



# Introduction

---

Alison PATOU

Patou.alison@gmail.com



# PRÉSENTATION



ALISON PATOU  
PATOU.ALISON@GMAIL.COM

# A savoir

- 2 notes pour ce module





# Programme



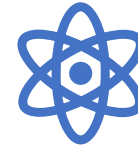
Introduction



What is data ?



Big Data



Machine Learning,  
Data Science, IA



Focus sur l'Intelligence  
Artificielle

1

# Introduction

Objectifs de ce cours

Savoir expliquer les concepts du Big Data et de l'IA

Comprendre les enjeux de l'IA

Convergence entre IA&Big Data

La chaine de traitement des données

# Cela se passe aujourd'hui...



24 PO (24 000 000 000 000 000 000) / j  
(environ 30 000 disques durs de PC)



10 000 000 photos / jour  
3 000 000 000 'like' / jour



2015 : 500 000 000 tweets / jour



800 000 000 utilisateurs / mois

# Cela se passe aujourd'hui...



24 PO (24 000 000 000 000 000 000) / j  
(environ 30 000 disques durs de PC)



10 000 000 photos / jour  
3 000 000 000 'like' / jour



2015 : 500 000 000 tweets / jour



800 000 000 utilisateurs / mois

## Explosion du volume des données numériques

Particuliers



Entreprises



Objets connectés



Open Data

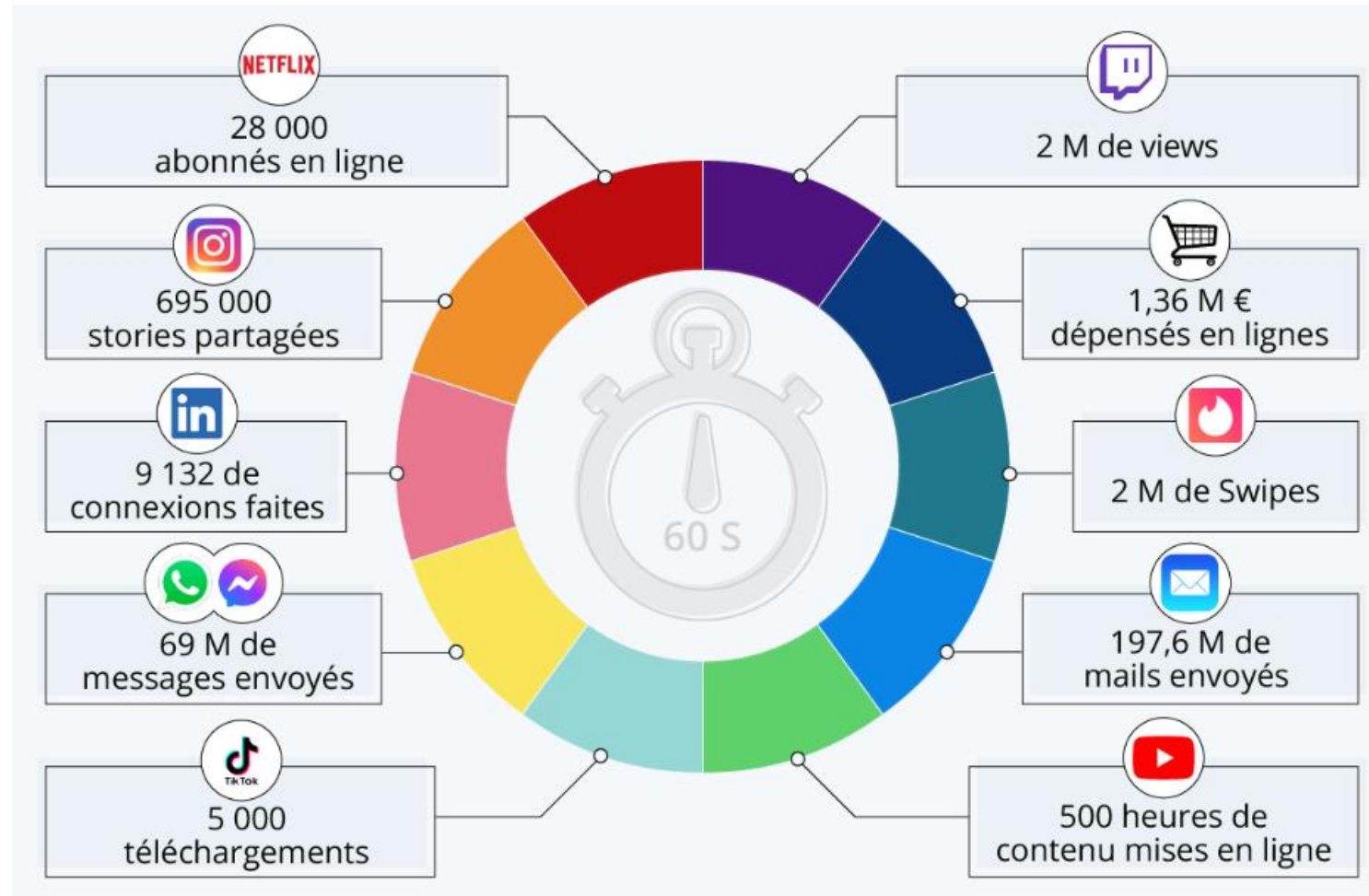


Internet





# Une minute sur Internet

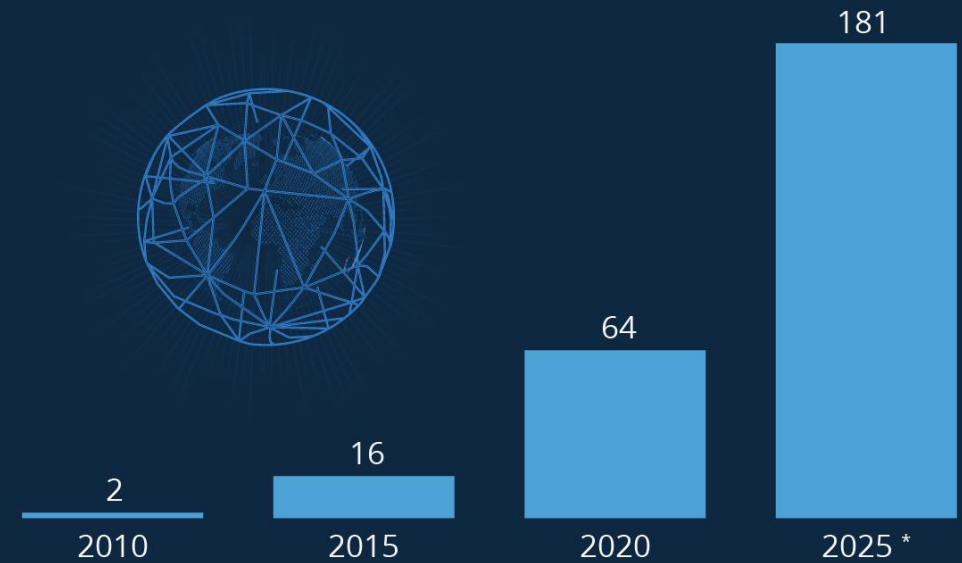


## Des faits :

- Plus de données ont été créées durant les deux dernières années que dans toute l'histoire de l'humanité auparavant
- En 2020, on est à 64 Zo de données
- On pense à faire du \*3 pour 2025

## Le Big Bang du Big Data

Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets



Un zettaoctet équivaut à mille milliards de gigaoctets.

