



# Story 3 - Introduction IA

## Contexte

Toute entreprise a une large marée de données et peut être considérée comme une source abondante, vitale et nécessaire pour sa survie. Nous pouvons avec une préparation adéquate utiliser de nouvelles méthodes et différentes applications de l'IA pour diriger les données brutes ou les images vers des informations significatives, ou directement vers des connaissances (modèles) pour aider les gestionnaires à prendre une décision.

## Mots clés

### ▼ Intelligence

C'est la capacité **d'apprendre** ou de **s'adapter** à des situations nouvelles.

Les capacités cognitives sont les processus que l'humain utilise pour **reconnaître, raisonner et s'adapter** son environnement.

L'intelligence est l'ensemble des processus retrouvés dans des systèmes, plus ou moins complexes, vivants ou non, qui permettent de comprendre, d'apprendre ou de s'adapter à des situations nouvelles.

L'intelligence peut être également perçue comme la capacité à traiter l'information pour atteindre des objectifs.

L'intelligence reste un concept encore mal défini sur le plan scientifique.

Il existe plusieurs sorte d'intelligence dont l'intelligence émotionnelle, l'empathie, l'intelligence artistique.

### ▼ Intelligence artificielle

Ensemble de théorie et de techniques (programme) capable de simuler l'intelligence humaine (capacité de raisonnement, de communication, de déduction, de perception). Elle pour cela capable de traiter un très grand nombre de données ~~et même dépasser les capacités humaines dans certains cas.~~

→ Une calculatrice est-elle plus intelligente qu'un homme ?

- Ensemble de théorie et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence.
- Caractérise l'élaboration de programmes informatiques capables de prendre en charge des tâches habituellement effectuées par des humains car demandant un apprentissage, une organisation de la mémoire et un raisonnement. L'objectif est de parvenir à transmettre à une machine des fonctions propres au vivant : rationalité, raisonnement, mémoire et perception.
- La construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique.
- L'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence.

### ▼ Modèle

Fichier qui a fait l'objet d'un traitement IA préalable (soumis à une base d'apprentissage) et qui peut être utilisé pour automatiser certaines tâches.

Le modèle d'IA est la construction mathématique générant une déduction ou une prédiction à partir de données d'entrée. Le modèle est estimé à partir de données annotées lors de la phase d'apprentissage (ou d'entraînement) du système d'IA.

#### ▼ **Système expert**

Programme capable de se comporter comme un humain sur une tâche donnée.

Ordinateur capable de se comporter comme un expert (humain), mais dans un domaine bien précis.

#### ▼ **Réseau de neurones**

Modèle dont la structure est identique au réseau de neurone du cerveau humain pour apprendre et reconnaître.

Utilisé en deep learning mais pas que, il s'agit d'un concept. Ils sont efficaces dans la modélisation des problèmes complexes et non linéaires.

Un réseau de neurones artificiels, ou réseau neuronal artificiel, est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones

biologiques, et qui par la suite s'est rapproché des méthodes statistiques.

Un réseau de neurones est un modèle informatique dont la structure en couches est similaire à la structure en réseau des neurones du cerveau, avec des couches de nœuds connectés. Un réseau de neurones peut apprendre à partir de données ; il peut ainsi être entraîné à reconnaître des tendances, classer des données et prévoir des événements à venir.

Un réseau de neurones décompose vos données en couches d'abstraction. Il peut être entraîné sur de nombreux exemples en vue de reconnaître des modèles au niveau de la parole ou des images, par exemple, à l'instar du cerveau humain. Son comportement est défini par la façon dont ses éléments individuels sont reliés et par la solidité (ou poids) de ces liaisons. Ces poids sont automatiquement ajustés au cours de l'entraînement selon une règle d'apprentissage spécifiée jusqu'à ce que le réseau de neurones exécute correctement la tâche souhaitée.

