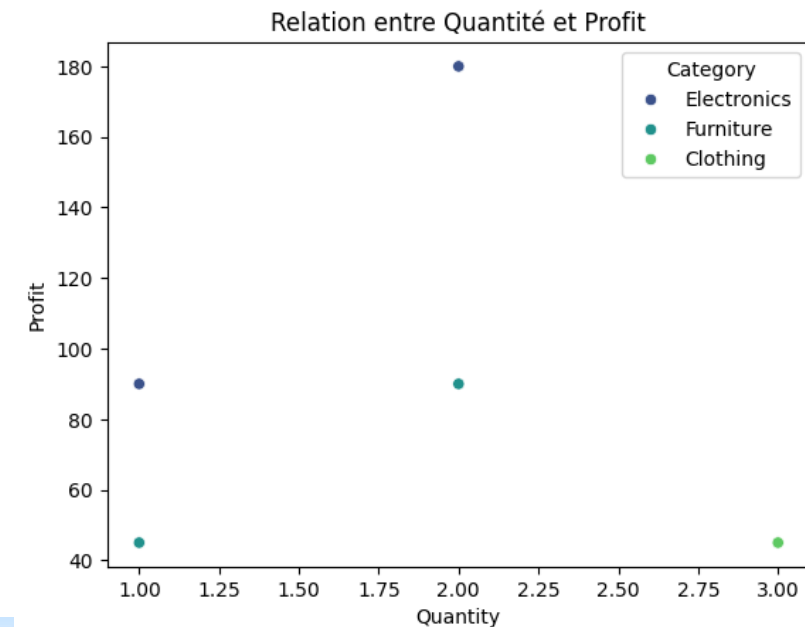


7 Insights et Interprétations :

3. Relation entre la quantité et le profit ?

```
sns.scatterplot(x='Quantity', y='Profit', hue='Category', data=df, palette="viridis")  
plt.title("Relation entre Quantité et Profit")  
plt.show()
```

Ce code utilise la bibliothèque **Seaborn** pour tracer un **scatter plot** (nuage de points) qui visualise la relation entre la quantité (*Quantity*) et le profit (*Profit*) dans le DataFrame **df**.



Résumé :

Ces exemples simples sont adaptés pour débiter en Data Science et apprendre à manipuler les bibliothèques **Pandas**, **NumPy**, **Matplotlib**, et **Seaborn**.

Les étapes précédentes montrent :

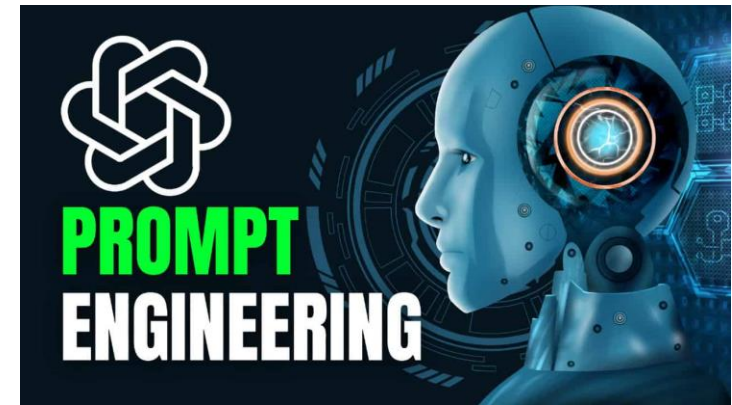
1. Comment explorer les données.
2. Comment les transformer pour ajouter des informations utiles.
3. Comment visualiser les relations et les tendances pour tirer des conclusions.

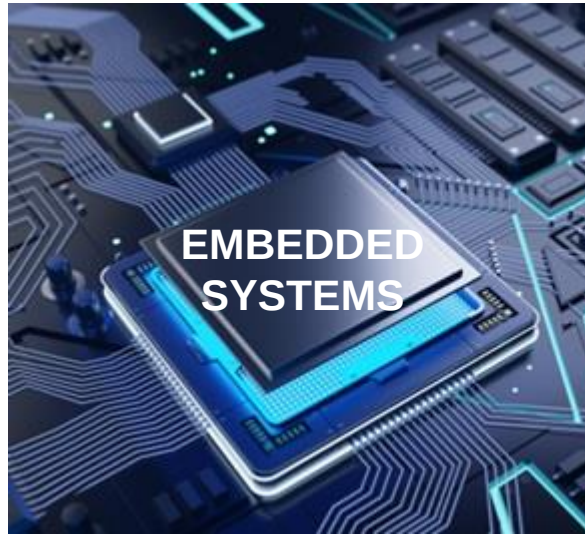
11. Le Prompt-Engineering

Le **Prompt Engineering** est l'art de formuler des requêtes (prompts) optimisées pour obtenir des **réponses précises et pertinentes** d'un modèle d'intelligence artificielle comme ChatGPT, Gemini, Claude, NotebookLM, Perplexity.ai et d'autres.

Principes clés du Prompt Engineering

1. **Clarté et précision** → Plus le prompt est clair, plus la réponse sera pertinente.
2. **Contexte détaillé** → Ajouter des détails pour orienter l'IA.
3. **Contraintes et formatage** → Spécifier la longueur, le ton, ou la structure de la réponse.
4. **Exemples** → Fournir des modèles de réponse attendus.





FIN !
Questions ?