\* Prévision en date de mars 2021.

Sources : IDC, Seagate, Statista



2

# Machine Learning, Data Science, IA

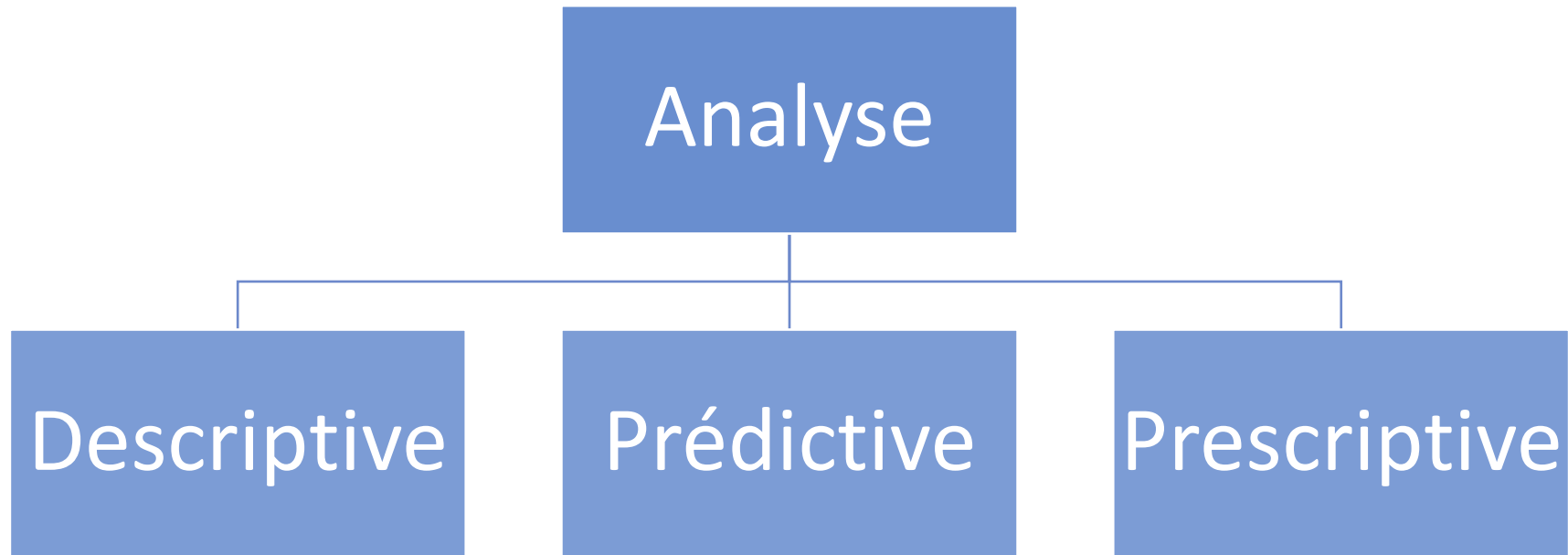
# L'analyse

Stocker des données de types différents, provenant de sources différentes sur des technologies différentes c'est bien

## MAIS

**Ce qui est encore plus intéressant est d'en tirer de la valeur grâce à l'analyse**

# Découpage de l'analyse



Ce qu'il s'est passé  
Ex : corrélations, moyenne, ...

Ce qu'il se passera  
Ex : Détection fraude,  
prédiction météo

Optimisation  
***Pourquoi cela se passera et  
recommandation d'actions pour  
tirer un bénéfice des prédictions***

# Des cas d'usages

## Cas d'usage 1

## Recommandation de produits en ligne

Exemple de recommandation sur Amazon

### Les clients ayant acheté cet article ont également acheté



Le degré zéro de l'écriture  
► Roland Barthes  
★★★★☆ (7)  
Poche  
EUR 6,27



Qu'est-ce que la littérature  
Jean-Paul Sartre  
★★★★☆ (3)  
Poche  
EUR 7,98



Figures I  
Genette  
Poche  
EUR 7,70



Littérature : 140 textes  
théoriques et critiques  
Jacques Vassevière  
★★★★☆ (8)  
Broché  
EUR 24,13

# Des cas d'usages

## Cas d'usage 1

## Recommandation de produits en ligne

- ✓ Historique des achats
- ✓ Autres produits achetés lors de la commande par d'autres utilisateurs
- ✓ Produits similaires basés sur des caractéristiques
- ✓ « Traduction » et analyse des clics sur le site d'achat en temps réel

Exemple de recommandation sur Amazon

**Les clients ayant acheté cet article ont également acheté**



**Le degré zéro de l'écriture**  
Roland Barthes  
★★★★☆ (7)  
Poche  
EUR 6,27



**Qu'est-ce que la littérature**  
Jean-Paul Sartre  
★★★★☆ (3)  
Poche  
EUR 7,98



**Figures I**  
Genette  
Poche  
EUR 7,70

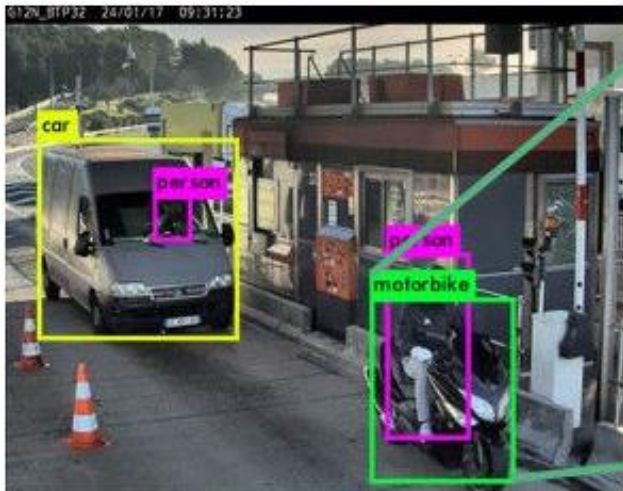


**Littérature : 140 textes théoriques et critiques**  
Jacques Vassevière  
★★★★☆ (8)  
Broché  
EUR 24,13

# Des cas d'usages

**Cas d'usage 2** Une société d'autoroute veut mieux classifier les véhicules au péage et éviter les interventions des agents lors des appels par interphone des usagers.