Les réseaux de neurones sont particulièrement bien adaptés à la reconnaissance de formes pour identifier et classer des objets ou des signaux dans les systèmes de la parole, de la vision et du contrôle. Ils peuvent également être utilisés pour la prédiction et la modélisation de séries chronologiques.

#### ▼ Machine learning (apprentissage automatique)

Technologie qui utilise des données pour apprendre seul et inférer des modèles d'application optimal dans un domaine spécifique.

C'est la technologie la plus ancienne et la plus simple. Elle s'appuie sur un algorithme qui adapte lui-même le système à partir des retours faits par l'humain. La mise en place de cette technologie implique l'existence de données organisées. Le système est ensuite alimenté par des données structurées et catégorisées lui permettant de comprendre comment classer de nouvelles données similaires. En fonction de ce classement, le système exécute ensuite les actions programmées. Il sait par exemple identifier si une photo montre un chien ou un chat et classer le document dans le dossier correspondant. Après une première phase d'utilisation, l'algorithme est optimisé à partir des feedbacks du développeur, qui informent le système des classifications erronées et lui indiquent les bonnes catégories.

Le Machine learning fonctionne à partir d'une base de données contrôlable, le **Deep learning a besoin d'un volume de données bien plus considérable**. Le système doit disposer de plus de 100 millions d'entrées pour donner des résultats fiables.

On donne aux ordinateurs l'accès à des données, puis on les laisse apprendre par eux-mêmes, sans intervention humaine ou reprogrammation logicielle. Ce qui leur permet de s'améliorer progressivement, de manière autonome. Et de dépasser ainsi les fonctions et les capacités initialement programmées. Certains algorithmes ne se contentent plus de reconnaître des images, mais se montrent capables de les produire et de donner des yeux aux machines.

le machine learning est un modèle d'intelligence artificielle qui apprend seul grâce à l'analyse d'un jeu de données. Pour cela, il utilise **des procédés automatiques issus des statistiques et des probabilités**. Son objectif ? Aider les humains à exploiter des volumes de données inenvisageables et leur faire gagner du temps dans des activités à faible valeur ajoutée. Par exemple : examiner des images médicales, segmenter une base de clients en marketing ou évaluer la probabilité qu'une machine tombe en panne.

Le Machine Learning utilise l'apprentissage supervisé ou non-supervisé.

#### ▼ Deep learning (apprentissage profond)

Le Deep learning est une technologie incluse dans le machine learning. Technologie qui n'a pas besoin de données structurées et qui utilise des réseaux de neurones **multicouches** pour apprendre et catégoriser de manière autonome les données.

La profondeur est liée aux couches cachées de ses réseaux neuronaux.

C'est une sous-catégorie du Machine Learning. Le Deep learning (apprentissage profond) n'a pas besoin de données structurées. Le système fonctionne à partir de plusieurs couches de réseaux neuronaux, qui combinent différents algorithmes en s'inspirant du cerveau humain. Ainsi, le système est capable de travailler à partir de données non structurées.

Cette approche est particulièrement adaptée pour les tâches complexes, lorsque tous les aspects des objets à traiter ne peuvent pas être catégorisés en amont. Le système du Deep learning identifie lui-même les caractéristiques discriminantes.

Important : avec le **Deep learning**, le système identifie lui-même les **caractéristiques discriminantes des données**, sans avoir besoin d'une catégorisation préalable. Le système n'a pas besoin d'être entraîné par un développeur. Il évalue lui-même le besoin de modifier le classement ou de créer des catégories inédites en fonction des nouvelles données.

Concrètement, toutes les tâches accomplies à l'aide du Machine learning peuvent être résolues avec le Deep learning. Il ne faut donc pas nécessairement opposer apprentissage profond et automatique.

Le **Deep learning mobilise beaucoup plus de ressources** et n'est donc pas un processus efficient. Les champs d'application des deux technologies sont donc en principe bien délimités : toute tâche que le Machine learning peut exécuter doit être traitée par cette même technique.

Il repose sur un réseau de neurones artificiels, qui imitent le fonctionnement de notre cerveau. Et ce système crée une machine virtuelle composée de milliers d'unités, chacune chargée de petits calculs simples.

## ▼ Data mining

Technique IA - Il s'agit d'extraire une connaissance (lien entre plusieurs informations - modèle - patterns) à partir d'une grande quantité de données.

L'**exploration de données**, connue aussi sous l'expression de **fouille de données**, **data mining**, ou encore **extraction de connaissances à partir de données**, a pour objet l'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques.

Elle se propose d'utiliser un ensemble d'algorithmes issus de disciplines scientifiques diverses telles que les statistiques, l'intelligence artificielle ou l'informatique, pour construire des modèles à partir des données, c'est-à-dire trouver des structures intéressantes ou des motifs selon des critères fixés au préalable, et d'en extraire un maximum de connaissances.

### ▼ Apprentissage supervisé

L'algorithme utilise un très grand jeu de données étiquetées (on lui dit les données les réponses qu'on attend) soit pour classer un objet dans une catégorie spécifiée (classification), ou attribuer une valeur à un objet (régression).

Dans le cadre du machine learning. Il s'agit de deux sous-catégories (avec l'apprentissage non supervisé) qui permettent de résoudre des problèmes. Par exemple, on « nourrit » un programme avec des milliers de photos de voitures, étiquetées. Après cet « entraînement », le programme peut reconnaître, seul, des voitures de tous types sur les nouvelles images qui lui seront présentées.

Qu'il s'agisse des types d'apprentissage supervisé ou non supervisé, **tout part d'un jeu de données très important**. Et quand on dit « très important », cela peut signifier jusqu'à plusieurs millions d'images pour la base ImageNet. C'est à partir de cette base que l'algorithme peut apprendre. Mais **dans le cadre de l'apprentissage supervisé, la machine connaît déjà les réponses qu'on attend d'elle**. Elle travaille à partir de données étiquetées.

Cette méthode permet de réaliser **deux types de tâches** :

- Des tâches de classification : Ces tâches consistent à **attribuer une classe à des objets**.
- Des tâches de régression : Ici, on n'attribue pas une classe mais **une valeur mathématique** : un pourcentage ou une valeur absolue.

### ▼ Apprentissage non-supervisé



Programme qui utilise un très grand jeu de données non étiquetées et crée ses propres réponses grâce au **clustering** (Comme la classification mais les classes sont inventées par le programme) ou au **filtrage collaboratif** (recommandation).

Dans le cadre du machine learning. Il s'agit de deux sous-catégories qui permettent de résoudre des problèmes.

La différence par rapport à l'apprentissage supervisé lorsqu'on parle du type d'apprentissage non-supervisé, c'est que **les réponses que l'on cherche à prédire ne sont pas disponibles dans les jeux de données**. Ici, l'algorithme utilise un jeu de données non étiquetées. On demande alors à la machine de créer ses propres réponses. Elle propose ainsi des réponses à partir d'analyses et de groupement de données.

- Des tâches de clustering : Ici, on demande à la machine de grouper des objets dans des ensembles de données les plus homogènes possible. Cette technique peut sembler proche de celle de la classification dans l'apprentissage supervisé, mais à la différence de cette dernière, les classes ne sont pas pré-remplies par un humain, c'est la machine qui "invente" ses propres classes, à un niveau de finesse pas toujours évident pour un humain. Une technique très utile dans le marketing pour **faire de la segmentation client** notamment.
- Des tâches de filtrage collaboratif : ici l'objectif est toujours de **personnaliser une expérience client**. C'est une technique utilisée par de très nombreuses plateformes, telles que Netflix, Spotify, etc. Leurs algorithmes étudient ce que vous avez regardé, aimé, mais aussi ce que des profils similaires au vôtre ont apprécié, pour vous faire des recommandations automatiques. Le modèle s'appuie sur des facteurs implicites inconscients (ce que l'utilisateur a fait), plutôt que des facteurs explicites (des critères remplis par l'utilisateur).