De l'analyse d'images aux Péages

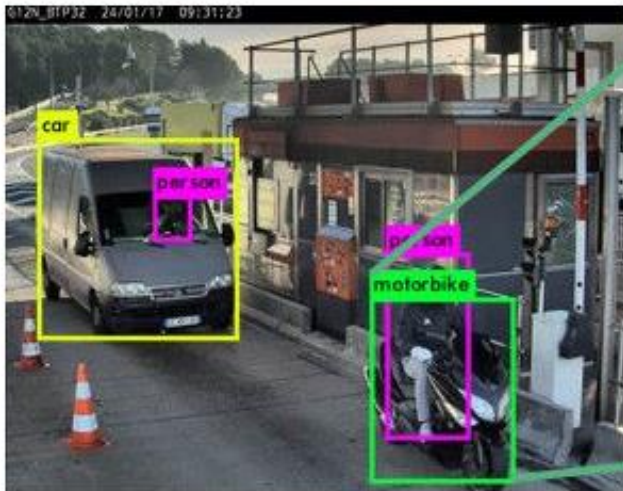




# Des cas d'usages

**Cas d'usage 2** Une société d'autoroute veut mieux classifier les véhicules au péage et éviter les interventions des agents lors des appels par interphone des usagers.

De l'analyse d'images aux Péages



- ✓ Données météorologiques
- ✓ Calendrier (vacances)
- ✓ Analyse du flux en temps réel
- ✓ Historique des images, passages au péage, ...

# Des cas d'usages

Cas d'usage 3 Une société de chauffeur privé souhaite ajuster ses prix en fonction de la demande et des temps de trajet.



# Des cas d'usages

Cas d'usage 3 Une société de chauffeur privé souhaite ajuster ses prix en fonction de la demande et des temps de trajet.



- ✓ Données de trafic
- ✓ Données d'évènements
- ✓ Analyse des réseaux de transport public des villes

# Des cas d'usages

## Cas d'usage 4 Compteur intelligent pour adapter l'offre, détecter les défaillances et le risque d'attrition



# Des cas d'usages

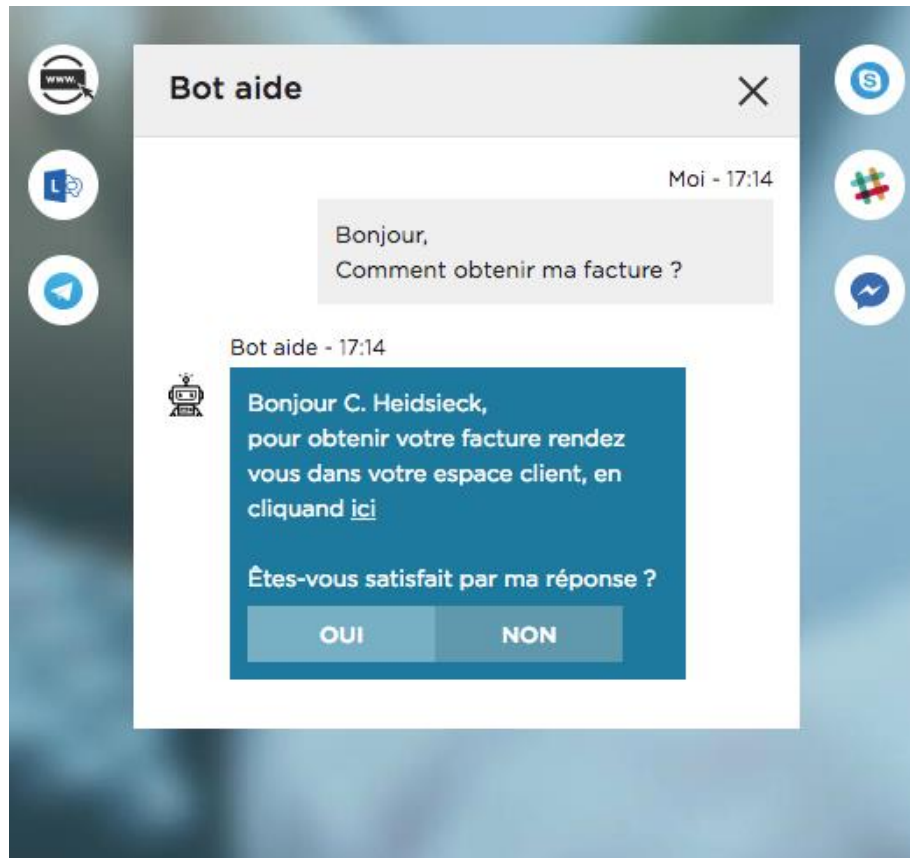
## Cas d'usage 4 Compteur intelligent pour adapter l'offre, détecter les défaillances et le risque d'attrition



- ✓ Données de consommation
- ✓ Données des clients
- ✓ 1 relevé par jour de ses 11 millions de clients