#### ▼ Apprentissage semi-supervisé

Cet algorithme utilise une petite quantité de données étiquetées comme le machine learning supervisé et une plus grande quantité de données non étiquetées comme le machine learning non supervisé pour entraîner les



algorithmes. Les données étiquetées peuvent être utilisées pour entraîner partiellement l'algorithme de machine learning, et cet algorithme partiellement entraîné trouve également des informations de manière organique.

### ▼ Apprentissage renforcé (reinforcement learning)

Le programme apprend par expérience en optimisant une récompense dans le temps.

Ici, la machine peut se montrer encore plus créative, car elle peut élaborer ses propres stratégies et s'adapter dans le temps et dans son environnement pour réaliser des tâches données. Elle élabore des réponses complexes. L'algorithme devient un agent autonome,

dont l'objectif est de réaliser une action au sein d'un environnement. S'il y parvient selon les critères établis par le développeur, il est récompensé. À lui d'élaborer la suite d'actions qui lui permet d'atteindre son objectif.

C'est cette méthode qui sera exploitée pour construire les algorithmes des voitures autonomes par exemple. Mais c'est aussi comme cela qu'on peut entraîner des machines capables de jouer à des jeux de stratégie.

L'apprentissage par renforcement consiste, pour un agent autonome (ex. : robot), à apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps.

### ▼ Classification

Technique de machine learning supervisée qui permet de classer des objets dans des classes(labels/groupe) prédéfinies.

→ 2 classes ou plus.

La classification est un type de Machine Learning supervisé dans lequel un algorithme « apprend » à classer de nouvelles observations à partir d'exemples de données étiquetées.

### ▼ Régression linéaire

Technique de machine learning supervisée qui permet de prédire une valeur réelle/continue à partir des entrées.

→ La régression logistique est une classification.

Les modèles de régression décrivent la relation entre une variable réponse (sortie) et une ou plusieurs variables prédictives (entrée).

#### ▼ Clustering

Technique permettant grâce à l'apprentissage non supervisé de trouver des modèles ou des groupements dans les données

Le clustering est une approche d'apprentissage non supervisée dans laquelle les réseaux de neurones peuvent être utilisés pour effectuer une analyse exploratoire des données afin de trouver des modèles cachés ou des regroupements dans les données. Ce processus implique de grouper les données par similarité. L'analyse de séquence génomique, l'étude de marché et la reconnaissance d'objets sont des exemples d'applications d'analyse de cluster.

#### ▼ IoT (Internet of Things) internet des objet

Réseau d'objets connectés qui communique sans interaction humaine.

#### ▼ Éthique

L'**éthique** est une discipline philosophique portant sur les jugements moraux et dont le concept est donc très proche de celui de la morale. C'est une réflexion fondamentale de tout peuple afin d'établir ses normes, ses limites et ses devoirs. Dans les pays laïcs, l'éthique est étroitement liée à la philosophie politique qui la sous tend, mais elle peut aussi être liée à la culture, la religion, les croyances et les traditions d'un pays, d'un groupe social, ou d'un système idéologique.

Ce sont les règles qui ne nuisent pas aux utilisateurs :  
transparence, responsabilité, équité, fiabilité, inclusion,  
confidentialité.

## Problématiques

*Comment transformer un océan de données en flux régulier d'information pertinente pour répondre à des attentes précises ?*

Grâce à différentes techniques d'intelligence artificielle, choisies en fonction du type de donnée et de ce que l'on souhaite en faire.

## Hypothèses (vraies ou fausses)

*L'IA est plus performante qu'un algorithme classique.*

Vrai

## Plan d'action

✓ Parcourir des ressources

<https://experiences.microsoft.fr/articles/intelligence-artificielle/comprendre-utiliser-intelligence-artificielle/>

<https://siecledigital.fr/2018/08/20/histoire-intelligence-artificielle/>

<https://datascientest.com/intelligence-artificielle-definition>

<https://www.cmsinfo.org/high-tech/domaines-application-intelligence-artificielle/>

<https://www.coe.int/fr/web/artificial-intelligence/history-of-ai>

<https://ellcie-healthy.com/lintelligence-artificielle-une-evolution-fulgurante/>

<https://datascientest.com/quelle-difference-entre-le-machine-learning-et-deep-learning>

<https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/>

Les miennes

<https://experiences.microsoft.fr/articles/intelligence-artificielle/apprentissage-supervise-et-non-supervise-quelles-differences/>

- ✓ Définir les mots-clés
- ✓ Répondre aux hypothèses
- ✓ Répondre à la problématique
- ✓ Livrable

## Histoire

**1950** : article d'Alan Turing intitulé "*Computing Machinery and Intelligence*". De cet article découlera ce que l'on appelle aujourd'hui le Test de Turing qui permet d'évaluer la capacité d'une machine à tenir une conversation humaine.

**1974** : C'est une période appelée le « *AI Winter* ». Beaucoup d'experts ne réussissent pas à faire aboutir leurs projets et les gouvernements britannique et américain réduisent leurs financements pour les académies. Ils préfèrent soutenir des idées ayant plus de chances de déboucher sur quelque chose de concret.

**1980s** : Le succès des systèmes experts permet de relancer les projets de recherche sur l'intelligence artificielle.

**1997** : Deep Blue créé par IBM a battu Garry Kasparov, alors champion du monde d'échec. Deep Blue s'appuyait sur un algorithme systématique de force brute, où tous les coups envisageables étaient évalués et pondérés.

**2003** : Geoffrey Hinton (de l'Université de Toronto), Yoshua Bengio (de l'Université de Montréal) et Yann LeCun (de l'Université de New York) décident de démarrer un programme de recherche pour remettre au goût du jour les réseaux neuronaux. Des expériences menées simultanément à Microsoft, Google et IBM avec l'aide du laboratoire de Toronto de Hinton ont alors démontré que ce type d'apprentissage parvenait à diminuer de moitié les taux d'erreurs pour la reconnaissance vocale. Des résultats similaires ont été atteints par l'équipe de Hinton pour la reconnaissance d'image.

**2011** : Watson d'IBM a battu en direct les deux plus grands champions de *Jeopardy!*

**2013** : Facebook ouvre le Facebook Artificial Intelligence Research (FAIR) dirigé par Yann Le Cun.

**2016** : AlphaGO (IA de Google spécialisée dans le jeu de Go) battra le champion d'Europe (Fan Hui) et le champion du monde (Lee Sedol) puis elle-même (AlphaGo Zero). Précisons que le jeu de Go a une combinatoire bien plus importante que les