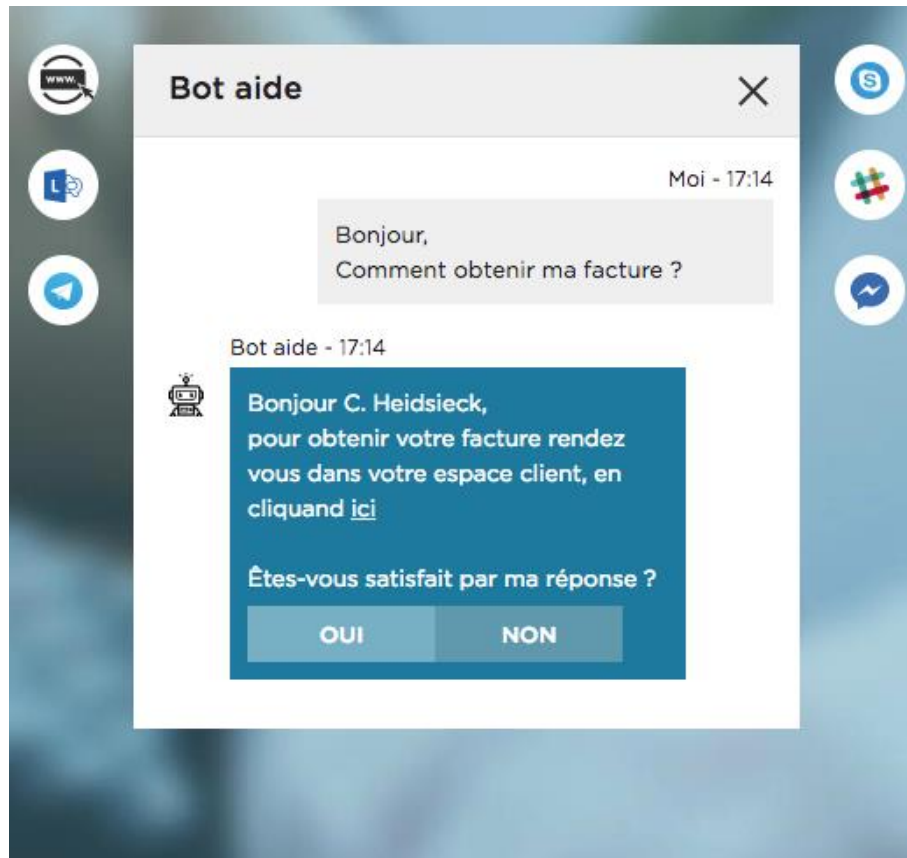
# Des cas d'usages

## Cas d'usage 5 Un chatbot conversationnel pour répondre aux problématiques des utilisateurs



# Des cas d'usages

## Cas d'usage 5 Un chatbot conversationnel pour répondre aux problématiques des utilisateurs



- ✓ Historique des demandes clients
- ✓ Analyse du langage
- ✓ Décrypte les mots-clés et apporte des réponses utiles et personnalisées en fonction des questions.

# Qu'est-ce que la Data Science ?

---



15 min

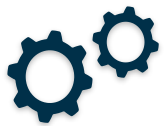


# La data science

## ➤ Définition de la Data Science

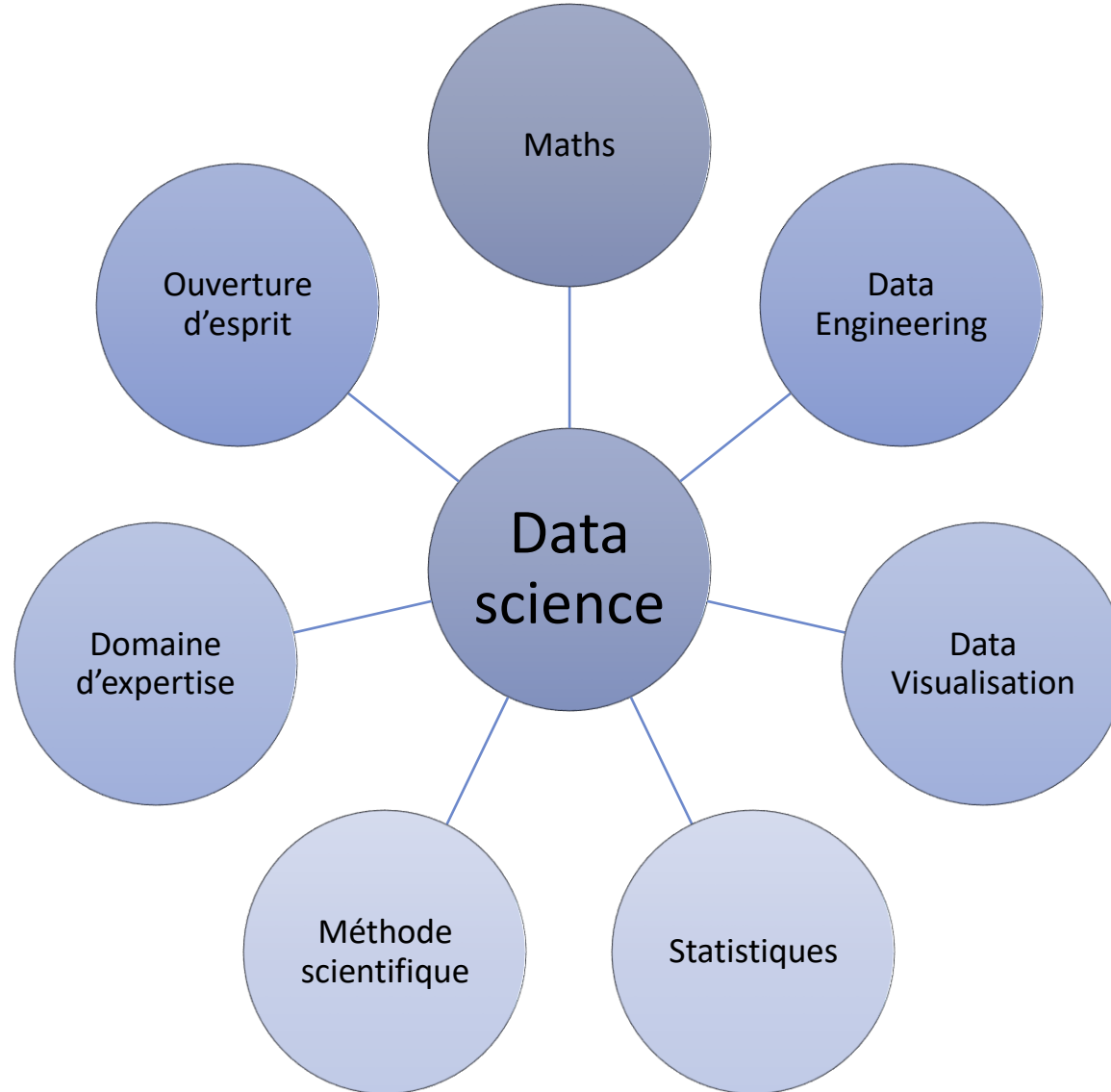
### Objectif : Extraire des connaissances d'ensembles de données

La Data Science est un **mélange disciplinaire** entre les statistiques, le développement d'algorithmes et la technologie, dont le but est la **résolution de problèmes analytiques complexes**.



En étudiant les données, le Data Scientist peut **découvrir et comprendre des tendances et des comportements complexes**. Il aide à extraire des informations permettant aux entreprises de **prendre des décisions** plus adaptées à leurs différentes problématiques.