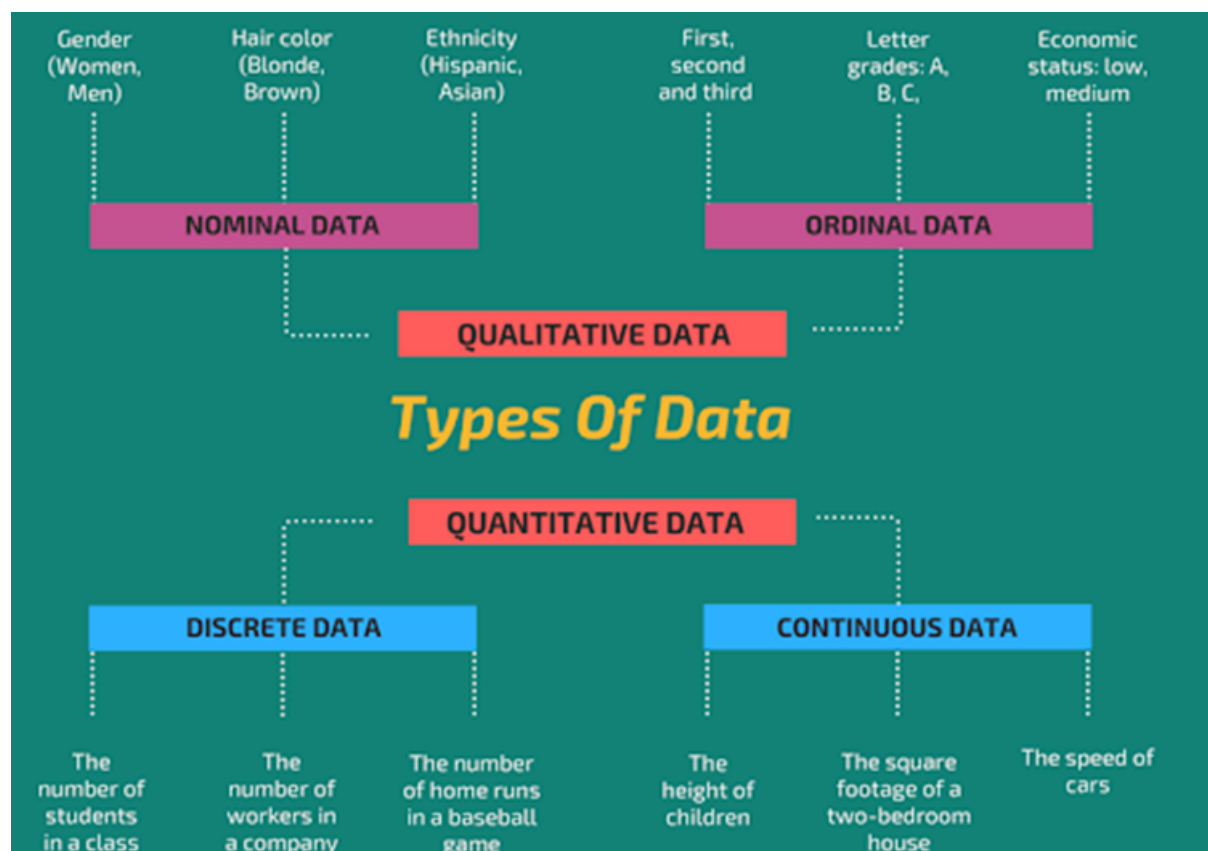
échecs (plus que le nombre de particules dans l'univers) et qu'il n'est pas possible d'avoir des résultats aussi significatifs en force brute (comme pour Deep Blue en 1997).

**La révolution actuelle** de l'intelligence artificielle et de la science qui en découle est rendue possible par « une combinaison de 3 facteurs ». Selon Harry Shum : « une vaste quantité de data ; une puissance informatique extraordinaire, notamment grâce au cloud ; et des algorithmes révolutionnaires, basés sur le deep-learning ».

## Les différents type d'IA

- Machine learning
- Deep learning

## Les différents type de données utilisées et leurs origines (qualitatives/quantitatives)



## Comparer les différents moyens d'optimiser les processus (tableau machine learning / deep learning )

	Machine Learning	Deep Learning
Base de données	Contrôlable	>1 millions de données
Entraînement	Entraînement par l'humain nécessaire	Apprentissage autonome
Algorithme	Algorithme modifiable	Réseau neuronal d'algorithmes
Champ d'application	Actions simples de routine	Tâches complexes











































## Avantages / inconvénients IA

Avantages	Inconvénients
	A besoin de données
Libère du temps en relayant les individus dans certaines tâches (nécessitant de la réflexion)	Sensible aux biais cognitif et sociétaux des programmeurs qui développent les applications IA
Permet de produire un traitement inédit de l'information, ne reproduisant pas strictement le processus de traitement humain	Le Deep Learning implique l'existence d'une boîte noire. Le processus de traitement des données qui a conduit au résultat ne peut être complètement restitué du fait de sa complexité et son caractère non traductible en langage humain.
	Sans cesse en évolution, requiert des ajustements permanents