# Qu'est-ce que la Data Science



# La data science

## Le Big Data

- ✓ Le mot tendance qu'on entend partout
- ✓ Englobe la partie architecture, data science et data visualisation

## Data Science

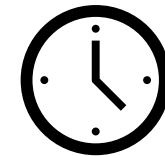
- ✓ Résolution de problématiques analytiques complexes
- ✓ Mélange disciplinaire de statistiques et d'algorithmie

## Le Machine Learning

- ✓ Un domaine de la Data Science
- ✓ Apprentissage par le biais d'exemples déjà présents dans la base de données

# Qu'est-ce que le Machine Learning ?

---



15 min

# La data science

La Data Science : La vraie valeur ajoutée du Big Data

## ➤ Qu'est ce que le Machine Learning ?

« **L'apprentissage automatique**, c'est la capacité d'un ordinateur à apprendre sans avoir été explicitement programmé. »

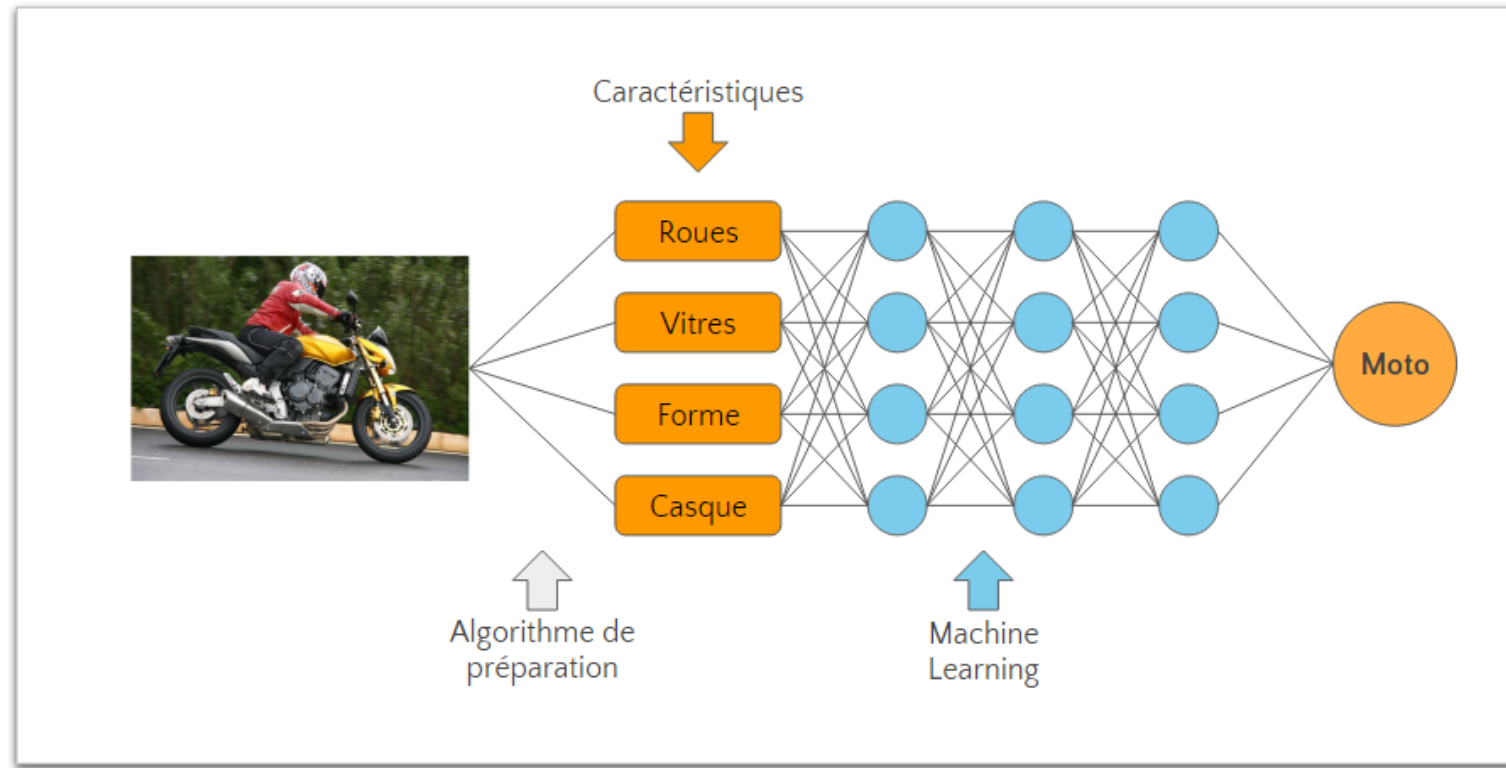
*Arthur Samuel, pionnier du Machine Learning*

C'est un outil complémentaire du Data Mining. Il consiste en un **apprentissage par le biais d'exemples** présents dans la base de données.

Cela consiste en la mise en place d'**algorithmes** ayant pour objectif d'obtenir une **analyse prédictive** à partir de données, dans un but précis.



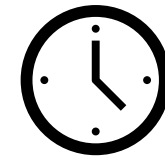
Principe : **Bâtir des modèles qui établissent les relations entre de nombreux facteurs** permettant l'évaluation des risques ou celle des opportunités associés à un ensemble spécifique de conditions, afin **d'orienter la prise de décision** vers telle ou telle action.