## Avantages / inconvénients Machine learning/Deep learning

	Avantages	Inconvénients
Machine Learning	+ Simple à mettre en oeuvre	+ Moins efficace
Deep Learning	+ Performance	+ Demande plus de ressources + Nécessite plus de données

## Tableau comparatif des modes d'apprentissage

	Supervisé	Non supervisé	Renforcé	Profond - Supervisé	Profond - Non Supervisé	Profond - Renforcé
Types de données en entrée	 Structurées	 Structurées	 Structurées	 Non structurées	 Non structurées	 Structurées et non structurées
Apprentissage en continu	 Non	 Non	 Oui	 Non	 Non	 Oui
Transfert d'apprentissage*	 Non	 Non	 Non	 Oui	 Oui	 Oui
Complexité	 Faible à moyenne	 Moyenne	 Élevée	 Très élevée	 Extrêmement élevée	 Extrêmement élevée
Durée de l'apprentissage	 Secondes à heures	 Minutes à jours	 En continu	 Minutes à jours	 Minutes à jours	 En continu
Transparence et explicabilité	 Très bonne	 Bonne	 Faible	 Très faible	 Nulle	 Nulle
Risque de biais	 Fort	 Faible à Moyen	 Faible à Moyen	 Fort	 Moyen	 Moyen
Type de ressources	CPU	CPU	CPU	CPU, TPU ou GPU	CPU, TPU ou GPU	CPU, TPU ou GPU

## Applications

### Santé

- Comprendre des radiographies mieux que les médecins
- Médecine préventive et prédictive



## Transport

- Véhicules autonome : Conduire des voitures ou aides à la conduite.
- optimisation des flux

## Tertiaire

- Chatbots
- Faire de la traduction
- Amplifier les capacités de vos collaborateurs en les déchargeant des tâches répétitives, à faible valeur ajoutée, vous assurer que vos clients, conducteurs d'automobiles, ne s'endorment pas au volant.
- Les forces de vente, elles gagnent, grâce aux algorithmes, la capacité de mieux cibler les bons prospects et de connaître les clients à choyer, grâce à l'analyse automatique des conversations.
- Informatique décisionnelle : le Machine learning peut aussi servir à visualiser les données importantes de l'entreprise et à rendre différentes prévisions compréhensibles pour les décideurs humains.
- Marketing en ligne

## Industrie

- Maintenance prédictive : repose sur quelques étapes clés : collecter les données à l'aide de capteurs ; les centraliser et les analyser ; modéliser des schémas de panne et déployer des algorithmes qui apprennent à reconnaître les signes avant-coureurs... Pour des gains opérationnels, des risques réduits et une maintenance améliorée.
- Gestion des stocks

## Numérique

- Sécurité informatique : Contrairement aux solutions basées sur le Machine learning, les systèmes IT et de cyber-sécurité qui s'appuient sur le Deep learning peuvent identifier aussi bien les dangers documentés que les risques

jusqu'alors inconnus grâce à leur capacité à détecter les anomalies dans les patterns connus du réseau neuronal. Le Deep learning démultiplie l'efficacité des mesures de sécurité.

- Reconnaissance vocale, faciale.

### Agriculture

- Des modèles prédictifs de rendements alimentaires sont actuellement en phase de test. Rien que dans ce secteur, les investissements relatifs à l'IA seront conséquents ce marché devrait passer de 520 millions de dollars à 2,5 milliards de dollars d'ici 2025.

### Défense

- Robots de soutiens intelligents, drones autonomes, détection de mouvement, de suivi, d'identification.
- Reconnaissance faciale.

### Artistique

- Imaginer une partie manquante d'une photographie
- Créer des musiques

## Métiers

### Data Analyst

Un analyste de données, est un quelqu'un qui est **capable d'interroger des sources de données** pour en faire des rapports et des visualisations graphiques (graphes camemberts, histogrammes etc...). Un Data Analyst a une **compréhension forte du domaine métier** dans lequel il opère. Ce qui lui permet de mieux communiquer avec les gens du métier.

Pour mieux explorer les données, un Data Analyst est généralement à l'aise avec les outils statistiques. Toutefois, il n'est pas forcément aussi "calé" techniquement qu'un *software engineer* pour traiter les grands volumes de données (Big Data).

**Compétences et outils** : Excel, Access, SQL, SPSS, Tableau, Statistiques...

## Business Intelligence Developer

Les développeurs B.I. (Business Intelligence / informatique décisionnelle) vont **mettre en place des outils de B.I. pour les besoins de l'entreprise**. Ces outils se présentent généralement sous forme de *Data warehouses*, *Datamart*, ainsi que des *bases de données multidimensionnelles* construits à partir d'agrégation de données en provenance de plusieurs bases de données. La construction des *Data warehouse* et les *bases OLAP* est généralement effectuée à travers des Job ETL (Extract, Transform, Load) en utilisant l'outil Talend par exemple.

Ces Bases de données multidimensionnelles et *Data warehouses* sont par la suite utilisées par les développeurs B.I pour construire des tableaux de bords (*Dashboards*) et des rapports utiles pour les managers et les décideurs.

Les développeurs de B.I. ont généralement une connaissance métier moindre que celle d'un Data Analyst. Cependant, ils sont plus "calés" techniquement pour s'interfacer avec les différentes sources de données.

**Compétences et outils** : SQL, OLAP, Data warehouses, Cubes, SSAS, SSIS, ETL (Talend...)

## Data Engineer

Un Data Engineer est quelqu'un ayant un background technique en développement logiciel. Il peut être un *Software Engineer* qui s'est reconverti dans le Big Data.

Les Data Engineers vont mettre en place des systèmes de Big Data pour traiter ces dernières. Ils opteront pour des outils de stockage performants comme les bases de données NoSQL

et se baseront sur Hadoop, Spark, Map/Reduce pour traiter convenablement ces grands volumes de données.

Les Data Engineer vont **collecter, transformer les données de différentes sources**. Ce travail préparatoire permettra d'avoir des données "propres", prêtes pour qu'on leur applique dessus des techniques de Machine Learning.

En d'autres termes, le travail d'un Data Engineer est de **préparer le terrain** pour qu'un Data Scientist puisse se servir des données propres pour en tirer des tendances (Insights).

**Compétences et outils** : SQL, NoSQL, Hadoop, Data Lake, Big Data, Spark, Software Engineering, Map/Reduce...

## Data Scientist

Un Data Scientist est un **profil pluridisciplinaire** qui aura pour mission première de tirer de l'information utile (*insights*) depuis des données brutes. Le métier du Data Scientist est à **l'intersection entre *Data Analyst* et de *Data Engineer***. Tout en ayant des connaissances métiers dans le domaine dans lequel il évolue.

Un *Data scientist* va **explorer et exploiter les gisements de données de l'entreprise** pour leur appliquer des techniques de machine learning. Il s'agit donc d'une forme de *Data Analysis* poussée sur de grands volumes de données. L'exposition au contexte Big Data exige qu'un Data Scientist soit familier avec des concepts comme **Map/Reduce, Hadoop, Data lake** etc...

L'information utile recherchée par un Data Scientist est **spécifique à une entreprise** et plus généralement à un domaine métier. Pour cela, un Data Scientist doit être à l'aise avec le domaine métier dans lequel il opère. Pour cela, il côtoiera les gens du métier pour creuser avec eux les différentes pistes de réflexion.

Finalement, un data scientist doit être **un bon communicant** pour mieux communiquer ses retournements. Il utilisera pour cela des différents supports de présentation comme les présentations *PowerPoint*, ainsi que des visualisations graphiques (histogrammes, camemberts...) plus parlantes aux décideurs.

**Compétences et outils** : SQL, NoSQL, Python, R, Machine Learning, Deep Learning, Statistiques, Software Engineering...

## Objectifs

- ☐ Comprendre les différentes applications de l'IA

## Remarques pertinentes