# Machine learning

# Qu'est-ce que le Deep Learning ?

---

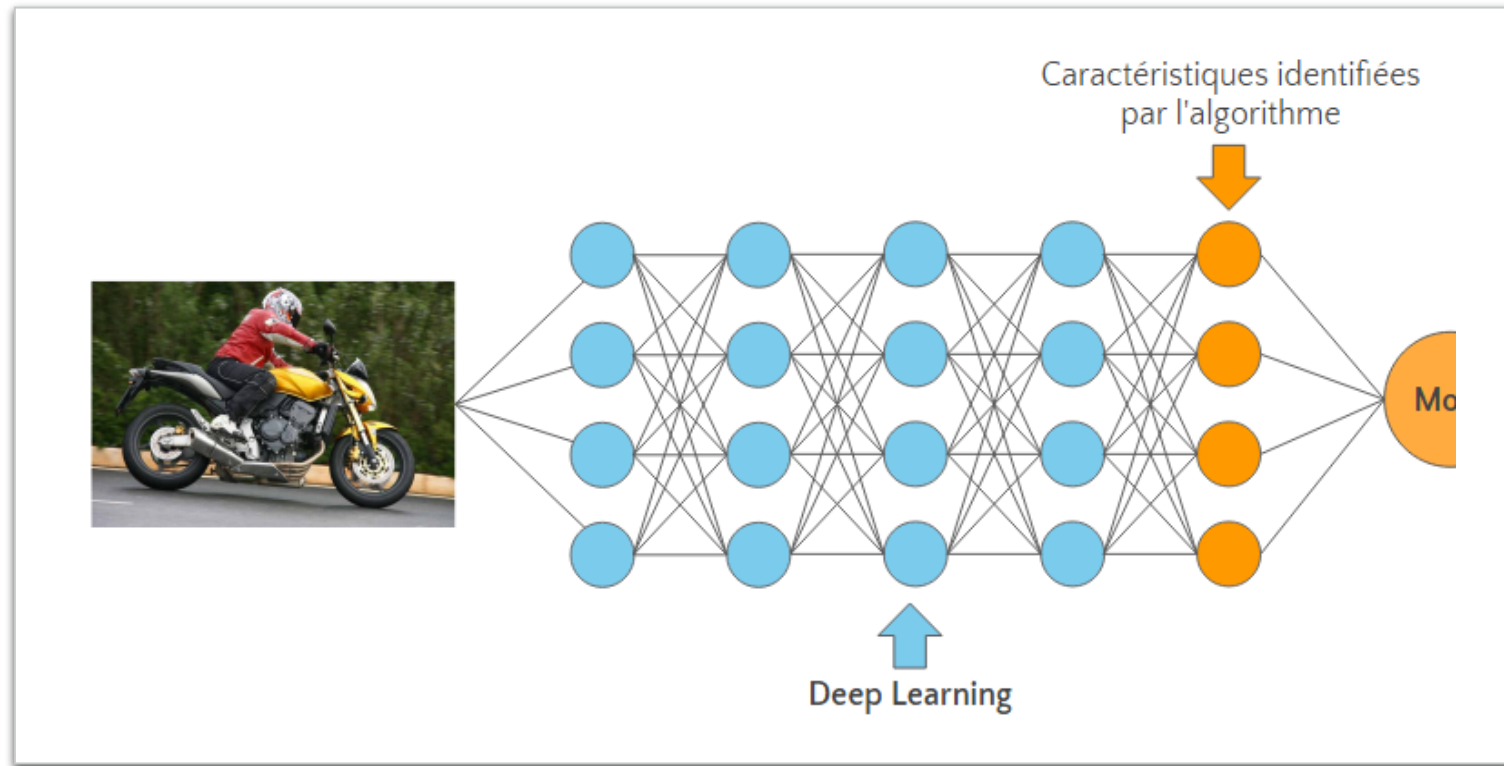


15 min

# Deep learning

- Les caractéristiques essentielles du traitement ne seront plus identifiées par un traitement humain dans un algorithme préalable, mais directement par l'algorithme de Deep Learning.
- L'algorithme construira, à partir des exemples à disposition, ses propres caractéristiques
- Le but est d'être le plus similaire à l'identification humaine





# Réseaux de neurones

# Réseaux de neurones

## *Chihuahua vs. Muffin*



# Deep learning

Visualisations souvent  
utilisées pour représenter l'IA  
(Intelligence Artificielle)

Derrière : Algorithme de  
réseaux de neurones

## *Chihuahua vs. Muffin*



chocolate cookie



fawn smooth Chihuahua



baked blueberry muffin



white chihuahua



fawn smooth Chihuahua



blueberry muffin



[unknown]



brown coated Chihuahua



baked muffin



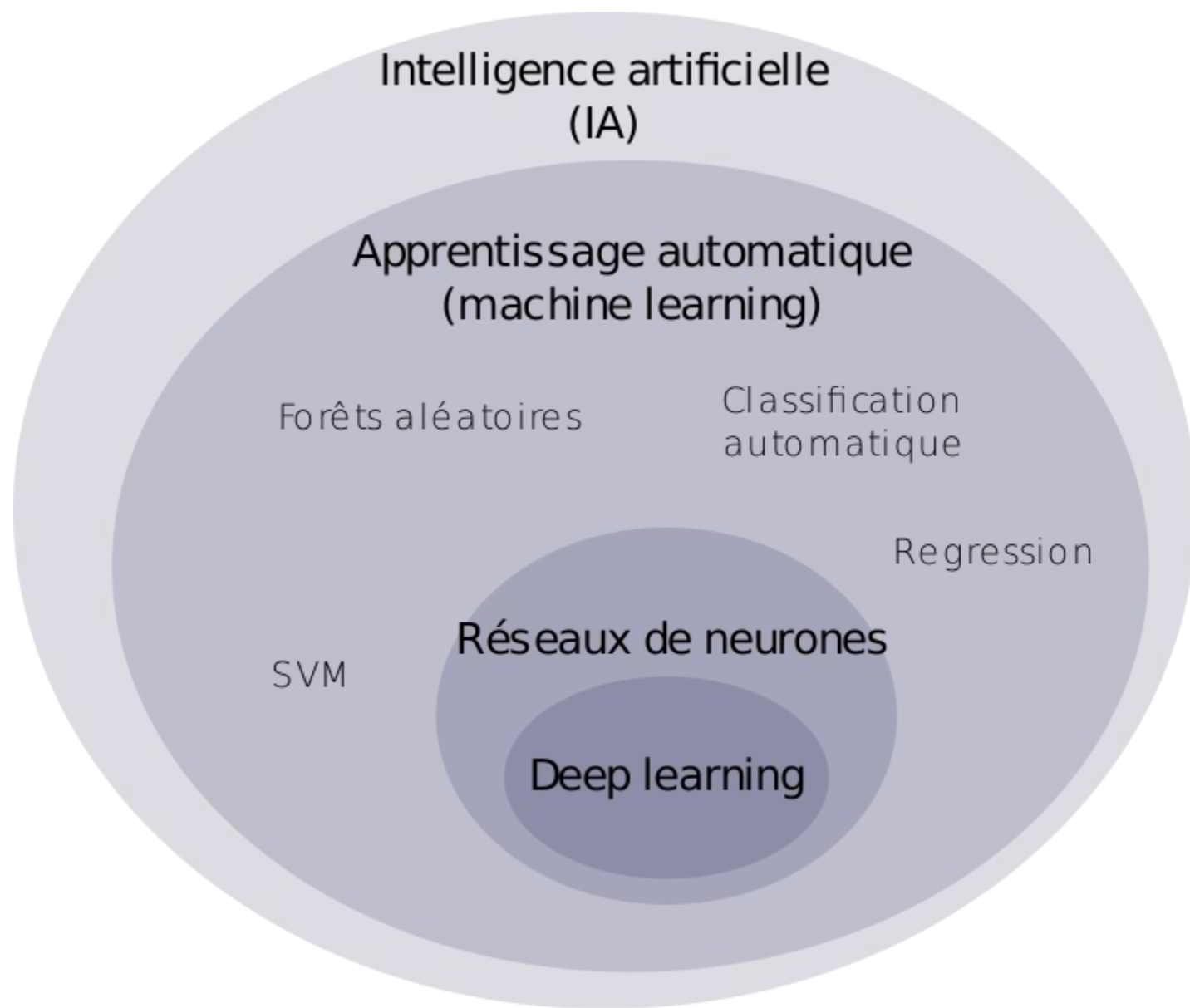
beige short coated puppy

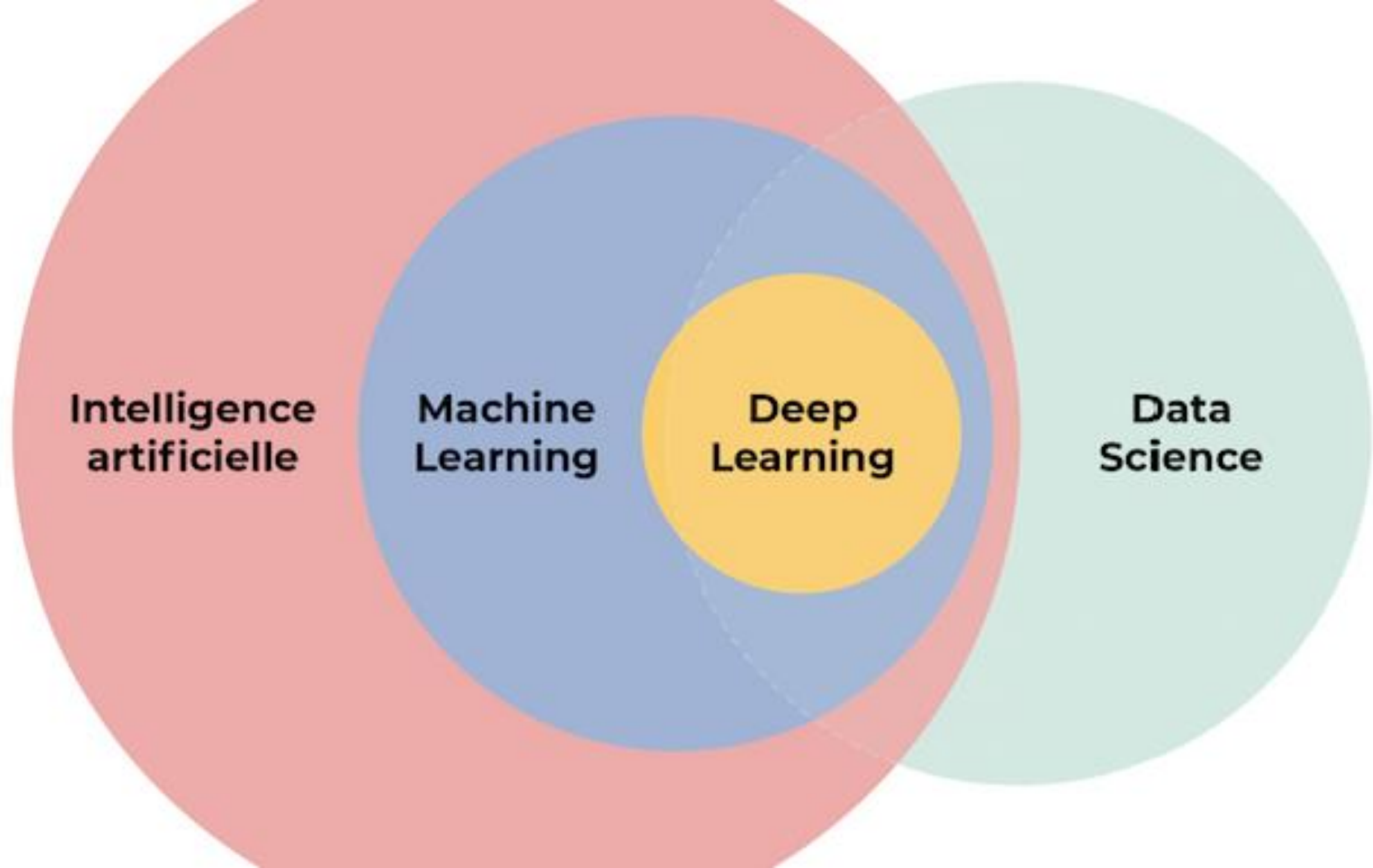


tan smooth Chihuahua puppy



blueberry cupcakes







3

Focus IA

# Test de Turing

Article de 1950 : Computing Machinery and Intelligence.

Idée : ne pas définir l'IA mais la tester.

- Protocole : un individu communique à l'aide d'un terminal d'ordinateur avec un interlocuteur invisible. Il doit décider si l'interlocuteur est un être humain ou un système d'IA imitant un être humain.
- Construire un cerveau artificiel en imitant moins la physiologie du modèle humain, que son comportement logique, les "états de pensée" étant équivalents aux instructions de la machine.
- Prédiction de Alan Turing : 30 % de réussite pour une machine discutant avec une personne non avertie pendant 5 minutes.
- Les principales composantes d'un système d'IA doivent être les connaissances, le raisonnement, la compréhension du langage naturel et l'apprentissage. **TURING**





# L'Intelligence Artificielle

- L'explosion de la puissance de calcul des machines et des données dont nous disposons ont fait basculer l'Intelligence Artificielle, dans les années 2010, d'une idée de science-fiction à une réalité de plus en plus proche, devenue enjeu scientifique majeur.



**L'Intelligence Artificielle** (IA, ou AI en anglais pour *Artificial Intelligence*) consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle. L'IA se retrouve implémentée dans un nombre grandissant de domaines d'application.

# L'IA est déjà partout

**Actuellement, on a déjà de l'intelligence artificielle un peu partout autour de nous**



I'm listening...

Hey Siri<sup>™</sup>  
tap to edit



Waze



12:32



240 m

Jefferson Ave

13:04PM 21mi. 32min. >





Internet est

internet est **lent**

internet est **mort**

internet est **il un espace privé**

internet est **lent windows 10**

internet est **un service du web**

internet est **dangereux**

internet est **un moyen de communication et de liberté**

internet est **la première chose que l'homme a créé sans la comprendre**

internet est **long**

internet est **attaqué**

Recherche Google

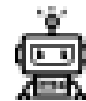
J'ai de la chance

[En savoir plus](#)

Moi - 17:14

Bonjour,  
Comment obtenir ma facture ?

Bot aide - 17:14



Bonjour C. Heidsieck,  
pour obtenir votre facture rendez  
vous dans votre espace client, en  
cliquand [ici](#)

Êtes-vous satisfait par ma réponse ?

OUI

NON



evebeauregard

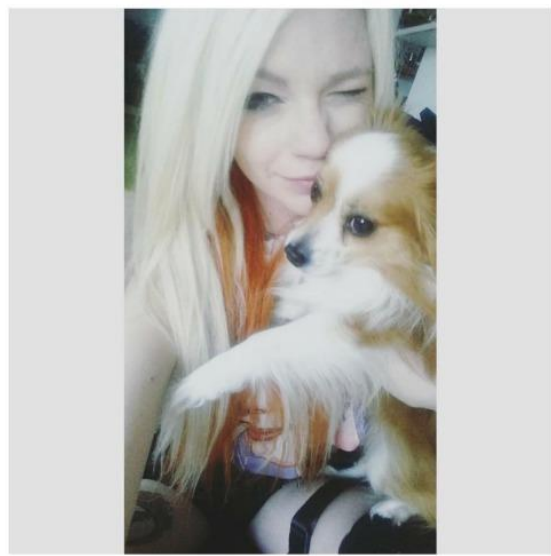


1,392 likes  
evebeauregard Quick snap of my new @ladykloves pencil skirt and sunglasses! The skirt is amazing - IT EVEN HAS POCKETS! Better photos when I'm not running out the dooo... more  
View all 14 comments  
rngeebus 10/10 would ask out for a coffee if bumped into on the street, absolutely stunning!  
kimmeeeee Looking fab 🍷  
1 HOUR AGO

rincityofficial



rincityofficial



133 likes  
rincityofficial ❤️  
View all 3 comments  
geekofawolf Eyyy soo cute!!! 😍  
renee.87 Does she have a name? @rincityofficial  
12 MINUTES AGO

sunkissed\_kris

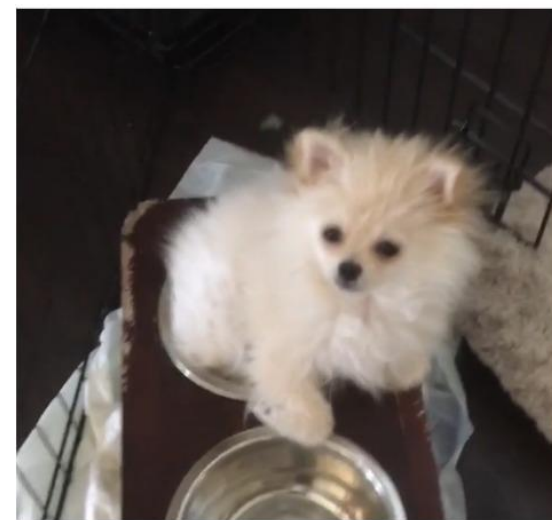


205 likes  
sunkissed\_kris Lace game strong 🍷💙📸  
@katrinanicolephotography... more  
View all 4 comments  
eric\_butterworth Killing it!!!  
zdoctore Love the look of this  
42 MINUTES AGO

sssniperwolf



sssniperwolf



7,019 views  
sssniperwolf Lumpy chillin in her food bowl 😊  
View all 68 comments  
crystalnicolee\_\_ She's so cute OMG!!! 😍😍  
chick\_101\_soccer Booiiii  
18 MINUTES AGO

**Ou encore :**

- ✓ **Les recommandations de musique sur Spotify**
- ✓ **La détection de spam dans les mails**
- ✓ **Autofocus de l'appareil photo de votre smartphone**
- ✓ **...**

***L'Intelligence Artificielle ce n'est pas que des robots, le mythe de l'Intelligence Artificielle qui remplace l'homme est un fantasme***



# Concrètement, c'est quoi l'IA ?

## ***Il existe deux types d'Intelligence Artificielle :***

- L'IA faible ou verticale :

- n'a pas vocation à évaluer
- ne va pas évoluer
- vise à effectuer une ou plusieurs tâches

**-> La machine apprend à reproduire un comportement humain, elle peut faire cette tâche mieux qu'un humain mais elle ne sait rien faire d'autres que la tâche qu'on lui aura apprise.**

- L'IA forte ou horizontale :

- approche qui cherche à comprendre et construire
  - \* Des systèmes de plus en plus autonomes
  - \* Des algorithmes capable de résoudre des problèmes
  - \* La programmation d'un apprentissage

**-> Reproduit toute l'intelligence humaine, avec sa capacité de conscience et de pouvoir réappliquer la tâche que l'on a appris à tout autre chose.**

Environnement technique adéquat



# Des algorithmes de Data Science innovants

Analyse de  
texte

VoiceMining

Reconnaissance  
d'image

Deep/Machine  
Learning

Théorie du  
signal

# IA, Machine Learning

**L'Intelligence Artificielle** est la capacité des machines à se comporter d'une manière que nous considérons comme «intelligente», notamment les fonctions cognitives.



**Le Machine Learning** utilise l'intelligence artificielle pour traiter d'importants volumes de données, permettre à la machine d'apprendre par elle-même et de s'améliorer avec l'expérience

# Objectifs à long terme de l'IA

## Construire des agents intelligents

- capable de résoudre des problèmes variés
- de façon autonome
- en s'adaptant au contexte rencontré

## Principales difficultés

- proposer des méthodes génériques
- sans trop d'hypothèses sur les problèmes à résoudre
- mais suffisamment précises pour pouvoir proposer des théories et des algorithmes

# En résumé

- Nous produisons chaque jour de nombreuses données comme les emails, les photos, etc.
- Le **Big Data**, c'est l'ensemble de ces données massives.
- L'**intelligence artificielle** et la **Data Science** sont deux disciplines qui sont utilisées conjointement, notamment pour mettre en place du **Machine Learning** ou du **Deep Learning** .
- Nous définissons l'IA comme "**toute technologie informatique algorithmique qui permet de résoudre des problèmes complexes qu'on aurait cru réservés à l'intelligence humaine, en simulant des capacités humaines comme la perception et le raisonnement.**"